

Nina Sturm

Problemhaltige Textaufgaben lösen

Einfluss eines Repräsentationstrainings
auf den Lösungsprozess
von Drittklässlern



Springer Spektrum

Landauer Beiträge zur mathematikdidaktischen Forschung

Reihe herausgegeben von

J. Roth, Landau, Deutschland

S. Schuler, Landau, Deutschland

In der Reihe werden exzellente Forschungsarbeiten zur Didaktik der Mathematik an der Universität Koblenz-Landau publiziert. Sie umfassen das breite Spektrum der Forschungsarbeiten in der Didaktik der Mathematik am Standort Landau, das in der einen Dimension von empirischer Grundlagenforschung bis hin zur fachdidaktischen Entwicklungsforschung und in der anderen Dimension von der Unterrichtsforschung bis hin zur Hochschuldidaktischen Forschung reicht. Dabei wird das Lehren und Lernen von Mathematik vom Kindergarten über alle Schulstufen und Schulformen bis zur Hochschule und zur Lehrerbildung beleuchtet. In jedem Fall wird konzeptionelle Arbeit mit qualitativen und/oder quantitativen empirischen Studien verbunden. In der Reihe erscheinen neben Qualifikationsarbeiten auch Publikationen aus weiteren Landauer Forschungsprojekten.

Weitere Bände in der Reihe <http://www.springer.com/series/15787>

Nina Sturm

Problemhaltige Textaufgaben lösen

Einfluss eines Repräsentationstrainings
auf den Lösungsprozess von
Drittklässlern

Mit einem Geleitwort von Prof. Dr. Renate Rasch

Nina Sturm
Landau, Deutschland

Dissertation der Universität Koblenz-Landau, 2017

Diese Arbeit ist zugleich eine Dissertation mit dem Originaltitel „Generierung und Nutzung externer Repräsentationen beim Lösen problemhaltiger Textaufgaben in der Grundschule“ am Fachbereich 7: Natur- und Umweltwissenschaften der Universität Koblenz-Landau.

Landauer Beiträge zur mathematikdidaktischen Forschung
ISBN 978-3-658-21397-8 ISBN 978-3-658-21398-5 (eBook)
<https://doi.org/10.1007/978-3-658-21398-5>

Die Deutsche Nationalbibliothek verzeichnetet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Spektrum

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Spektrum ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Geleitwort

Die Befähigung von Kindern zum Problemlösen, das Heranführen an eigenständige Problemlöseaktivitäten ist im Mathematikunterricht der Grundschule noch immer nicht selbstverständlich. Zum einen fehlt es an geeigneten Aufgaben, zum anderen konzentrieren sich Lehrende in der knappen Unterrichtszeit verständlicherweise auf Hauptkomponenten des Mathematikunterrichts – das Rechnen und andere grundlegende Fähigkeiten. Mit den Bildungsstandards ist das Problemlösen stärker in den Mittelpunkt gerückt. Problemhafte Textaufgaben sind eine Aufgabendomäne, die sich zur Entwicklung von Problemlösefähigkeiten bei Grundschulkindern eignet. Ein zentrales Element hierbei sind externe Repräsentationen. Schon Grundschulkinder können lernen, diese zur Lösungsunterstützung zu nutzen. Wie und in welchem Maß dies bei Grundschulkindern gelingen kann, ist Gegenstand der vorliegenden Publikation.

Es gelingt Nina Sturm überzeugend auf der Grundlage einer umfassenden Untersuchung im Mathematikunterricht der dritten und vierten Klassenstufe das Potenzial zur Entwicklung von Problemlösekompetenzen durch Textaufgaben aufzuzeigen. Die Autorin widmet sich den Problemlösefähigkeiten aus der Sicht des Generierens externer Repräsentationen durch Grundschulkinder. Im Rahmen einer Pilotstudie galt es zunächst herauszufinden, welche Darstellungsformen nutzen Grundschulkinder, um das Verstehen und Lösen problemhaftriger Textaufgaben zu stützen. Wird vorwiegend mit Rechnungen gearbeitet, sind es skizzenhafte Darstellungen oder Herangehensweisen in Tabellenform, die insbesondere das Denken über Zuordnungen abbilden? Ein interessantes Untersuchungsziel, dass sich daran anschloss hat die Aufgabenabhängigkeit der Darstellungsformen, die die Schülerinnen und Schüler nutzten, zum Gegenstand. Die wichtigste Frage blieb allerdings, wie erfolgreich lässt sich der Einsatz von verschiedenen externen Repräsentationsformen beim Lösen trainieren, erfolgreich im Hinblick auf den Lösungserfolg und auf die Entwicklung von Problemlösekompetenzen.

Zunächst werden in der vorliegenden Publikation die am Lösungserfolg beteiligten Prozesse auf der Grundlage eines umfangreichen Quellenstudiums herausgearbeitet. Die theoretischen Ausführungen sind in zwei Kapitel gegliedert. Das erste Kapitel ist dem Problemlösen gewidmet. Ausgewählte problemhafte Textaufgaben, mögliche Repräsentationsformen sowie zugrunde liegende Prozesse werden vorgestellt und charakterisiert. Die Autorin denkt von den beiden grundlegenden Modellen aus, dem gestaltpsychologischen und dem informationsverarbeitenden Ansatz. Es wird herausgearbeitet, dass eine erfolgreiche Problemlösung davon abhängt, ob es gelingt, eine Repräsentation des Anfangszustandes zu konstruieren und darauf aufbauend die Problemstruktur immer vollständiger zu durchdringen. Nina Sturm verweist darauf, dass eine geeignete Repräsentation die Lösenden bei den notwendigen Umstrukturierungen

unterstützt. Es wird deutlich, dass die Anforderungen beim Problemlösen, sich von bisherigen Lösungserfahrungen zu trennen und Problemrepräsentationen zu verändern, vor allem für Grundschulkinder, die ja gerade erst Lösungserfahrungen gewonnen haben, mit Anstrengung verbunden ist. Es wird in Anlehnung an vorangegangene Studien im Zusammenhang mit den eigenen Untersuchungszielen darauf verwiesen, wie sich die Problemlösenovizen unterstützen und fördern lassen.

Im Zentrum des zweiten Kapitels stehen die argumentativen Tätigkeiten von Grundschulkindern, durch die Lösungsprozesse beeinflusst und im günstigen Fall vorangebracht werden können. Die Autorin verweist darauf, dass kommunikative Anlässe das Verankern von Wissen unterstützen sowie Argumentationsfähigkeiten und damit die Entwicklung eines subjektiven Beweisbedürfnisses ermöglichen. Mathematische Entdeckungen beim Lösen werden als inhaltliche Basis der individuellen Lösungsprozesse gesehen. Beim Austausch mit den Gleichaltrigen können Lösungsschritte, mentale Bilder und Denkvorgänge stärker bewusst werden als das ohne Austausch möglich wäre.

Mit Schlussfolgerungen aus den theoretischen Ausführungen wird im dritten Kapitel der empirische Teil der Untersuchung eingeleitet. Bevor Fragestellungen formuliert und Hypothesen zugeordnet werden, wird die eigene Studie in bisherige Forschungen zur Thematik eingebettet. Belege für die methodisch sorgfältig durchdachte Untersuchung findet der Leser/die Leserin im Kapitel vier.

Eindrucksvoll werden durch die Ergebnisse und die sich anschließende Diskussion in den Kapiteln fünf und sechs wissenschaftliche Belege zum Erfolg des von der Autorin konzipierten Repräsentationstrainings erbracht. Es kann nachgewiesen werden, dass sich Grundschulkinder durch ein Training zu Nutzung von Repräsentationsformen animieren lassen, eigene lösungsunterstützende Repräsentationen zu erzeugen. Die Autorin konnte weiterhin zeigen, dass sich mit den eingesetzten Aufgaben Fähigkeiten im Problemlösen fördern lassen.

Problemhaltige Textaufgaben sind Aufgaben, die auch nach mehrwöchigem Training das Denkpotenzial der Kinder immer wieder aufs Neue herausfordern. Bei einem Training kann es also immer nur darum gehen, dass die trainierte Personengruppe lernt, Unterstützungskomponenten, die den Lösungserfolg wahrscheinlicher machen, zu entwickeln. Dieses Experiment gelingt Nina Sturm eindrucksvoll beim Training von externen Repräsentationsformen für Textaufgaben. Dabei macht die Autorin berechtigterweise darauf aufmerksam, dass eine gewisse Bescheidenheit ratsam ist, wenn es um die Ergebnisse bezüglich des Lösungserfolges geht. Das gilt auch für die argumentativen Prozesse, die das Lösen begleiten - im Zusammenhang mit der Zuversicht, dass noch zahlreiche solcher Entwicklungsmöglichkeiten vor dieser Altersgruppe liegen.

Danke

Viele lieben Menschen haben mich auf meinem bisherigen Forschungsweg und bei der Entstehung dieser Arbeit begleitet und unterstützt.

Mein besonderer Dank gilt Frau Prof. Dr. Renate Rasch. Ihre Begeisterung und Leidenschaft für problemhafte Textaufgaben, ihr fundiertes Wissen sowie ihre langjährige Forschungserfahrung inspirierten mich immer wieder aufs Neue. In wertvollen Gesprächen und Beratungen gelang es ihr in besonderer Weise, mir das Vertrauen und die Freiheit zu geben, eigene Wege zu gehen, aber auch mich einzufangen und zu bremsen, wann immer es nötig war. Ich danke ihr für die gewinnbringenden Impulse und Ratschläge zu dieser Arbeit. Es macht mit Stolz, Teil ihrer Arbeitsgruppe gewesen zu sein.

Mein Dank gilt auch Herrn Prof. Dr. Wolfgang Schnott für seine hilfreichen Anregungen und die Betreuung während meiner Promotion. Die gemeinsamen Gespräche, in denen fachdidaktische und psychologische Sichtweisen symbiotisch genutzt wurden, eröffneten neue Perspektiven und brachten die Weiterentwicklung des Projektes voran.

Frau Prof. Dr. Marianne Grassmann danke ich sehr herzlich für die Begutachtung meiner Arbeit.

Ein spezieller Dank geht an alle Lehrkräfte und ehemaligen Kolleginnen und Kollegen, die mir die Umsetzung meines Projektes trotz des langen Untersuchungszeitraumes ermöglichten. Sie sorgten für die entsprechenden Rahmenbedingungen und eine gute Atmosphäre. Auch den Dritt- und Viertklässlern gilt mein besonderer Dank. Ihre Begeisterungsfähigkeit, ihre Motivation und ihre vielfältigen, ausgeklügelten Herangehensweisen erfreuten und beeindruckten mich.

Als Promovendin des DFG-geförderte Graduiertenkollegs UpGrade hatte ich die Möglichkeit an vielen nationalen und internationalen Tagungen teilzunehmen. Für diese Möglichkeit sowie die angebotenen Austausch- und Beratungsformate und die gewinnbringenden Gespräche im Rahmen der interdisziplinären Forschungskolloquien möchte ich mich bei Dr. Heidrun Ludwig und dem gesamten UpGrade-Team bedanken. Auch an alle Hilfskräfte, die mich in dieser Zeit unterstützt und begleitet haben, geht mein Dank.

Die vielen aufmunternden Gespräche und Aktivitäten mit meinen lieben Kolleginnen und Kollegen werden mir in Erinnerung bleiben. Mit Kerstin Sitter denke ich an eine gutherzige Bürokollegin zurück, mit der ich viel gelacht habe, aber auch fachliche und ernste Gespräche führen konnte. Ihren moralischen Beistand und ihre unermüdliche Unterstützung weiß ich sehr zu schätzen. Ich freue mich darauf, beides bald zurückgeben zu dürfen. Mit Caroline Bhowmik habe ich eine aufregende Zeit im Graduierten-

kolleg verbracht. Wir haben viel gemeinsam erlebt, internationale Tagungen besucht, uns ausgetauscht, gegenseitig aufgebaut und Mut zugesprochen. Trotz der Entfernung, die heute zwischen uns liegt, gelingt es uns, den Kontakt zu halten. Ich freue mich sehr, mit beiden Kolleginnen freundschaftlich verbunden zu sein.

Meinen Freundinnen und Freunden, die mich in dieser Zeit mental gestärkt und getragen haben, möchte ich für ihre Geduld, ihr Verständnis, ihren Rückhalt aber auch für ihre Spontanität ganz herzlich danken.

Dankbar und tief verbunden bin ich meinem Partner, Tobias Juraschek, für sein Verständnis bei der Fertigstellung der Arbeit sowie seine Geduld und Unterstützung in dieser beschwerlichen Zeit.

Mein ganz besonderer Dank gilt meiner Familie, meinen Eltern und meiner Schwester Nicole, die mich über die gesamten Jahre unermüdlich unterstützt haben und denen ich diese Arbeit widme. Sie haben alle Höhen und Tiefen miterlebt und damit verbunden, die Abgabe der Arbeit herbeigesehnt. Auch wenn die Familienzeit zu kurz kam, meine Nerven oft blank lagen, erfuhr ich immer Verständnis, Beistand und Halt. Es tut gut zu wissen, dass ich auf deren Unterstützung immer zählen kann.

Zusammenfassung

Das Lösen problemhaltiger Aufgaben sollte Grundschulkindern am Ende der vierten Jahrgangsstufe mithilfe ihrer erworbenen mathematischen Kenntnisse, Fertigkeiten und Fähigkeiten erfolgreich gelingen. Wenn man jedoch die empirischen Befunde berücksichtigt, dann stellt die anspruchsvolle mathematische Struktur und Komplexität von Problemaufgaben für viele Grundschulkinder eine Barriere dar, die sie mit eigenen Kräften kaum bewältigen können. Die mangelnden Problemlösekompetenzen und ausbleibenden Erfolge könnten zum einen ein Indiz dafür sein, dass Problemaufgaben für Grundschulkinder Handlungsaufforderungen außerhalb ihrer Zone der nächsten Entwicklung darstellen, welche von ihnen *noch* nicht zu bewältigen sind. Dagegen spricht jedoch, dass aus didaktischer Perspektive der Einsatz problemhaltiger Aufgaben in der Grundschule befürwortet wird. Zum anderen geben die Befunde Hinweise darauf, dass es ungeübten Problemlösenden letztlich an Erfahrung mangelt und sie weder auf ein Wissensnetz noch auf Werkzeuge zurückgreifen können, die sie zum Lösen der Aufgaben befähigen. Aus diesem Grund geht die vorliegende Arbeit der Frage nach, ob die Problemlösekompetenzen von Grundschulkindern durch eine mehrwöchige Intervention gefördert werden können. Aus psychologischer und mathematikdidaktischer Sicht werden externe Repräsentationen als Werkzeuge angesehen, mit deren Hilfe ProblembARRIEREN erfolgreich überwunden werden können. Zudem bilden sie eine Kommunikationsgrundlage, die insbesondere für Grundschulkinder hilfreich und unterstützend sein kann. Unter der Annahme, dass Unterricht, der das Problemlösen, Darstellen und Kommunizieren vernetzt, der Problematik entgegenwirken kann, dass ungeübte, erfolglose Lösende hauptsächlich versuchen die Lösung im Kopf zu produzieren, sollte die gezielte Förderung der Generierung und Nutzung externer Repräsentationen positive Effekte auf den Lösungserfolg und die Problemlösekompetenzen haben.

Im empirischen Teil der Arbeit wird eine Interventionsstudie evaluiert, die das Externalisieren der Denkwege sowie das Kommunizieren und Reflektieren darüber anregte. Den Ergebnissen zufolge, sind positive Effekte auf den Lösungserfolg und die Problemlösekompetenzen zu verzeichnen. Die erwirkte Leistungssteigerung zeigt einerseits, dass problemhaltige Aufgaben für die Grundschule geeignet sind und Grundschulkinder der Handlungsaufforderungen nachkommen können. Andererseits zeigt die positive Entwicklung, dass die Generierung und Nutzung externer Repräsentationen trainiert werden kann und externe Repräsentationen für Grundschulkinder zum Denk- und Erkenntniswerkzeuge werden können. Als unterrichtspraktische Konsequenz lässt sich ableiten, dass ein Unterricht, der dem explorierenden Ansatz folgt und die Kinder dafür sensibilisiert, ihren Lösungsgedanken und -ideen eine Struktur zu

geben, die Lernenden Schritt für Schritt zum selbstständigen und erfolgreichen Problemlösen führen kann. Die Förderung externer Repräsentationen sollte bereits in der Grundschule stärker forciert werden.

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	XVII
Tabellenverzeichnis	XIX
Abbildungsverzeichnis.....	XXIII
Einleitung.....	1
Theoretischer Teil	5
1 Mathematisches Problemlösen in der Grundschule	7
1.1 Mathematisches Problemlösen in den Bildungsstandards der Grundschule	7
1.2 Was ist ein (mathematisches) Problem?	9
1.3 (Mathematisches) Problemlösen aus psychologischer Sicht	10
1.3.1 Gestaltpsychologischer Ansatz.....	11
1.3.2 Informationsverarbeitender Ansatz	13
1.3.3 Representational Change Theory	14
1.3.3.1 Grundlage der Konzeptualisierung	14
1.3.3.2 Problemlösen aus Sicht der Representational Change Theory	16
1.4 Mathematisches Problemlösen auf Basis problemhaltiger Textaufgaben ...	22
1.4.1 Charakteristika problemhaltiger Textaufgaben	23
1.4.2 Empirische Erkenntnisse	24
1.4.3 Schlussfolgerung	26
1.5 Mathematisches Problemlösen mit externen Repräsentationen.....	28
1.5.1 Repräsentationen – Formen, Merkmale, Funktionen	28
1.5.2 Verstehensprozess beim Lösen problemhaltiger Textaufgaben ...	30
1.5.3 Schwierigkeiten im Verstehensprozess beim Lösen problemhaltiger Textaufgaben.....	34
1.5.3.1 Sprachlich und mathematisch bedingte Schwierigkeiten.....	34
1.5.3.2 Repräsentationsbedingte Schwierigkeiten	36
1.5.3.3 Arbeitsgedächtnis.....	38
1.5.3.4 Vor- und Erfahrungswissen	38
1.5.4 Ableitungen aus den Schwierigkeiten im Verstehensprozess	39
1.5.5 Potenzial externer Repräsentationen	43

1.5.5.1	Bedeutung für den Produzenten.....	44
1.5.5.2	Bedeutung für die Interaktion unter Gleichaltrigen....	45
1.5.5.3	Bedeutung für Kommunikations- und Reflexionsphasen im Mathematikunterricht.....	47
1.5.5.4	Schlussfolgerungen und Konsequenzen für den Unterricht	50
1.5.6	Formen externer Repräsentationen im Mathematikunterricht.....	52
1.5.6.1	Zeichnungen.....	55
1.5.6.2	Rechnungen.....	57
1.5.6.3	Tabellen.....	60
1.5.6.4	Listen.....	62
1.5.6.5	Schriftsprachliche Beschreibungen.....	63
1.5.7	Externe Repräsentationen in den Bildungsstandards der Grundschule.....	64
2	Argumentative Tätigkeiten im intersubjektiven Austausch.....	67
2.1	Relevanz argumentativer Tätigkeiten in der Grundschule.....	68
2.2	Argumentieren – Begründen – Beweisen	70
2.2.1	Argumentieren auf Basis externer Repräsentationen	70
2.2.2	„Wann fängt das Beweisen an?“	71
2.2.3	Argumentieren als experimentelles Beweisen	74
2.2.4	Argumentieren als inhaltlich-anschauliches Beweisen	76
2.2.5	Begründen als übergeordnete Kompetenz.....	79
2.3	Argumentieren im Rahmen der Untersuchung	80
2.4	Potenzial intersubjektiven Austauschs	81
2.4.1	Voneinander- und Miteinander-Lernen	82
2.4.2	Metakognitive Aktivitäten.....	83
2.4.3	Prozesse des Helfens und Kollaborierens.....	84
	Empirischer Teil.....	87
3	Förderung der Generierung und Nutzung externer Repräsentationen	89
3.1	Schlussfolgerungen aus den theoretischen Ausführungen.....	89
3.2	Schlussfolgerungen aus dem Stand der Forschung	90
3.3	Forschungsfragen und Hypothesen	93
3.3.1	Ziele des Trainingsexperiments.....	93
3.3.2	Forschungsfragen.....	94
4	Methode.....	105

4.1	Methodische Überlegungen und Entscheidungen	105
4.1.1	Unterrichtskonzept.....	105
4.1.1.1	Rolle der Lehrkraft.....	110
4.1.1.2	Rolle der Lernenden.....	112
4.1.2	Kriterien zur Aufgabenauswahl.....	113
4.2	Erkenntnisse und Konsequenzen aus der Pilotstudie	116
4.2.1	Monitoring des Unterrichtskonzepts	117
4.2.1.1	Trainingskomponenten.....	117
4.2.1.2	Kommunikative Settings.....	119
4.2.2	Wahl der Aufgaben.....	123
4.2.2.1	Aufgaben der Pilotstudie.....	123
4.2.2.2	Trainingsaufgaben.....	133
4.2.3	Kategorisierung externer Schülerrepräsentationen	138
4.2.3.1	Entwicklung des Kategoriensystems	138
4.2.3.2	Kategoriensystem zur Kodierung des Datenmaterials	145
4.2.4	Ableitungen des organisatorischen Rahmens	153
4.2.4.1	Wahl der Altersgruppe	153
4.2.4.2	Wahl der Trainingsdauer.....	157
4.3	Design der experimentellen Interventionsstudie.....	158
4.4	Das Training	162
4.4.1	Trainingsrationale	163
4.4.2	Ausgangslage des Trainings	164
4.4.3	Zentrale Ziele des Trainings	165
4.4.4	Aufbau des Trainings	166
4.5	Variablen und ihre Operationalisierung	169
4.5.1	Abhängige Variablen.....	169
4.5.1.1	Wahl der abhängigen Variablen.....	170
4.5.1.2	Grundgedanken zur Operationalisierung	173
4.5.1.3	Textaufgaben-Performance-Test.....	179
4.5.1.4	Bildung der Testscores	187
4.5.2	Unabhängige Variablen: Repräsentationsmerkmale	188
4.5.3	Kontrollvariablen.....	192
4.5.3.1	Intelligenz.....	193
4.5.3.2	Mathematische Basiskompetenzen	196
4.5.3.3	Sprachliche Fähigkeiten und Leseverständnis	198
4.5.3.4	Metakognitive Fähigkeiten	200

4.5.3.5 Attitudes und Beliefs.....	213
4.5.3.6 Mehrsprachigkeit	218
4.5.3.7 Geschlecht.....	219
4.6 Versuchsdurchführung	220
4.7 Einordnung der Versuchsteilnehmer.....	223
4.7.1 Die Trainingsgruppen.....	225
4.7.2 Die Nicht-Trainingsgruppen.....	225
4.8 Statistische Methoden	226
4.8.1 Multivariate Varianzanalyse bei der Testung der Gleichverteilung der Leistungsvoraussetzungen.....	226
4.8.2 Univariate Varianzanalyse bei der Testung der Gleichverteilung der Prätest-Leistungen	227
4.8.3 Testung der Verletzung der Unabhängigkeitsannahme.....	229
4.8.4 Univariate Varianzanalysen mit Messwiederholung bei der Testung der Merkmalsausprägungen zu verschiedenen Messzeitpunkten	229
4.8.5 Aufgabenspezifische Analyse der Effekte auf den Lösungserfolg	230
4.8.6 Analyse des Effekts selbstgenerierter externer Repräsentationen auf den Lösungserfolg bzw. die Problemlösekompetenzen.....	231
5 Ergebnisse	241
5.1 Vor-Analysen zur Vergleichbarkeit der Stichprobe.....	242
5.1.1 Test der Gleichverteilung der Leistungsvoraussetzungen für die vier experimentellen Bedingungen.....	242
5.1.2 Test der Gleichverteilung der Prätest-Leistungen für die vier experimentellen Bedingungen	245
5.1.3 Testung der Verletzung der Unabhängigkeitsannahme.....	246
5.1.4 Test der Einflüsse des Geschlechts und der Mehrsprachigkeit ..	247
5.2 Vor-Analyse zur Wirksamkeitshypothesenprüfung.....	248
5.3 Analyse der abhängigen Variablen	253
5.3.1 Forschungsfrage 1: Effekt der Interventionsmaßnahme	254
5.3.2 Forschungsfrage 2: Effekt des Repräsentationstrainings	263
5.3.3 Forschungsfrage 3: Effekt der Kommunikation	273
5.3.4 Forschungsfrage 4: Effekt unterschiedlicher Gruppen.....	281
5.3.5 Zusammenfassung bisheriger Ergebnisse.....	293

5.3.6	Forschungsfrage 5: Effekt selbstgenerierter externer Repräsentationen	296
6	Diskussion	307
6.1	Ziele und Fragestellung der Interventionsstudie	307
6.2	Forschungsfrage 1: Effekt der Interventionsmaßnahme	310
6.3	Forschungsfrage 2: Effekt des Repräsentationstrainings	314
6.4	Forschungsfrage 3: Effekt der Kommunikation	319
6.5	Forschungsfrage 4: Effekt der Gruppe	324
6.6	Forschungsfrage 5: Effekt selbstgenerierter externer Repräsentationen ...	327
6.7	Methodische Überlegungen	331
6.8	Implikationen für die Praxis	334
6.9	Zusammenfassung und Ausblick	337
	Literatur	343

Abkürzungsverzeichnis

T+ Training

T- kein Training

K+ kommunikative Zweiersettings

K- keine kommunikativen Zweiersettings

T+K+ Training, kommunikative Zweiersettings

T+K- Training, keine kommunikativen Zweiersettings

T-K+ kein Training, kommunikative Zweiersettings

T-K- kein Training, keine kommunikativen Zweiersettings (Kontrollgruppe)

Prä Prätest

Post Posttest

Fo Follow-up-Test

Tabellenverzeichnis

Tabelle 3.1. Hypothesen zu dem Lösungserfolg, den Problemlösekompetenzen und der Anzahl selbstgenerierter externer Repräsentationen trainierter und untrainierter Lernender	98
Tabelle 3.2. Hypothesen zu dem Lösungserfolg, den Problemlösekompetenzen und der Zahl selbstgenerierter externer Repräsentationen der Gruppe mit und ohne Kommunikation.....	100
Tabelle 3.3. Hypothesen zu dem Lösungserfolg, den Problemlösekompetenzen und der Anzahl selbstgenerierter externer Repräsentationen der vier experimentellen Bedingungen	101
Tabelle 4.1. Ergebnisse der Itemanalyse für die 11 problemhaltigen Textaufgaben der Pilotstudie	127
Tabelle 4.2. Kategorien von Zeichnungen auf Basis einer inhaltlich strukturierenden Inhaltsanalyse	146
Tabelle 4.3. Kategorien von Rechnungen auf Basis einer inhaltlich strukturierenden Inhaltsanalyse	148
Tabelle 4.4. Kategorien von Tabellen auf Basis einer inhaltlich strukturierenden Inhaltsanalyse.....	150
Tabelle 4.5. Kategorien zur Erfassung schriftsprachlicher Beschreibungen.....	152
Tabelle 4.6 Zentrale Trainingsbausteine der Reflexionsphase	167
Tabelle 4.7. Instrument zur Analyse der individuellen Problemlöseprozesse.....	172
Tabelle 4.8. Textaufgaben-Performance-Test: Reliabilitäten der Skalen.....	181
Tabelle 4.9. Faktorladungen der Hauptachsenanalyse für die Repräsentationsmerkmale zum Zeitpunkt des Posttests	190
Tabelle 4.10. Verteilungskennwerte der CPM-Normierungsstichprobe nach Altersgruppen auf Rohwert-Basis.....	195
Tabelle 4.11. Item-Skala-Statistik zur Trennschärfe der 8 Lehrerfragebogenitems der metakognitiven Fähigkeiten der Lernenden	202
Tabelle 4.12. Metakognitive Fähigkeiten: Reliabilitäten der Skalen	205
Tabelle 4.13. Analyse der Schülerfragebogenitems metakognitiver Fähigkeiten zum Prätest (PAC) und Posttest (RAC)	205
Tabelle 4.14. Korrelations- und Kovarianzmatrix der vier Faktoren zum Posttest....	208
Tabelle 4.15. Faktorladungen der Hauptachsenanalyse für die Schülerfragebogenitems metakognitiver Fähigkeiten zum Zeitpunkt des Posttests (RAC).....	209
Tabelle 4.16. Metakognitive Fähigkeiten: Reliabilitäten der Skalen	212
Tabelle 4.17. Attitudes und Beliefs: Reliabilitäten der Skalen.....	215

Tabelle 4.18. Analyse der Schülerfragenbogenitems Attitudes und Beliefs zum Prätest (PAII) und zum Posttest (RAII)	215
Tabelle 4.19. Ablauf der Studie mit den einzelnen Erhebungszeitpunkten.....	221
Tabelle 4.20. Blockbildung der 20 Klassen in Units durch das Parallelisierungsverfahren auf Basis der Intelligenz, der mathematischen Basiskompetenzen und des Leseverständnisses	224
Tabelle 4.21. Übersicht über die Verteilung der Probanden in den Trainingsgruppen T+K+ und T+K-	225
Tabelle 4.22. Übersicht über die Verteilung der Probanden in den Nicht-Trainingsgruppen T-K+ und T-K-	225
Tabelle 4.23. Einfluss der Störvariablen auf das Kriterium Lösungserfolg	232
Tabelle 4.24. Schrittweise multiple Regression anhand der Tabellenmerkmale zur Aufspürung von Suppressoreffekten	236
Tabelle 5.1. Ausgangsvoraussetzungen auf Gruppenebene nach dem Parallelisierungsverfahren zum Messzeitpunkt 0	243
Tabelle 5.2. Deskriptive Ergebnisse der Trainingsgruppen mit und ohne Kommunikation in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten	249
Tabelle 5.3. Deskriptive Ergebnisse der Trainingsgruppen mit und ohne Kommunikation in der Anzahl selbstgenerierter externer Repräsentationen zu zwei Messzeitpunkten	252
Tabelle 5.4. Deskriptive Ergebnisse der Trainings- und Nicht-Trainingsgruppe im Lösungserfolg zu allen Messzeitpunkten.....	263
Tabelle 5.5. Deskriptive Ergebnisse der Trainings- und Nicht-Trainingsgruppe in den Problemlösekompetenzen zu allen Messzeitpunkten.....	266
Tabelle 5.6. Deskriptive Ergebnisse der Trainings- und Nicht-Trainingsgruppe in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten	270
Tabelle 5.7. Deskriptive Ergebnisse der Kommunikationsgruppe und der Nicht-Kommunikationsgruppe im Lösungserfolg zu allen Messzeitpunkten	273
Tabelle 5.8. Deskriptive Ergebnisse der Kommunikationsgruppe und der Nicht-Kommunikationsgruppe in den Problemlösekompetenzen zu allen Messzeitpunkten	276
Tabelle 5.9. Deskriptive Ergebnisse der Kommunikationsgruppe und der Nicht-Kommunikationsgruppe in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten	279
Tabelle 5.10. Deskriptive Ergebnisse der vier experimentellen Bedingungen im Lösungserfolg zu allen Messzeitpunkten.....	282
Tabelle 5.11. Deskriptive Ergebnisse der vier experimentellen Bedingungen in den Problemlösekompetenzen zu allen Messzeitpunkten.....	286

Tabelle 5.12. Deskriptive Ergebnisse der vier experimentellen Bedingungen in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten	290
Tabelle 5.13. Übersicht über Haupt- und Interaktionseffekte der 3x2x2-Varianzanalyse mit Messwiederholung auf dem ersten Faktor für alle drei abhängigen Variablen	294
Tabelle 5.14. Korrelationsmatrix der Kontrollvariablen und der abhängigen Variablen Lösungserfolg.....	296
Tabelle 5.15. Hierarchische Regressionsanalyse von Kontrollvariablen und Repräsentationsmerkmalen auf den Lösungserfolg.....	297
Tabelle 5.16. Korrelationsmatrix der Kontrollvariablen und der abhängigen Variablen Problemlösekompetenzen	302
Tabelle 5.17. Hierarchische Regressionsanalyse von Kontrollvariablen und Repräsentationsmerkmalen auf die Problemlösekompetenzen	303

Abbildungsverzeichnis

Abbildung 1.1. Integratives Modell des Text- und Bildverständnisses nach Schnotz.....	31
Abbildung 4.1. Ablaufmodell der inhaltlich strukturierenden Inhaltsanalyse nach Mayring	142
Abbildung 4.2. Experimentelle Bedingungen des Trainingsexperimentes.	159
Abbildung 4.3. Vierstufige Ratingskala zur Vorhersagefähigkeit (A) und Evaluierungsfähigkeit (B).....	211
Abbildung 4.4. Residuenanalyse des Regressionsmodells mit 13 Prädiktoren.....	237
Abbildung 4.5. Residuenanalyse des Regressionsmodells mit 37 Prädiktoren.	239
Abbildung 5.1. Vergleichbarkeit der vier experimentellen Bedingungen hinsichtlich der Leistungsvoraussetzungen (Intelligenz, mathematische Basiskompetenzen, Leseverständnis) nach dem Parallelisierungsverfahren zum Messzeitpunkt 0.	244
Abbildung 5.2. Mittlere Anzahl selbstgenerierter externer Repräsentationen für die Klassen mit kommunikativen Zweiersettings (K+) und ohne kommunikative Zweiersettings (K-) zu drei Messzeitpunkten.....	250
Abbildung 5.3. Mittlere Anzahl von Zeichnungen, Rechnungen, Listen, schriftsprachlichen Beschreibungen und Tabellen für die Trainingsgruppen mit und ohne kommunikative Zweiersettings zu zwei Messzeitpunkten.	251
Abbildung 5.4. Mittlerer Testscore des Lösungserfolgs zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.	255
Abbildung 5.5. Mittlere Aufgabenscores des Lösungserfolgs zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.	256
Abbildung 5.6. Mittlerer Testscore der Problemlösekompetenzen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.	258
Abbildung 5.7. Mittlere Aufgabenscores der Problemlösekompetenzen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.	259
Abbildung 5.8. Mittlerer Testscore der Anzahl selbstgenerierter externer Repräsentationen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.	261
Abbildung 5.9. Mittlere Aufgabenscores der Anzahl selbstgenerierter externer Repräsentationen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.	262

Abbildung 5.10. Mittlerer Testscore des Lösungserfolgs der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.	264
Abbildung 5.11. Mittlere Aufgabenscores des Lösungserfolgs der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.	265
Abbildung 5.12. Mittlerer Testscore der Problemlösekompetenzen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.	267
Abbildung 5.13. Mittlere Aufgabenscores der Problemlösekompetenzen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.	268
Abbildung 5.14. Mittlerer Testscore der Anzahl selbstgenerierter externer Repräsentationen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.	270
Abbildung 5.15. Mittlere Aufgabenscores der Anzahl selbstgenerierter externer Repräsentationen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.	272
Abbildung 5.16. Mittlerer Testscore des Lösungserfolgs der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte.	274
Abbildung 5.17. Mittlere Aufgabenscores des Lösungserfolgs der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte.	275
Abbildung 5.18. Mittlerer Testscore der Problemlösekompetenzen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppen (K-) über drei Messzeitpunkte.	277
Abbildung 5.19. Mittlere Aufgabenscores der Problemlösekompetenzen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte.	278
Abbildung 5.20. Mittlerer Testscore der Anzahl selbstgenerierter externer Repräsentationen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte.	279
Abbildung 5.21. Mittlere Aufgabenscores der Anzahl selbstgenerierter externer Repräsentationen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte.	280
Abbildung 5.22. Relative Häufigkeiten des Lösungserfolgs-Testscores zum Zeitpunkt des Posttests.....	282
Abbildung 5.23. Verteilung des mittleren Lösungserfolgs für die vier experimentellen Bedingungen.	283
Abbildung 5.24. Grafische Darstellung der Interaktion 2. Ordnung (Zeit x Training x Kommunikation) für den Lösungserfolg.	284
Abbildung 5.25. Grafische Darstellung der Interaktion 3. Ordnung (Zeit x Training x Kommunikation x Aufgabentyp) für den Lösungserfolg.....	285

Abbildung 5.26. Verteilung der mittleren Problemlösekompetenzen für die vier experimentellen Bedingungen.	287
Abbildung 5.27. Grafische Darstellung der Interaktion 2. Ordnung (Zeit x Training x Kommunikation) für die Problemlösekompetenzen.	288
Abbildung 5.28. Grafische Darstellung der Interaktion 3. Ordnung (Zeit x Training x Kommunikation x Aufgabentyp) für die Problemlösekompetenzen.	289
Abbildung 5.29. Relative Häufigkeit der Anzahl selbstgenerierter externer Repräsentationen zum Posttest.	291
Abbildung 5.30. Grafische Darstellung der Interaktion 2. Ordnung (Zeit x Training x Kommunikation) für die Anzahl selbstgenerierter externer Repräsentationen.	292
Abbildung 5.31. Grafische Darstellung der Interaktion 3. Ordnung (Zeit x Training x Kommunikation x Aufgabentyp) für die Anzahl selbstgenerierter externer Repräsentationen.	293
Abbildung 6.1. Mittelwertsunterschiede für den Lösungserfolg, die Problemlösekompetenzen und die Anzahl selbstgenerierter externer Repräsentationen für alle experimentellen Bedingungen.	310



Einleitung

„Ein Schüler, von dem nie etwas verlangt wird, was er nicht kann, wird auch nie alles leisten, wozu er fähig ist.“

John Stuart Mill

Hätten Kinder als Mitglieder einer Schulbuchjury die Möglichkeit, ihr Mathematikschulbuch für das nächste Schuljahr mitzubestimmen, so wäre vermutlich ein Auswahlkriterium, dass möglichst wenige Textaufgaben enthalten sein sollen. Wenn es nach vielen von ihnen gehen würde, wären Textaufgaben kein Bestandteil des Mathematikunterrichts mehr, eher längst abgeschafft, auch wenn sie fest im Curriculum verankert sind (KMK 2005b, S. 9). Es handelt sich um Aufgaben, die sie zum einen herausfordern, die sie vielleicht auch nicht auf Anhieb lösen können, weil sie etwas von ihnen verlangen, was sie scheinbar noch nicht können, aber zu lernen fähig sind. Zum anderen sind es Aufgaben, die in ihrer Zone der nächsten Entwicklungen liegen, Möglichkeiten der Nachahmung schaffen und sie über sich hinauswachsen lassen (Vygotskij, 2002, S. 331, 335).

Dass problemhaltige Textaufgaben in der Zone der nächsten Entwicklung von Grundschulkindern liegen, geht insbesondere aus zurückliegenden Ist-Stand-Analysen hervor. Die von Grundschulkindern erzielten Lösungsraten fallen gering und nicht zufriedenstellend aus (Groß, 2013, S. 114–133; Hohn, 2012, S. 106). Die als mühselig empfundene Lösungssuche scheitert häufig schon beim Verstehen des Sachverhalts und dessen Übersetzung in die mathematische Sprache. „Was muss ich den jetzt rechnen?“, „Plus, minus, mal oder durch?“, „Kannst du mir das sagen?“ sind Fragen, die Lehrkräfte im problemlösenden Mathematikunterricht begegnen. Viele empirische Studien zeigen, dass Problemlösenovizen größtenteils die Lösung mental produzieren, sie nach Schlüsselwörtern suchen, die ihnen Hinweise auf die richtige Rechenoperation liefern, oder aber nach einer Verbindung zu dem aktuellen Lernstoff. Wird gerade das Multiplizieren gelernt, so muss aus Sicht der Kinder die Multiplikation in der Textaufgabe auch relevant sein. Nicht selten favorisieren sie schon in der Grundschule einen rechnerischen Zugang (R. Rasch, 2009, S. 69) und verknüpfen gegebene Informationen miteinander, wenn auch teilweise willkürlich und unreflektiert (Booker, Bond, Sparrow, & Swan, 2014, S. 76; Muir, Beswick, & Williamson, 2008, S. 231). Was aber, wenn die Textaufgabe zusätzlichen Problemcharakter aufweist, ihr eine ungewohnte mathematische Struktur zugrunde liegt, welche die Suche nach Schlüsselwörtern oder nach Parallelen zum aktuellen Unterrichtsinhalt erfolglos macht? Welche

Handlungsmöglichkeiten können dann beim Lösen von Textaufgaben unterstützend herangezogen werden?

Wenn man lernpsychologische Erkenntnisse heranzieht, dann gelingt die Aktivierung erforderlicher Operationen, wenn die geschilderte Problemsituation adäquat repräsentiert wird (Kintsch & Greeno, 1985, S. 109; Mayer & Hegarty, 1996, S. 34; Stern, 1998, S. 118). Gelingt dies nicht auf direktem Wege, so können aus Sicht der *Representational Change Theory* repräsentationale Veränderungen den Problemlöseprozess voranbringen (Gick & Lockhart, 1995, S. 215; Ohlsson, 1984a, S. 125). Eine weitere Unterstützungsmaßnahme wird im Austausch mit externen Repräsentationen und über externe Repräsentationen gesehen (Dörfler, 2006, S. 215).

An dieser Stelle setzt die vorliegende Untersuchung an. Es wird versucht, durch die Entwicklung und Implementierung eines Repräsentationstrainings, das schlechte Abschneiden der Grundschulkinder beim Problemlösen, zu kompensieren. Ausgangspunkt war, das Potenzial der Grundschulkinder aufzugreifen, in den Mittelpunkt zu stellen und zum Lern- und Unterrichtsgegenstand zu machen. Die Kinder zum Externalisieren ihrer Gedankengänge und Ideen anzuregen, ohne sie in ihrer Kreativität und Ideenvielfalt einzuschränken, war ein Ziel der Untersuchung. Ein Training auf Basis ihrer eigenen individuellen Ressourcen sollte sie beim Generieren und Nutzen externer Repräsentationen unterstützen und stärken sowie den mathematischen Erkenntnisgewinn voranbringen. Dabei wurde fokussiert, welche Heran- und Vorgehensweisen die Lernenden von sich wählen, welche repräsentationalen Veränderungen sie in Betracht ziehen, aber auch, welche Vor- und Nachteile die externen Repräsentationen im konkret zu lösenden Fall mit sich bringen. Die Kinder sollten vielfältige Lösungserfahrungen sammeln und sich am Konstruieren geeigneter Repräsentationen erproben. Dies geht einher mit der Tatsache, dass sich externe Repräsentationen eignen, mathematisches Wissen und zugrunde liegende Strukturen öffentlich und für alle anschauungsgestützt zugänglich zu machen (Duval, 1999, S. 3; E. C. Wittmann, 2014, S. 227). Von Interesse ist zudem, ob ein Austausch unter Gleichaltrigen mit und über externe Repräsentationen ein Gewinn für die Grundschulkinder darstellt, ob es ihnen gelingt, Fehlerquellen aufzuspüren, auf gleicher Stufe voneinander zu lernen und sich gegenseitig zu helfen.

Ein Prätest-Posttest-Kontrollgruppen-Design soll Aufschluss geben, zu welchen Effekten bei Lernenden die Interventionsmaßnahme auf Basis externer Repräsentationen in Verbindung mit kommunikativen Settings im Mathematikunterricht der Grundschule führt.

Hierbei sind folgende Fragestellungen von zentraler Bedeutung:

1. Können Drittklässler durch die Interventionsmaßnahme ihren Lösungserfolg, ihre Problemlösekompetenzen und die Anzahl selbstgenerierter externer Repräsentationen erhöhen?
2. Unterscheiden sich trainierte und untrainierte Drittklässler nach der Interventionsmaßnahme in ihren Leistungen hinsichtlich des Lösungserfolgs, des Problemlösekompetenzzuwachses und der Anzahl selbstgenerierter externer Repräsentationen?
3. Unterscheiden sich Drittklässler, die sich unter Gleichaltrigen austauschen durften, in ihren Leistungen hinsichtlich des Lösungserfolgs, des Problemlösekompetenzzuwachses und der Anzahl selbstgenerierter externer Repräsentationen von Drittklässlern, die sich nicht unter Gleichaltrigen austauschen durften?
4. Welchen Einfluss haben die experimentellen Bedingungen aus den Faktoren Repräsentationstraining und kommunikative Zweiersettings auf die Leistungen (Lösungserfolg, Problemlösekompetenzen, Anzahl selbstgenerierter externer Repräsentationen) von Drittklässlern?
5. Welchen Einfluss kann selbstgenerierten externen Repräsentationen auf den Lösungserfolg bzw. die Problemlösekompetenzen zugesprochen werden und welche Repräsentationen lassen sich als gute Prädiktoren zur Vorhersage des Lösungserfolgs bzw. der Problemlösekompetenzen identifizieren?

Aus der Untersuchung werden vielfältige Implikationen für die Unterrichtspraxis erwartet. Es liegt die Vermutung nahe, dass die regelmäßige Implementierung problemhaltiger Textaufgaben eine Steigerung der Leistungen bewirkt, was deren Einsatz in der Grundschule legitimieren würde. Wenn es gelingt, mithilfe des Trainings die Vorteile der Generierung externer Repräsentationen für die Kinder erfahrbar zu machen, dann wird davon ausgegangen, dass die Kinder häufiger Denk- und Erkenntniswerkzeuge konstruieren und in ihren Lösungsprozess integrieren. Sie erleben die eigene Kompetenz, welche mit der externen Repräsentationen in Verbindung gebracht wird und nachhaltig ist (Heinrich, Bruder, & Bauer, 2015, S. 284). Somit wäre die Konstruktion externer Repräsentationen bereits in der Grundschule trainierbar. Ferner könnte ein Zugang über unterschiedliche Vor- und Herangehensweisen im Hinblick auf die heterogene Lerngruppe innerhalb einer Grundschulklasse ein Zugewinn für jeden Einzelnen sein. Jedes Kind könnte entsprechend seiner Fähigkeiten und Fertigkeiten, entsprechend seiner Zone der nächsten Entwicklung, lernen und nachahmen. In kommunikativen Settings könnten sie zusätzlich voneinander und miteinander lernen und ihre Kompetenzen ausbauen.

Theoretischer Teil

Die vorliegende Arbeit gliedert sich in zwei Hauptbereiche, den theoretischen und den empirischen Teil. Im Erstgenannten wird auf einer theoretischen Ebene ausführlich das mathematische Problemlösen beleuchtet. Neben der Darstellung, was in der vorliegenden Arbeit unter einem *mathematischen Problem* verstanden wird, wird in Kapitel 1 ausführlich aufgezeigt, wie es aus psychologischer Sicht Problemlösenden gelingen kann, ProblembARRIEREN erfolgreich zu überwinden und zur Einsicht zu gelangen. Dass hierbei insbesondere externe Repräsentationen lösungsunterstützend sein können, an denen repräsentationale Veränderungen und Umstrukturierungen vollzogen werden können, wird neben dem Herstellen von Bezügen und Verknüpfungen zur Fachdidaktik in diesem Kapitel dargelegt.

Die Rolle intersubjektiven Austauschs, dessen Relevanz und Potenzial wird in Kapitel 2 behandelt. Inwiefern Grundschulkinder in der Lage sind kommunikativ und argumentativ tätig zu werden und inwiefern dies beim Problemlösen gewinnbringend sein kann, wird in diesem Zusammenhang auf theoretischer Ebene beleuchtet.

Es sei an dieser Stelle darauf hingewiesen, dass die Quellen vorrangig im Text ausgewiesen werden. Im Theorienteil werden jedoch aus Gründen der Übersichtlichkeit und besseren Lesbarkeit die Quellen am Seitenende angegeben, wenn mehr als vier Quellen unterschiedlicher Autoren als Beleg angeführt sind.



1 Mathematisches Problemlösen in der Grundschule

Dem Problemlösen werden in der Mathematikdidaktik seit den 1980er Jahren, verstärkt jedoch seit der Jahrtausendwende mit den NCTM-Standards und dem schlechten Abschneiden deutscher Schüler in der PISA-Studie, besondere Aufmerksamkeit und Bedeutung zugesprochen.¹ Es wird die Ansicht geteilt, dass gerade die Mathematik prädestiniert sei, Problemlösekompetenzen anzubahnen und zu fördern (Heinze, 2007, S. 4; Kilpatrick, 1985, S. 3), weshalb Problemlöseprozesse zum festen Bestandteil des Mathematikunterrichts gehören sollten.² Hierbei ist explizit die Grundschulzeit eingeschlossen (Pehkonen, Näveri, & Laine, 2013, S. 19; Schipper, 2005, S. 358).

Das Lösen von Problemen soll weniger als Tätigkeit, sondern vielmehr als Fähigkeit aufgefasst werden, die es zu fördern gilt (Heinze, 2007, S. 4 in Anlehnung an Schoenfeld, 1985). Die Lernenden sollen dadurch befähigt werden, „intentional“ zu handeln, indem sie ihren Lösungsprozess selbstbestimmt wählen und steuern sowie ihre Handlungen bewusst und zielorientiert hinterfragen und begründen (A. Meyer & Fischer, 2013, S. 184; M. Miller, 1986, S. 141). Dies wird zum Grundsatz der vorliegenden Untersuchung gemacht und nachfolgend erörtert.

Das Kapitel wird eröffnet mit der Relevanz des mathematischen Problemlösens aus Sicht der Bildungsstandards (Abschnitt 1.1), gefolgt von der Problemdefinition, der die Untersuchung folgt (Abschnitt 1.2). Um aufzuzeigen, welche Prozesse das Problemlösen charakterisieren und wie es letztlich zu intentionalem Handeln kommen kann, ist es erforderlich, eine psychologische Sichtweise heranzuziehen (Abschnitt 1.3.1–1.3.2). Die der Studie zugrunde liegende Theorie leitet sich aus den vorherigen Ansätzen ab (Abschnitt 1.3.3) und wird im Anschluss daran in Bezug zum Lösen problemhaltiger Textaufgaben gesetzt (Abschnitt 1.4). Die Frage danach, welche Rolle dabei Externalisierungen zugesprochen wird, rundet das Kapitel ab (Abschnitt 1.5).

1.1 Mathematisches Problemlösen in den Bildungsstandards der Grundschule

Die Bildungsstandards für das Fach Mathematik akzentuieren die Dualität von inhaltsbezogenen und prozessbezogenen Kompetenzen und heben ausdrücklich hervor, dass „nicht nur Kenntnisse und Fertigkeiten, sondern stets auch inhaltsübergreifende Fähigkeiten in den Blick zu nehmen sind“ (Selter, 2005, S. 366; vgl. Schipper, 2005, S. 356;

¹ Bruder und Collet (2011b, S. 8), Heinze (2007, S. 3), Käpnick (2014, S. 20), NCTM (2000), Reiss und Törner (2007, S. 435, 437), Schipper (2005, S. 358)

² Bruder und Collet (2011b, S. 7), Heinze (2007, S. 3), Pehkonen (2001, S. 70), Pehkonen, Näveri und Laine (2013, S. 19), Reiss und Törner (2007, S. 431), Ruwisch (2014, S. 35), Winter (1975, S. 108, 1995, S. 42)

Walther, 2004, S. 2). Das *Problemlösen* zählt neben dem *Argumentieren*, *Kommunizieren*, *Modellieren* und *Darstellen* zu den allgemeinen mathematischen Kompetenzen, welche auch als prozessbezogene Kompetenzen bezeichnet werden (KMK 2005b, S. 7–8). Sie sind maßgeblich an der Entwicklung inhaltsbezogener mathematischer Kompetenzen beteiligt.

Die mathematische Grundbildung für Schülerinnen und Schüler hängt also wesentlich davon ab, in welchem Maße im Unterricht Anlässe geschaffen werden, selbst oder gemeinsam *Probleme mathematisch zu lösen*, über das Verstehen und das Lösen von Aufgaben zu *kommunizieren*, über das Zutreffen von Vermutungen oder über mathematische Zusammenhänge zu *argumentieren*, Sachsituationen in der Sprache der Mathematik zu *modellieren* und für die Bearbeitung von Problemen *geeignete Darstellungen zu ersinnen* oder auszuwählen. (Walther, Selter, & Neubrand, 2008, S. 20, Hervorhebung im Original)

Schipper (2005, S. 358) bündelt dies im Ziel guten Mathematikunterrichts: Lernende sollen im Mathematikunterricht die Möglichkeit haben, „Mathematik aktiv zu betreiben“ und sich prozessbezogene Kompetenzen zu eigen machen zu können. Dies kann insbesondere beim selbstständigen Lösen problemhaltiger Aufgaben vollzogen werden (Heinrich et al., 2015, S. 280; 2005b, S. 7; s. a. Wegerich, 2015, S. 262).

Das Problemlösen wird in folgenden Bildungsstandards zusammengefasst (KMK 2005b, S. 7):

- mathematische Kenntnisse, Fertigkeiten und Fähigkeiten bei der Bearbeitung problemhaltiger Aufgaben anwenden;
- Lösungsstrategien entwickeln und nutzen (z. B. systematisch probieren);
- Zusammenhänge erkennen, nutzen und auf ähnliche Sachverhalte übertragen.

Lernende greifen beim Lösen eines Problems auf ihre individuellen Ressourcen wie mathematische Kenntnisse, Fertigkeiten und Fähigkeiten zurück (Schoenfeld, 1985, S. 17). Sie stehen dann vor der Herausforderung, mögliche Herangehensweisen und Strategien zu durchdenken, zu adaptieren oder auch zu entwickeln, Zusammenhänge aufzuspüren und zu verknüpfen, um vorliegende Probleme zu lösen (Heinrich et al., 2015, S. 285; KMK 2005b, S. 7; Schukajlow, 2011, S. 80; Stahl, 1975, S. 301).

„Die durch eigenständiges und erfolgreiches Problemlösen gewonnenen Einsichten, Wertehaltungen und Erkenntnisse gelten als besonders nachhaltig im Sinne von motivierend, weil eigene Kompetenz erlebt wurde, und sie sind nachhaltig im Sinne von dauerhaft verfügbar und durchaus auch flexibel einsetzbar“ (Heinrich et al., 2015, S. 284; vgl. Bruder & Collet, 2011b, S. 35; Süllwold, 1960, S. 106). Bevor der Prozess des Problemlösens konkreter aufgeschlüsselt wird, erfolgt die Klärung des Begriffs *Problem*.

1.2 Was ist ein (mathematisches) Problem?

Beschäftigt man sich mit der Frage danach, wann es sich genau um ein Problem handelt und wie es sich von einer Aufgabe abgrenzt, stößt man sowohl in der psychologischen als auch in der fachdidaktischen Literatur auf verschiedene Definitionen.

Nach Brander, Kompa und Peltzer (1989, S. 111) gilt die Beschreibung nach Duncker als allgemeine Grundlage heutiger Problemdefinitionen:

Ein „Problem“ entsteht z. B. dann, wenn ein Lebewesen ein Ziel hat und nicht „weiß“, wie es dieses Ziel erreichen soll. Wo immer der gegebene Zustand sich nicht durch bloßes Handeln (Ausführen selbstverständlicher Operationen) in den erstrebten Zustand überführen lässt, wird das Denken auf den Plan gerufen. Ihm liegt es ob, ein vermittelndes Handeln allererst zu konzipieren. (Duncker, 1935, S. 1, Hervorhebung im Original)

Bereits Duncker betont in seinen Ausführungen die Notwendigkeit, Operatoren aufzuspüren, die fern von einer reinen Reproduktion die Überführung in den Zielzustand ermöglichen. Das „Ausführen selbstverständlicher Operatoren“ und somit „bloßes Handeln“ reichen – um es in den Worten Dunckers zu formulieren – nur aus, um Aufgaben zu bewältigen, nicht jedoch, um Problemzustände ineinander zu überführen (vgl. Winter, 1992, S. 363, 2000, S. 7). Seine Problembeschreibung hebt die gestalt-psychologische Annahme des produktiven Denkens vom reproduktiven Denken ab. Auch in späteren Definitionen lassen sich gleiche Schwerpunkte wiederfinden:

Ein Individuum steht einem Problem gegenüber, wenn es sich in einem inneren oder äußerem Zustand befindet, den es aus irgendwelchen Gründen nicht für wünschenswert hält, aber im Moment nicht über die Mittel verfügt, um den unerwünschten Zustand in den wünschenswerten Zielzustand zu überführen. Ein Problem ist also gekennzeichnet durch drei Komponenten:

1. Unerwünschter Anfangszustand s_a
2. Erwünschter Endzustand s_o
3. Barriere, die die Transformation von s_a in s_o im Moment verhindert. (Dörner, 1976, S. 10, in Anlehnung an Klix, 1971, S. 640 und Süllwold, 1960, S. 96)

Diese Definition expliziert stärker die Barriere und setzt das Ziel nicht zwingend als gegeben voraus (Brander et al., 1989, S. 111). Klix (1971, S. 640) führt ergänzend zur dritten Komponente aus,

daß keine mit der Merkmalsbildung schon gegebene Relation hinreicht, um den Lösungszustand zu erzeugen. Vielmehr müssen die vom Zielzustand geforderten Relationen erst gesucht bzw. der gegebene Zustand so verändert werden, daß über die Bildung neuer Relationen und ihrer Verknüpfung der Lösungszustand erreichbar wird.

Sowohl Dörner (1976, S. 10–11) als auch Klix (1971, S. 639–640) stellen in ihren Definitionen heraus, dass die individuellen Ressourcen, welche die Problemlösenden mitbringen, bekannt sein müssen. Nur wenn berücksichtigt wird, mit welchem Wissen, welchen Fähigkeiten und Werkzeugen die Lösenden starten, kann nahezu sicherge-

stellt werden, dass ein Problem auch tatsächlich nicht eine Aufgabe oder reine Reproduktion von Wissen darstellt.³

The difficulty with defining the term *problem* is that problem solving is relative. The same task that call for significant efforts from some students may well be routine exercises for others, and answering them may just be a matter of recall for a given mathematician. Thus being a „problem“ is not property inherent in a mathematical task. Rather, it is a particular relationship between the individual and the task that makes the task a problem for that person. The word *problem* is used here in this relative sense, as a task that is difficult for the individual who is trying to solve it. Moreover, that difficulty should be an intellectual impasse rather than a computational one. (Schoenfeld, 1985, S. 74, Hervorhebung im Original)

An dieser Stelle könnte noch eine Vielzahl an Definitionen und Ausführungen aus der Psychologie oder Fachdidaktik aufgegriffen werden.⁴ Hiervon wird jedoch abgesehen, da sie im Kern den bisherigen entsprechen. Stattdessen werden die der Untersuchung zugrunde liegenden zentralen Aspekte auf Basis der Definitionen zusammengetragen.

Mathematische Problemstellungen werden in Anlehnung an A. Meyer und Fischer (2013, S. 185 in Bezug auf Bandura, 2001, S. 2) als Handlungsaufforderungen aufgefasst (vgl. Bruder & Collet, 2011b, S. 12). Die Lernenden können bei deren Bewältigung keinen bekannten Weg reproduzieren, sondern stehen vor der Herausforderung mithilfe ihrer individuellen Ressourcen die Transformation des Anfangs- in den Zielzustand zu bestreiten und schließlich die Barriere(n) mit eigenen Mitteln zu überwinden.

Um erklären zu können, welche kognitiven Prozesse beim Problemlösen konkret stattfinden, wodurch sie beeinflusst, aber auch wie sie sich zunutze gemacht werden können, welche Handlungen über Erfolg, Misserfolg oder ein erstes Scheitern entscheiden und wie es letztlich dem Problemlösenden gelingen kann, die auftretende(n) Barriere(n) zu überwinden, werden nachfolgend die Erkenntnisse der Psychologie herangezogen.

1.3 (Mathematisches) Problemlösen aus psychologischer Sicht

Dass das Problemlösen auch ein vielbeforschter Forschungsschwerpunkt der Psychologie darstellt, dessen Erkenntnisse sich die Didaktik zunutze machen kann bzw. sogar zunutze machen muss, darf an dieser Stelle nicht vernachlässigt werden (Jörissen &

³ Brander et al. (1989, S. 118), Bransford und Stein (1993, S. 6), Heinze, (2007, S. 5), R. Rasch, (2001b, S. 27), Schoenfeld (1985, S. 17), Verschaffel, Greer und de Corte (2000, S. XIII), Winter (1992, S. 363)

⁴ Z. B. Aebli (1980, S. 26–27), Bransford und Stein (1993, S. 7), Franke und Ruwisch (2010, S. 65), Frensch und Funke (1995, S. 18), Heinrich et al. (2015, S. 280), Heinze (2007, S. 5–6), Hussy (1993, S. 20), Mayer (1983, S. 4), Reiss und Törner (2007, S. 431), Robertson (2001, S. 4), Winkelgren (1974, S. 10–12), Winter (1992, S. 363)

Schmidt-Thieme, 2015, S. 389; Reiss & Törner, 2007, S. 432; Schnottz, Baadte, Müller, & Rasch, 2011, S. 205–216). Im Wesentlichen lassen sich aus psychologischer Sicht zwei Richtungen unterscheiden, der gestaltpsychologische und der informationsverarbeitende Ansatz (Reiss & Törner, 2007, S. 432–435; Schnottz et al., 2011, S. 205–216). Führt man diese zusammen, lassen sich essenzielle Rückschlüsse über erfolgreiches Überwinden von Problembarrieren ziehen (Knoblich, Ohlsson, Haider, & Rhenius, 1999; Ohlsson, 1984b, 1992). Aufgrund dessen werden nachfolgend beide Ansätze kurz aufgezeigt (Abschnitt 1.3.1 und 1.3.2) und deren Präsenz für die vorliegende Studie anhand ihrer Zusammenführung abgeleitet (Abschnitt 1.3.3).

1.3.1 Gestaltpsychologischer Ansatz

Folgt man der Gestaltpsychologie, so hängt eine erfolgreiche Problemlösung davon ab, ob der Problemlösende in der Lage ist, eine Repräsentation des Anfangszustandes zu konstruieren (Schnottz et al., 2011, S. 205; vgl. Bandura, 2001, S. 6; Heinze, 2007, S. 7). Er steht vor der Aufgabe, seine Wahrnehmung umzustrukturieren, sprich strukturelle Beziehungen in der Situation zu verändern (z. B. Duncker, 1935, S. 39)⁵, um die Problemstruktur vollständig zu durchdringen und somit ein tieferes Verständnis des Problems zu erlangen (Knoblich & Öllinger, 2006, S. 42). Das Verharren in einer Sichtweise bzw. in einer „inadæquate[n] Ansicht der Lage“ (Wertheimer, 1964, S. 226) blockiert die Wahrnehmung der vorliegenden Problemstruktur und letztlich die Überwindung der Barriere.

An important contribution of the Gestaltists is the idea that people get stuck solving problems because they cannot change their problem-solving set – since they cannot look at the situation in a new way, they cannot see a new way to fit the elements together. (Mayer, 1983, S. 33–34)

Um erfolgreich zu sein, darf bei der Lösungssuche weder „strukturell blind“ vorgegangen, noch an nicht zielführenden Lösungsmustern festgehalten werden (Robertson, 2001, S. 54; Wertheimer, 1964, S. 159). Hingegen muss der Lösende der Problemstellung mit einer neuen, unvoreingenommenen und „weiten“ Betrachtungsweise gegenübertreten (Mayer, 1983, S. 34; Ohlsson, 1984b, S. 65; Wertheimer, 1964, S. 227). Die Konstruktion einer geeigneten Repräsentation unterstützt ihn dabei und letztlich beim Vollziehen der Umstrukturierung (Anderson, 1996, S. 257–258; Duncker, 1935, S. 52, 132; Schnottz et al., 2011, S. 206–209). Erst dann, wenn es gelingt, die Problemsituation und ihre kennzeichnenden Strukturen ganzheitlich wahrzunehmen und zielführend umzustrukturieren, kann die Lösung unmittelbar an der konstruierten Repräsentation abgelesen werden (Mayer, 1983, S. 34; Duncker, 1935, S. 44; Wertheimer, 1964, S. 228). Das heißt, durch ein Verändern, Umordnen, Gruppieren oder Restrukturieren

⁵ Humphrey (1963, S. 154), Mayer (1983, S. 34), Ohlsson (1984b, S. 65), Wertheimer (1964, S. 66)

gegebener Daten können Beziehungen, die zuvor nicht offensichtlich waren, aufgespürt werden und zu einer Veränderung der mentalen Repräsentation führen (Schnotz et al., 2011, S. 209).⁶

The process of problem solving is a search to relate one aspect of a problem situation to another, and it results in *structural understanding* – the ability to comprehend how all the parts of the problem fit together to satisfy the requirements of the goal. This involves reorganizing the elements of the problem situation in a new way so that they solve the problem. (Mayer, 1983, S. 35, Hervorhebung im Original)

Die auf strukturellen Veränderungen basierende Einsicht wird von dem Problemlösen den subjektiv als „plötzlich, unerwartet und überraschend“ erlebt (Knoblich & Öllinger, 2006, S. 3, S. 43).⁷ Die Lösung wird evident; ihre Gewinnung gleicht mehr einem Gedankenblitz und einem „Aha-Erlebnis“ als einer schrittweisen Annäherung, z. B. durch Ausprobieren mehrere Lösungsmöglichkeiten.⁸

Während in der Gestaltpsychologie schwerpunktmäßig die Umstrukturierung und Einsicht fokussiert wird, nahm Duncker zusätzlich das schrittweise Problemlösen in den Blick (Duncker, 1935, S. 96–98, § 7; vgl. Knoblich & Öllinger, 2006, S. 42): Er brachte hervor, dass Problemlösende auch mit einem „Suchbereich“ konfrontiert sind, den es zu durchlaufen gilt. Aus der Sicht Knoblichs und Öllingers (2006, S. 42) ebnete er hiermit den Weg für die Informationsverarbeitungpsychologie.

Es bleibt festzuhalten, dass dem Problemlösenden aus Sicht der Gestaltpsychologie die Lösung dann offensichtlich wird, wenn er es schafft, die gegebene Problemsituation mit der Konstruktion einer lösungsunterstützenden Repräsentation umzustrukturieren und neu wahrzunehmen (Ohlsson, 1984a, S. 117, 1992, S. 16; Schnotz et al., 2011, S. 206; Wertheimer, 1964, S. 224).⁹ Wie und warum es letztlich zur plötzlichen Einsicht kommen kann, wird von der Gestalttheorie weder hinterfragt noch aufgeschlüsselt und bleibt daher ungeklärt (Robertson, 2001, S. 59).

⁶ Vgl. Humphrey (1963, S. 156–159), Ohlsson (1984a, S. 119–120), Wertheimer (1964, S. 47–48)

⁷ Vgl. Humphrey (1963, S. 158–159), Koffka (1935, S. 641), Mayer (1983, S. 36), Schnotz et al. (2011, S. 206–209)

⁸ Siehe u. a. Bühler (1908, S. 13), Duncker (1935, S. 35), Guilford (1950, S. 451), Humphrey (1963, S. 158), van der Waerden (1968, S. 1–2), Wertheimer (1964, S. 55, 79)

⁹ Für eine tiefere Einsicht in den gestaltpsychologischen Ansatz sowie konkrete Beispiele sei an dieser Stelle auf Duncker (1935, S. 37–54), Mayer (1983, S. 37–66), Schultz (1969, S. 273–307) sowie Schnotz et al. (2011, S. 205–209), Wertheimer (1964), Andersson (1996, S. 257–260) und Heinze (2007, S. 7–8) verwiesen. Ergänzende Informationen zu der Kritik, die dem Ansatz entgegengebracht wird, sind z. B. in Ohlsson (1984b, S. 70–76), Robertson (2001, S. 60–65), Schnotz et al. (2011, S. 214) und Schultz (1969, S. 304–305) zu finden.

1.3.2 *Informationsverarbeitender Ansatz*

Blickt man mit dem informationsverarbeitenden Ansatz auf das Problemlösen, dann stehen nicht die Umstrukturierung und die Einsicht, sondern die Suche nach dem richtigen Weg durch „einen Raum alternativer Lösungswege“, den „space of alternatives“, im Vordergrund (Ohlsson, 1984a, S. 118, 1984b, S. 65).¹⁰ Konkret bedeutet dies, dass der Problemlösende vor der Aufgabe steht, den Zielzustand zu identifizieren und korrekt zu repräsentieren (Dörner, 1976, S. 26–27; Schnottz et al., 2011, S. 209–210) sowie die vom Zielzustand geforderten Relationen aufzuspüren (Klix, 1971, S. 640). Dies impliziert, dass die geschilderte Problemsituation verstanden werden muss (Newell & Simon, 1972, S. 809; Ohlsson, 1992, S. 9). Auf dieser Wissensbasis aufbauend, müssen Operationen und Relationen entdeckt und miteinander verknüpft werden, um den Anfangszustand in den Zielzustand überführen zu können.¹¹ Dabei muss es dem Lösenden gelingen, „einen Raum mit alternativen Wegen“, den sogenannten „Problemraum“, zu konstruieren (Newell & Simon, 1972, S. 59, S. 809; Schnottz et al., 2011, S. 210). Hierbei sind die unterschiedlichsten Wege denkbar und zulässig, sofern sie Zustandstransformationen bedingen (Anderson, 1996, S. 235). Werden Zwischenzustände als Folge von Teilschritten, d. h. einer „Sequenz von Operatoren“ (Anderson, 1996, S. 236; vgl. Heinze, 2007, S. 6), gewonnen, so können diese eine Annäherung an die Lösung sowie eine schrittweise Überwindung der Barriere bewirken (Heinze, 2007, S. 6; Schnottz et al., 2011, S. 210; vgl. Brander et al., 1989, S. 115).

Alle denkbaren Zwischenzustände, die aufgrund individuell ausgewählter Operatoren während des Lösungsprozesses erreicht werden können (Ohlsson, 1984a, S. 118), bilden mit dem Anfangs- und Zielzustand den sogenannten „Problemraum“.¹² Newell und Simon (1972, S. 809) bezeichnen ihn als einen Raum, „(1) where problem solving takes place and (2) that contains not only the actual solution but possible solutions that the problem solver might consider“. In der Literatur wird der Problemraum mit einem „Labyrinth von Zuständen“ verglichen, wobei die Operatoren als Brücke zwischen zwei Zuständen zu verstehen sind (Newell, Shaw, & Simon, 1962, S. 70–74; Hayes, 1981, S. 51–52).

Eine erfolgreiche Problemlösung hängt davon ab, ob der Zielzustand Teil des Problemraums ist. Vorausgesetzt, dies trifft zu, dann kann durch eine Suchraumeinschränkung der richtige Weg zur Lösung gefunden werden (Klix, 1971, S. 644, S. 646; Wes-

¹⁰ Vgl. Anderson (1996, S. 235), Klix (1971, S. 640), Newell und Simon (1972, S. 809, 823), Knoblich und Öllinger (2006, S. 46)

¹¹ Anderson (1996, S. 241), Dörner (1976, S. 15, 26), Kaplan und Simon (1990, S. 377), Klix (1971, S. 640), Knoblich und Öllinger (2006, S. 46)

¹² Siehe u. a. Anderson (1996, S. 237–238), Klix (1971, S. 644), Knoblich und Öllinger (2006, S. 46–47), Newell und Simon (1972, S. 809–811), Ohlsson (1984a, S. 118)

sells, 1990, S. 363–364). Liegt der Zielzustand jedoch außerhalb des Problemraums, so kann das Ziel erst erreicht werden, wenn der Problemraum so lange erweitert wird, bis der Zielzustand inkludiert ist (Klix, 1971, S. 644, S. 647). Dies schließt ein, dass der Problemraum während des Problemlöseprozesses mehrfach verändert oder modifiziert werden kann/muss (Newell & Simon, 1972, S. 809), vor allem dann, wenn die ausgewählten Operatoren den Problemlöseprozess nicht voranbringen (Kaplan & Simon, 1990, S. 377). Unter diesen Voraussetzungen gilt es, ein Problem schrittweise, mit einer Sequenz von Schritten, zu lösen: „To solve a problem is to proceed step-wise through a space of alternatives, until an action sequence is found which leads from the problem to its solution“ (Ohlsson, 1984b, S. 65, vgl. 1984a, S. 117–118).

Nicht nur, weil bei der Lösungssuche die kürzeste Operatorsequenz angestrebt wird (Anderson, 1996, S. 237; Klix, 1971, S. 644, S. 706), wird die Suche nach einem ziel führenden Weg durch Heuristiken nachweislich unterstützt.¹³ Im Idealfall gelingt es damit, den Problemraum einzuzgrenzen und zu reduzieren (Kaplan & Simon, 1990, S. 379). Gerade wenn der Lösende vor der Herausforderung steht, seinen bisherigen Problemraum umzustrukturieren, und er dabei auf der Metaebene nach möglichen Problemräumen sucht, ist eine selektive Suche vorteilhaft (Kaplan & Simon, 1990, S. 377). Heuristische Vorgehensweisen verhelfen ihm den Problemraum einzuschränken und einen passenden nächsten Schritt auszuwählen (Schnotz et al., 2011, S. 211).

Zusammenfassend besteht aus Sicht der Informationsverarbeitungspsychologie die Lösung eines Problems darin, den richtigen Weg durch den Problemraum zu finden, weshalb sie auch als Problemraumtheorie bezeichnet wird (Knoblich & Öllinger, 2006, S. 45–49).

1.3.3 *Representational Change Theory*

Die *Representational Change Theory* vereint den gestaltpsychologischen und den informationsverarbeitenden Ansatz. Ihre Bedeutung für die vorliegende Untersuchung wird in den folgenden beiden Abschnitten aufgezeigt.

1.3.3.1 *Grundlage der Konzeptualisierung*

Eine Gegenüberstellung beider Ansätze macht deutlich, dass deren Problemlösekonzepte unterschiedliche, keinesfalls jedoch gegensätzliche Schwerpunkte akzentuieren: Während die Informationsverarbeitungspsychologie den Problemlöseprozess und damit die Suche nach dem richtigen Weg durch den Problemraum hervorhebt, führt die

¹³ Heinze (2007, S. 6), Knoblich und Öllinger (2006, S. 47), Newell und Simon (1972, S. 98–105), Ohlsson (2012, S. 105–109), Papert (1993, S. 87), Schnotz et al. (2011, S. 211), Wessels (1990, S. 360–366)

Gestaltpsychologie die (plötzliche) Einsicht und somit die erfolgreiche Problemlösung auf Umstrukturierungen zurück und betont dadurch stärker die Problemstruktur (Ohlsson, 1984b, S. 77; Schnotz et al., 2011, S. 212).

Mit dem Fokus, herauszufinden, was erfolgreiches Problemlösen tatsächlich ausmacht, kann das Potenzial beider Sichtweisen nur ausgeschöpft werden, wenn sie nicht nebeneinanderstehend, sondern vernetzt betrachtet werden. Die Verknüpfung von Struktur und Suchprozess ist Gegenstand der *Representational Change Theory* (Ohlsson, 1984a, S. 117, 1984b, S. 65). Sie gilt als Erweiterung der Informationsverarbeitungstheorie (Knoblich & Öllinger, 2006, S. 50; Ohlsson, 1992, S. 19, 2012, S. 117; Schnotz et al., 2011, S. 212). Für diese Erweiterung spricht, dass zum einen plötzliche und unerwartete Einsichten aus Sicht der Problemraumtheorie nicht existieren (Knoblich & Öllinger, 2006, S. 50; Ohlsson, 1992, S. 17). Zum anderen werden Gründe dafür, dass Lösungsprozesse ins Stocken kommen, es nicht gelingt, die Distanz vom Anfangs- zum Zielzustand zu reduzieren, und Problemlösungen verwehrt bleiben, in der unüberschaubaren Größe des Problemraums gesehen (Kaplan & Simon, 1990, S. 379–381; Öllinger, Jones, & Knoblich, 2014, S. 266). Folglich bleibt ohne Erweiterung nicht nachvollziehbar, warum Probleme mit einem kleinen Problemraum schwer lösbar sein können, wenn doch davon ausgegangen wird, dass die Anzahl der notwendigen Schritte die Schwierigkeit der Aufgabe bestimmt (Johnson-Laird & Byrne, 1991, S. 163; Knoblich, Ohlsson, & Raney, 2001, S. 1000; Knoblich & Öllinger, 2006, S. 46, S. 50).

Darüber hinaus bleibt ungeklärt, warum die Suche nach dem richtigen Weg durch den Problemraum nicht immer zum Erfolg führt (Knoblich & Öllinger, 2006, S. 50; Ohlsson, 1984b, S. 77, 1992, S. 17; Schnotz et al., 2011, S. 214), obwohl die Lösenden über die erforderlichen Kompetenzen verfügen (Knoblich et al., 2001, S. 1000; Ohlsson, 1992, S. 4–6). Unangemessene Verwendungen von Heuristiken (Chronicle, MacGregor, & Ormerod, 2004, S. 15–16; Öllinger et al., 2014, S. 266) bleibt ebenso unbeachtet wie kognitive Prozesse, die es ermöglichen, den Problemraum mithilfe einer neuen Repräsentation umzustrukturieren (Kaplan & Simon, 1990, S. 379–380; Öllinger et al., 2014, S. 266; s. a. Schnotz et al., 2011, S. 214). Eine anfänglich inadäquat konstruierte Repräsentation wird aus informationsverarbeitender Sicht nicht als mögliche Fehlerquelle in Betracht gezogen (Kaplan & Simon, 1990, S. 378). Dass bewusst ablaufende repräsentationale Veränderungen von enormer Bedeutung sind, um Repräsentationen zu optimieren, wird mehrfach herausgestellt (z. B. Kaplan & Simon, 1990, S. 377–381; Knoblich et al., 2001, S. 1001; Knoblich & Öllinger, 2006, S. 56).

Eine Erweiterung der Problemraumtheorie wird aus vorstehenden Gründen vielfach gefordert: Warum stoßen Problemlösende auf ProblembARRIEREN und geraten dabei ins Stocken, wenn sie doch eigentlich in der Lage sind, das Problem zu lösen? Wie gelingt

es ihnen, ProblembARRIEREN zu überwinden? Welche Prozesse sind daran beteiligt? Was passiert, nachdem die ProblembARRIERE überwunden ist (z. B. Kaplan & Simon, 1990, S. 378; Knoblich & Öllinger, 2006, S. 50; Ohlsson, 1984b, S. 65, 1992, S. 6)?

Ohlsson (1984a, S. 119–125, 1992, S. 9–25) widmet sich in seiner *Representational Change Theory* der Klärung dieser Fragen, indem er die Informationsverarbeitungstheorie um die Kernannahme der Umstrukturierung aus der Gestaltpsychologie erweitert.

1.3.3.2 Problemlösen aus Sicht der *Representational Change Theory*

ProblembARRIEREN und deren Überwindung stehen im Fokus der *Representational Change Theory* (Ohlsson, 1992, S. 6; Öllinger, Jones, & Knoblich, 2008, S. 271; Öllinger et al., 2014, S. 266–267). Ohlsson (1992, S. 19) stellt die Hypothese auf, dass unbewusst ablaufende kognitive Prozesse repräsentationale Veränderungen hervorrufen, die zu vollständiger Einsicht und somit zur Überwindung von Barrieren führen können. Dabei geht er davon aus, dass die Lösungssuche ins Stocken gerät oder ganz zum Stilstand kommt, wenn sie sich auf eine inadäquate Repräsentation stützt (Knoblich et al., 2001, S. 1000; Ohlsson, 1992, S. 10; Öllinger et al., 2008, S. 271). Ohlsson (1992, S. 20) betont, dass der Konfrontation mit einer ProblembARRIERE eine Lösungssuche vorausgegangen sein muss, was mit der Tatsache konform geht, dass die „Hauptarbeit des bewussten Denkens“ vor der Einsicht erfolgt (van der Waerden, 1968, S. 8).

Entstehung von Barrieren

Was der Problemlösende letztlich als Anfangssituation repräsentiert bzw. nicht repräsentiert, wird von seinem Vorwissen und von seinen Wahrnehmungsprozessen bestimmt.¹⁴ Alle individuell verfügbaren Wissenselemente bilden den Probletraum ab (Knoblich et al., 1999, S. 1535; van der Waerden, 1968, S. 8). Welche Handlungsmöglichkeiten der Lernende in Betracht zieht, welches Wissen er mit der Problemsituation assoziiert und in Verbindung bringt, bestimmt letztlich unbewusst seine mentale Repräsentation.¹⁵ Diese konstruierte Repräsentation fungiert dann als „Gedächtnis-Trigger“ (Ohlsson, 1992, S. 10; vgl. Schnottz et al., 2011, S. 213) und aktiviert Kenntnisse aus vorherigen Problemlösungen sowie eigene Erfahrungen, die hilfreich, aber auch hemmend sein können.¹⁶ Erstere können zur Generierung adäquater Repräsentationen

¹⁴ Gick und Lockhart (1995, S. 215, 221), Heinze (2007, S. 8), Ohlsson (1992, S. 24), Öllinger et al. (2008, S. 271), Schukajlow (2011, S. 78)

¹⁵ Knoblich, Ohlsson und Raney (2001, S. 1001), Knoblich und Wartenberg (1998, S. 209), Ohlsson (1992, S. 23–24), Robertson (2001, S. 66), Schukajlow (2011, S. 78)

¹⁶ Heinze (2007, S. 9), Knoblich et al. (1999, S. 1534), Knoblich und Öllinger (2006, S. 51), Ohlsson (1992, S. 12)

tionen verhelfen, die wiederum Wissen aktivieren, welches direkt für die Lösungsfindung genutzt werden kann (Schnotz et al., 2011, S. 213). Letztere haben zur Folge, dass der Lösende basierend auf seiner (inadäquaten) Repräsentation des Anfangszustandes Wissen aktiviert, welches er nicht für die gegenwärtige Problembewältigung nutzen kann und ihm folglich die Konstruktion einer geeigneten Repräsentation verwehrt.¹⁷ Auch wenn der Fokus zielgerichtet ist, können auf Basis einer fehlerhaften Repräsentation weder die notwendigen Operatoren abgerufen noch der Suchprozess am Laufen gehalten werden (Knoblich & Öllinger, 2006, S. 51; Ohlsson, 1992, S. 10). Dem Lösenden fehlt in diesem Moment eine passende Lösungsidee, alle Möglichkeiten erscheinen zunächst als ausgeschöpft und das Problem scheint unlösbar zu sein (Knoblich et al., 1999, S. 1534, 2001, S. 1000; Ohlsson, 1992, S. 4; Öllinger et al., 2008, S. 271). Knoblich et al. (2001, S. 1007) konnten in einer Eye-Movement-Studie zeigen, dass dies häufig dann auftritt, wenn Lernende die Problemstellung länger fixieren und Augenbewegungen ausbleiben.

Überwindung von Barrieren

Um die Situation des Nichtweiterwissens hinter sich zu lassen, muss sich der Lösende von seinen bisherigen Lösungserfahrungen trennen und die Problemsituation umstrukturieren. Konkret bedeutet dies, dass durch Veränderungen der Problemrepräsentationen ProblembARRIEREN überwunden werden können.¹⁸ Diese Veränderungen sind an der Repräsentation sowohl des Anfangszustandes als auch des Endzustandes denkbar (Knoblich & Öllinger, 2006, S. 51). Sie gelten im Allgemeinen als „Schlüssel zum Erfolg“ (Öllinger et al., 2008, S. 271).

Ziel ist es, den Raum möglicher Lösungen durch die Veränderung der Problemrepräsentation derart umzuwandeln, einzugrenzen oder zu erweitern, dass die Lösung Teil des „neuen“ Problemraums wird.¹⁹ Die „neue“, modifizierte Problemrepräsentation soll bisher inaktive, lösungsrelevante Wissenselemente aktivieren (Knoblich et al., 1999, S. 1535, 2001, S. 1001; Ohlsson, 1984a, S. 121, 1992, S. 12; Robertson, 2001, S. 75), sodass mithilfe von Heuristiken eine effiziente Suche durch den modifizierten Problemraum ermöglicht wird (Gick & Lockhart, 1995, S. 215, 220–223; Öllinger et al., 2014, S. 267). Hierbei ist entscheidend, inwiefern der Problemlösende in der Lage ist, den Problemraum mental vorausschauend nach möglichen Lösungswegen zu durchsuchen, bzw. ab wann ihn seine Gedächtniskapazitäten daran hindern und brem-

¹⁷ Knoblich et al. (1999, S. 1535), Ohlsson (1992, S. 10), Öllinger et al. (2008, S. 271), Robertson (2001, S. 50)

¹⁸ Gick und Lockhart (1995, S. 215, 218), Knoblich et al. (1999, S. 1553), Ohlsson (1984a, S. 125), Öllinger et al. (2008, S. 271), Öllinger et al. (2014, S. 267), Robertson (2001, S. 51), Schnotz et al. (2011, S. 213)

¹⁹ Knoblich et al. (1999, S. 1535), Ohlsson (2011, S. 117), Öllinger et al. (2014, S. 267)

sen, die noch erforderlichen Schritte zu „sehen“ („mental look-ahead“, Ohlsson, 1984a, S. 124, 2011, S. 113–114). Eine repräsentationale Veränderung kann demnach zur Einsicht führen, wenn es gelingt, den aktuellen Wissensstand dem Zielzustand anzunähern und die noch erforderlichen Schritte bis zur Lösung zu reduzieren (Ohlsson, 1984a, S. 124).

Überwindung von Barrieren auf Basis vollständiger Einsicht (full insight)²⁰

Führt eine einfache und kurze Abfolge von Schritten, die im Kopf vollzogen werden kann, bereits zur Einsicht, so kann die Lösung direkt und vollständig an der mentalen Repräsentation abgelesen werden (Kershaw & Ohlsson, 2004, S. 3; Ohlsson, 1992, S. 23, 2011, S. 114; Segal, 2004, S. 147). Es werden kaum Kapazitäten des Arbeitsgedächtnisses beansprucht, sodass die Lösungskonstruktion in diesem Moment unbewusst abzulaufen scheint.²¹ Folgende Zitate beschreiben das Phänomen des Unbewussten sehr treffend: „Dass die Lösung eines Problems gefunden wurde, wird dem Problemlöser erst in dem Moment klar, in dem sie vor seinem geistigen Auge erscheint“ (Knoblich & Öllinger, 2006, S. 63). „Das Unbewusste hat, wie ein gewissenhafter Archivar, etwas gefunden, das die Bedingungen erfüllt, also kann es das Gefundene getrost vorlegen und sagen: Das ist die Lösung“ (van der Waerden, 1968, S. 9). Da die Lösungsfindung in wenigen Schritten und somit schnell und unauffällig vollzogen ist (Stern, 1998, S. 44), kann man bei vollständiger Einsicht davon ausgehen, dass das gestaltpsychologische Kriterium der Plötzlichkeit erfüllt ist (Knoblich & Öllinger, 2006, S. 51; Ohlsson, 1984a, S. 125), auch wenn es sich bei der Plötzlichkeit nach Ohlsson (z. B. 2011, S. 114) mehr um eine Illusion handelt.

Überwinden von Barrieren auf Basis von Teileinsichten (partial insights)²²

In Situationen, in denen Lösende nicht mehr weiter wissen, ist es bei komplexen Problemstellungen häufig erforderlich, Problemlösungen schrittweise zu vollziehen, gerade wenn die Kapazitäten des Arbeitsgedächtnisses ausgeschöpft sind (Murray & Byrne, 2013, S. 211–212; Ohlsson, 1984a, S. 124). Der Lösende kann sich dann in kleinen Schritten an die Lösung herantasten, kann mehrere einzelne Umstrukturierungen vornehmen und Teileinsichten erlangen (Koffka, 1935, S. 631; Ohlsson, 1984b, S. 76, 1992, S. 3–5, 17–18, 21–23). Dass eine Problemlösung in einer „Serie von Einsichten“ realisiert werden kann, wird nicht als Nachteil interpretiert (Koffka, 1935, S. 631; Ohlsson, 1992, S. 5). Im Gegenteil, bei komplexen Problemstellungen ist meist mehr

²⁰ In Anlehnung an Koffka (1935, S. 645).

²¹ Knoblich et al. (2001, S. 1000), Ohlsson (1992, S. 17, 2011, S. 114), Segal (2004, S. 147), van der Waerden (1968, S. 9)

²² In Anlehnung an Koffka (1935, S. 631).

als eine Hürde zu meistern, weshalb die vollständige Einsicht und die Lösung des Problems auch über mehrere Teileinsichten erreicht werden kann (Murray & Byrne, 2013, S. 211; Ohlsson, 1992, S. 5).

Demzufolge kann ein Problemlösender auch erfolgreich sein und Einsicht erlangen, wenn sein unmittelbares, spontanes, augenblickliches Umstrukturieren als Reaktion auf Problembarrieren in mehreren Schritten und über Umwege erfolgt (Ohlsson, 1992, S. 5). Der Problemlöseprozess ist dann von vielen, kleineren repräsentationalen Veränderungen und Teileinsichten geprägt, die den Weg durch den Problemraum ebnen können. Allein die Tatsache, dass das Problem durch die Umstrukturierung neu wahrgenommen wird und somit ein einfach strukturierter Problemraum erzeugt wird, reicht jedoch nicht aus (Ohlsson, 1992, S. 16–17; Schnotz et al., 2011, S. 215). Umstrukturierungen gelten nur als „Schlüssel zum Erfolg“, wenn ausgehend von der neu konstruierten Problemrepräsentation ein passender Weg durch den Problemraum hin zum Zielzustand gefunden wird (Gick & Lockhart, 1995, S. 209; Schnotz et al., 2011, S. 215). Dies schließt ein, dass nicht jede repräsentationale Veränderung automatisch eine Lösungsunterstützung darstellt und immer zur Überwindung der Barriere führen muss (Ohlsson, 1992, S. 22–23, 1984a, S. 123; vgl. Knoblich & Öllinger, 2006, S. 51; Öllinger et al., 2014, S. 267). Ist der Zielzustand mit der gegenwärtigen Repräsentation erreicht, wurde eine adäquate Repräsentation konstruiert (Gick & Lockhart, 1995, S. 212). War er noch nicht erreichbar, sind weitere Schritte erforderlich (Ohlsson, 1992, S. 21).

Whether a representational change will bring the goal state within the range of mental look-ahead or not depends on three factors: (a) which representation the problem solver is switching to, (b) what the distance between the current knowledge state and the goal state is in the new representation, and (c) how severely the problem solver's capacity limitation restricts his mental look-ahead. If the distance to the goal state is larger in the new representation than the distance to the horizon, insight will not occur. (Ohlsson, 1984a, S. 125, Hervorhebung im Original)

Möglichkeiten repräsentationaler Veränderungen

Die Veränderung einer inadäquaten Repräsentation kann auf unterschiedliche und vielfältige Weise erfolgen.²³ Schwerpunktmäßig werden in der Literatur vier zentrale Prozesse unterschieden: *Elaboration* (Kaplan & Simon, 1990, S. 389; Ohlsson, 1992, S. 13), *Re-Interpretation* („re-encoding“, Ohlsson, 1992, S. 13–14, 1990, S. 243), *Löckerung von Bedingungen* („constraint relaxation“, Ohlsson, 1992, S. 14–16) und die *Zerlegung von Gruppierungen* („chunk decomposition“, z. B. Knoblich et al., 1999,

²³ Knoblich et al. (1999, S. 1535–1536), Knoblich und Öllinger (2006, S. 52–56), Knoblich und Wartenberg (1998, S. 210–216), Ohlsson (1992, S. 12–16), Robertson (2001, S. 67–68, 74)

S. 1535–1536). Die beiden letztgenannten Möglichkeiten werden aufgrund ihrer Relevanz für die Untersuchung konkretisiert.²⁴

Repräsentationen des Zielzustandes können durch das *Lockern von Bedingungen* modifiziert werden (Knoblich & Öllinger, 2006, S. 54–46; Luo, Niki, & Knoblich, 2006, S. 431; Ohlsson, 1992, S. 14). Dies wird unumgänglich, wenn zu Beginn des Lösungsprozesses eine zu enge mentale Zielrepräsentation generiert wurde, die erst angereichert werden muss, um die Barriere überwinden zu können (Knoblich et al., 1999, S. 1535; Knoblich & Öllinger, 2006, S. 54–46; Ohlsson, 1992, S. 14). In anderen Worten müssen die implizit auferlegten Bedingungen, die den Problemraum eingrenzen, gelockert werden (Knoblich & Öllinger, 2006, S. 54; Luo et al., 2006, S. 430; Öllinger et al., 2008, S. 271–272), sodass der gegenwärtige Problemraum im nächsten Schritt in einen neuen, erweiterten Problemraum überführt werden kann und die notwendigen Operatoren abgerufen werden können (Luo et al., 2006, S. 430; Öllinger et al., 2014, S. 266; Schnotz et al., 2011, S. 214). Lernende sind unbewusst gewillt, Zieleinschränkungen vorzunehmen, wenn sie diese in früheren Problembewältigungen als zielführend kennengelernt haben und sich just an sie erinnern (Knoblich et al., 1999, S. 1535; Knoblich & Öllinger, 2006, S. 54; Öllinger et al., 2008, S. 272). Werden sie mit der Problemstellung konfrontiert und stellen Vermutungen an, welche Lösungen als Problemlösung infrage kommen und welche nicht, dann aktivieren sie automatisch Wissen, das bestimmten Bedingungen unterliegt (Knoblich et al., 1999, S. 1535). Sie können zielführend sein oder aber einer Lockerung bedürfen (Knoblich et al., 1999, S. 1535). Im Rahmen des Textaufgabenlösens wird davon ausgegangen, dass die Konstruktion mentaler Repräsentationen durch die jeweilige semantische Struktur des vorliegenden Problems beeinflusst und erschwert wird, die es zu lockern und zu verstehen gilt (Nesher, Greeno, & Riley, 1982, S. 373–379, 384; Riley, Greeno, & Heller, 1983, S. 162–165, vgl. Abschnitt 4.1.2).

Darüber hinaus bildet die *Zerlegung von Gruppierungen*, auch als *chunk decomposition* (z. B. Knoblich et al., 1999, S. 1535–1536) bezeichnet, eine weitere Möglichkeit, Repräsentationen des Zielzustandes zu verändern. Es stellt keine Ausnahme dar, dass Lernende, die mit einer Problemstellung konfrontiert sind, automatisch gegebene Problemelemente gruppieren und als Einheit wahrnehmen (Knoblich & Öllinger, 2006, S. 53; Luo et al., 2006, S. 431; Öllinger et al., 2008, S. 271). Diese Einheiten stellen Muster, sogenannte *chunks*, dar, „die aus verschiedenen Elementen und den dazugehörigen Beziehungen zusammengesetzt sind, und die in verschiedenen Problemen immer wieder auftreten“ (Heinze, 2007, S. 7). Daher wird die mentale Repräsentation hauptsächlich unbewusst durch vorausgehende Erfahrungen manipuliert (Ander-

²⁴ Für eine detaillierte Darstellung der ersten beiden Veränderungsmöglichkeiten wird auf die zuvor aufgeführte Literatur verwiesen.

son, 1996, S. 119; Knoblich & Öllinger, 2006, S. 53). Haben Lösende ganze Lösungsprozeduren aus vorherigen Bearbeitungen abgespeichert, so kann deren Aktivierung das Gedächtnis entlasten (Heinze, 2007, S. 7). Der Prozess ist dann mit Schwierigkeiten verbunden, wenn der Gruppierung Veränderungen einzelner Elemente vorausgehen müssen (Knoblich & Öllinger, 2006, S. 53; Luo et al., 2006, S. 431). Wird dies übersehen oder ignoriert, können auf Basis inadäquater Gruppierungen keine zielführenden Operatoren aktiviert werden (Knoblich & Öllinger, 2006, S. 53). Die Gruppierung ist im konkreten Fall dem Erkenntnisgewinn hinderlich bzw. blockiert ihn, sodass der Lösende vor der Aufgabe steht, seine mentale Repräsentation in einzelne, sinnvolle Bestandteile zu zerlegen (Anderson, 1996, S. 121; Knoblich et al., 1999, S. 1536; Luo et al., 2006, S. 431; Öllinger et al., 2008, S. 271). Unter Berücksichtigung der Voraussetzungen gilt es dann, diese einzelnen Komponenten neu zu gruppieren (Luo et al., 2006, S. 431). Es liegt nahe, dass Experten bei einer Problembewältigung auf ein umfangreicheres Wissen im Sinne von chunks zurückgreifen können, als Novizen dies möglich ist (Anderson, 1996, S. 119–121; Heinze, 2007, S. 7; vgl. Bransford, Brown, & Cocking, 2000, S. 36–42). Die Individuen unterschieden sich hier in ihrem Wissen, das sie abrufen können.

Fazit und Bezug zur vorliegenden Untersuchung

Die *Representational Change Theory* nach Ohlsson hat sich in vielerlei Studien bewährt, um komplexe Problemlöseprozesse unter Berücksichtigung repräsentationaler Veränderungen zu entschlüsseln und zu erklären.²⁵ Ihre Gültigkeit wurde vielfach auch in mathematischen Kontexten und explizit dem Lösen von Textaufgaben bestätigt.²⁶ Im Vordergrund stehen immer Problembewältigungen, bei denen die Lösenden auf (mindestens) eine Barriere stoßen, die es ihnen verwehrt, die Lösung aus dem Gedächtnis abzurufen sowie ohne repräsentationale Veränderungen zur Einsicht zu gelangen (siehe Abschnitt 1.2). Insbesondere beim Lösen komplexer Probleme werden unweigerlich Wege eingeschlagen, die nicht immer (direkt) zur richtigen Lösung führen (z. B. Hohn, 2012, S. 106–108; Knoblich & Wartenberg, 1998, S. 216; R. Rasch, 2001b, S. 89).

Ist dies der Fall, gibt der Lösende entweder auf oder aber er bestreitet einen anderen Weg. Nach Ohlsson (2012, S. 122–123) wird die Suche nach einem richtigen Weg

²⁵ Exemplarisch kann auf folgende Untersuchungen verwiesen werden: Kershaw und Ohlsson (2004), Knoblich et al. (1999), Knoblich et al. (2001), Knoblich und Wartenberg (1998), Murray und Byrne (2013), Öllinger, Jones, Faber und Knoblich (2013), Thevenot und Oakhill (2008)

²⁶ Im Kontext der Mathematik lassen sich folgende Untersuchungen ergänzen: Hegarty, Mayer und Monk (1995, S. 21–22), Lewis und Mayer (1987, S. 370), Riley et al. (1983, S. 165–169), Schnottz et al. (2011, S. 213–216, 236–240), Thevenot und Oakhill (2005, S. 1320–1321, 2006, S. 759–770, 2008, S. 316)

durch die kognitiven Prozesse der eigenen Problemwahrnehmung, des Abrufens und Auswählen von Handlungen, der Evaluation des Ergebnisses oder Zwischenergebnisses und des Setzens von Zielen und Teilzielen bestimmt. Kommt der Problemlöseprozess zum Stillstand, muss die geschilderte Situation neu wahrgenommen werden, indem die Repräsentation umstrukturiert wird. Dies hat zur Folge, dass der Problemraum verändert wird und neue, im Idealfall zielführende Handlungen abrufbar werden. Der Kern des Problemlösens stellt für Ohlsson (2012, S. 122–123) daher die heuristische Suche im Problemraum dar, wobei der Problemraum durch die fünf kognitiven Prozesse bestimmt wird.

Darüber hinaus konnte empirisch belegt werden, dass Lösende bei der Konfrontation mit Textaufgaben bekannte Schemata aktivieren, welche die Konstruktion ihrer Repräsentation beeinflusst. Schemata werden als Wissensstrukturen verstanden, „in denen aufgrund von Erfahrungen typische Zusammenhänge eines Realitätsbereichs repräsentiert sind“ (Steiner, 1988, S. 124; vgl. Novick & Bassok, 2005, S. 335–336). Sie resultieren aus im Arbeitsgedächtnis verarbeiteten Informationen, die im Langzeitgedächtnis abgespeichert werden (Sweller, van Marriënboer, & Paas, 1998, S. 289) und sind insbesondere beim Verstehen, Planen, Durchdenken und Ausführen von Tätigkeiten sowie dem schlussfolgernden Denken von Bedeutung (Stern, 1998, S. 28).

Führte die Aktivierung solcher Schemata nicht zum Erfolg, waren auch in diesen Fällen Umstrukturierungen nötig (Lewis & Mayer, 1987, S. 370; Riley et al., 1983, S. 165–167). Dies lässt den Schluss zu, dass Textaufgaben nur erfolgreich gelöst werden können, wenn die geschilderte Problemsituation bzw. die zugrunde liegende Struktur adäquat repräsentiert wird und auf dieser Basis die erforderlichen Operationen aktiviert werden können.²⁷ Robertson (2001, S. 70) bewertet die Erkenntnisse der *Representational Change Theory* als „important consequences for our understanding of how we solve all kinds of problems“ und hebt in diesem Zusammenhang explizit die Bearbeitung von Textaufgaben hervor.

1.4 Mathematisches Problemlösen auf Basis problemhaltiger Textaufgaben

Als Hauptgegenstand der Untersuchung werden problemhaltige Textaufgaben in mathematisches Problemlösen eingebettet. Sie repräsentieren eine Aufgabendomäne, die nicht mit der Reproduktion von Wissen zu bewältigen ist, sondern den Lösenden krea-

²⁷ Dieser Schluss stützt sich auf vielfältige Untersuchungen. Exemplarisch wird auf folgende verwiesen: Hohn (2012, S. 138, 140), Kintsch und Greeno (1985, S. 109–110), Lewis und Mayer (1987, S. 370), Mayer und Hegarty (1996, S. 34–35), Novick und Bassok (2005, S. 342), Riley et al. (1983, S. 165–186), Robertson (2001, S. 75), Schnottz et al. (2011, S. 216), Stern (1998, S. 118), Thevenot und Oakhill (2005, S. 1311, 2006, S. 756)

tives Denken abverlangt (Guilford, 1950, S. 452–454; Mayer, 1995, S. 3). Der Problemcharakter der ausgewählten Aufgabendomäne wird im folgenden Abschnitt expliziert.

1.4.1 *Charakteristika problemhaltiger Textaufgaben*

Problemhaltige Textaufgaben grenzen sich von Textaufgaben ab, in denen Signalwörter Hinweise auf auszuführende Rechenoperationen geben (Winter, 1992, S. 364, 2000, S. 7). Stattdessen zeichnen sie sich durch ihre anspruchsvolle mathematische Struktur aus (z. B. R. Rasch, 2001b, S. 26), die vor allem jüngere, ungeübte Lösende nicht ohne Weiteres entschlüsseln können. Auch wenn problemhaltige Textaufgaben Alltagssituationen widerspiegeln, sind die zugrunde liegenden mathematischen Zusammenhänge in einer für die Lösenden unbekannten Form geschildert: „solche, über die das Grundschulkind in der Regel bisher noch nicht nachgedacht hat“ (R. Rasch, 2001b, S. 26). Weitere Schwierigkeiten bestehen darin, wenn mehrere Bedingungen parallel berücksichtigt oder die geschilderte mathematische Formulierungen verstanden werden müssen (Charles, Lester, & O'Daffer, 1987, S. 7; R. Rasch, 2001b, S. 26, 2015a, S. 205). Baumert und Lehmann (1997, S. 233) sprechen solchen Aufgaben eine große Bedeutung zu und fordern deren stärkere Fokussierung im Unterricht. Sie ermöglichen es, den in TIMSS und PISA identifizierten Schwächen deutscher Schüler entgegenzuwirken, gerade weil sie flexibles Denken, das Vernetzen von Informationen, aber auch die Berücksichtigung mehrerer Bedingungen fordern (vgl. Bruder & Collet, 2011b, S. 21).

Die Problemstellungen haben etwas Rätselhaftes, Andersartiges gemein, das die Lösenden erst durchschauen und verstehen müssen (vgl. Abschnitt 1.3.3.2).²⁸ Die Lösung kann nicht ad hoc „aus dem Gedächtnis abgerufen werden“ (Hussy, 1993, S. 20; vgl. Bransford & Stein, 1993, S. 8). Auch eine reine Reproduktion bekannter Lösungsschemata hilft den Lösenden nicht weiter (R. Rasch, 2015a, S. 204, 2015b, S. 33). Da es ihnen aufgrund ihres Entwicklungs- und Wissensstandes verwehrt bleibt, die notwendigen Transformationen in ein Gleichungssystem zu überführen, muss es ihnen gelingen, die Problemsituation neu wahrzunehmen (R. Rasch, 2001b, S. 27, 2008a, S. 6; vgl. Bruder, 2000, S. 2; Gürtler, Perels, Schmitz, & Bruder, 2002, S. 225–226).

Die Komplexität der Aufgabendomäne fordert die Lösenden heraus, neue Wege zu gehen, Entdeckungen zu wagen, indem sie lösungsrelevante Informationen aus der

²⁸ Charles, Lester und O'Daffer (1987, S. 8–9), Franke und Ruwisch (2010, S. 66), Gürtler, Perels, Schmitz und Bruder (2002, S. 225–226), Reusser und Reusser-Weyeneth (1994, S. 20)

geschilderten Problemsituation lösen, vorhandene Beziehungen und Zusammenhänge aufspüren und miteinander verknüpfen.²⁹

Um die Lösungssuche bestreiten zu können, müssen sich die Lernenden gedanklich ein Bild von der Situation machen, eine mentale Repräsentation generieren (vgl. Abschnitt 1.5.2).³⁰ Diese ist Teil der „Verstehensarbeit“ (Reusser & Reusser-Weyeneth, 1994, S. 21) und kann lösungsunterstützend sein, wenn es z. B. gilt, die eigenen Gedankengänge umzukehren, das Problem in seine Komponenten zu zerlegen oder gegebene Problemelemente zu verändern und umzustrukturieren (vgl. Abschnitt 1.3.3.2).³¹ Konkret impliziert dies, dass sich mithilfe von Heuristiken an die Lösung herangetastet werden kann.³² Die selbstgenerierte Repräsentation muss so lange verändert werden, bis sie adäquat den Raum möglicher Lösungen widerspiegelt und das Ablesen der Lösung ermöglicht (Cox, 1999, S. 354; Franke & Ruwisch, 2010, S. 73; Winter, 1997, S. 63–64). Um erfolgreich zu sein, müssen sich die Lernenden bei der Lösungssuche sowohl geistig beweglich (z. B. Bruder, 2000, S. 5)³³ als auch kreativ zeigen (z. B. Guilford, 1950, S. 452–453)³⁴.

Die Lernenden stehen beim Problemlösen vor der Herausforderung, ihr bereits vorhandenes Wissensnetz zu aktivieren und an die jeweilige Problemsituation anzupassen, neu zu durchdenken, zu ergänzen oder auch umzustrukturieren.³⁵ Dabei ist stets der Fokus auf das Erreichen des Ziels gerichtet (R. Rasch, 2015a, S. 204). Ferner darf nicht vernachlässigt werden, dass sprachliche Kompetenzen erforderlich sind, um die Zusammenhänge, die in Textform dargeboten sind, zu erfassen.³⁶

1.4.2 Empirische Erkenntnisse

Empirische Studien zeigen, dass vor allem jüngere Kinder die anspruchsvollen Anforderungen beim Lösen problemhaltiger Textaufgaben und die sich vor ihnen auftür-

²⁹ Bruder (2000, S. 3), Charles et al. (1987, S. 8), Gürler et al. (2002, S. 225), Klix (1971, S. 640–659), Pehkonen (2001, S. 62), R. Rasch (2015a, S. 205), Schukajlow (2011, S. 80), Winter (1992, S. 366, 369, 2000, S. 11)

³⁰ Franke und Ruwisch (2010, S. 73–74), Mayer und Hegarty (1996, S. 36), R. Rasch (2009, S. 69), Verschaffel et al. (2000, S. 169), Winter (1997, S. 57)

³¹ Bruder (2000, S. 5), Charles et al. (1987, S. 8–9), Franke und Ruwisch (2010, S. 66), Gürler et al. (2002, S. 225–226), Reusser und Reusser-Weyeneth (1994, S. 20)

³² Charles et al. (1987, S. 8), Franke und Ruwisch (2010, S. 67–70), Verschaffel et al. (2000, S. XiV), Winter (1997, S. 63–68, 2000, S. 10–11, 2003, S. 181–182)

³³ Bruder (2003, S. 7), Hasdorff (1976, S. 16), Lompscher (1975, S. 36), R. Rasch (2008b, S. 147), Winter (1997, S. 59)

³⁴ Franke und Ruwisch (2010, S. 66), Mayer (1995, S. 3), Winter (2001, S. 289–290)

³⁵ Bransford, Brown und Cocking (2000, S. 42–44), R. Rasch (2001b, S. 27, 2015a, S. 204), Verschaffel et al. (2000, S. XIV), Winter (2000, S. 8, 10)

³⁶ Franke und Ruwisch (2010, S. 73–74), Hegarty et al. (1995, S. 18, 29), Hohn (2012, S. 142), Kintsch und van Dijk (1978, S. 375), Rezat und Rezat (2012, S. 36), Schnottz (1994a, S. 20)

mende Barriere nicht ohne Weiteres bewältigen konnten (z. B. Groß, 2013, S. 114–133; Hohn, 2012, S. 106–108, 151; Perfetti, 2001, S. 12801; R. Rasch, 2001b, S. 89; Stahl, 1975, S. 279; Thevenot & Oakhill, 2005, S. 1317; Winter, 2000, S. 7). Situationen des Stockens und Nichtweiterwissens sind sodann als Barrieren anzusehen, die ohne Unterstützung sehr häufig in Misserfolg münden. Studien von Groß (2013) und Hohn (2012) konnten diese Ergebnisse (nicht nur) für die Grundschule bestätigen: Drittklässler lösten im Mittel $M = 0.45$ ($SD = 0.69$), Viertklässler $M = 0.83$ ($SD = 1.11$), Sechstklässler $M = 2.16$ ($SD = 0.96$) und Neuntklässler $M = 2.61$ ($SD = 1.04$) von 5 problemhaltigen Textaufgaben richtig (Hohn, 2012, S. 106). Prozentual gesehen schwankte der Lösungserfolg in Abhängigkeit von der Aufgabe bei Drittklässlern zwischen 0 % und 15 %, bei Viertklässlern zwischen 0 % und maximal 30 % (Groß, 2013, S. 114–133).

Ferner geht mit dem geringen Lösungserfolg einher, dass ungeübte Lösende selten von sich aus ihre Lösungsgedanken oder auch -ideen notieren (z. B. Fricke, 1987, S. 64)³⁷ und ihnen die Konstruktion passender Darstellungen nicht vertraut ist. Aus der Studie von Hohn (2012, S. 112) geht hervor, dass beim Bearbeiten problemhaltiger Textaufgaben mit zunehmender Klassenstufe das Externalisieren signifikant steigt. Während Drittklässler im Mittel nur bei einer von fünf Problemaufgaben ($M = 1.07$, $SD = 1.41$) externe Repräsentationen einbezogenen, nutzten sie Viertklässler bei $M = 2.25$ ($SD = 1.55$) Problemaufgaben.³⁸ In Bezug auf die Grundschuljahrgänge wurde herausgestellt, dass Drittklässler signifikant seltener externe Repräsentationen konstruierten als Viertklässler und sich darin signifikant weniger flexibel zeigten als ihre Jahrgangskollegen (Hohn, 2012, S. 113–117). Depiktionale Herangehensweisen, wie beispielsweise das Erstellen einer Zeichnung, nahmen im Vergleich zu deskriptionalen, wie beispielsweise dem Erstellen einer Rechnung, eine untergeordnete Stellung ein (Hohn, 2012, S. 110). Dies geht konform mit der Tatsache, dass Lernende vermehrt arithmetische Zugänge wählen (R. Rasch, 2009, S. 69). R. Rasch (2009, S. 84) konnte in ihren Untersuchungen beobachten, dass, wenn junge, ungeübte Löser etwas zu Papier brachten, es sich meist nur um das Ergebnis in Form der „Lösungszahl“ als Resultat der geleisteten Kopfarbeit handelte.

Es liegt nahe, dass „offene Formen von Rechennotizen, wie sie zur Darstellung von mathematischen Gedanken zu Textaufgaben mitunter notwendig sind, (...) scheinbar nicht kultiviert“ werden (R. Rasch, 2009, S. 84). Ein Scheitern oder Aufgeben blieb daher nicht aus, wenn die Lösungssuche ausschließlich im Kopf bewältigt wurde oder

³⁷ Hohn (2012, S. 154), R. Rasch (2009, S. 82, 84), Schreiber (2014, S. 6), Schumacher (2014, S. 1124)

³⁸ Weitere Unterschiede bezüglich der Sekundarstufe I können in Hohn (2012, S. 112–114) eingesehen werden.

es nicht gelang, den geschilderten Sachverhalt und dessen Struktur adäquat zu externalisieren.³⁹ Nicht zuletzt konnte dies auf Unsicherheiten bei der Konstruktion oder fehlendes Vor- und Erfahrungswissen zurückgeführt werden.

Darüber hinaus schienen die individuell zur Verfügung stehenden Kapazitäten schnell ausgeschöpft zu sein (siehe Abschnitt 1.5.3.3). Studien von Lompscher (1975, S. 36–38) und Stahl (1975, S. 294) konnten zeigen, dass vor allem jüngere Lernende unflexibel, meist konzeptlos und unsystematisch beim Lösen komplexer Probleme vorgehen und den Kern von Problemsituationen nicht exakt erfassen, was sich mehrfach replizieren ließ (z. B. Bransford et al., 2000, S. 41; Hegarty, Mayer, & Monk, 1995, S. 18, 24, 29; Hohn, 2012, S. 117). Es resultierten, teilweise gestützt auf Schlüsselwörter, vorschnelle und unüberlegte Rechnungen, die das Ziel verfehlten.⁴⁰

Die Kontrolle des Ergebnisses wird trotz ihrer Wichtigkeit von Grundschulkindern selten vollzogen. Im Schnitt äußerte in der Untersuchung von Groß (2013, S. 142) auf Nachfrage jeder vierte Drittklässler, sein Ergebnis kontrolliert zu haben. Eine erfolgreiche Ergebniskorrektur gelang nur knapp 6 % dieser Probanden. In der vierten Jahrgangsstufe stieg der Anteil durchgeföhrter Kontrollen auf maximal 50 % an, führte aber auch nur in 7 % aller Fälle zum Erfolg.

Lernende benötigen Unterstützung sowie Strategien und Techniken zur Hand, mit deren Hilfe sie einerseits geeignete Repräsentationen konstruieren können (R. Rasch, 2001b, S. 47; Schütte, 2008a, S. 136), aber auch solche, durch die sie erste, konstruierte Repräsentationen zielführend verändern und umstrukturieren können, um vorliegende Probleme zu bewältigen.⁴¹ Denn Erfolgserlebnisse stärken das Vertrauen in sich und die eigenen Fähigkeiten (Bruder & Collet, 2011b, S. 35; Heinrich et al., 2015, S. 284; Süllwold, 1960, S. 106).

1.4.3 *Schlussfolgerung*

Die Auseinandersetzung mit problemhaltigen Textaufgaben fordert ein hohes Maß geistiger Tätigkeit und initiiert komplexe Denkprozesse (Hussy, 1993, S. 18; R. Rasch, 2009, S. 69). Sie stellen für Grundschulkinder herausfordernde Aufgaben dar, die zum Knobeln anregen, individuelle Wege zulassen und dadurch das Erreichen der Zone der

³⁹ De Corte, Verschaffel und de Win (1985, S. 467–469), Hegarty et al. (1995, S. 19), Hohn (2012, S. 112–113, S. 138), Lewis (1989, S. 527), Mayer und Hegarty (1996, S. 31)

⁴⁰ Booker, Bond, Sparrow und Swan (vgl. 2014, S. 76), Krägeloh und Prediger (2015, S. 140), Lorenz und Radatz (1993, S. 143), Muir et al. (2008, S. 231)

⁴¹ Bruder und Collet (2011b, S. 36), Hohn (2012, S. 154), Stahl (1975, S. 295–296), Winter (1992, S. 366–367, 2003, S. 181–182)

nächsten Entwicklung ermöglichen.⁴² Lernende sind mit der Herausforderung konfrontiert, eine mentale Repräsentation zu konstruieren, die ihnen hilft, den Anfangszustand in den Zielzustand zu überführen und das Problem zu enträteln. Es wurde herausgestellt, dass dies nur mit *adäquaten* Repräsentationen vollzogen werden kann.

Der geringe Erfolg beim Lösen problemhaltiger Textaufgaben kann darauf zugrückgeführt werden, dass Grundschulkinder entwicklungsbedingt noch nicht fähig sind, abstrakt-logisch zu denken, sondern vielmehr ihr Denken und Operieren auf Bilder stützen (Lorenz, 2004, S. 53; Piaget, 1972, S. 32). Während Erwachsene und auch schon ältere Kinder eher in der Lage sind, sich Problemstrukturen mental vorzustellen und daran verschiedene Handlungen zu vollziehen, müssen Grundschulkinder dies erst lernen (Lorenz, 2004, S. 54). Dies impliziert, dass das mentale Herstellen von Beziehungen zwischen Mengen sowie das mentale Vollziehen von Handlungsabläufen für junge Lernende Herausforderungen darstellen, die mit zunehmender Komplexität des Problems wachsen (Lorenz, 2004, S. 54; Stern, 1998, S. 84, 100).

Beim Lösen problemhaltiger Textaufgaben ist nicht auszuschließen, dass das von den Lösenden konstruierte Bild, die mentale (interne) Repräsentation, nur unzutreffend oder schlimmstenfalls gar nicht den Kern der Problemmerkmale abbildet. Sind sie in diesem Fall nicht in der Lage, gedanklich die erforderlichen repräsentationalen Veränderungen vorzunehmen, so bleibt ihnen das Ziel verwehrt (vgl. Abschnitt 1.3.3.2). Alternativ ist denkbar, dass ein mentales Bild konstruiert wurde, welches nicht oder noch nicht ausreicht, um den aktuellen Wissensstand dem Zielzustand anzunähern. Grundschulkinder schaffen es dann eher nur vereinzelt, die notwendigen, noch ausstehenden Umstrukturierungen ausschließlich im Kopf zu vollziehen. Liegt eine adäquate, mentale Repräsentation vor, dann muss die Lösung lediglich noch *abgelesen* werden.

In der Summe kann festgehalten werden, dass Grundschulkinder als Novizen des Problemlösens anzusehen sind. Erkenntnisse aus der Textverarbeitungspychologie werden herangezogen, um eine fundierte Vorstellung zu erhalten, welche Prozesse im Kopf problemlösender Kinder ablaufen. Lassen sich Ursachen identifizieren, die den Verstehensprozess und somit auch die Lösungsfindung der Kinder fördern, behindern oder diese sogar verhindern, können Rückschlüsse für mögliche Förderungen gezogen werden.

⁴² Bruder und Collet (2011b, S. 145), Deseniss (2015, S. 84–85), R. Rasch (2012b, S. 133), Vygotskij (2002, S. 242)

1.5 Mathematisches Problemlösen mit externen Repräsentationen

Bevor die Verstehensprozesse beim Lösen problemhaltiger Textaufgaben (Abschnitt 1.5.2) im Fokus stehen, soll, des Verständnisses wegen, kurz aufgezeigt werden, was unter Repräsentationen verstanden wird und welche Formen unterschieden werden (Abschnitt 1.5.1). Schwierigkeiten, die sich in Verstehensprozessen ergeben (Abschnitt 1.5.3), und wie diesen Schwierigkeiten im Unterricht entgegengewirkt werden kann (Abschnitt 1.5.4), werden im Anschluss daran aufgezeigt. In diesem Zusammenhang wird auch die Rolle selbstgenerierter externer Repräsentationen herausgestellt und positioniert (Abschnitt 1.5.5). Welche externen Repräsentationen explizit im Mathematikunterricht der Grundschule zu erwarten sind und welche Merkmale diese auszeichnen, werden im Abschnitt 1.5.6 erörtert. Abschließend werden die Bildungsstandards (Abschnitt 1.5.7) herangezogen, um vorhandene Vernetzungen prozessbezogener Kompetenzen (Darstellen, Kommunizieren und Argumentieren) zu unterstreichen.

1.5.1 Repräsentationen – Formen, Merkmale, Funktionen

Schnottz und Bannert (1999, S. 219, Hervorhebung durch Autorin) verstehen eine Repräsentation als „ein Objekt oder ein Ereignis, das *für etwas anderes steht*, es repräsentiert“.⁴³ Eine Repräsentation wird daher als Modell verstanden, das den Inhalt bzw. die geschilderten Objekte widerspiegelt (Palmer, 1978, S. 262).

Demnach wird beim Lösen einer problemhaltigen Textaufgabe durch das Lesen oder Hören der Textaufgabe eine Assoziation ausgelöst, die *für den geschilderten Sachverhalt steht*. Zum einen kann das mit dem Text Assoziierte im Kopf konstruiert und daher mental (intern) repräsentiert werden. Zum anderen kann der Lösende seine Gedanken und Ideen – seine Sicht der Dinge – zu Papier bringen und somit extern repräsentieren (Krauthausen & Scherer, 2014, S. 154; Schnottz & Bannert, 1999, S. 219; Steiner, 1988, S. 100). Beides hat gemein, dass eine Repräsentation eine selektive Darstellung der eigenen Perspektive widerspiegelt, d. h. ein Individuum konstruiert seine an die Anforderungen angepasste individuelle Repräsentation und nimmt dadurch eine selektive Auswahl und Reduktion von Merkmalen vor, die seiner Meinung nach lösungsrelevant sind.⁴⁴ Stern (1998, S. 30, Hervorhebung im Original) spricht in diesem Zusammenhang von einem „auf das *Handlungsziel* bezogene[n] Ausschnitt aus der Realität“ (vgl. Norman, 1993, S. 48). Darüber hinaus wird eine weitere Unterschei-

⁴³ Vgl. Hörmann (1994, S. 18), Norman (1993, S. 47), Palmer (1978, S. 262), Peterson (1996, S. 7), Schnottz (1994a, S. 145), Schnottz et al. (2011, S. 216)

⁴⁴ Bruner (1971, S. 31), Norman (1993, S. 47–49), Palmer (1978, S. 262), Schnottz (1994a, S. 145), Stern (1998, S. 30)

dung vorgenommen: Das Repräsentierte kann entweder deskriptionale oder depiktionale Form annehmen (Schnotz, 2014, S. 76; Schnotz & Bannert, 1999, S. 220).

Die bekannteste deskriptionale Repräsentationsform stellen Texte dar (Schnotz, 2014, S. 76). Mathematische oder physikalische Formeln zählen ebenfalls zu dieser Gruppierung. Nach Peirce (1933, S. 360) handelt es sich bei deskriptionalen Repräsentationen um Symbole, die als Zeichen mit einer beliebig festgelegten Struktur verstanden werden können (Schnotz, 2002, S. 66, 2014, S. 76; Schnotz et al., 2011, S. 217; Schnotz & Bannert, 2003, S. 143). Eine Textaufgabe besteht demnach aus Symbolen in Form von Worten und Sätzen. Symbole an sich haben aber keine Ähnlichkeit mit dem Repräsentierten (Peirce, 1933, S. 360; Schnotz et al., 2011, S. 218). Ihnen liegt vielmehr eine Konvention zugrunde, die deren Beziehung zum Repräsentierten bestimmt.⁴⁵ Beispielsweise haben die Worte ‚Klassen‘ und ‚Computerraum‘ keine Ähnlichkeit mit realen Klassen bzw. einem realen Computerraum, sie stehen in einer Textaufgabe als Symbol für ein Objekt oder ein Ereignis. Verben und Präpositionen wie ‚gehen‘ und ‚in‘ verdeutlichen zugrunde liegende Zusammenhänge, und zusätzliche Adjektive können besondere Merkmale erkennbar werden lassen.⁴⁶ Zusammen ermöglichen sie es dem Leser, den Text zu erfassen und zu verstehen.

Dem gegenüber stehen depiktionale Repräsentationen wie beispielsweise Bilder und Diagramme (Schnotz, 2014, S. 76; Schnotz et al., 2011, S. 218). Sie unterscheiden sich von deskriptionalen Repräsentationen dadurch, dass sie ikonische Zeichen verwenden, die durch ihre Strukturmerkmale und räumliche Anordnung konkrete Rückschlüsse auf das Repräsentierte zulassen.⁴⁷ Dabei müssen die repräsentierten Merkmale nicht zwingend mit den Merkmalen des zu repräsentierenden Objektes oder Ereignisses übereinstimmen (Schnotz, 2002, S. 66). Deren Nähe zum repräsentierten Objekt oder Ereignis kann über eine konkrete Form der strukturellen Übereinstimmung, im Falle von Bildern, bis hin zur abstrakten Form über Analogien, im Falle von Diagrammen, charakterisiert sein (Schnotz, 2002, S. 66, 2014, S. 7; vgl. Kosslyn, 1994, S. 5). Deskriptionale Repräsentationen können vorliegende Relationen konkret entnommen werden, bei depiktionalen Repräsentationen ist dies nur implizit möglich.⁴⁸

Die Differenzierung zwischen depiktional und deskriptional legt nahe, dass sich Repräsentationen nicht nur in ihrer Form, sondern letztlich auch in ihrer Anwendung und

⁴⁵ Hörmann (1994, S. 18), Peirce (1906, S. 495), Schnotz (2014, S. 76), Schnotz et al. (2011, S. 217)

⁴⁶ Kosslyn (1994, S. 5), Peirce (1906, S. 496–497), Schnotz (2014, S. 76), Schnotz et al. (2011, S. 218), Schnotz und Bannert (2003, S. 143)

⁴⁷ Kosslyn (1994, S. 5), Peirce (1933, S. 359), Schnotz (2002, S. 66, 2014, S. 76), Schnotz et al. (2011, S. 218)

⁴⁸ Johnson-Laird (1983, S. 161–162), Johnson-Laird und Garnham (1980, S. 373), Kosslyn (1994, S. 324–325), Schnotz (2014, S. 76–77), Schnotz et al. (2011, S. 218), Schnotz und Bannert (1999, S. 220)

Benutzung je nach Zweck voneinander unterscheiden. Depiktionale Repräsentationen haben den Vorteil, einen Sachverhalt informell vollständig und greifbar darstellen zu können, sodass erforderliche Informationen im Idealfall einfach abzulesen sind (Kosslyn, 1994, S. 5), während der Vorteil deskriptionaler Repräsentationen darin liegt, allgemeines, abstrakteres Wissen darzustellen.⁴⁹

Bisher unerforscht ist, welchen Einfluss externe Repräsentationen auf das Lösen problemhaltiger Textaufgaben nehmen.

1.5.2 *Verstehensprozess beim Lösen problemhaltiger Textaufgaben*

Werden Verstehensprozesse rekonstruiert, so darf nicht davon ausgegangen werden, dass Lösende mit dem Lesen einer Textaufgabe automatisch deren Bedeutung und Inhalt erfasst und verinnerlicht haben. Vielmehr stellt der vorliegende Text die Grundlage für den Aufbau von Wissenskonstruktionen im Verstehensprozess dar (Schnottz, 1994a, S. 20). Die Textaufgabe fungiert als externe Repräsentation, die es zu verstehen und zu lösen gilt (z. B. Franke & Ruwisch, 2010, S. 73; Graesser, Millis, & Zwaan, 1997, S. 184; Schnottz & Bannert, 1999, S. 219). Der Verstehensprozess beginnt in dem Augenblick, in dem der Lösende die Problemaufgabe liest oder hört. Es besteht unangefochtene Einigkeit, dass konkret das Lesen und Verstehen der Problemaufgabe die Konstruktion dreier mentaler Repräsentationen auslöst, die sich reziprok beeinflussen: Textoberflächenrepräsentationen, propositionale Repräsentationen sowie mentale Modelle.⁵⁰ Nur wenn es gelingt die geschilderte Situation und die zugrunde liegenden Beziehungen zu repräsentieren, kann die Textaufgabe verstanden und gelöst werden (van Dijk & Kintsch, 1983, S. 337 vgl. Abschnitt 1.3.3.2 und 1.4).

Textoberflächenrepräsentationen

Zunächst soll davon ausgegangen werden, dass Lösende die Textaufgabe lesen und nicht hören (siehe Abschnitt 1.5.4). Nach dem *integrativen Modell des Text- und Bildverstehens* nach Schnottz wird mit dem Lesen und der Informationsaufnahme eine mentale Repräsentation der Textoberflächenstruktur konstruiert (siehe Abbildung 1.1).

⁴⁹ Johnson-Laird (1983, S. 146–147), Johnson-Laird und Byrne (1991, S. 35), Kosslyn (1994, S. 324–325), Schnottz (2014, S. 76–77), Schnottz et al. (2011, S. 219), Schnottz und Bannert (1999, S. 220)

⁵⁰ Kintsch (1998, S. 49), Kintsch, Welsch, Schmalhofer und Zimny (1990, S. 140–141), McNamara, Ozuru, Best, & O'Reilly (2007, S. 466), Perfetti (2001, S. 12801), Schnottz (2014, S. 8), Schnottz und Bannert (1999, S. 222), van Dijk und Kintsch (1983, S. 10–19)

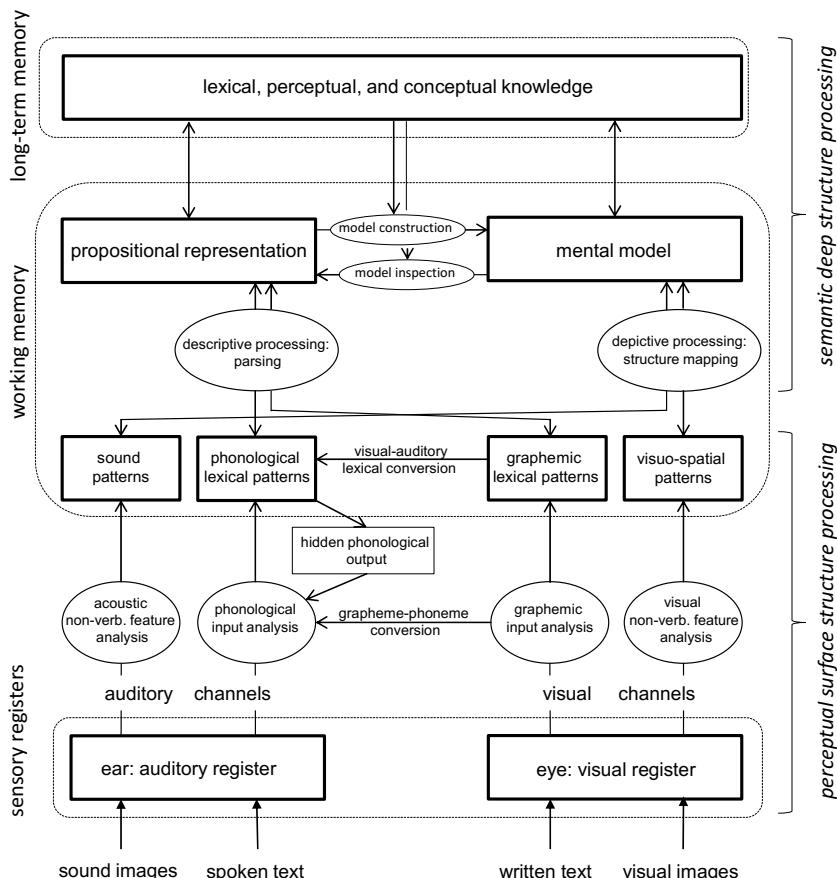


Abbildung 1.1. Integratives Modell des Text- und Bildverständnisses nach Schnotz (2014, S. 83). Es stellt eine Erweiterung des Modells von Schnotz und Bannert (1999, S. 222) dar.

Dies impliziert, dass die einzelnen sprachlichen Textbausteine visuell erfasst und mental repräsentiert, aber auch erinnert werden (Schnotz, 1994a, S. 179). Nach Kürschner und Schnotz (2007, S. 50) sind hierfür „die Formulierungen, die lexikalische und die syntaktische Konstruktion“ des Textes entscheidend (vgl. Schnotz, 1994a, S. 179; Schnotz & Bannert, 1999, S. 220).

Es ist nicht gesagt, dass all das Identifizierte letztlich für die Lösung von Bedeutung ist. Die Textoberflächenrepräsentation offenbart auch nicht, ob das Problem verstan-

den wurde, sie stellt jedoch Weichen für die Konstruktion weiterer, mentaler Repräsentationen (Schnotz, 2014, S. 8; Schnotz et al., 2011, S. 223) und hat demnach hauptsächlich eine Unterstützungsfunction inne (Schnotz, 1994a, S. 46).

Propositionale Repräsentationen

Der Leser konstruiert auf der Basis der Textoberflächenrepräsentation eine propositionale Repräsentation. Losgelöst von der Formulierung oder der im Text vorgefundenen Position der Wörter (*Symbole*), werden die einzelnen Textbausteine konzeptuell organisiert sowie zueinander in Beziehung gesetzt und es werden vorliegende Relationen herausgestellt (Kürschner & Schnotz, 2007, S. 50; Schnotz, 2002, S. 70, 2014, S. 8; Schnotz & Bannert, 1999, S. 221). Eine Proposition spiegelt die entdeckte Relation vorhandener Textbausteine wider und kann demnach als „komplexes Symbol“ verstanden werden, das nach bestimmten Regeln mehrere, simplere Symbole miteinander verknüpft.⁵¹ Sie wird daher als „zusammengesetzte Einheit“ (Schnotz & Bannert, 1999, S. 221; van Dijk & Kintsch, 1983, S. 113) oder „Textbasis“ (Kintsch & Greeno, 1985, S. 116; Kintsch & van Dijk, 1978, S. 365) bezeichnet, die konstruiert werden muss, um einen Text verstehen zu können (van Dijk & Kintsch, 1983, S. 127).

Da sich Problemaufgaben aus mehreren Sätzen zusammensetzen, und jeder einzelne Satz in Propositionen überführt werden kann, müssen die Zusammenhänge auch in mehreren, miteinander verknüpften Propositionen geschildert werden.⁵² Um Textaufgaben verstehen und lösen zu können, müssen Propositionen verstanden und passend zueinander in Beziehung gesetzt werden (Kintsch & Greeno, 1985, S. 114–116; van Dijk & Kintsch, 1983, S. 366–369; vgl. Mayer & Hegarty, 1996, S. 38). Im Falle der YU-GI-OH![®]-Kartenaufgabe von R. Rasch (2016, S. 59), *Lukas und Jonas haben zusammen 30 YU-GI-OH![®]-Karten. Lukas hat 6 mehr als Jonas. Wie viele Karten hat Lukas? Wie viele Karten hat Jonas?*, bedeutet dies, dass die Kinder die Textbausteine „haben“, „hat“, „zusammen“ und „mehr als“ und deren zugrunde liegende Relationen verstehen müssen, um die Aufgabe bewältigen zu können.

Es entsteht somit eine „innere, mentale Sprache“, die sehr stark textorientiert konzipiert ist (Schnotz, 1988, S. 311, 1994a, S. 155; vgl. Kosslyn, 1994, S. 5). Propositionale Repräsentationen werden demnach, ebenso wie die Textoberflächenrepräsentationen, dem deskriptionalen Repräsentationszweig zugeordnet, wobei die Erfassung und Untersuchung vorliegender Strukturen im Vordergrund steht (Schnotz, 2002, S. 70, 2014, S. 10).

⁵¹ Kosslyn (1994, S. 5), Perfetti (2001, S. 12801), Schnotz (2002, S. 67), Schnotz et al. (2011, S. 222), Schnotz und Bannert (1999, S. 221), van Dijk und Kintsch (1983, S. 114–115)

⁵² Johnson-Laird (1980, S. 97), Mayer und Hegarty (1996, S. 38), Schnotz (1994a, S. 151), van Dijk (1980, S. 25), van Dijk und Kintsch (1983, S. 113, 119, 371)

Unbestritten hängt dieser Prozess und letztlich die Konstruktion der propositionalen Repräsentation von dem Wissen ab, das der Lösende aus seinem Langzeitgedächtnis abzurufen fähig ist.⁵³ Ist der Lösende in der Lage, eine adäquate proposionale Repräsentation zu konstruieren, so kann er recht präzise den vorliegenden Sachverhalt intern wiedergeben (Schnottz, 1994a, S. 155).

Mentale Modelle

Nach dem *integrativen Modell des Text- und Bildverständens* wird auf Grundlage der Textbasis eine depiktionale Repräsentation, ein mentales Modell der geschilderten Sachsituation generiert.⁵⁴ Dabei wird die Struktur des zu repräsentierenden Objekts oder Ereignisses nicht eins zu eins, sondern über Analogien repräsentiert (Johnson-Laird, 1980, S. 98, 1983, S. 156, 2005, S. 186; Johnson-Laird & Byrne, 1991, S. 43; Schnottz & Bannert, 1999, S. 221, 2003, S. 147). Darunter wird verstanden, dass analog zu der beschriebenen Problemsituation ein Modell, ein „internes Quasi-Objekt“ (Schnottz, 1994a, S. 158) konstruiert wird, dessen Manipulation ein Ablesen der gewünschten Informationen ermöglicht, welche wiederum Rückschlüsse auf das repräsentierte Objekt oder Ereignis zulassen.⁵⁵ Es darf nicht als originalgetreues Abbild der Problemsituation verstanden werden, wie es bei einer visuellen Wahrnehmung und Vorstellung der Fall ist (Schnottz & Bannert, 1999, S. 221). Vielmehr lassen sich die beiden depiktionalen Repräsentationsformen durch ihren unterschiedlichen informativen Gehalt voneinander abgrenzen: Einerseits können mentale Modelle durch die Aktivierung von Vor- und Erfahrungswissen um Informationen angereichert werden, die nicht explizit Teil der Problemstellung sind und demnach der Perzeption verborgen bleiben (Franke & Ruwisch, 2010, S. 73; Schnottz & Bannert, 1999, S. 221–223, 2003, S. 143, 147; van Dijk & Kintsch, 1983, S. 337–338). Andererseits können lösungssirrelevante Informationen ausgeblendet werden und es kann dadurch eine Selektion von Informationen vorgenommen werden (Franke & Ruwisch, 2010, S. 73; Schnottz & Bannert, 2003, S. 147). Demzufolge formen allgemeines Wissen und Erfahrungswissen des Lernenden sowie die proposionale Repräsentation das mentale Modell und regen die Generierung neuer Erkenntnisse an.⁵⁶ Gerade die Interaktion zwischen der deskriptionalen (propositionalen Repräsentation) und der depiktionalen Repräsentation

⁵³ Duarte, Gogolin und Kaiser (2011, S. 44), Kintsch und Greeno (1985, S. 116), Schnottz (2014, S. 12, 16), van Dijk und Kintsch (1983, S. 127)

⁵⁴ Franke und Ruwisch (2010, S. 72–74), Garnham (1981, S. 560), Perfetti (2001, S. 12801), Schnottz (2014, S. 8), Schnottz und Bannert (1999, S. 223, 2003, S. 145–146), van Dijk und Kintsch (1983, S. 371)

⁵⁵ Johnson-Laird (1983, S. 157), Schnottz (1994a, S. 147, 158), Schnottz und Bannert (1999, S. 221), van Dijk und Kintsch (1983, S. 339)

⁵⁶ Garnham (1981, S. 560), Johnson-Laird (1980, S. 104), Schnottz und Bannert (1999, S. 220), van Dijk und Kintsch (1983, S. 338, 344)

(mentales Modell) ist für das Verständnis und den Erkenntnisgewinn von großer Bedeutung.⁵⁷

Informationen, die mithilfe des mentalen Modells gewonnen werden, werden wiederum der propositionalen Repräsentation hinzugefügt, sodass sich die Prozesse der Modellkonstruktion und Modellinspektion wechselseitig bedingen (Schnotz, 2002, S. 68; Schnotz & Bannert, 2003, S. 146–147). Vollständiges Problemverständnis kann demnach durch die Konstruktion eines adäquaten mentalen Modells erreicht werden, wenn alle lösungsrelevanten Informationen berücksichtigt und aktiviert werden, sodass die Lösung einfach abgelesen werden kann.⁵⁸

1.5.3 Schwierigkeiten im Verstehensprozess beim Lösen problemhaltiger Textaufgaben

Zieht man an dieser Stelle ein Resümee, so bleibt festzuhalten, dass sich die Erkenntnisse und Ansichten der Problemlöseforschung mit denen der Textverständnisforschung vereinen lassen. Sie haben gemein, dass eine adäquate Repräsentation als Voraussetzung für vollständiges Verständnis und die Lösungsfindung gilt. Das *integrative Modell des Text- und Bildverständnisses* gibt ergänzend Aufschluss darüber, wovon die Konstruktion einer adäquaten Repräsentation abhängig ist bzw. welche Faktoren sie (positiv wie negativ) beeinflussen. Demzufolge stehen Problemlösende, insbesondere Problemlöseanfänger, in ihrem Verstehens- und Lösungsprozess vor gewissen Schwierigkeiten und Hürden, die es zu überwinden gilt. Entscheidend dabei ist, dass sich Lernende in ihren Voraussetzungen, Leistungsfähigkeiten sowie auch in ihrem Textverständnis voneinander unterscheiden (van Dijk & Kintsch, 1983, S. 339–340). Exemplarisch werden nachfolgend ohne Anspruch auf Vollständigkeit die zentralen Schwierigkeiten erörtert.⁵⁹

1.5.3.1 Sprachlich und mathematisch bedingte Schwierigkeiten

Mit dem Lesen der Aufgabenstellung ist direkt zu Beginn der Problembewältigung die erste Hürde zu nehmen: Gerade bei Grundschulkindern und Lernenden mit migrationsbedingter Mehrsprachigkeit beeinflussen (bildungs-)sprachliche Kompetenzen und

⁵⁷ Kintsch und Greeno (1985, S. 128), Mayer und Hegarty (1996, S. 38), Schnotz und Bannert (1999, S. 223)

⁵⁸ Vgl. Duarte et al. (2011, S. 41–43), Johnson-Laird (2005, S. 185), Khemlani und Johnson-Laird (2013, S. 145), King (2007, S. 268), McNamara et al. (2007, S. 466), Schnotz (1994a, S. 180), Schnotz und Bannert (2003, S. 146), van Dijk und Kintsch (1983, S. 337–338, 341)

⁵⁹ Darüber hinaus sei auf mögliche Schwierigkeiten hingewiesen, die sich aufgrund gehäufter Misserfolge aufbauen können, die sich aber auch durch mangelnden Bezug zum geschilderten Sachkontext oder dem Prozess des Mathematisierens ergeben (vgl. Deseniss, 2015, S. 99–101).

Lesekompetenzen den individuellen Verstehensprozess.⁶⁰ Unter Ersterem wird in Anlehnung an Gogolin (2009) verstanden, dass die Lernenden über eine sich vom Alltäglichen abhebende deutsche Sprache verfügen sollen (Prediger, 2013b, S. 26). Letzterem wird in Anlehnung an Schnotz und Dutke (2004, S. 63) ein erweitertes Verständnis von Lesekompetenz zugrunde gelegt, das sich vom traditionellen Verständnis abhebt: „Lesekompetenz ist vielmehr als Fähigkeit anzusehen, schriftliche Dokumente zu verstehen, in denen sowohl verbale Informationen in Form von Schriftzeichen (graphemisch) als auch piktoriale Informationen in Form von Bildzeichen (grafisch) enthalten sind.“

Hinzu kommt die Tatsache, dass losgelöst von Alter und Herkunft nicht jeder Leser automatisch das Gleiche mit dem Gelesenen assoziiert (z. B. Graumann, 1960, S. 5; Krauthausen & Scherer, 2014, S. 251; van Dijk & Kintsch, 1983, S. 339). Die letztlich individuell erzeugte Textrepräsentation wird sowohl von den individuellen (bildungs-)sprachlichen Voraussetzungen (Duarte et al., 2011, S. 44) als auch von Zielen, Interessen und der jeweiligen Motivation beeinflusst (Duarte et al., 2011, S. 36; van Dijk & Kintsch, 1983, S. 339–340).

Werden der gelesene Text, einzelne Sätze oder die Bedeutung einzelner Wörter nicht verstanden (Leiss, Schukajlow, Blum, Messner, & Pekrun, 2010, S. 124; Prediger, 2013b, S. 27–28), dann hat dies zur Folge, dass der geschilderte Zusammenhang und dessen Strukturen nicht als Ganzes erfasst werden, dass inadäquate mentale Repräsentationen sowie ein oberflächliches oder falsches Verständnis erzeugt werden.⁶¹ Empirische Studien⁶² konnten zeigen, dass insbesondere Lernende mit migrationsbedingter Mehrsprachigkeit mit solchen Schwierigkeiten zu kämpfen haben. Gleichermaßen konnte für ungeübte Problemlöser nachgewiesen werden (Schoenfeld & Herrmann, 1982, S. 486–492). Es zeigte sich, dass gerade das Verstehen und Deuten von Präpositionen und Komposita, der sogenannte Strukturwortschatz eines Textes, neben komplexen Verbformen und Satzverbindungen als Hürden des Verständnisses anzusehen sind.⁶³

Darüber hinaus wurde herausgefunden, dass Kinder mit mangelndem Operationsverständnis oftmals raten, welche Operation es in der vorliegenden Textaufgabe anzu-

⁶⁰ Duarte et al. (2011, S. 35), Krauthausen und Scherer (2014, S. 206), Magliano, Millis, Ozuru und McNamara (2007, S. 111), Perfetti (2001, S. 12800–12801, 2007, S. 358), Pöhler (2014, S. 925), Prediger (2013b, S. 26, 28), Prediger, Renk, Büchter, Gürsoy und Benholz (2013, S. 54–55)

⁶¹ Franke und Ruwisch (2010, S. 73, 85–89), King (2007, S. 268), Magliano et al. (2007, S. 111), Oakhill und Cain (2007, S. 47–48, 64), Perfetti (2001, S. 12800–12801), van Dijk und Kintsch (1983, S. 370)

⁶² Deseniss (2015, S. 304–317, 327), Duarte et al. (2011, S. 44), Gogolin et al. (2004, S. 135, 139, 151)

⁶³ Deseniss (2015, S. 65), Duarte et al. (2011, S. 44), Kaiser und Schwarz (2009, S. 68–69), Prediger (2013b, S. 30)

wenden gilt (Kittel, 2013, S. 110; Mayer, Larkin, & Kadane, 1984, S. 265–266). Es ist häufig zu beobachten, dass die gegebenen Zahlen aufgrund mangelnden Verständnisses willkürlich miteinander verknüpft werden. Nicht selten halten sie sich auch an Zählhandlungen konkreter Vorstellungsbilder fest (Schipper, 2013, S. 301). Trainierte Löser versuchen dagegen, das geschilderte Problem zu verstehen (Bransford et al., 2000, S. 49), und sind eher in der Lage, Strukturmerkmale wahrzunehmen (Schipper, 2013, S. 301; Schoenfeld & Herrmann, 1982, S. 486–492).

Um die in Problemstellungen geschilderten, komplexen Beziehungen identifizieren zu können, reicht es nicht aus, wenn einzelne Textbausteine fokussiert und dekodiert werden (Franke & Ruwisch, 2010, S. 73; Mayer & Hegarty, 1996, S. 39; Novick & Bassok, 2005, S. 336). Für das Erkennen von Sinnzusammenhängen sind vielmehr Fähigkeiten erforderlich, die der Problemlöser nicht einfach abrufen und reproduzieren kann, sondern entwickeln muss (Kendeou, van den Broek, White, & Lynch, 2007, S. 31; Oakhill & Cain, 2007, S. 64). Insbesondere das Lösen von Textaufgaben erfordert ausgeprägte sprachliche Kompetenzen, um geschilderte Sachsituationen richtig erfassen und zergliedern zu können, sowie um die mathematischen Strukturen aufspüren zu können (Renkl & Stern, 1994, S. 29). Anhand der exemplarisch aufgezeigten Propositionen der YU-GI-OH!®-Kartenaufgabe bedeutet dies, dass die Lernenden das vorliegende Problem nur lösen, wenn sie die Bedeutung von ‚haben‘, ‚hat‘, ‚zusammen‘ und ‚mehr als‘ kennen, und wenn sie fähig sind, die implizit gegebenen Vergleiche zu deuten sowie lösungsförderliches Wissen zu aktivieren (van Dijk & Kintsch, 1983, S. 370). Demnach stoßen Problemlösende gerade bei der Bearbeitung von Textaufgaben auf sprachliche Barrieren und Verständnisschwierigkeiten.

1.5.3.2 Repräsentationsbedingte Schwierigkeiten

Wenn Lernende Schwierigkeiten haben, den ihnen vorliegenden Text bzw. einzelne Textbausteine zu verstehen, dann wird dies in deren mental konstruierten Repräsentationen deutlich (Krägeloh & Prediger, 2015, S. 140; Prediger, 2013b, S. 28). Ihnen bleibt die Aktivierung lösungsunterstützender Handlungsschemata, die sie zu adäquaten Repräsentationen und zur vollständigen Einsicht führen, verwehrt, wenn sie ausschließlich den Kontext oder andere irrelevante Oberflächenmerkmale fokussieren.⁶⁴ Sie generieren stattdessen fehlerhafte oder unvollständige Textoberflächenrepräsentationen und propositionale Repräsentationen, welche die Lösungsfindung erschweren oder vollständig blockieren (Mayer & Hegarty, 1996, S. 39; Novick & Bassok, 2005, S. 336; Perfetti, 2001, S. 12804). Bleiben für die Lösung essenzielle Textbausteine oder zugrunde liegende Relationen unberücksichtigt oder werden sie falsch gedeutet

⁶⁴ Bransford et al. (2000, S. 38), Krägeloh und Prediger (2015, S. 140), Perfetti (2001, S. 12802), Vitale und Romance (2007, S. 77)

(vgl. Krägeloh & Prediger, 2015, S. 140), ist ein enormer kognitiver Mehraufwand nötig, um die einzelnen Textbausteine zu ergänzen oder die vorliegenden Beziehungen zu korrigieren (Schnotz, 1994a, S. 156).

Die Generierung adäquater Textoberflächenrepräsentationen und propositionaler Repräsentationen allein reicht ebenfalls noch nicht aus. Wenn die Lösenden auf diesem Stand innehalten und kein mentales Modell konstruieren, dann sind sie nur in der Lage, den Inhalt des Textes wiederzugeben, und erzielen lediglich oberflächliches Verständnis, lösen jedoch nicht die Problemstellung (King, 2007, S. 268). Tiefes Verständnis bleibt ihnen verwehrt.

Letztlich kann auch die Entwicklung eines adäquaten mentalen Modells langwierig und aufwendig verlaufen. Novizen fällt es nicht immer leicht, ihr vorhandenes mentales Modell zu manipulieren und so lange gedanklichen Veränderungen zu unterziehen, bis die Lösung abgelesen werden kann (Johnson-Laird, 2005, S. 188). Repräsentationen von Experten und Novizen unterscheiden sich dahin gehend, dass es Experten eher gelingt, lösungsrelevante Strukturelemente aufzugreifen und zu repräsentieren, während Novizen an einzelnen Wörtern festhalten und dazu tendieren, lösungsirrelevante Informationen zu repräsentieren (Mayer & Hegarty, 1996, S. 45–46; Novick & Bassok, 2005, S. 336). Hierbei ist entscheidend, wie sie die gegebenen Informationen aufnehmen, verarbeiten, abspeichern, aber auch, welches Wissen sie erinnern, abrufen und aktivieren können (siehe Schoenfeld & Herrmann, 1982, S. 485; Steiner, 1988, S. 114). Ferner sind metakognitive Aktivitäten involviert, die bestimmen, inwiefern Lernende in der Lage sind, ihr Lösungsverhalten zu steuern und zu reflektieren (Schoenfeld, 1985, S. 139–140, 1992, S. 349). Ein Festhalten an inadäquaten Repräsentationen kann auch durch mangelnde metakognitive Aktivitäten hervorgerufen werden (Bannert, 2007, S. 29; Schoenfeld, 1985, S. 139–140, 1992, S. 349). Insbesondere ungeübte Problemlösende überprüfen und reflektieren kaum unaufgefordert ihr Lösungsvorgehen. Sie behalten verstärkt erste, spontan aufgegriffene Handlungen bei und erhalten sie aufrecht. Wird ein Ergebnis erzielt, so wird dies meist angenommen und nicht hinterfragt. In den seltensten Fällen verspüren Lernende von sich aus ein Beweisbedürfnis.⁶⁵ Fehlendes subjektives Beweisbedürfnis sowie Schwierigkeiten des Versprachlichens werden auf den Unterricht zurückgeführt, der Begründungen noch nicht als „antizipiertes schriftsprachliches Handlungsmuster“ vorsieht (Neumann, Beier, & Ruwisch, 2014, S. 123). Ein vertieftes Verständnis für eine Thematik entwickelt sich demnach erst, wenn der Unterricht vielfältige Begründungssituationen forciert (vgl. Kapitel 2).

⁶⁵ Jahnke und Ufer (2015, S. 341), Krauthausen (2001, S. 104), Krumsdorf (2009, S. 11), Neumeister und Vogt (2012, S. 569)

1.5.3.3 Arbeitsgedächtnis

Nach dem *integrativen Modell des Text- und Bildverständens* wird beim Lesen die geschilderte Problemsituation über die visuellen Sinneskanäle aufgenommen, im sensorischen Gedächtnis verarbeitet und an das Arbeitsgedächtnis übermittelt (siehe Abbildung 1.1, Abschnitt 1.5.2). Hierfür stehen jedoch nur begrenzte Kapazitäten zur Verfügung (Schnottz, 2014, S. 80, 83). Sowohl die Kapazitäten der Sinneskanäle, um Textinformationen zu verarbeiten und an das Arbeitsgedächtnis zu übermitteln, als auch die Kapazitäten des Arbeitsgedächtnisses sind begrenzt.⁶⁶ Hinzu kommt, dass jeder Lerner über ein Arbeitsgedächtnis verfügt, das unterschiedlich stark ausgeprägt ist (Wiley, Sanchez, & Jaeger, 2014, S. 600–608, 615). Das Arbeitsgedächtnis junger Lernender ist enorm beansprucht, oft auch überbeansprucht, wenn sie beim Lösen komplexer problemhaltiger Textaufgaben die anspruchsvolle mathematische Struktur aufspüren, ungewohnte Formulierungen entschlüsseln oder mehrere Bedingungen gleichzeitig berücksichtigen müssen (Stern, 1998, S. 40 vgl. Abschnitt 1.4). Sind zusätzlich noch Umstrukturierungen von Problemrepräsentationen, bewusste Anwendungen von Lernstrategien oder eine Suche nach Mustern vorzunehmen, so werden hierfür weitere Kapazitäten des Arbeitsgedächtnisses beansprucht (Schnottz & Kürschner, 2007, S. 496, siehe Abschnitt 1.3.3.2). Es ist daher nicht abzuweisen, dass die Überwindung der Barriere(-n) aufgrund von Überschreitung der Arbeitsspeicherkapazitäten scheitern kann (Stern, 1998, S. 40), insbesondere dann, wenn die Anforderungen der Problemstellungen derart hoch sind und kaum freie Kapazitäten übrig lassen (Schnottz & Kürschner, 2007, S. 497).

1.5.3.4 Vor- und Erfahrungswissen

Wie bereits in den Abschnitten 1.3.3.2, 1.4.2 und 1.5.2 aufgezeigt, hängt die mentale Verarbeitung des Textes auch von dem Vor- und Erfahrungswissen des Einzelnen ab (Duarte et al., 2011, S. 43; Schnottz, 1994a, S. 42–44, 216, 2014, S. 88; Vitale & Romance, 2007, S. 75). Lernende unterscheiden sich diesbezüglich sowie in der Fähigkeit, passendes Wissen zu aktivieren (Bransford et al., 2000, S. 38, 43), was zur Folge hat, dass jeder individuelles Wissen und bisherige Erfahrungen aus dem Langzeitgedächtnis abrupt und in seine Repräsentationskonstruktion einbezieht (Schnottz & Bannert, 2003, S. 223, 233; Unsworth & Engle, 2005, S. 213–215, 218–219; van Dijk & Kintsch, 1983, S. 337–338). Vorrangig erfahrenen und geübten Lernenden gelingt es, strukturelle Parallelen zu zeitlich vorgesetzten Problemlösungen aufzuspüren und diese Wissensstrukturen aufzugreifen (Novick & Bassok, 2005, S. 335–337; R. Rasch,

⁶⁶ Baddeley (1997, S. 29–31, 39), Chandler und Sweller (1992, S. 233), Craik und Lockhart (1972, S. 672), G. A. Miller (1956, S. 95–96), Sweller (2005, S. 19), Sweller und Paas (2014, S. 34), van Dijk und Kintsch (1983, S. 348–349), Wiley et al. (2014, S. 599–600)

2001b, S. 125–128). Für Novizen stellt dies eine enorme Herausforderung dar (Bransford et al., 2000, S. 49; Novick & Bassok, 2005, S. 335–336; R. Rasch, 2001b, S. 129). Zuvor erworbene Strukturen müssen nicht nur erkannt, sondern auch in modifizierter Form auf die neue Situation übertragen werden.

Die Wahrscheinlichkeit, ein adäquates mentales Modell konstruieren zu können, steigt, wenn auf passendes Vor- und Erfahrungswissen zurückgegriffen werden kann (Schnitz, 2014, S. 88; van Dijk & Kintsch, 1983, S. 352). Dadurch können nicht explizit gegebene Informationen ergänzt (Schnitz, 1994a, S. 34–35, 2014, S. 82) sowie die gestellten Anforderungen mithilfe vorheriger Problemlöseprozeduren bewältigt werden, und es kann auch das Arbeitsgedächtnis entlastet werden (Adams, Bell, & Perfetti, 1995, S. 318; van Dijk & Kintsch, 1983, S. 352 vgl. Abschnitt 1.3.3.2). Darüber hinaus können geringe Arbeitsgedächtniskapazitäten und Defizite in propositionalen Repräsentationen durch ausreichendes Vor- und Erfahrungswissen aufgefangen werden (Schnitz, 2014, S. 82; Schnitz et al., 2011, S. 225).

1.5.4 Ableitungen aus den Schwierigkeiten im Verstehensprozess

Mit dem Ziel, Lernende beim Verstehen und Lösen problemhaltiger Textaufgaben zu unterstützen, muss den genannten Schwierigkeiten (siehe Abschnitt 1.5.3) entgegengewirkt werden. Es kristallisiert sich heraus, dass es neben dem Textverständnis der Lernenden vorrangig gilt, die Überführung des Gelesenen in eine adäquate Repräsentation zu unterstützen und zu fördern:⁶⁷

1. Auf textlicher Ebene können Verständnisschwierigkeiten im Vorfeld reduziert werden, indem sprachliche Anforderungen der Aufgabenstellungen an die Zielgruppe angepasst sind (vgl. Duarte et al., 2011, S. 44–49; Eichler, 2015, S. 54; Rösch & Paetsch, 2011, S. 73).
2. Werden die Textaufgaben nebst dem eigenständigen Lesen noch vorgelesen, dann wird der Sachverhalt nicht nur visuell, sondern auch auditiv verarbeitet und zusätzlich über die auditiven Sinneskanäle an das Arbeitsgedächtnis übermittelt (Baddeley, 1997, S. 26; Mayer, 2014, S. 66; Schnitz, 2014, S. 80). D. h., die Informationen können sowohl deskriptional als auch depiktional verarbeitet werden, was die Wahrscheinlichkeit, eine adäquate propositionale Repräsentation zu konstruieren, erhöht (Schnitz, 2014, S. 83–84). Die zur Verfügung stehenden Verarbeitungskapazitäten werden dadurch optimiert (Sweller et al., 1998, S. 289). Das Vorlesen hilft sowohl leistungsstarken und leistungsschwachen Kindern als auch Kindern mit migrationsbedingter Mehrsprachigkeit, da es

⁶⁷ Krauthausen und Scherer (2014, S. 247–248), Mayer und Hegarty (1996, S. 48), Prediger (2013b, S. 27–28), R. Rasch (2001b, S. 48)

einen weiteren Zugang zur Problemaufgabe eröffnet. Losgelöst von den sprachlichen Voraussetzungen des Einzelnen werden nahezu gleiche Ausgangsbedingungen geschaffen (R. Rasch, 2009, S. 80). Darüber hinaus konnte gezeigt werden, dass das Hören von kurzen Texten zu besseren visuellen Repräsentationen führt als das Lesen (Baddeley, 1997, S. 30; Kürschner & Schnotz, 2007, S. 60). Es empfiehlt sich daher, das Vorlesen des Textes zusätzlich aufzugreifen.

3. Gemeinsam über die Textaufgabe zu sprechen und Textbausteinen aufzugreifen, kann als verbale Unterstützung angesehen werden, um sprachliche Unklarheiten auf Wort-, Satz- oder Textebene zu klären (Deseniss, 2015, S. 292–293, 327–328; Prediger, 2013b, S. 29–30, 33; Roos, 2014, S. 29), aber auch der Entwicklung inadäquater Repräsentationen frühzeitig entgegenzuwirken. In diesem Rahmen bietet sich eine Thematisierung lösungsunterstützender Werkzeuge an (Bransford et al., 2000, S. 48–49; Krägeloh & Prediger, 2015, S. 140–141). Fachübergreifende Lesestrategien (wie z. B. „Suche nach Schlüsselwörtern“) sollen ebenso wie das Thematisieren isolierter Wörter, ohne Berücksichtigung des jeweiligen Sachzusammenhangs, vermieden werden (Krägeloh & Prediger, 2015, S. 140–141; Prediger, 2013b, S. 29–30), weil sie die Hürden auf Text-, Satz- und Wortebene forcieren könnten.
Deseniss (2015, S. 339–343) konnte zeigen, dass gerade die anfängliche Auseinandersetzung mit der Problemstellung den Problemlöseprozess entscheidend beeinflusst. Demzufolge ist der Fokus des Unterrichts darauf zu richten, die Bedeutung lösungsrelevanter Aspekte zu klären und deren Verständnis abzusichern (Deseniss, 2015, S. 343). Dabei sollen neben sprachlichen und kontextbedingten Aspekten, vor allem auch mathematische Aspekte Berücksichtigung finden (Deseniss, 2015, S. 344). Nur wenn Verständnisbarrieren thematisiert werden und deren Überwindung unterstützt wird, können Fehlerquellen, die aus mangelnder Sprachkompetenz resultieren, reduziert werden (Deseniss, 2015, S. 341, 347).
- 4a. Sprachlich und mathematisch bedingte sowie repräsentationsbedingte Schwierigkeiten, die sich beim Lösen von Problemaufgaben ergeben, lassen sich häufig auf mangelnde Erfahrungen und unzureichendes Wissen der Lernenden zurückführen. Dies ist keinesfalls verwunderlich, denn Grundschulkinder gelten als Problemlösenovizen. Unter Berücksichtigung ihres Vor- und Erfahrungswissens ihre Problemlösekompetenzen können durch wiederholtes Lösen von Problemen Problemlösekompetenzen erworben und weiterentwickelt werden.⁶⁸ Letztlich sind gerade eigene Problemlöseerfahrungen notwendig und hilfreich, um

⁶⁸ Eichler (2015, S. 57), Fehse (2001, S. 51), Heinrich et al. (2015, S. 284), Kilpatrick (1985, S. 8), Törner und Zielinski (2013, S. 259), van Essen und Hamaker (1990, S. 306)

die gewonnenen Erkenntnisse und Einsichten zu sichern (Heinrich et al., 2015, S. 284–285). Darüber hinaus kann eine intensive Auseinandersetzung mit einer Problemstellung eine höhere Anstrengungsbereitschaft auslösen, die in neuen mathematischen Erkenntnissen münden kann (Heinrich et al., 2015, S. 285). Alles was sich die Kinder selbst erarbeitet und angeeignet haben, steht ihnen mit erhöhter Wahrscheinlichkeit langfristig zur Verfügung und kann im Idealfall flexibel auf andere Kontexte übertragen werden. Von Vorteil ist, dass Gelerntes in Form von Informationen, die im Arbeitsgedächtnis verarbeitet wurden, als Schema im Langzeitgedächtnis abgespeichert und als solches immer wieder abgerufen werden kann (Sweller et al., 1998, S. 289).

4b. Ferner können Lücken im Vor- und Erfahrungswissen durch die Bereitstellung von Repräsentationen als zusätzliche Informationsquellen geschlossen werden. Vielfältige Untersuchungen konnten in diesem Zusammenhang zeigen, dass Lernende mit einem hohen Maß an Vorwissen eher in der Lage sind, ohne zusätzliche Informationen ein passendes mentales Modell zu konstruieren, als Lernende mit einem geringen Maß an Vorwissen⁶⁹ (Mayer, 1997, S. 14–15; Schnottz & Bannert, 1999, S. 233, 2003, S. 153–154; vgl. auch Bransford et al., 2000, S. 38). Die Tatsache wird in zwei konträre Richtungen geschärft: a) Lernenden, den Vor- und Erfahrungswissen fehlt, fällt es schwer, ausschließlich mithilfe des Textes ein mentales Modell aufzubauen (Mayer, 1997, S. 14–16; Schnottz, 2014, S. 88; Schnottz & Bannert, 2003, S. 153); b) zusätzlich zur Verfügung gestellte Repräsentationen sind für Lernende mit hohem Vorwissen sogar eher hinderlich als förderlich („expertise reversal effect“, Kalyuga, Ayres, Chandler, & Sweller, 2003, S. 23–31; „redundancy effect“, Kalyuga & Sweller, 2014, S. 248).

Während die ersten vier Schwierigkeitsreduktionsmöglichkeiten (1–4a) leicht umsetzbar sind (vgl. Abschnitt 4.1), bringt die letzte (4b) Unstimmigkeiten mit sich. Es besteht in der Ergänzung eines depiktionalen Zweiges zwar der Vorteil, dass die Lernenden auf unterschiedlichen (deskriptionalen und depiktionalen) Wegen zur Modellkonstruktion gelangen können, was häufig für leistungsschwächere Lernende lernförderlich ist (Schnottz & Bannert, 1999, S. 233). Nachteilig ist jedoch, dass durch eine zusätzliche Bereitstellung adäquater depiktionaler Repräsentationen die Problemhaftigkeit problemhaltiger Textaufgaben verloren ginge, wenn alle lösungsrelevanten Informationen herausgelesen werden können (Reuter, 2016, S. 84). Es wäre dann zu erwarten, dass Lernende die Informationen direkt der depiktionalen Repräsentation entnehmen und den Text vernachlässigen, weil redundante Informationen bereits ausge-

⁶⁹ Analoges konnte für schlechte Leser gezeigt werden (Schnottz, 2014, S. 88).

blendet sind (Ayres & Sweller, 2014, S. 220). Die Hinzugabe einer weiteren Informationsquelle ist nur nützlich, wenn sie ergänzend zur Lösungsfindung herangezogen werden müssen (Ayres & Sweller, 2014, S. 220; Chandler & Sweller, 1991, S. 323, 330, 1996, S. 166–168). Demnach wirkt sich eine gesplittete Aufmerksamkeit insbesondere für leistungsstarke Lerner nachteilig aus, wenn zur Lösungserarbeitung bereits eine der zur Verfügung gestellten Informationsquellen ausreicht (vgl. Aspekt 4b, *redundancy effect*).

Eine Studie von Reuter (2016, S. 118, 122) deckte auf, dass Lernende, die zusätzlich zum Problemtext adäquate, aber auf mehreren Stufen abstrahierte unvollständige Repräsentation zur Verfügung gestellt bekamen, die problemhaltigen Textaufgaben nicht signifikant besser lösten als Lernende, die keine vorgegebene Repräsentation erhielten.

Zusammenfassend kann festgehalten werden, dass zusätzliche Informationsquellen in Form von depiktionalen Repräsentationen für die Lösungsfindung von Leistungsschwächeren lernförderlich oder nahezu unentbehrlich, wiederum für den Lösungsprozess von Leistungsstärkeren eher lernhinderlich sein können. Aufgrund dessen muss eine Möglichkeit gesucht werden, von der möglichst viele Lernende profitieren. Dies veranlasste, in den Blick zu nehmen, welche Vorteile es für den Lernprozess mit sich bringt, wenn Lernende ihre Lösungsunterstützung in Form externer Repräsentation selbst generieren und mit ihr arbeiten.⁷⁰ Die individuelle Sicht der Dinge und somit das mentale Modell können durch das Externalisieren konkretisiert sowie präzisiert, manipuliert und weiterentwickelt werden (Krauthausen & Scherer, 2014, S. 245–246). Das Gedachte nimmt äußere Gestalt an, wird dadurch nachvollziehbar und zeigt auf, welche zu repräsentierenden Inhalte der Erzeuger im Kopf hat (Norman, 1993, S. 47–49 siehe Abschnitt 1.5.1). Es konnte empirisch gezeigt werden, dass bereits Grundschulkinder in der Lage sind, vielfältige externe Repräsentationen in Problemlöseprozessen zu generieren (z. B. Hardy, 2007, S. 13; R. Rasch, 2001b, 2008a, 2012b). Ferner konnte gezeigt werden, dass sie in der Lage sind, externe Repräsentationen auf unterschiedliche Weise zu deuten und sich darüber auszutauschen (z. B. Söbbeke & Steenpaß, 2010, S. 222–223).

Das *integrative Modell des Text- und Bildverständens* kann auch in diesem Zusammenhang herangezogen werden, wenn es gilt, kognitive Prozesse in umgekehrter Richtung zu analysieren:

Ein Individuum kann spezifische Ideen im Kopf haben, übereinstimmende propositionale Repräsentationen und mentale Modelle in seinem Arbeitsgedächtnis entwerfen und diese Repräsentationen dann durch das Schreiben eines entsprechenden Textes oder das Zeichnen eines externen Bildes oder Diagramms externalisieren. Das Individuum kann seinen eigenen Text dann

⁷⁰ Bruner (1971, S. 33), Franke und Ruwisch (2010, S. 79), Krauthausen und Scherer (2014, S. 116), Papert (1993, S. 142–143), R. Rasch (2001b, S. 47), Selter (1994, S. 64–67)

erneut lesen und sein eigenes Bild oder Diagramm überdenken, mentale Repräsentationen des Inhaltes rekonstruieren und dann seine mentalen Repräsentationen noch tiefgehender ausarbeiten, über die vorherigen Repräsentationen hinaus. (Schnottz et al., 2011, S. 225)

Das Potenzial externer Repräsentationen wird in der Unterstützung und Entlastung mental ablaufender Denk- und Lösungsprozesse einerseits sowie deren Weiterentwicklung andererseits gesehen (siehe Abschnitt 1.5.5). Ferner können Lernende von ihnen profitieren, wenn sie Lerngegenstände explorieren oder darüber kommunizieren und reflektieren (z. B. Fetzer, 2009, S. 21–23; Selter, 1994, S. 31–35). In der Mathematikdidaktik werden sie unter anderem als „heuristische Hilfsmittel“ (Bruder & Collet, 2011b, S. 45–67), „Werkzeuge des Lernens“ (Krauthausen & Scherer, 2014, S. 242), „Denk- und Erkenntniswerkzeuge“ (Söbbeke & Steenpaß, 2014, S. 13)⁷¹ oder „Prototypen“ im Sinne „dauerhafter Denkmittel“ (Dörfler, 1989, S. 81–82) bezeichnet. In Anlehnung an König (1992, S. 24) wird darunter verstanden, „daß sie vom konkreten Inhalt der zu lösenden Aufgabe weitgehend unabhängig sind und daß sie den Aufgabenlöser zum Aufbau von Suchräumen für effektive Lösungsvarianten bei beliebigen Aufgaben befähigen“ (König, 1992, S. 24, Hervorhebung im Original).

In der vorliegenden Arbeit wird der Bezeichnung Denk- und Erkenntniswerkzeuge gefolgt, da diese Bezeichnung am treffendsten deren holistische Bedeutung für den Lösungsprozess zum Ausdruck bringt. Es geht daraus hervor, dass sie für die Prozesse sowohl des Verstehens, Planens und Lösens als auch des Kontrollierens unterstützend herangezogen werden können, um mathematisch tätig zu sein und Zusammenhänge autonom zu erschließen (Steenpaß, 2014, S. 20; vgl. Söbbeke, 2005, S. 65). Ihre kontinuierliche und ganzheitliche Präsenz zeichnet sich in dieser Benennung mehr als in der Bezeichnung Hilfsmittel ab.

Nachfolgend werden deren zentrale Vorteile und Bedeutung für jeden Einzelnen sowie den intersubjektiven Austausch aufgezeigt.

1.5.5 Potenzial externer Repräsentationen

Wenn man berücksichtigt, dass externe Repräsentationen insbesondere dann lösungsunterstützend sind, wenn sie von den Lösenden selbstgeneriert werden, dann stellt sich die Frage danach, in welchen Unterrichtsphasen oder auch Sozialformen sie bevorzugt einzusetzen sind und welche Vorteile daraus hervorgehen. Nachfolgend wird deren Rolle in der explorativen Phase (Abschnitt 1.5.5.1), der kommunikativen Phase (Abschnitt 1.5.5.2) und der reflexiven Phase (Abschnitt 1.5.5.3) des Unterrichts aufgezeigt.

⁷¹ Vgl. Kautschitsch (1984, S. 61), Söbbeke (2005, S. 64–65, 98), Steenpaß (2014, S. 20)

1.5.5.1 Bedeutung für den Produzenten

Das Abbilden mentaler Prozesse bringt den Vorteil mit sich, dass sich der Produzent fortwährend an seinen eigenen Notizen orientieren und sukzessive zur Lösung vorarbeiten kann (Lorenz, 2004, S. 54; Steiner, 1988, S. 114). Er konstruiert seine eigene Gedächtnissstütze (Lorenz, 2004, S. 48), durch die nahezu sichergestellt werden kann, dass in Situationen des Überlegens keine wichtigen Gedanken verloren gehen (Krauthausen & Scherer, 2014, S. 116; Rathgeb-Schnierer, 2005, S. 173). Externalisiertes kann, gleich ob auf depiktionaler oder auf deskriptionaler Ebene, individuelle Denk- und Lösungsprozesse stützen.⁷² Ferner schafft es Raum für individuelle Entdeckungen (E. C. Wittmann, 1998, S. 158). Der Lernende generiert seine individuelle Auslegung dargebotener Informationen, die sich auf eine dynamische Auseinandersetzung mit externen und mentalen Modellen bezieht (Cox, 1999, S. 347; Reisberg, 1987, S. 288; vgl. Hohn, 2012, S. 31). Entdeckerhefte, Lerntagebücher oder Forscherhefte gelten als etabliertes Medium, um solche individuellen Auseinandersetzungen mit der Problemstellung zu dokumentieren.⁷³

Ferner wird das *Arbeitsgedächtnis* des Produzenten entlastet.⁷⁴ Die individuellen Denk- und Handlungsprozesse sowie die einzelnen Aufgabenbedingungen müssen nicht im Arbeitsgedächtnis präsent gehalten werden (Fehse, 2001, S. 38; Franke & Ruwisch, 2010, S. 79; Steiner, 1988, S. 116). Im Vergleich zum Lösen im Kopf ist der kognitive Verarbeitungsaufwand durch das Externalisieren erheblich geringer (Kirsh, 2010, S. 443; Schnotz, 1994b, S. 113). Die frei werdenden Kapazitäten können für die Zielfindung genutzt werden (Fehse, 2001, S. 52–53; Sweller, 1994, S. 299).

Der Produzent hat dadurch während seines gesamten Denk- und Lösungsprozesses seine Sicht der geschilderten Situation vor Augen, er kann sie immer wieder einbeziehen, neu durchdenken und wenn nötig verändern, systematisch variieren, umstrukturieren, er kann aber auch bereits vollzogene Schritte widerrufen.⁷⁵ Das Umsetzen von Vermutungen oder Eingehen von Wagnissen ist im Vergleich zur mentalen Verarbeitung leichter zu simulieren und auch schneller wieder rückgängig zu machen. Externe Repräsentationen haben den Vorteil höherer Stabilität gegenüber mentaler Repräsentation.

⁷² Amarel (1966, S. 113), Franke und Ruwisch (2010, S. 75, 79), Kirsh (2010, S. 441, 443), Lorenz (2004, S. 54), Norman (1993, S. 49), Reisberg (1987, S. 282), Simon (1996, S. 132), Steiner (1988, S. 114)

⁷³ Z. B. Berthold, Nückles und Renkl (2007, S. 565), Philipp (2013, S. 48), Rathgeb-Schnierer (2005, S. 174), Selter (1994, S. 39–49)

⁷⁴ Cox (1999, S. 348), Fehse (2001, S. 38), Norman (1993, S. 51), Schnotz et al. (2011, S. 217), Steiner (1988, S. 114), Sweller und Paas (2014, S. 34), Sweller, van Marriënboer und Paas (1998, S. 289), Thevenot und Oakhill (2008, S. 319)

⁷⁵ Reisberg (1987, S. 286), Schnotz et al. (2011, S. 225), Söbbeke und Steenpaß (2014, S. 10), E. C. Wittmann (2014, S. 229)

tionen (Kürschner & Schnottz, 2008, S. 140; Lorenz, 2005, S. 156; Schnottz et al., 2011, S. 217). Da sie stetig verändert, durch neue Informationen ergänzt, überarbeitet und angepasst werden können, bieten sie dem Individuum ein hohes Maß an Flexibilität und Beweglichkeit bei der Umstrukturierung und Einschränkung des Problemraumes (Fehse, 2001, S. 44, 50, 97; vgl. Newell & Simon, 1972, S. 809; Reisberg, 1987, S. 286). Darüber hinaus werden die Strukturierung der Denkprozesse und die individuelle Wissensaneignung durch externe Repräsentationen unterstützt und erleichtert (Reisberg, 1987, S. 281–282; Steiner, 1988, S. 116). Sie führen den Lernenden zur Lösung, wenn es gelingt, lösungsrelevante Merkmale zu repräsentieren und irrelevante zu selektieren (Franke & Ruwisch, 2010, S. 145; Norman, 1993, S. 49) sowie auf dieser Basis das erforderliche Vorwissen zu aktivieren (Steiner, 1988, S. 114, s. Abschnitt 1.3.3.2).

Folglich ermöglichen externe Repräsentationen, in Kommunikation mit sich selbst zu treten (Cox, 1999, S. 347–348; Rathgeb-Schnierer, 2005, S. 166; Schnottz et al., 2011, S. 217, 225). Rathgeb-Schnierer (2005, S. 174) spricht in diesem Zusammenhang von „schriftlichem Kommunizieren“. Der Produzent steht vor der Aufgabe, seine Lösungsgedanken in Worte oder Bilder zu fassen (Reisberg, 1987, S. 281). In Momenten des Nachdenkens, Strauchelns oder Zurückblickens hat er immer die Möglichkeit, in die Rolle des Empfängers zu schlüpfen und sein eigenes, bisheriges Produkt von außen zu betrachten und zu reflektieren (Renkl & Nückles, 2006, S. 145; Schnottz et al., 2011, S. 217, 225). Von Vorteil ist, dass dieser Prozess nicht im Kopf, sondern, das Arbeitsgedächtnis entlastend, an der externen Repräsentation vollzogen werden kann (Schnottz et al., 2011, S. 217, 225; van Essen & Hamaker, 1990, S. 302). Der Produzent gewinnt eine Distanz zu seinem Denk- und Lösungsvorgehen, was sein Bewusstsein schärfen und erweitern kann (Selter, 1994, S. 46). Der Rollenwechsel eignet sich, um eventuell erforderliche Umstrukturierungen oder repräsentationale Veränderungen aufzuspüren, welche im Anschluss daran direkt ausgeführt werden können (van Essen & Hamaker, 1990, S. 302). In ihrer Funktion als Denk- und Erkenntniswerkzeuge stellen externe Repräsentationen „eine wichtige Brücke zwischen dem konkreten Tun (dem Zeichnen, Abdecken, Zählen, Zeigen, etc.), dem Denken des Kindes und der abstrakten mathematischen Struktur dar“ (Söbbeke & Steenpaß, 2014, S. 10).

1.5.5.2 Bedeutung für die Interaktion unter Gleichaltrigen

Externe Repräsentationen lassen mental vollzogene Handlungen und individuelle Entdeckungen sichtbar und real werden, wovon nicht nur der Produzent selbst profitieren kann.⁷⁶ Sie stärken und untermauern bei komplexen Problembewältigungen auch

⁷⁶ Hahn und Janott (2012, S. 39), Maier (2000, S. 10–11), Rathgeb-Schnierer (2005, S. 173), van Essen und Hamaker (1990, S. 302)

Kommunikationen unter Gleichaltrigen, den Peers, oder Kommunikationen mit der Lehrkraft, insbesondere dann, wenn Lernende sie während des Austauschs vor sich liegen haben.⁷⁷ Sie fungieren in diesem Zusammenhang als Gesprächs- und Argumentationsstütze (Norman, 1993, S. 48; vgl. Krauthausen & Scherer, 2014, S. 28) bzw. als Gesprächsanker im Sinne einer „gemeinsamen Referenz“ (Fehse, 2001, S. 20, 75; vgl. Baker, 1994, S. 245).

Von Vorteil ist, dass der Produzent im Gespräch seine Gedankengänge verbal und non-verbal an seinen Notizen transparent machen, durch Zeigen Bezüge herstellen sowie seine Argumentationen auf das Externalisierte stützen kann.⁷⁸ Dies geht konform mit der Tatsache, dass Grundschulkinder als Novizen des Kommunizierens und Argumentierens gelten und deshalb ein Medium benötigen, auf das sie ihre Verständigung stützen können (Fetzer, 2011, S. 42; Lorenz, 2005, S. 155). Externe Repräsentationen können als ein solches Medium fungieren, indem durch das Veranschaulichen der eigenen Lösungsgedanken Strukturen sichtbar werden, die für Dritte nachvollziehbar werden (Fetzer, 2011, S. 48; Lorenz, 2005, S. 156). Sie haben das Potenzial, sprachliche Ungenauigkeiten zu überwinden, wenn sie kindliches Denken widerspiegeln (Lorenz, 2005, S. 163). Auf dieser Basis kann der Interaktionspartner (Rezipient) den mental vollzogenen Lösungsprozess seines Gegenübers (Produzenten) leichter rekonstruieren, als wenn er diesem nur verbal folgen muss.⁷⁹ Gelingt es Lernenden, den Kommunikationsprozess an den selbstkonstruierten Repräsentationen auszurichten, dann gewinnen sie an kognitiven Ressourcen, die sie für den Austausch von Argumenten und die Auseinandersetzung mit den Repräsentationen ihrer Mitschüler nutzen können (Fehse, 2001, S. 75; Lorenz, 2005, S. 160). Sinnstiftende Diskussionen können angeregt werden, in denen alternative Vorgehensweisen erörtert, reflektiert und diskutiert, aber auch notwendige Umstrukturierungen, Veränderungen und Ergänzungen gemeinsam ausgehandelt werden.⁸⁰ Solche Prozesse können die Verarbeitung des Lerninhalts vertiefen, den Lernprozess intensivieren und in tieferem Verständnis münden (Fehse, 2001, S. 21; Renkl & Nückles, 2006, S. 142). Von Repräsentationen, auf die sich im Gespräch direkt bezogen wird, gehen wichtige Impulse für die Gewinnung neuer Einsichten aus (Gysin, 2013, S. 133; Steinbring & Nührenbörger, 2010, S. 169).

⁷⁷ Cox (1999, S. 347), Fetzer (2007, S. 264), Franke und Ruwisch (2010, S. 76), Gallin (2010, S. 6), Gysin (2013, S. 137), Norman (1993, S. 49–51), Roos (2014, S. 38), Schnotz et al. (2011, S. 217)

⁷⁸ Fetzer (2011, S. 44, 47), Hahn und Janott (2012, S. 39), Lorenz (2005, S. 156), Söbbeke und Steenpaß (2014, S. 10), Verboom (2014a, S. 42), E. C. Wittmann (1998, S. 158)

⁷⁹ Bransford und Johnson (1972, S. 720), Fetzer (2011, S. 48), Hörmann (1994, S. 471), Norman (1993, S. 49), Schnotz et al. (2011, S. 217)

⁸⁰ Vgl. Baker (1994, S. 245), Fetzer (2007, S. 264), Steinbring und Nührenbörger (2010, S. 169), Webb (1989, S. 29)

Es kann festgehalten werden, dass selbstgenerierte Repräsentationen in Kommunikationen unter Gleichaltrigen nicht nur den Produzenten stärken, sondern auch den Rezipienten, der auf ihrer Basis leichter lernen kann (Fehse, 2001, S. 67–72; Plötzner & Fehse, 1998, S. 203–216). Analoges kann auch auf mehrere Rezipienten, wie beispielsweise eine Gruppe bzw. eine Klasse, übertragen werden, sodass ein Miteinander- und Voneinander-Lernen möglich wird (z. B. Gysin, 2013, S. 137; Selter, 1994, S. 26–28).

1.5.5.3 Bedeutung für Kommunikations- und Reflexionsphasen im Mathematikunterricht

Das Externalisieren mentaler Modelle schließt implizit ein, dass das Verständnis, die Kenntnisse sowie das Vorwissen und Fehlvorstellungen der Lernenden transparent werden (Lorenz & Radatz, 1993, S. 58; Roos, 2014, S. 38; E. C. Wittmann, 2003, S. 30, vgl. Kapitel 1.5.2). Es gilt, an dem individuellen Wissen der Lernenden anzuknüpfen und dessen Weiterentwicklung zu fördern (M. Meyer & Prediger, 2012, S. 6). Es liegt daher nahe, die Denk- und Erkenntniswerkzeuge der Lernenden nicht nur unter Gleichaltrigen, sondern auch im Klassenverband zum Kommunikations- und Reflexionsanlass zu machen.⁸¹

In Anlehnung an Selter (1994, S. 30–31) werden in der vorliegenden Studie die von den Lernenden individuell konstruierten externen Repräsentationen als Eigenproduktionen aufgefasst. Es handelt sich dabei um Lösungen, die „ohne Einfluss durch Unterricht zustande gekommen sind“ (E. C. Wittmann, 2003, S. 33). Selbst wenn die Eigenproduktionen der Lernenden implizit eine Begründung mitliefern und der Lösungsprozess selbsterklärend scheint, kann letztlich nicht gesichert werden, dass ein intensives Nachdenken sowie eine vertiefte Durchdringung des Lösungsprozesses seitens der Rezipienten erfolgt (Maier, 2000, S. 13; Rathgeb-Schnierer, 2005, S. 166). Kommunikationen mit und über Repräsentationen müssen deshalb (nicht nur unter Gleichaltrigen) angeregt werden (Dörfler, 2006, S. 215; Roos, 2014, S. 38–39).

Der Blick zurück auf Lösung, Lösungsweg und individuelle Repräsentationen ist in mehrfacher Hinsicht unentbehrlich (Eichler, 2015, S. 56):

Die klassenöffentliche Diskussion *unterschiedlicher Lösungswege* sowie die *Beurteilung ihrer Eignung* tragen dazu bei, dass sich die Kinder der verschiedenen Strategien und Methoden des Lösen von Sachaufgaben überhaupt erst *bewusst werden* und es lernen, diese *gezielt auszuwählen* und einzusetzen: Ob nun Skizzieren, Tabellieren, systematisches Probieren, das für die Mathematik charakteristische Zurückführen neuer Aufgaben auf bereits bekannte Aufgaben oder Anderes: Erfolgreich angewandte Arbeitsweisen werden als solche gekennzeichnet und

⁸¹ Cox (1999, S. 347), Gallin (2010, S. 8), Hahn und Janott (2012, S. 28), M. Meyer und Prediger (2012, S. 6), Philipp (2013, S. 48), Rathgeb-Schnierer (2005, S. 173), Söbbeke und Steenpaß (2014, S. 13)

immer wieder ausprobiert. Dabei erwerben die Kinder Lösungserfahrungen, die es ihnen erleichtern, neue Aufgaben zu lösen. (Eichler, 2015, S. 56–57, Hervorhebungen durch Autorin)

Hieraus kann abgeleitet werden, dass die Reflexion problemhaltiger Textaufgaben auf Basis externer Repräsentationen erfolgen kann (vgl. Selter, 1994, S. 31). Selbstgenerierte Repräsentationen haben Modellcharakter und verfolgen das Ziel, den mathematischen Erkenntnisgewinn voranzubringen (Dörfler, 2006, S. 212; Steenpaß, 2014, S. 24). Sie ermöglichen es, in der gemeinsamen Kommunikation und Reflexion „unsichtbares“ mathematisches Wissen und identifizierte Strukturen öffentlich und für alle Lernenden anschauungsgestützt zugänglich zu machen.⁸² Wurde eine falsche oder unvollständige Lösung erzielt, so kann die gemeinsame Kommunikation genutzt werden, um mögliche Fehlerquellen oder Fehlinterpretationen aufzuspüren sowie Korrektur- oder Ergänzungsvorschläge zu sammeln und zu diskutieren.⁸³

Alle folgenden, in diesem Abschnitt kursiv gekennzeichneten Textbausteine beziehen sich auf die Hervorhebungen der Autorin im Zitat Eichlers (2015, S. 56–57), mit dem Ziel, den jeweiligen Bezug herauszustellen und mit zusätzlichen Forschungsergebnissen zu untermauern.

Der Einbettung unterschiedlicher externer Repräsentationen im Sinne der Berücksichtigung *unterschiedlicher Lösungswege* wird hierbei eine große Bedeutung zugesprochen.⁸⁴ Der „kreative Wildwuchs“ der Kinder (M. Meyer & Voigt, 2009, S. 14) soll Ausgangspunkt der Reflexion sein, indem an „informelle Lösungsstrategien und selbst erfundene Notationsformen der Schüler“ angeknüpft wird (Selter, 1994, S. 33 in Anlehnung an Streefland, 1990, S. 39). Es lassen sich alternative Denk- und Erkenntniswerkzeuge aufzeigen (Dörfler, 2006, S. 212–213; Krauthausen & Scherer, 2014, S. 250) sowie *bewusst* machen (Söbbeke & Steenpaß, 2014, S. 13) und es lässt sich ein breites Spektrum an Denk- und Herangehensweisen offerieren (Schreiber, 2014, S. 9). Von Letzterem ist auszugehen, weil ein und dieselbe Problemstellung sehr unterschiedliche Eigenproduktionen hervorrufen kann, die alle „Ausdruck wie Folge des [individuellen] Situationsverständnisses“ sind (Franke & Ruwisch, 2010, S. 147; vgl. Kuntze, 2013, S. 20). Werden diese in den Unterricht implementiert und wird über deren Eignung in der konkret vorliegenden Situation nachgedacht und diskutiert, wird

⁸² Duval (1999, S. 3), Söbbeke und Steenpaß (2014, S. 13), Steenpaß (2014, S. 20), E. C. Wittmann (2014, S. 227)

⁸³ Freesemann (2014, S. 114), Freesemann und Breucker (2014, S. 11), Maier (2000, S. 11), Steinbring und Nührenbörger (2010, S. 169)

⁸⁴ Brenner et al. (1997, S. 667), Dienes und Golding (1970, S. 44–45), Guberman und Leikin (2012, S. 37), Jörissen und Schmidt-Thieme (2015, S. 396), Kuntze (2013, S. 20–21), Lengnink (2013, S. 217), Rittle-Johnson und Star (2009, S. 541–542), Verboom (2014a, S. 43)

ein Voneinander-Lernen auf gleicher Stufe möglich.⁸⁵ Ihre Unterschiedlichkeit kann genutzt werden, um den lernförderlichen Charakter der jeweiligen Repräsentationen explizit herauszustellen, Fehlvorstellungen zu überwinden sowie verschiedenartige Zugänge darzubieten, um heterogenen Lerngruppen gerecht zu werden (Dienes & Golding, 1970, S. 44). Das Ignorieren individueller Vorgehensweisen würde „verschenktem Potential“ gleichen (Lengnink, 2013, S. 217).

Diese unterschiedlichen Lösungswege sollen aber keinesfalls für sich nebeneinanderstehend thematisiert werden. Ein derartig isoliertes Abarbeiten, ohne vorhandene Bezüge herzustellen, ist kontraproduktiv und mündet in Unverständnis und fehlender Einsicht (Thies, 2002, S. 86; Verboom, 2014a, S. 43). Stattdessen sollen Gemeinsamkeiten im Klassengespräch akzentuiert, diskutiert und reflektiert werden.⁸⁶ Es soll gemeinsam thematisiert werden, warum die jeweilige Repräsentationsform zur geschilderten Problemstellung passt, aber auch, warum alle Repräsentationsformen zueinanderpassen und wie sie sich übersetzen lassen (Freesemann, 2014, S. 111–115; Prediger, Freesemann, Moser Opitz, & Hußmann, 2013, S. 14; Schukajlow & Blum, 2011, S. 258–259). Es gilt, Argumente auszutauschen, nachzuvollziehen, sowie zu hinterfragen und zu prüfen (KMK 2005b, S. 8; Krauthausen, 2001, S. 103; Hartmann, 2008, S. 152–153).

Das Hauptaugenmerk der Reflexion muss dabei auf transparent werdende Struktureigenschaften und nicht auf Oberflächenmerkmale gerichtet sein (Krauthausen & Scheerer, 2014, S. 250; Söbbeke & Steenpaß, 2014, S. 11).

Die Struktur der Darstellung ermöglicht damit eine Einsicht in eine mathematische Gesetzmäßigkeit; aber sie ist nicht direkt ablesbar oder konkret sinnlich erfahrbar; sie muss aktiv von dem Kind in die Darstellung hineinge deutet werden. Ein Verständnis solch abstrakter Beziehungen und Strukturen entwickelt sich nicht automatisch und spontan, indem Kinder Anschauungsmittel im Unterricht benutzen, es muss ganz explizit erarbeitet werden. (Söbbeke & Steenpaß, 2010, S. 217; vgl. Gysin, 2013, S. 137; Jörissen & Schmidt-Thieme, 2015, S. 396; Lorenz, 2006b, S. 45; Schipper, 2013, S. 297)

Werden Repräsentationen lediglich in den Unterricht implementiert, ohne sie tiefer gehend zu beschreiben, zu charakterisieren und zu analysieren, besteht die Gefahr, dass sie von den Kindern unterschiedlich gedeutet werden und die tatsächliche Struktur nicht gesehen bzw. nicht verstanden wird.⁸⁷ Diese besteht erst recht, wenn ausschließlich die erzielten Lösungszahlen bzw. Ergebnisse in den Blick genommen wer-

⁸⁵ Eichler (2015, S. 57), Lorenz (2005, S. 154), R. Rasch (2001b, S. 72), Rathgeb-Schnierer (2005, S. 166), Selter (1994, S. 26–28), Winter (1983, S. 81–82)

⁸⁶ Dreher (2013, S. 219), Lorenz (2006b, S. 44–45), Schukajlow und Blum (2011, S. 260), Thies (2002, S. 86), Verboom (2014a, S. 43), Winkel (2011, S. 901)

⁸⁷ Brenner et al. (1997, S. 666), Söbbeke (2009, S. 115), Steenpaß (2014, S. 7), Verboom (2014a, S. 43)

den (Steinweg, 2013, S. 82). Prozesse des Deutens, insbesondere des richtigen Deutens müssen langfristig gefördert werden.⁸⁸

Aus diesem Grund wird empfohlen, das Erkennen von Strukturen, Mustern und Beziehungen auf Basis der Repräsentationen *immer wieder* zum Gegenstand der Reflexionen zu machen (Söbbeke & Steenpaß, 2014, S. 11, 13; vgl. Dreher, 2013, S. 219; E. C. Wittmann, 2014, S. 229). Dabei gilt es, neu erworbenes Wissen in Zusammenhang mit dem Umgang externer Repräsentationen in neuen Zusammenhängen anzuwenden und zu erproben, um es sicher, flexibel und Verständnis voranbringend nutzen zu können (Guberman & Leikin, 2012, S. 37; E. C. Wittmann, 1993, S. 394, 1998, S. 157). Mit zunehmenden *Lösungserfahrungen* werden ein Strukturblick (Söbbeke & Steenpaß, 2010, S. 217) und das Aufspüren und Entschlüsseln zugrunde liegender Operationen geschärft (Söbbeke & Steenpaß, 2014, S. 13). Wenn dies gelingt, können auf dieser Basis Vor- und Nachteile einzelner Bearbeitungen dargelegt und verglichen sowie die individuellen Ansätze entsprechend ihrer *Eignung beurteilt* und bewertet werden (Kuntze, 2013, S. 20–21; Lester, 1985, S. 43; Schreiber, 2014, S. 9). Die Lernenden erleben auf der Basis vielfältiger Problemlösungen, wann und warum sich welche Repräsentationsformen eignen, und bauen nach und nach ein individuelles Repräsentationsrepertoire auf, das sie aufgabenbezogen zu einer *gezielten Auswahl* befähigt.⁸⁹ Schwächere Lernende benötigen für solche Abstrahierungen mehr Lösungserfahrungen als schnelle und fitte Lerner (Dienes & Golding, 1970, S. 45; Schipper, 2013, S. 298). Es ist daher nicht abzuweisen, dass sich Lernende in ihrem zur Verfügung stehenden Repertoire beträchtlich unterscheiden können (Cox & Brna, 1995, S. 282).

Ferner wird das Reflektieren über das eigene Denken angeregt, wenn sich aktiv mit den Lösungen Gleichaltriger beschäftigt und auseinandergesetzt wird (Lorenz, 2005, S. 152; Winter, 1983, S. 91). Dem Reflektieren wird daher eine metakognitive Funktion zugesprochen, welches die Lernenden befähigt, ihr eigenes Vorgehen zu optimieren, und langfristig zum Bewältigen von Problemaufgaben verhilft (vgl. Abschnitt 2.4.2).⁹⁰

1.5.5.4 Schlussfolgerungen und Konsequenzen für den Unterricht

Wird berücksichtigt, dass die Förderung von Problemlösekompetenzen einen langfristigen Prozess darstellt, der mit vielfältigen Lösungserfahrungen einhergehen muss, so darf keinesfalls übergegangen werden, dass Grundschulkinder diese Forderung selten

⁸⁸ Dedekind (2012, S. 8), Eichler (2015, S. 57), Franke und Ruwisch (2010, S. 154), Schipper (2013, S. 297), van Essen und Hamaker (1990, S. 306)

⁸⁹ Cox (1999, S. 360), Cox und Brna (1995, S. 274, 282), Fehse (2001, S. 53, 141), Hahn und Janott (2012, S. 28), Kuntze (2013, S. 20), Lester (1985, S. 43)

⁹⁰ Z. B. Bezold (2009, S. 87–88), Eichler (2015, S. 56), Lorenz (2005, S. 152, 154), Schoenfeld (1985, S. 139–140)

erfüllen können. Sie haben zu Beginn ihrer Schulzeit nur gering ausgeprägte Problemlöse- und Repräsentationskompetenzen vorzuweisen, da es ihnen an den geforderten Erfahrungen mangelt. Werden Probleme in den Mathematikunterricht implementiert, so muss dies zwingend mit einer Unterstützung der Lernenden einhergehen.

Ob das Aufgreifen unterschiedlicher Lösungswege dabei von Vorteil ist, sieht Lorenz (2005, S. 157, 2013, S. 9) kritisch. Aus seiner Sicht bringen unterschiedliche Lösungswege eine nicht zu vernachlässigende Gefahr mit sich: Es darf nicht davon ausgegangen werden, dass alle aufgegriffenen Repräsentationen für sich sprechen und direkt von den Lernenden nachvollzogen und verstanden werden.⁹¹ Hinzu kommt, dass das Übertragen von Handlungen von einer auf die andere Repräsentation anspruchsvolle Übersetzungsprozesse erfordert, die (insbesondere schwächere) Lernende herausfordern (Lorenz, 2005, S. 157, 2013, S. 9; Schukajlow & Blum, 2011, S. 259–260). Es darf dabei nicht unberücksichtigt bleiben, dass unterschiedlichen Repräsentationen divergente Handlungen und unterschiedliche kindliche Denkprozesse zugrunde liegen (Kuntze, 2013, S. 20; Lorenz, 2005, S. 152; Selter, 1994, S. 23). Sie müssen als neuer Lernstoff verstanden werden, der erst gelernt werden muss (E. C. Wittmann, 1993, S. 394).

Es sind enorme kognitive Fähigkeiten und Anstrengungen erforderlich, um sich in die vollzogenen Denk- und Lösungsweisen anderer hineinzudenken, sie verstehen zu wollen und vorhandene Strukturen hineinzudeuten (z. B. Lorenz, 2006a, S. 21). Erschwendend kommt hinzu, das Externalisiertes in den seltensten Fällen einer Eins-zu-eins-Übersetzung der Gedankengänge entspricht (Lorenz, 2005, S. 156–157). Vorhandene Lösungsideen werden erst in die „Sprache der Veranschaulichungsmittel“ transformiert (Lorenz, 2005, S. 156), weshalb die Schülerwerke nur eingeschränkt erkennbar werden lassen, was sich die Kinder tatsächlich gedacht haben (Fetzer, 2007, S. 264).

Um diesen Problematiken entgegenzuwirken, muss sich in der Reflexionsphase des Unterrichts verstärkt deren Deutung und Interpretation zugewandt werden, sodass sie die entsprechende Würdigung erfahren (z. B. Eichler, 2015, S. 57; Hahn & Janott, 2011, S. 18; Lorenz & Radatz, 1993, S. 58). Nur wenn es die Eigenproduktionen nachzuvollziehen, zu verstehen, zu analysieren und zu vergleichen gilt, besteht die Möglichkeit, dass sie als Denk- und Erkenntniswerkzeuge aufgenommen und in folgenden Bearbeitungen verwendet werden (siehe Abschnitt 1.5.5.3). Eine Beschränkung auf eine geringe Anzahl zentraler, exemplarischer Eigenproduktien wird empfohlen (Bauer, 1993, S. 79; Thies, 2002, S. 86; Verboom, 2014a, S. 43). Nur wenn es gelingt, in der Kommunikation die repräsentierten Strukturen aufzuzeigen, sodass sie nachvollziehbar

⁹¹ Vgl. Gerster und Schultz (2004, S. 351–352), Lorenz und Radatz (1993, S. 58), Schukajlow und Blum (2011, S. 259), Steinweg (2013, S. 82)

und verstanden werden, kann eine Steigerung der Argumentationskompetenzen erzielt werden (Bezold, 2009, S. 101; Söbbeke & Steinbring, 2007, S. 65).

Darüber hinaus konnte vielfach gezeigt werden, dass die Konstruktion und die Nutzung externer Repräsentationen lösungsunterstützend und lernförderlich sind.⁹² Auch hier darf nicht ausgeklammert werden, dass Grundschulkinder die Konstruktion adäquater, externer Repräsentationen erst erlernen und trainieren müssen, um sie als Denk- und Erkenntniswerkzeuge nutzen zu können.⁹³ Konkret benötigen sie Unterstützung beim Externalisieren ihrer mentalen Modelle. Die Konstruktion und Nutzung mathematischer Repräsentationen sollte im Problemlöseanfangsunterricht, mit dem Ziel, das Verständnis zu fördern, arithmetischen und algebraischen Prozeduren zeitlich vorgeordnet sein (Brenner et al., 1997, S. 664; Zhe, 2012, S. 64). Brenner und Kollegen warnen davor, vorschnell mit dem Rechnen zu beginnen, ohne sichergestellt zu haben, dass das geschilderte Problem verstanden wurde (Brenner et al., 1997, S. 666; vgl. Hahn & Janott, 2011, S. 18).

Es bleibt herauszufinden, ob die Konstruktion und Nutzung externer Repräsentationen beim Lösen problemhaltiger Textaufgaben von Grundschulkindern trainiert werden kann, sodass sowohl Produzenten als auch Rezipienten profitieren können. Bevor sich dieser Fragestellung angenommen wird, wird aufgezeigt, welche externen Repräsentationen im Mathematikunterricht der Grundschule zu erwarten sind und welche Eigenschaften sie charakterisieren.

1.5.6 Formen externer Repräsentationen im Mathematikunterricht

In der Mathematikdidaktik werden häufig auch nach Inhaltenbereichen vielfältige externe Repräsentationen unterschieden und diskutiert. Im Rahmen des Sachrechnens zählen Zeichnungen bzw. Skizzen zu den zentralen Denk- und Erkenntniswerkzeugen (Abschnitt 1.5.6.1). Neben Zeichnungen haben auch Rechnungen (Abschnitt 1.5.6.2) und Tabellen ihren Stellenwert (Abschnitt 1.5.6.3).

Vereinzelt wird dem Erstellen von Listen (Abschnitt 1.5.6.4) und dem Verfassen von schriftsprachlichen Beschreibungen (Abschnitt 1.5.6.5) Relevanz zugesprochen. In der Literatur werden unterschiedliche Bezeichnungen für die letztgenannte Repräsentationsform verwendet: „verbale Darstellungen“ (Bayrhuber, Leuders, Bruder, & Wirtz, 2010, S. 29), „mündliche oder schriftliche verbale Beschreibungen“ (Hahn & Janott, 2011, S. 16), „sprachliche Äußerungen“ (Winkel, 2013, S. 31), „verbale Repräsentati-

⁹² Brenner et al. (1997), Duncker (1935), Goldin und Shteingold (2001), Kilpatrick, Swafford und Findell (2001), Koerber (2000), R. Rasch (2001b, 2009)

⁹³ Cox (1999, S. 360), Dörfler (2006, S. 210), Fehse (2001, S. 50), Prediger (2013a, S. 172), R. Rasch (2001b, S. 47), Renkl und Nückles (2006, S. 143)

onen“ (Brenner et al., 1997, S. 667; Zhe, 2012, S. 64). Alle vereinen die Tatsache, dass neben Zeichnungen, Tabellen, Gleichungen und Graphen auch Repräsentationen in Form von Wort (gesprochen) und Schrift (geschrieben) existieren.

Wenngleich sie verstärkt der Sekundarstufe I zugeordnet werden, zählen Terme und Gleichungen (z. B. Bruder & Collet, 2011b, S. 67–68; Hembree, 1992, S. 256) ebenso wie Graphen und Diagramme (z. B. Brenner et al., 1997, S. 679; Hahn & Janott, 2011, S. 16) zur Gruppe externer Repräsentationen. Graphen und Diagramme werden nachfolgend nicht berücksichtigt, weil sie sich schwerpunktmäßig im Rahmen funktionaler Zusammenhänge in der Sekundarstufe verorten lassen.⁹⁴ Das Lesen und Deuten von Graphen und Diagrammen ist zwar bereits Gegenstand des Mathematikunterrichtes der Grundschule, weniger jedoch deren eigenständige Konstruktion (Stern et al., 2003, S. 194).

Per se kann jedoch nicht gesagt werden, dass es *die* Repräsentationsform gibt, die Einsicht und Erfolg garantiert (Cox, 1999, S. 344). Vielmehr müssen die als Denk- und Erkenntniswerkzeuge dienenden externen Repräsentationen zu dem jeweils zu lösenden Problem und dem Lösenden selbst passen (Cox, 1999, S. 343–344; Jörissen & Schmidt-Thieme, 2015, S. 395; R. Rasch, 2001b, S. 47). Ersteres tritt ein, wenn die Strukturen der Problemstellung adäquat abgebildet und Handlungen aktiviert werden, die einfach und in wenigen Schritten auszuführen sind (vgl. Abschnitt 1.4.2). Letzteres wird vom Repräsentationswissen des Lösenden und dementsprechend von seinem individuell zur Verfügung stehenden Repräsentationsrepertoire, von seinen individuellen Präferenzen und dem als notwendig empfundenen Umfang des Externalisierens bestimmt (Cox, 1999, S. 355–359). Hierunter wird verstanden, dass es Lernende gibt, die keine externe Repräsentation, wiederum andere minimal oder aber voll ausgearbeitete Repräsentationen benötigen. Wenn das Externalisieren als Zwang oder als zusätzliche Last empfunden wird, wirkt es sich negativ auf Motivation und Leistungsbereitschaft aus.

Jede Repräsentationsform weist Vor- und Nachteile auf, wenn die erforderlichen Zusammenhänge darzustellen oder herauszulesen sind (Cox, 1999, S. 354; Jörissen & Schmidt-Thieme, 2015, S. 395; R. Rasch, 2001b, S. 47). Diese können von Produzent zu Produzent, aber auch von Rezipient zu Rezipient unterschiedlich wahrgenommen werden, weshalb Problemlösenovizen an vielfältigen Situationen erfahren und erlernen

⁹⁴ Bayrhuber et al. (2010, S. 28), KMK (2004, S. 8, 11–12), Stern, Aprea und Ebner (2003, S. 194), Vollrath (1989, S. 5)

müssen, wie aufgabenbedingt nach adäquaten Repräsentationsformen gesucht, deren jeweilige Generierung angestrebt und umgesetzt werden kann.⁹⁵

Bevor nachfolgend für das Problemlösen in der Grundschule relevante Repräsentationsformen unter dem Fokus des aktuellen Forschungsstandes erörtert und diskutiert werden, sollen vorab zwei zentrale Merkmale herausgestellt werden: kontrollierendes und systematisches Vorgehen. Gleich, ob eine zeichnerische, eine tabellarische oder eine rechnerische Heran- und Vorgehensweise gewählt wird, gelten sie übergeordnet als zielführend und effektiv. Um Wiederholungen zu vermeiden, werden diese beiden Merkmale jedoch nicht bei jeder einzelnen Repräsentationsform aufgeführt, sondern vorweg erörtert.

Kontrollierendes und systematisches Vorgehen als übergeordnete Repräsentationsmerkmale

Das Anfertigen externer Repräsentationen führt den Lernenden nicht automatisch zum Ziel, sondern vielmehr das Reflektieren darüber (Schipper, 2013, S. 291). Metakognitive Fähigkeiten erweisen sich als tragfähige Einflussgröße für den Lernerfolg (vgl. Abschnitt 4.5.3.4).⁹⁶ Empirische Studien konnten deren Bedeutung für das Problemlösen bestätigen. Lernende, die ihr Ergebnis überwachen und kontrollieren, zählen zu den erfolgreicheren Problemlösern.⁹⁷ Metakognitive Fähigkeiten müssen jedoch erst erworben werden (R. Rasch, 2012b, S. 132; Schütte, 2004, S. 142), weshalb es sich zu überprüfen und aufzugreifen lohnt, inwiefern sie sich in den externen Repräsentationen der Lernenden widerspiegeln. Es ist nützlich, zu „erleben, dass Probieren, Irren und Korrigieren völlig normal beim Lösen von Aufgaben und beim Lernen von Mathematik ist“ (Eichler, 2015, S. 63 Hervorhebung im Original). Dass diese Kompetenzen trainierbar sind, konnte empirisch bereits mehrfach belegt werden.⁹⁸

Darüber hinaus gilt ein systematisches gegenüber unsystematischem, willkürlichem, teilweise auch planlosem Vorgehen als erfolgreiche Vorgehensweise beim Problemlösen, das Barrieren zu überwinden ermöglicht (Käpnick, 1998, S. 254; König, 1992, S. 28; Krägeloh & Prediger, 2015, S. 140). Grundschulkinder hantieren in der Regel eher spontan und auf spielerischer Ebene mit Zahlen, sodass ein systematisches und

⁹⁵ Brenner et al. (1997, S. 684), Eichler (2015, S. 69), Franke und Ruwisch (2010, S. 147, 154), Fricke (1987, S. 64), Jörissen und Schmidt-Thieme (2015, S. 395), R. Rasch (2001b, S. 47), Schipper (2013, S. 297), Wessells (1990, S. 340)

⁹⁶ Z. B. Kramarski, Weisse und Kololshi-Minsker (2010, S. 185–186), Wang, Haertel und Walberg (1993, S. 272, 278, 1997, S. 1), Winkel (2012, S. 9–18)

⁹⁷ Desoete, Roeyers und Buysse (2001, S. 445), Hembree (1992, S. 256), Kramarski et al. (2010, S. 185–186), Lester und Kroll (1990, S. 57), Muir et al. (2008, S. 230)

⁹⁸ Z. B. Desoete (2008, S. 203), Desoete, Roeyers, & De Clercq (2003, S. 195), Efkides, Papadaki, Papantoniu und Kiosseoglou (1997), Kramarski et al. (2010, S. 185)

geschicktes Agieren nur vereinzelt zu beobachten ist (Fuchs, 2006, S. 91; Käpnick, 1998, S. 255). Käpnick (1998, S. 254) und Fuchs (2006, S. 91) ordnen diese Vorgehensweise dem sachbetonten Problemlösen zu, vorzugsweise auf formaler Ebene, das vorrangig leistungsstärkeren Kindern gelingt (vgl. R. Rasch & Schütte, 2008, S. 72). Systematisches Agieren zeichnet sich unter anderem dadurch aus, dass basierend auf einer Regel alle denkbaren Zahlenkombinationen gesucht werden und dabei eine bestimmte Ordnung eingehalten wird (Fuchs, 2006, S. 91). Das Konstanthalten einer Größe oder des Ergebnisses kann als eine solche Regel systematischen Vorgehens verstanden werden (vgl. M. Link, 2012, S. 125, 274; R. Rasch & Schütte, 2008, S. 71; Steinweg, 2013, S. 155–162). Ein ausgeprägter „Zahlenblick“ kann die Kinder dazu befähigen, Beziehungen auf Anhieb zu sehen und zu nutzen bzw. die Zahlen wohl durchdacht zu zerlegen und wieder neu zusammensetzen (Schütte, 2004, S. 143). Um rein schematisches Lösen zu verhindern, gilt es auch beim Umgang mit externen Repräsentationen, den Zahlenblick der Lernenden langfristig zu schulen (vgl. R. Rasch & Schütte, 2008, S. 84).

Die Darstellung der für die Grundschule repräsentativen externen Repräsentationen erfolgt vor dem Hintergrund der späteren Bezugnahme als Aufzählung, ohne eine Ordnung zu suggerieren (vgl. Abschnitt 4.2.3).

1.5.6.1 Zeichnungen

Das Erstellen von Zeichnungen zählt mit zu den wichtigsten heuristischen Hilfsmitteln beim Lösen problemhaltiger Textaufgaben (Bruder & Collet, 2011b, S. 46–56; Eichler, 2015, S. 61; Franke & Ruwisch, 2010, S. 69, 147). Der Heurismus wird von vielen Schulbüchern aufgegriffen und umgesetzt: *Mache dir zuerst ein Bild von der Sache* oder *Erstelle eine Skizze* sind typische Arbeitsaufträge beim Lösen von Textaufgaben.

- Skizzen oder Zeichnungen unterscheiden sich dahin gehend, welcher Zustand konkret dargestellt ist (Eichler, 2015, S. 63; vgl. Bruder & Collet, 2011b, S. 55; König, 1992, S. 35). Wird eine vollständige Rekonstruktion des Lösungsprozesses möglich, weil der Anfangs- und/oder Zielzustand abgebildet ist und sichtbar manipuliert wird? Oder wird die Manipulation gedanklich vollzogen (Steinweg, 2013, S. 29)?
- Wenn geschilderte Informationen zeichnerisch visualisiert werden, können Strukturmerkmale sowie zugrunde liegende Beziehungen sichtbar werden.⁹⁹ Eine Meta-Analyse von Hembree (1992, S. 256–257) ergab, dass die Nutzung voll-

⁹⁹ Eichler (2014, S. 5), Franke und Ruwisch (2010, S. 149), Fricke (1987, S. 65), M. Meyer und Prediger (2012, S. 6), Pantziara, Gagatsis und Elia (2009, S. 40, 43), Schipper (2013, S. 290), Uesaka, Manalo und Ichikawa (2007, S. 328), van Essen und Hamaker (1990, S. 302), (2014a, S. 42)

ständiger, adäquater Zeichnungen den größten Einfluss auf den Lösungserfolg ausübt.

- Der Konstruktionsprozess setzt eine systematische Problemanalyse und Abstraktionsprozesse in Gang, die dem Aufspüren der Problemstruktur sowie dem Planen und Ausarbeiten der Lösung förderlich sein können.¹⁰⁰ Die Zeichnung gibt Aufschluss was der Produzent abstrahiert (Gerhard & Glaser, 2014, S. 19). Dies zeigt sich beispielsweise darin, ob Dinge oder Personen aus dem Sachkontext abstrakt und ohne Kontextbezug oder originalgetreu mit direktem Kontextbezug gezeichnet werden.
- Kinder stellen sich Situationen häufig bildlich vor (Lorenz, 2004, S. 53, 56), was sich darin zeigt, dass sich ihr Zeichnen an dem geschilderten Geschehen des Problems orientiert, aber auch an es erinnert.¹⁰¹ Die im Text dargestellte Situation kommt dann oftmals illustrativ in Kinderzeichnungen zum Ausdruck (Gerhard & Glaser, 2014, S. 19; Ott, 2015, S. 172, 179; R. Rasch, 2012b, S. 131, 2015b, S. 212). Es wird demnach transparent, ob Lernende in der Lage sind, *objektiv Wesentliches* aus dem Text herauszulesen und darzustellen (Eichler, 2014, S. 5).

Gelingt es die im Text geschilderten, für die Lösungsfindung essenziellen Informationen und ihre Relationen adäquat zu veranschaulichen, so fördert dies grundlegendes Verständnis und eine tiefere Informationsverarbeitung.¹⁰² Ein weiterer Vorteil des Zeichnens besteht darin, dass implizit vorhandene Informationen und Beziehungen sichtbar und irrelevante Informationen ausgeblendet werden (Bruder & Collet, 2011b, S. 47; Cox, 1999, S. 353; van Essen & Hamaker, 1990, S. 302).

Zeichnungen gelten empirisch gestützt als Denk- und Erkenntniswerkzeuge, insbesondere dann, wenn sie selbst generiert werden.¹⁰³ Doch die empirische Beweislage ist inkonsistent. Weitere, zahlreiche empirische Studien konnten die den Zeichnungen zugesprochene lösungsunterstützende Funktion nicht bestätigen: Die alleinige Konstruktion von Zeichnungen wird per se noch nicht als Lösungsgarant angesehen.¹⁰⁴

¹⁰⁰ De Bock, Verschaffel, Janssens, van Dooren und Claes (2003, S. 446), Eichler (2008, S. 29), Fricke (1987, S. 69), Hahn und Janott (2012, S. 39), König (1992, S. 35), M. Meyer und Prediger (2012, S. 4, 6), van Essen und Hamaker (1990, S. 302–303)

¹⁰¹ Eichler (2015, S. 58), Fricke (1987, S. 64), Krauthausen und Scherer (2014, S. 117), Ott (2015, S. 179)

¹⁰² Cox (1999, S. 353–354), Eichler (2015, S. 58), Hembree (1992, S. 256–257), Pantziara et al. (2009, S. 43), Renkl und Nückles (2006, S. 135), van Meter und Garner (2005, S. 299–305)

¹⁰³ Exemplarisch wird verwiesen auf: Cox (1999, S. 353), Csíkos, Szitányi und Kelemen (2012, S. 61), Dörfler (2006, S. 214), Hembree (1992, S. 256–257), Renkl und Nückles (2006, S. 145), Stern et al. (2003, S. 201–202), van Essen und Hamaker (1990, S. 309–310, Exp. 2), van Meter und Garner (2005, S. 309–311, 322)

¹⁰⁴ Exemplarisch wird verwiesen auf: De Bock et al. (2003, S. 445–446, 459), Pantziara et al. (2009, S. 55), Uesaka et al. (2007, S. 332), van Essen und Hamaker (1990, S. 305–306, Exp. 1)

Gründe hierfür gehen mit der Tatsache einher, dass der Prozess des visuellen Repräsentierens für Problemlösenovizen mit erheblichen Schwierigkeiten verbunden ist. Zum einen ist der Konstruktionsprozess mit einer enormen kognitiven Belastung verbunden, sodass die zur Verfügung stehenden Kapazitäten hierfür benötigt werden und dem weiteren Lösen des Problems nicht mehr zur Verfügung stehen (Renkl & Nückles, 2006, S. 143). Zum anderen mangelt es ihnen an nötigem Vor- und Erfahrungswissen, um damit vertraut zu sein, was eine adäquate Zeichnung aus- und kennzeichnet. Oft wird von Novizen die Konstruktion auf Oberflächenmerkmale oder ein falsches Verständnis der Problemsituation gestützt, sodass inadäquate Zeichnungen und falsche Ergebnisse die Folge sind (z. B. Gagatsis & Elia, 2004, S. 453).

Es besteht Konsens dahin gehend, dass vor allem Problemlösenovizen ausreichend Gelegenheiten benötigen, um ihre Fähigkeiten im Umgang mit Zeichnungen systematisch entwickeln und erlernen zu können (siehe Abschnitt 1.5.5.3).¹⁰⁵ Van Essen und Hamaker (1990, S. 310) konnten in ihrer Untersuchung zeigen, dass Erst- und Zweitklässler von einem Training, das sie zum Erstellen von Zeichnungen anregte, nicht profitieren konnten, Fünftklässler dagegen schon. Dass Erstere nicht von dem Training profitieren konnten, kann dergestalt verstanden werden, dass eine signifikante Steigerung weder der Anzahl generierter Zeichnungen noch des Lösungserfolgs zu verzeichnen waren. Dies führten sie darauf zurück, dass Zeichnungen erst zu Denk- und Erkenntniswerkzeugen werden, wenn die Lernenden vielfältige Erfahrungen mit ihnen als Medium des Probierens und Tüftelns, aber auch Korrigierens und Revidierens erlebten (vgl. Eichler, 2015, S. 63). Sie müssen sie explizit als Lösungshilfe wahrnehmen, um ihren Einsatz und den dadurch erforderlichen Mehraufwand als gerechtfertigt anzusehen (Eichler, 2008, S. 29–30, 2015, S. 63). In diesem Zusammenhang wird die Bedeutung der Reflexion über die Zeichnungen betont (de Bock, Verschaffel, Janssens, van Dooren, & Claes, 2003, S. 459; Franke & Ruwisch, 2010, S. 147; Schipper, 2013, S. 291).

1.5.6.2 Rechnungen

Das Beschreiben und Erfassen von Zusammenhängen ist auf symbolisch-numerischer Ebene wie z. B. das Aufstellen von Multiplikationstermen aufgrund der fehlenden anschaulichkeit für Grundschulkinder mit einem sehr hohen Abstraktionsanspruch verbunden.¹⁰⁶ Sie stehen vor der Herausforderung, „Sachprobleme in die Sprache der Mathematik [zu] übersetzen, innermathematisch [zu] lösen und diese Lösungen auf die

¹⁰⁵ Csíkos et al. (2012, S. 42), Eichler (2015, S. 61), van Essen und Hameker (1990, S. 305–310), van Meter und Garner (1990, S. 305–310)

¹⁰⁶ Bayrhuber (2010, S. 29), Bruder und Collet (2011b, S. 67), Jörissen und Schmidt-Thieme (2015, S. 386), M. Meyer und Prediger (2012, S. 4), R. Rasch und Schütte (2008, S. 73), Verboom (2014b, S. 5)

Ausgangssituation [zu] beziehen“ (KMK 2005b, S. 8). Trotz des hohen Anspruches wird häufig beobachtet, dass Lernende Problemstellungen direkt auf arithmetischer Ebene zu lösen versuchen (R. Rasch, 2009, S. 69). Weil rechnerische Herangehensweisen als Konsequenz ihrer Entwicklung angesehen werden und demnach zu dem Lösungsrepertoire der Lernenden gehören, müssen sie im Unterricht aufgegriffen und vertieft werden (R. Rasch, 2009, S. 69).

- Wenn ungeübte Grundschulkinder rechnend an Problemstellungen herangehen und versuchen, sie auf arithmetische Weise zu lösen (R. Rasch, 2009, S. 69), dann ist dabei ihr Vorgehen weniger mit Aktivitäten des Verstehens und Repräsentierens verknüpft. Gelingt es ihnen, sich vom Rechnen zu lösen und die Repräsentationsebenen zu wechseln, indem sie z. B. zusätzlich eine Zeichnung oder eine Tabelle erstellen, wird das Verstehen von Zusammenhängen und Erkennen von Strukturen unterstützt.¹⁰⁷ Die Problemstruktur kann anschaulich erschlossen werden und in eine (oder mehrere) Rechnung(-en) überführt werden (Eichler, 2015, S. 60–61; Gerster & Schultz, 2004, S. 388). Ferner gelingt es gut, mit der Hilfe von Rechnungen bereits gefundene Ergebnisse abzuleiten oder zu kontrollieren (Lauter, 1994, S. 6). Ersteres schließt ein, dass beispielsweise Terme in Skizzen oder Zeichnungen hineinzusehen sind (Steinweg, 2013, S. 57–62). Sie können als Rechtfertigung gefundener Strukturen interpretiert werden (Selter, 1994, S. 253).
- Wird angestrebt, eine innermathematische Lösung auf numerischem Wege zu finden, ist es erforderlich, die gegenständlichen Informationen vom Problemkontext zu lösen und in eine oder auch mehrere Rechnungen zu überführen (R. Rasch & Schütte, 2008, S. 72–73; vgl. KMK 2005b, S. 8). Schreiber (2014, S. 8) spricht in diesem Zusammenhang von „Darstellungen mit Zahlen“. Sie gelten als kontextunabhängig und haben den Vorteil, den Rechenprozess kurz und präzise darzustellen (vgl. Bruder & Collet, 2011b, S. 68; Steinweg, 2013, S. 190). Dabei können die Rechenprozesse in mehreren Schritten vollzogen oder geschickt in einer Rechnung vereint werden (R. Rasch, 2009, S. 86).
- Rechnungen eignen sich, um operative Zusammenhänge aufzuspüren (Schreiber, 2014, S. 9; Steinweg, 2005, S. 17). Veränderungen, wie beispielsweise, eine Zahl durch eine andere Zahl auszutauschen, lassen solche sichtbar werden. Dies verkörpert die Auffassung Dörfliers (2006, S. 213), Rechnen sei ein Manipulieren von Zahlzeichen oder Figuren. Hierbei sind ein immerwährender Rückbezug auf

¹⁰⁷ Fuchs (2006, S. 92), Käpnick (1998, S. 254–255), Lorenz (2013, S. 13), R. Rasch und Schütte (2008, S. 74)

den Kontext (z. B. Steinweg, 2013, S. 190) sowie eine Kontrolle des Ergebnisses unverzichtbar.

- Grundschulkinder legen beim Niederschreiben ihrer Rechenprozesse weniger den Fokus auf Exaktheit (R. Rasch, 2009, S. 86; Schütte, 2004, S. 139–140). Dies spiegelt sich oftmals in ihrer geschriebenen mathematischen Sprache wider, wenn Rechenschritte hintereinander hängend notiert werden und einzelne Gedankengänge transparent werden lassen, losgelöst davon, ob die Gleichheitsbeziehung erfüllt ist. Björns Rechnung verdeutlicht dies exemplarisch: $2 \cdot 9 = 18 + 9 = 27$ (R. Rasch, 2001b, S. 201). Das Gleichheitszeichen wird in diesem Zusammenhang als Rechenoperationen bzw. Rechenimpuls und nicht als Symbol einer Beziehung verwendet (vgl. Gerster & Schultz, 2004, S. 361; Steinweg, 2013, S. 97).
- Wie bereits im Rahmen des systematischen Vorgehens in Abschnitt 1.5.6 aufgezeigt, können systematische Veränderungen durch das Konstanthalten einer Größe oder des Ergebnisses von Vorteil sein (vgl. M. Link, 2012, S. 125, 274; R. Rasch & Schütte, 2008, S. 71; Steinweg, 2005, S. 16, 2013, S. 153–162). Ein ausgeprägter „Zahlenblick“ ist dabei Erfolg versprechend (Schütte, 2004, S. 143).
- Untersuchungen konnten zeigen, dass Schulanfänger bevorzugt additive Rechenzügänge nutzten, um Textaufgaben zu lösen (R. Rasch, 2009, S. 75–81; vgl. Selter, 1994, S. 117, 121, 125–128, 132–133; Steinweg, 2013, S. 102). Nahm die Komplexität notwendiger Rechenprozesse zu, so stieg auch die Schwierigkeit, den geschilderten Sachzusammenhang in die mathematisch-symbolische Ebene zu transformieren (R. Rasch, 2009, S. 78). Moser Opitz (2013, S. 203–204) konnte zeigen, dass es insbesondere rechenschwachen Kindern schwerfällt, Problemaufgaben zur Division zu lösen. Vielen gelang es nicht, passende Rechenausdrücke aufzustellen.

Leistungsstärkeren Lernenden fällt es im Gegensatz zu leistungsschwächeren leichter, Zusammenhänge zu erfassen und in eine oder mehrere Rechnungen zu überführen (R. Rasch & Schütte, 2008, S. 73). Sie sind eher fähig, systematisch nach Rechenwegen zu suchen und auf Aufgabenebene abzubilden sowie sich dem Ziel schrittweise anzunähern (R. Rasch & Schütte, 2008, S. 72; vgl. Lorenz & Radatz, 1993, S. 52). Leistungsschwächere tendieren dagegen zur Suche nach Schlüsselwörtern, um die gegebenen Zahlangaben dementsprechend miteinander zu verknüpfen – teilweise aus dem Grund, schnell rechnen zu wollen.¹⁰⁸ Es mangelt ihnen an Verständnis zugrunde liegender

¹⁰⁸ Booker et al. (2014, S. 76), Hegarty et al. (1995, S. 18, 24, 29), Krägeloh und Prediger (2015, S. 140), Lester und Kroll (1990, S. 57), Lorenz und Radatz (1993, S. 143), Muir et al. (2008, S. 231), Rösch und Paetsch (2011, S. 66–67)

Konzepte, was ihnen verwehrt, adäquate Rechnungen aufzustellen.¹⁰⁹ Hinzu kommt, dass mit zunehmender Komplexität der erforderlichen Rechenprozesse auch die Schwierigkeit steigt, den geschilderten Sachzusammenhang auf die mathematisch-symbolische Ebene zu transformieren (Lorenz, 2013, S. 12; R. Rasch, 2009, S. 78). In diesem Falle können Wechsel der Repräsentationsebenen beim Aufbau fehlenden Verständnisses unterstützen.¹¹⁰

Schütte (2004, S. 139) ließ eine weitere Schwierigkeit erkennen: Dass Kinder ihre Rechenwege nicht adäquat oder unvollständig niederschreiben, heißt nicht automatisch, dass sie nicht in der Lage sind, die Aufgabe rechnerisch, also auf der symbolisch-nummerischen Ebene, zu lösen (vgl. R. Rasch, 2009, S. 79). Grundschulkinder teilen nicht all ihre Gedanken explizit mit (Fetzer, 2011, S. 42). Es darf deshalb nicht angenommen werden, dass ihnen begleitendes oder nachträgliches Externalisieren leichtfällt (R. Rasch, 2009, S. 84; Schütte, 2004, S. 139), insbesondere dann nicht, wenn ihnen die Übung darin fehlt. Das Reflektieren und für Dritte nachvollziehbare Darstellen der eigenen Lösungsgedanken stellen anspruchsvolle Prozesse dar. Die Rechenwegnotation lässt daher nur bedingt Rückschlüsse über den wirklich stattgefundenen Rechenprozess zu (Schütte, 2004, S. 140). Hinzu kommt, dass sich das Denken und Konstruieren der Lösung im Kopf von dem Niedergeschriebenen unterscheiden kann, da angenommen wird, „dass sich das Denken im Prozess der Notation noch einmal umstrukturiert“ (Schütte, 2004, S. 140). Viele der Lösungsschritte finden „zwischen den Zeilen“ statt und werden gar nicht erst notiert (Schütte, 2004, S. 137, 140; vgl. Düll, 2009a, S. 28). Nach R. Rasch (2009, S. 86) wird das Rechnen „über Schritte transportiert, die für junge Grundschulkinder nicht zwangsläufig einen Rechensatz erkennen lassen“. Es wird herausgestellt, dass das Notieren von Zwischenschritten sowohl lernförderlich als auch lernhinderlich sein kann (Schütte, 2004, S. 132).

Im Vergleich zum Erstellen von Zeichnungen oder Tabellen wird in Rechenaktivitäten ein enormes Fach- und Strategiewissen benötigt, weshalb Bruder und Collet (2011b, S. 68) den Lösungsunterstützungscharakter von Rechnungen vorrangig in Mathematisierungsprozessen sehen.

1.5.6.3 Tabellen

Beim Lösen von Textaufgaben wird neben den Zeichnungen den Tabellen ein vielfältiges Potenzial zugesprochen (Franke & Ruwisch, 2010, S. 150–154; Fricke, 1987, S. 75–88).

¹⁰⁹ Freesemann (2014, S. 49–50), Scheld und Hahn (2011, S. 32), Stern et al. (2003, S. 192–193), Prediger, Hußmann, Leuders und Barzel (2011, S. 20)

¹¹⁰ Fuchs (2006, S. 92), Käpnick (1998, S. 254–255), Lorenz (2013, S. 13), R. Rasch und Schütte (2008, S. 74)

- Tabellen eröffnen dem Produzenten die Möglichkeit, gegebene Informationen zu sammeln, zu fokussieren, selektiert und übersichtlich darzustellen, zu strukturieren, aber auch zueinander in Beziehung zu setzen und gegenüberzustellen.¹¹¹ Dabei können mit dem Fokus auf die Spalten horizontale Beziehungen zwischen zwei Größen, mit dem Fokus auf die Zeilen vertikale Beziehungen innerhalb der Größen sprachentlastend erkundet und manifestiert werden.¹¹² Funktionale Zusammenhänge lassen sich mithilfe geeigneter Tabellen bereits von Grundschulkindern aufspüren und erschließen (Franke & Ruwisch, 2010, S. 152; Fricke, 1987, S. 84; Wälti, 2009, S. 34–36). Eine anspruchsvolle Denkleistung besteht dann vor allem darin, die Spalten und Zeilen mithilfe der im Text gegebenen Informationen treffend zu beschriften (Bruder & Collet, 2011b, S. 59).
- Ihr hoher Grad an möglicher Strukturiertheit und ihre Übersichtlichkeit in der Darbietung machen Tabellen insbesondere für eine Annäherung an die Lösung in Form einer Auflistung absolvieter Suchprozesse attraktiv.¹¹³ Tabellen können sowohl durch systematisches als auch intuitives Probieren den Weg zu vollständiger Einsicht ebnen (Franke & Ruwisch, 2010, S. 153; Fricke, 1987, S. 79–84). Cooper und Warren (2008, S. 25) bezeichnen sie als abstrakte Repräsentation von Veränderungen. Tabellarische Auflistungen können aber auch als Wegeprotokoll Anwendung finden, indem sich Stufe für Stufe unter Berücksichtigung vorheriger an die Lösung herangetastet wird.¹¹⁴ Eine übersichtliche, strukturierte Darbietung kann dann fortschreitend genutzt werden, um die Lösung direkt abzulesen (z. B. Düll, 2009b, S. 121) oder auch Rechnungen anzubahnen (Bruder & Collet, 2011b, S. 59; Franke & Ruwisch, 2010, S. 151; König, 1992, S. 36).
- Zusätzliche Spalten ermöglichen es, einzelne, vollzogene Schritte zu überwachen und zu kontrollieren (Steinweg, 2013, S. 33–34; vgl. Fricke, 1987, S. 76, 86–87; Düll, 2009b, S. 121–125): Zum einen könnte in einer zusätzlichen Spalte die Anzahl vollzogener Schritte dokumentiert werden, wenn deren Summe von Interesse ist. Zum anderen ist denkbar, die gesuchte Größe (Gesamtanzahl, Differenz etc.) ergänzend zu kontrollieren.
- Bei Tabellen kann ebenso wie bei Rechnungen davon ausgegangen werden, dass systematische Veränderungen gesteuert werden können, indem eine Größe oder das Ergebnis konstant gehalten wird (vgl. M. Link, 2012, S. 125, 274; R. Rasch &

¹¹¹ Bruder und Collet (2011b, S. 56), Düll (2009a, S. 28–29, 2009b, S. 45), Franke und Ruwisch (2010, S. 150, 152–153)

¹¹² Brenner et al. (1997, S. 679), Franke und Ruwisch (2010, S. 152–153), König (1992, S. 35–36), Prediger (2013b, S. 33), Vollrath (1989, S. 5)

¹¹³ Bruder und Collet (2011b, S. 57), Dedekind (2012, S. 18), Franke und Ruwisch (2010, S. 153–154), Fricke (1987, S. 76, 79–84)

¹¹⁴ Düll (2009a, S. 28–29, 2009b, S. 120–121), Franke und Ruwisch (2010, S. 150), Fricke (1987, S. 86–87), Krauthausen und Scherer (2014, S. 117)

Schütte, 2008, S. 71; Steinweg, 2005, S. 16, 2013, S. 153–162). Inhaltliche Überlegungen verhelfen dem Lernenden einerseits, mathematische Zusammenhänge und Beziehungen aufzuspüren, andererseits, sich strukturiert an die Lösung heranzutasten, quasi schrittweise vorzuarbeiten (Fuchs, 2006, S. 91; Käpnick, 1998, S. 252).

Ein wesentlicher Vorteil von Tabellen besteht darin, dass durch die Dokumentation der vollzogenen Schritte oder aller gefundenen Möglichkeiten eine Entlastung des Arbeitsgedächtnisses erfolgt. Die bisher ausgeführten und sich als noch nicht zielführend erwiesenen Lösungsmöglichkeiten müssen nicht im Arbeitsgedächtnis präsent gehalten werden. Vielmehr sind sie sichtbar vor Augen und können jederzeit einbezogen werden, sodass die freien Kapazitäten für die weitere Lösungssuche genutzt werden können.

Bruder und Collet (2011b, S. 61) stellen heraus, dass Tabellen vorrangig planende und weniger intuitive Problemlösende entlastet und stützt. Dies stützen sie auf die Tatsache, dass der Konstruktion von Tabellen sowie der Zuordnung gegebener Größen auf Spalten und Zeilen ein gewisses Maß an Strukturbewusstsein zugrunde liegen muss (Bruder & Collet, 2011b, S. 59; vgl. Düll, 2009b, S. 54). Ihre Generierung und ihre Nutzung entsprechen keinesfalls trivialen Tätigkeiten, sondern erfordern über die Strukturierung des Sachverhaltes hinaus vorausschauende, überblickende, kontrollierende und planerische Fähigkeiten (Bruder & Collet, 2011b, S. 61), gerade dann, wenn es vonnöten ist, abzuwägen, welche Lösungsmöglichkeiten bzw. welche Schritte denkbar und ausführbar sind.

1.5.6.4 Listen

Listen sind als heuristische Hilfsmittel in der fachdidaktischen Literatur nur vereinzelt zu finden (z. B. Düll, 2009b, S. 45–50; Muir et al., 2008, S. 229). Ein Grund hierfür wird in ihrer Nähe zu den Tabellen gesehen.

- Lösungsgedanken und -ideen können in Listen organisiert und systematisch notiert werden (Charles & Lester, 1982, S. 18, 23, 28, 40; Düll, 2009b, S. 46–47; Muir et al., 2008, S. 229).
- Nach Fricke (1987, S. 96–97) kann das Anfertigen einer Liste eine reduzierte, auch nummerierte Darstellung gegebener und gesuchter Informationen sein, die dem Lösenden ein Herantasten an den Lösungsplan ermöglicht.
- In ihrer Funktion als Gedächtnisstütze erinnern sie an Einkaufslisten, Notizzettel oder Datenblätter (Cox & Brna, 1995, S. 242; Düll, 2009b, S. 50).

- Sie finden vorrangig Anwendung, wenn mögliche Kombinationen oder Abläufe zu sammeln, übersichtlich darzustellen bzw. aufzulisten sind (Düll, 2009b, S. 46; Muir et al., 2008, S. 237).
- Die Kontrolle des Ergebnisses soll auch beim Erstellen von Listen nicht außer Acht gelassen werden (Fricke, 1987, S. 98–99).

Ihre Nähe zu Tabellen zeigt sich darin, dass gegebene Informationen horizontal oder vertikal angeordnet werden können (z. B. Düll, 2009b, S. 47; Selter, 1994, S. 268–270). Listen grenzen sich jedoch von Tabellen dahin gehend ab, dass ein Tabellenkreuz kein Bestandteil von Listen darstellt (Muir et al., 2008, S. 235–237; vgl. Selter, 1994, S. 269). Einerseits kann es implizit hineingelesen werden, auch wenn es explizit nicht vorhanden ist (Cox & Brna, 1995, S. 294; Düll, 2009b, S. 47). Andererseits ist es denkbar, dass die aufgelisteten Zwischenschritte oder Zwischenergebnisse im Kopf jeweils einer zweiten Größe zugeordnet, aber nicht explizit in einer zweiten Spalte notiert werden (Selter, 1994, S. 269).

1.5.6.5 Schriftsprachliche Beschreibungen

Problemsituationen können auch auf sprachlicher, beschreibender Ebene repräsentiert werden.¹¹⁵ Darunter wird in Anlehnung an Ruwisch (2014, 43) verstanden, dass Begründungsprozesse im Unterricht immer zuerst versprachlicht werden, sei es in mündlicher, sei es in schriftlicher Form, bevor sie in symbolischen Notationen, den formal-sprachlichen Begründungen, münden. In der vorliegenden Untersuchung stehen die schriftsprachlichen Tätigkeiten im Vordergrund. Sie können einerseits genutzt werden, um sich inhaltliche Klarheit zu verschaffen, indem der geschilderte Sachverhalt in der eigenen Sprache, der „Sprache des Verstehens“ (Wagenschein, 1968, S. 122), oder auch abgekürzt aufgeschrieben wird (Cox, 1999, S. 351; Zhe, 2012, S. 64). Andererseits können sie verwendet werden, um die eigenen mathematischen Entdeckungen, Ideen und Vermutungen zu versprachlichen sowie für sich und andere nachvollziehbar und verständlich zu machen (Brenner et al., 1997, S. 667; Ruwisch, 2014, S. 34; Zhe, 2012, S. 64). Oft treten sie in Kombination mit anderen Repräsentationen auf (Brenner et al., 1997, S. 667; Zhe, 2012, S. 64–65).

Verbales oder schriftliches Versprachlichen des individuellen, mathematischen Prozesses kann das Beschreiben, Hinterfragen und Begründen mathematischer Entdeckungen einschließen.¹¹⁶ In diesen Fällen werden die Deskriptionen der Lernenden zu Argumentationen (vgl. Abschnitt 2.2). Aufgrund argumentativer Anteile lassen sie sich

¹¹⁵ Bayrhuber et al. (2010, S. 28–29), Brenner et al. (1997, S. 667), Cox (1999, S. 345), Hahn und Janott (2011, S. 16), Winkel (2013, S. 31), Zhe (2012, S. 64)

¹¹⁶ Bezold (2009, S. 37–38), Hahn und Janott (2011, S. 16), Prediger (2013a, S. 167), Ruwisch (2014, S. 34), Selter (1994, S. 48)

von reinen Antwortsätze abgrenzen (Hoffman & Powell, 1989, S. 55). Auf schriftsprachlicher Ebene kann zwischen reinen Beschreibungen, Beschreibungen, die begründende Ansätze enthalten, und vollständigen, widerspruchsfreien Begründungen differenziert werden (Neumann et al., 2014, S. 118). Verschiedene Abstraktionsstufen lassen sich auch auf mathematischer Ebene unterscheiden: Zum einen werden mathematische Entdeckungen lediglich benannt oder beschrieben, zum anderen aber auch begründet, was mathematisch verallgemeinernde Aspekte inkludiert (Neumann et al., 2014, S. 25).

Grundlage einer wachsenden Argumentationskompetenz ist unter anderem die Fähigkeit, mathematische Phänomene und Zusammenhänge zunehmend treffender zu beschreiben. Diese Kompetenz entwickelt sich im Prozess des Schreibens.... Das Schreiben und Diskutieren kann seine Wirkkraft auf den Aufbau von Argumentationskompetenzen erst in der Wiederholung entfalten. Ein einmaliger Versuch muss scheitern. (Fetzer, 2009, S. 23, Hervorhebung im Original)

Unter dem Slogan „Schreib Mathe und sprich darüber!“ betont Fetzer (2009, S. 22–23) das „starke Duo: Schreiben und Diskutieren“, dessen Potenzial sie in der regelmäßigen Wiederholung sieht (vgl. Maier, 2000, S. 13). Gezielte Fragestellungen wie beispielsweise „Was fällt dir auf?“, „Warum ist das so?“, „Gilt das immer?“ können schriftliches Argumentieren initiieren und aktivieren (Fetzer, 2011, S. 46; Rehm, 1990, S. 95), garantieren jedoch nicht, dass die Lernenden tatsächlich argumentativ tätig werden (Fetzer, 2009, 22).

1.5.7 Externe Repräsentationen in den Bildungsstandards der Grundschule

Aus Gründen der Übersichtlichkeit und besseren Lesbarkeit werden nachfolgende Zitate aus den Bildungsstandards kursiv abgebildet und es wird auf Anführungszeichen verzichtet. Grammatikalische Anpassungen wurden vorgenommen.

Die Generierung und Nutzung externer Repräsentationen folgen den allgemeinen mathematischen Kompetenzen der Bildungsstandards (KMK 2005b, S. 8): Im Kompetenzbereich des *Darstellens* wird gefordert, *für das Bearbeiten mathematischer Probleme geeignete Darstellungen zu entwickeln, auszuwählen und zu nutzen*. Werden diese anschließend unter Gleichaltrigen oder im Klassengespräch diskutiert und reflektiert, stehen die Lernenden vor der Herausforderung, *eine Darstellung in eine andere zu übertragen sowie Darstellungen miteinander zu vergleichen und zu bewerten*. Diese Forderungen werden insbesondere dann gestellt, wenn sie von Lernenden als Denk- und Erkenntniswerkzeuge genutzt werden sollen (vgl. Abschnitt 1.5.5).

Darüber hinaus verdeutlichen schriftsprachliche Beschreibungen einerseits sowie die Bedeutung externer Repräsentationen für Kommunikations- und Reflexionsphasen im Mathematikunterricht andererseits die Verknüpfung zu zwei weiteren allgemeinen ma-

thematischen Kompetenzen: dem *Kommunizieren* und dem *Argumentieren*¹¹⁷ (KMK 2005b, S. 7–8). Mit der Verbindlichkeit allgemein mathematischer Kompetenzen wird über alle Leitideen hinweg das Versprachlichen mathematischer Zusammenhänge in kommunikativen wie argumentativen Lernsituationen betont und eingefordert (Neumann & Ruwisch, 2015, S. 317). Unter dem Oberbegriff des Argumentierens werden demnach in den Bildungsstandards folgende Aspekte zusammengefasst (vgl. KMK 2005b, S. 8):

- mathematische Aussagen hinterfragen und auf Korrektheit prüfen,
- mathematische Zusammenhänge erkennen und Vermutungen entwickeln,
- Begründungen suchen und nachvollziehen.

Mathematisches Begründen ist nicht als eigenständige prozessbezogene Kompetenz ausgewiesen, vielmehr wird es dem *Argumentieren* untergeordnet. Das Beweisen wird nicht explizit deklariert. Das Beschreiben mathematischer Entdeckungen, welches in einigen Argumentationsmodellen als Voraussetzung des Argumentierens gilt (siehe Abschnitt 2.2), wird in den Bildungsstandards dem *Kommunizieren* untergeordnet. Konkretisiert wird dies in Unterrichtssituationen, in denen die Kinder *eigene Vorgehensweisen beschreiben, Lösungswege anderer verstehen und gemeinsam darüber reflektieren* (KMK 2005b, S. 8). Dass Voneinander- und Miteinander-Lernen maßgeblich eigene Vorgehensweisen voranbringen sowie das eigene Repertoire erweitern können, bleibt unbestritten (Walther et al., 2008, S. 30).

Beschreibende, begründende und argumentative Tätigkeiten lassen sich nicht nur in den Bildungsstandards für das Fach Mathematik, sondern auch in den Kompetenzbereichen *Sprechen und Zuhören* sowie *Schreiben* der Bildungsstandards für das Fach Deutsch wiederfinden (KMK 2005a, S. 7–10): Der Deutschunterricht soll Grundschulkinder befähigen, *funktionsangemessen zu anderen zu sprechen*, indem sie *erzählen, informieren, argumentieren, appellieren*. In Gesprächen sollen sie sowohl *Inhalte zuhörend verstehen und gezielt nachfragen* als auch ihre *Anliegen und Konflikte gemeinsam mit anderen diskutieren und klären*. Darüber hinaus sollen sie in der Lage sein, über ihr Lernen zu sprechen, indem sie *Beobachtungen wiedergeben, Sachverhalte beschreiben, Begründungen und Erklärungen geben, Lernergebnisse präsentieren, über Lernerfahrungen sprechen und andere in ihren Lernprozessen unterstützen*. Das Schreiben von Texten soll *verständlich, strukturiert, adressaten- und funktionsgerecht* erfolgen. Es steht dabei im Vordergrund, *Lernergebnisse geordnet festzuhalten und diese zum Gegenstand von Veröffentlichungen zu machen*.

¹¹⁷ Detaillierte Ausführung, dass das Argumentieren bereits für die Primarstufe von essentieller Bedeutung ist, finden sich neben den Bildungsstandards auch in Brunner (2014a, S. 31–33, 62–63, 2014b, S. 231–232), Peterßen (2012, S. 30–34) sowie Reiss und Ufer (2009, S. 159).

Stellt man die Bildungsstandards der Fächer Mathematik und Deutsch für den Primarbereich vergleichend gegenüber, bleibt festzuhalten, dass beide der Förderung kommunikativer Kompetenzen einen besonderen Stellenwert zusprechen. Darüber hinaus sind losgelöst vom jeweiligen Unterrichtskontext die individuellen mündlichen und schriftlichen Sprachkompetenzen der Kinder zu fördern, sodass sprachlich-kommunikative Fähigkeiten im Mathematikunterricht in den Fokus rücken müssen (Neumann & Ruwisch, 2015, S. 317). Im konstruktiven, aktiv-entdeckenden Unterricht werden diese Fähigkeiten sowohl beim mündlichen als auch beim schriftlichen Versprachlichen der Lösungsgedanken angestrebt und postuliert (Neumann & Ruwisch, 2015, S. 317; Ruwisch, 2014, S. 43).

Zusätzlich sollen die Lernenden befähigt werden, sich mit Äußerungen anderer konstruktiv auseinanderzusetzen. In kommunikativen Settings sollen sie ihren Standpunkt argumentativ vertreten können, indem sie beschreibend, erklärend und begründend agieren, aber auch die Äußerungen ihrer Gesprächspartner aufgreifen, hinterfragen und konstruktiv diskutieren. Die Unterrichtspraxis soll demnach zwei Prozesse wechselseitig miteinander verknüpfen: „Verstehen und Verständnis mitteilen“ (Meissner, 1979, S. 307, vgl. Abschnitt 2.2.5).¹¹⁸

Kommunikative und argumentative Tätigkeiten begleiten und beeinflussen sowohl das Problemlösen als auch das Darstellen von Zusammenhängen (G. Wittmann, 2014, S. 90). Die Vernetzung des Darstellens, Kommunizierens und Argumentierens wird im folgenden Abschnitt aufgezeigt, wenn Prozesse des Argumentierens und ihre Facetten akzentuiert werden.

¹¹⁸ Bauersfeld (2002, S. 11–12), Hußmann und Leuders (2006, S. 84), Neumann et al. (2014, S. 115)



2 Argumentative Tätigkeiten im intersubjektiven Austausch

Um das Lernen auf eigenen Wegen, weg vom Gleichschritt, zu ermöglichen, muss der Unterricht geöffnet werden und stärker den Prozess fokussieren.¹¹⁹ Lernen im Allgemeinen gleicht nach Wittmann (1993, S. 394) einer „konstruktiven Aufbauleistung, die der Lernende ausgehend von seinem Vorwissen in aktiv-entdeckender Auseinandersetzung mit dem Stoff und in sozialer Interaktion zu erbringen hat“. Die Initiierung individueller Lernmöglichkeiten reicht an dieser Stelle nicht aus (Lengnink, 2013, S. 213). Es müssen sowohl regelmäßige Schreibanlässe (vgl. Abschnitt 1.5.6.5)¹²⁰ als auch Austauschphasen integriert werden, um prozessbezogene Kompetenzen fördern zu können (vgl. Abschnitt 1.5.7).¹²¹ Es gilt daher, monologisches und dialogisches Lernen zu vereinen, sodass konstitutive elementare Lernschritte möglich werden (M. Miller, 1986, S. 21, 140).

Unterrichtsprozesse, in denen das mathematische Argumentieren, Kommunizieren und Darstellen gefordert und gefördert werden sollen, müssen demnach in einen Unterricht implementiert werden, der nicht die Instruktion akzentuiert, sondern auf Kommunikation ausgerichtet ist (vgl. Brunner, 2014a, S. 94; M. Meyer & Prediger, 2009, S. 9). Dabei wird unter Kommunikation in Anlehnung an Rathgeb-Schnierer (2005, S. 166) nicht nur der reine Austausch über die Ergebnisse und deren Vergleich verstanden. Kommunikationen sollen nicht auf ein „einheitliches Endprodukt“ abzielen, sondern vielmehr individuelle Vor- und Herangehensweisen zum festen Bestandteil werden lassen (Rathgeb-Schnierer, 2005, S. 168).

Das heterogene Potenzial einer Lerngruppe, wie es in der Grundschule vorzufinden ist, ist geradezu prädestiniert, um ein Voneinander- und Miteinander-Lernen zu ermöglichen.¹²²

Um Mathematik als Prozess zu initiieren und ihm als Produkt gerecht zu werden, bedarf es auch der Phasen des Austauschs. Einerseits müssen die Lernenden individuelle Zugänge entwickeln und auf der Basis ihrer eigenen Erfahrungen konstruieren. Die Eigenproduktionen können sie im geschützten Rahmen mit Partnern austauschen und sich dort gemeinsam weiterentwickeln. (Lengnink, 2013, S. 214)

Es gilt, individuelle, subjektive Denkweisen handlungsorientiert zu wecken, deren Externalisierung anzuregen und zum Gegenstand intersubjektiven Austauschs zu ma-

¹¹⁹ Gallin und Ruf (1993, S. 5–9), Lengnink (2013, S. 212), Lorenz (2004, S. 7–15), Schipper (2013, S. 34), E. C. Wittmann (2003, S. 22–23)

¹²⁰ Fetzer (2009, S. 21), Maier (2000, S. 13), Selter (1994, S. 30–53)

¹²¹ KMK (2005b, S. 8), Lengnink (2013, S. 214), E. C. Wittmann (2003, S. 23), Hartmann (2008, S. 156)

¹²² Hengartner (1992, S. 25–27), Lengnink (2013, S. 214), Freudenthal (1974, S. 166)

chen.¹²³ „Das Reden auf der Grundlage der selbst verfassten Werke sowie das Besprechen und Diskutieren alternativer Vorgehensweisen sind die Aktivitäten, die für den Kompetenzaufbau des Erklärens und Argumentierens den ausschlaggebenden Beitrag leisten“ (Fetzer, 2009, S. 22). Solche Kommunikationsanlässe ermöglichen das Verankern von Wissen, das Anbahnen von Argumentationen sowie das Erzeugen eines subjektiven Beweisbedürfnisses.¹²⁴ Darüber hinaus sei nicht ausgeschlossen, dass Lernende bereits (in Ansätzen) beweisend tätig werden (Rathgeb-Schnierer, 2005, S. 167; Winter, 1983, S. 81–92). Nach Winter (1983, S. 92) ist das Initiiieren von Beweisakten „der einzige effektive Weg, Beweisen lernen zu lassen“.

In diesem Zusammenhang stellt sich die Frage danach, wie sich das Argumentieren vom Beweisen unterscheidet und welche Rolle dabei dem Begründen zugesprochen werden kann. Neben der Klärung ihrer Relevanz für die Grundschule (Abschnitt 2.1) wird eine theoretische Einbettung und Abgrenzung der Begrifflichkeiten erforderlich, welche in Abschnitt 2.2 vorgenommen wird. Der Fokus ist dabei stets auf Tätigkeiten und Prozesse gerichtet, die auf externen Repräsentationen basieren (Abschnitt 2.2.1). Die Frage danach, wann das Beweisen konkret beginnt (Abschnitt 2.2.2) und welche Formen – experimentelles Beweisen (Abschnitt 2.2.3) und inhaltlich-anschauliches Beweisen (Abschnitt 2.2.4) – das Beweisen annehmen kann, wird kritisch erörtert und von formal-deduktiven Beweisen abgegrenzt. In Abschnitt 2.2.5 wird die Rolle des Begründens und wie es in Zusammenhang zum Argumentieren und Beweisen steht, diskutiert. Basierend auf dem theoretischen Hintergrund wird das der Untersuchung zugrunde gelegte Begriffsverständnis abgeleitet (Abschnitt 2.3) sowie das Potenzial intersubjektiven Austauschs (Abschnitt 2.4) im Mathematikunterricht dargelegt.

2.1 Relevanz argumentativer Tätigkeiten in der Grundschule

Auch wenn mathematisches Argumentieren und Begründen den höchsten Kompetenzstufen zugeordnet werden und Schulanfängern einen hohen Anspruch abverlangen, sind diese Kompetenzen bereits in der Primarstufe anzubahnen und gezielt zu fördern.¹²⁵ Winter (1975, S. 109) propagiert: „Vom ersten Schultag an müssen die Wörter ‚warum‘, ‚wieso‘, ‚denn‘, ‚also‘, ‚wenn-dann‘, ‚weil‘, ‚obwohl‘ usw. eine hervorragende Rolle spielen“. Die Grundschulzeit ist nach Krauthausen (2001, S. 100) gerade-

¹²³ Siehe auch Gallin (2010, S. 4–9), Hengartner (1992, S. 22), Hirt, Wälti und Wollring (2010, S. 13–14), KMK (2005b, S. 6–8), M. Meyer (2007a, S. 29), Rathgeb-Schnierer (2005, S. 166)

¹²⁴ Vgl. Bezold (2009, S. 37), Bruder und Weigand (2005, S. 8), Selter (1994, S. 66–67), Winter (1983, S. 84), E. C. Wittmann (2003, S. 30)

¹²⁵ Vgl. Bezold (2009, S. 59), Fetzer (2011, S. 28), Jahnke und Ufer (2015, S. 349), KMK & IQB (2013, S. 13), KMK (2005b, S. 7–8), Krelle (2014, S. 25–26), Rehm (1990, S. 95), Walther, Selter und Neubrand (2008, S. 16–41), Winter (1983, S. 91), E. C. Wittmann (2014, S. 214)

zu prädestiniert, um solche *Erstbegegnungen* zu schaffen, da Kinder dieses Alters mit großer Neugierde und Unbefangenheit den Anforderungen gegenübertreten.

Wittmann (2014, S. 214–215) betont ebenso, Bezug nehmend auf Vygotskij, dass (operatives) Beweisen (siehe Abschnitt 2.2.2–2.2.4) ab der ersten Jahrgangsstufe möglich ist, wenn die Anforderungen in der Zone der nächsten Entwicklung liegen und Arbeitsmittel oder Repräsentationen verwendet werden, die der Altersgruppe vertraut sind (vgl. Krelle, 2014, S. 25). Empirische Studien¹²⁶ konnten zeigen, dass Grundschulkinder durchaus in der Lage sind, argumentativ tätig zu werden, wenn Begründungssituationen zum festen Bestandteil des Mathematikunterrichts werden. Untersuchungen, die sich Argumentationsprozessen im vorschulischen Bereich widmeten, konnten analoge Ergebnisse nachweisen (Krelle, 2014, S. 25; vgl. Nussbaumer, 1995, S. 33; Kline, 1998, S. 368). „In sum, within this age group children display a basic competence in argumentation, in their ability to engage in reason-giving, produce coherent and appropriate responses to others' opposition, and create varied responses that either extend or resolve the dispute“ (Kline, 1998, S. 368).

Den Ergebnissen nationaler und internationaler Studien zufolge muss das Argumentieren in den Grundschulen stärker verankert werden (Prenzel et al., 2007, S. 30; Reiss, Hellmich, & Thomas, 2002, S. 51).¹²⁷ Nicht nur Schulanfänger zeigen Schwierigkeiten beim Bearbeiten von Aufgaben, die argumentative Tätigkeiten inkludierten.¹²⁸ In ihrer qualitativen Interviewstudie „Kinder auf dem Weg zum flexiblen Rechnen“ konnte Rathgeb-Schnierer (2006, S. 281, 288) beobachten, dass den Lernenden sowohl das Versprachlichen als auch das Verschriftlichen ihrer Rechenwege Probleme bereitete, wobei Letzteres eindeutig als schwieriger identifiziert wurde (vgl. R. Rasch, 2009, S. 84). Es wird eine veränderte Unterrichtskultur gefordert, die mathematisches Verständnis und Argumentieren hervorhebt und der fehlenden Erfahrung und Übung seitens der Lernenden entgegenwirkt (z. B. Prenzel et al., 2007, S. 30; Rathgeb-Schnierer, 2006, S. 298–299).

¹²⁶ Vgl. Bezold (2009), Fetzer (2011), Neumann et al. (2014), Rathgeb-Schnierer (2006), Stein (1999)

¹²⁷ Für eine detaillierte Ableitung der Fördernotwendigkeit des Argumentierens aus den Ergebnissen internationaler und nationaler Studien wird auf Bezold (2009, S. 40–57) sowie Brunner (2014a, S. 33–35) verwiesen. Während Bezold ausführlich die Schlüsse aufzeigt, die sich aus TIMSS, PISA, IGLU, IGLU-E, den Orientierungs- und Vergleichsarbeiten VERA für das Argumentieren ergeben, berücksichtigt Brunner ausschließlich die TIMSS und PISA-Studie.

¹²⁸ Bezold (2009, S. 39), Heinze und Reiss (2004, S. 241), R. Rasch (2009, S. 74), Rathgeb-Schnierer (2006, S. 294)

2.2 Argumentieren – Begründen – Beweisen

Es besteht in der mathematischen Community zwar Einigkeit darüber, dass argumentative Aktivitäten fester Bestandteil des Mathematikunterrichts sein müssen, jedoch werden die Begriffe Argumentieren, Begründen und Beweisen nicht trennscharf verwendet (Brunner, 2014a, S. 27, 2014b, S. 230; Fetzer, 2011, S. 28; Reiss & Ufer, 2009, S. 157). Erschwerend kommt hinzu, dass das Verhältnis zwischen Argumentieren und Begründen und vor allem deren Beziehung zum Beweisen kontrovers interpretiert werden (Brunner, 2014a, S. 29).

Schwerpunktmaßig wird die Beziehung des Argumentierens zum Beweisen zwiespältig diskutiert (Brunner, 2014b, S. 231), weshalb zunächst deren Abgrenzung theoretisch erfolgt. Einerseits werden in der Literatur die Begriffe Argumentieren und Beweisen nahezu synonym verwendet bzw. es wird deren (enge) Verbindung betont.¹²⁹ Andererseits werden beide Aktivitäten als grundlegend voneinander abweichende Tätigkeiten angesehen:

We do consider that argumentation and mathematical proof are not of the same nature: The aim of the argumentation is to obtain the agreement of the partner in the interaction, but not in the first place to establish the truth of some statement. As a social behavior it is an open process, in other words it allows the use of any kind of means; whereas, for mathematical proofs, we have to fit the requirement for the use of some knowledge taken from a common body of knowledge on which people (mathematicians) agree. As outcomes of argumentation, problem's solutions are proposed by nothing is ever definitive. (Balacheff, 1991, S. 188–189, Bezug nehmend auf Perelman, 1970, S. 41; Originalquelle konnte nicht recherchiert werden).

Es wird auch die Meinung geteilt, dass sozialer Austausch und explizit das Argumentieren das Beweisen behindern können:

Our point is that in some circumstances social interaction might become an obstacle, when students are eager to succeed, or when they are not able to coordinate their different points of view, or when they are not able to overcome their conflict on a scientific basis. (Balacheff, 1991, S. 188; vgl. Duval, 1991, S. 233)

Mit der Absicht, das Argumentieren im Mathematikunterricht zu positionieren, werden nachfolgend anhand ausgewählter relevanter Positionen die unterschiedlichen Sichtweisen aufgezeigt, ohne Anspruch auf Vollständigkeit zu erheben.

2.2.1 Argumentieren auf Basis externer Repräsentationen

Vollrath (1980, S. 28) akzentuiert eine enge Verbindung zwischen dem Argumentieren und dem Begründen: „Man benutzt diesen Terminus [argumentieren] meist im Sinne von ‚begründen‘.“ Aus seiner Sicht kann dem Argumentieren ein breites Spektrum an

¹²⁹ Vgl. Brunner (2014a, S. 29), Krummheuer (2003a, S. 247), Krummheuer und Fetzer (2005, S. 29), Winter (1983, S. 91)

Aktivitäten zugeordnet werden, die sich nicht auf die „mathematisch eingeengte Form des Beweisens“ begrenzen lassen.

Dies hat zur Folge, dass unter Einbezug externer Repräsentationen auf unterschiedlichen Niveaustufen argumentiert werden kann. Nach Vollrath (1980, S. 34–35, 39) zeigt sich das Spektrum vorrangig in Argumentationen der Lernenden, die auf Einsichten beruhen. Diese Argumentationen können sich auf externe Repräsentationen stützen, wie es auch bei Argumentationen, die auf Erfahrungen und konkreten Beobachtungen beruhen, eintritt. Verweise auf Einsicht unterscheiden sich jedoch dahin gehend, dass mit den externen Repräsentationen operiert wird und resultierende Wirkungen ablesbar werden. Werden sie aus anschaulichen Objekten bzw. externen Repräsentationen gewonnen, die auf andere Fälle übertragbare Strategien aufzeigen, so können diese durchaus auch als Beweise akzeptiert werden (Vollrath, 1980, S. 29, in Anlehnung an Griesel, 1973, S. 151).

Krummheuers Argumentationsverständnis kann in ähnliche Richtung interpretiert werden: Das Argumentieren beschränkt sich nicht kategorisch auf mathematisches Beweisen (Krummheuer & Brandt, 2001, S. 19). In Anlehnung an Toulmin (2008, S. 116) unterscheidet Krummheuer (2003b, S. 124–126) zwischen analytischem und substantiellem Argumentieren.¹³⁰ Ersterem wird mathematischem Beweisen im streng deduktiven Sinn zugeordnet. Letzteres subsumiert Argumentationen, bei denen immer noch eine gewisse Unsicherheit bleibt, ob die gezogene Schlussfolgerung zulässig ist. Die Aussage „Hier gibt es dann immer noch etwas zu argumentieren“ (Krummheuer, 2003b, S. 125) kann derart interpretiert werden, dass dem substantiellen gegenüber dem analytischen Argumentieren kein Vollständigkeitsanspruch in der Argumentationskette zugrunde liegt (vgl. Brunner, 2014a, S. 7). Es handelt sich vielmehr um Argumentationen, die beispielsweise auf ikonischen Repräsentationen oder Objekten der Realität beruhen, und (nur) mit hoher Wahrscheinlichkeit das Ergebnis bzw. die Aussage bestätigen, jedoch deren Wahrheit nicht garantieren (Krummheuer, 2003b, S. 125).

2.2.2 „Wann fängt das Beweisen an?“

Wenn man sich dem Konsens anschließt, dass Argumentieren mit externen Repräsentationen eine (Vor-)Form des Beweisens darstellt (siehe Abschnitt 2.2.1), kommt zweifelsohne die Frage auf, *wann* das Beweisen tatsächlich beginnt und *ob* Grundschulkinder diesen Anspruch zu erfüllen fähig sind.

¹³⁰ Für eine detaillierte Darstellung und Diskussion der Struktur von Argumentationen aus Sicht der Toulmins wird neben Toulmin (1996, 2008) auf Fetzer (2007, 2011), Krummheuer (2003a, 2003b) und Schwarzkopf (2000) verwiesen.

Für Freudenthal (1979, S. 187) beginnt es „jedenfalls, ehe es einen Namen hat“. Er betont, dass dem Beweisen das Konstruieren, sei es geometrisches, arithmetisches oder mentales, vorausgeht (Freudenthal, 1979, S. 190). Er unterstützt und bekräftigt die Sichtweise der genetisch orientierten Mathematikdidaktik, die dem experimentellen Erkunden bereits ein gewisses Maß an Strenge und somit Beweischarakter zuspricht (Freudenthal, 1973a, S. 141, 1979, S. 200). Das Reflektieren verhilft dabei, Konstruktionen in einen „regelrechten“ Beweis zu überführen, es soll „Attitüde“ werden (Freudenthal, 1979, S. 194, 198):

Es fängt bei jeder neuen Aufgabe mit Probieren an, mit versuchsweisen Konstruktionen, und hat man es dann irgendwie gefunden, so reflektiert man über den Weg dahin, man schneidet Umwege ab, man eliminiert Überflüssiges, vollzieht Sichtwechsel, ordnet um, bezieht das Neue auf eher Gefundenes, spezialisiert und generalisiert. Und all das ist eine Tätigkeit, die man nicht nur ausübt, sondern über die man auch wieder reflektieren kann, soll. (Freudenthal, 1979, S. 198–199)

Beweisen kann nach Freudenthal mit geeignetem Material, genügender Selbsttätigkeit und ausreichender Übung erlernt werden: „Wie das Einsplus eins und Einmaleins soll der Schüler auch das Beweisen von neuem erfinden“ (Freudenthal, 1979, S. 197). Demzufolge beweisen Kinder ihre individuellen Entdeckungen ganz unterschiedlich, sodass Beweise diverser Niveaus resultieren. Diese Beweise sind wiederum Grundlage, um plausibel und nachvollziehbar zu erklären, warum eine bestimmte Aussage zutrifft (Hanna, 1989a, S. 21–22, 1989b, S. 49; vgl. Krumsdorf, 2009, S. 10–11). Hanna (1989b, S. 47, 50), die in diesem Zusammenhang zwischen „proofs that prove and proofs that explain“ unterscheidet, plädiert im Mathematikunterricht für eine stärkere Fokussierung von solch erklärenden Beweisen, da sie nicht nur zeigen, *dass* eine Aussage richtig ist, sondern auch, *warum* sie richtig ist.

Es kann hervorgehoben werden, dass Lernende, die ihr Vorgehen erklären, argumentieren, nachweisen oder begründen, schon beweisend tätig sind (Krauthausen, 2001, S. 99–101). Wenn man den Beweisbegriff ausschließlich auf das Prinzip der Deduktion reduziert, indem Behauptungen aus mathematischen Gesetzen oder Definitionen abgeleitet werden (Bezold, 2009, S. 74; Blum, Drüke-Noe, Hartung, & Köller, 2006, S. 36; Philipp, 2013, S. 8), dann besteht Konsens, dass nicht das gesamte Beweisspektrum erfasst wird (z. B. Blum & Kirsch, 1989, S. 208; Freudenthal, 1979, S. 200; Winter, 1983, S. 70; E. C. Wittmann & Müller, 1988, S. 254).

Im Kern ist „das Ideal des ‚absolut strengen‘ Beweises nicht aufrechtzuerhalten“ (E. C. Wittmann & Müller, 1988, S. 250).¹³¹ Vielmehr müssen im schulischen Kontext Beweistypen berücksichtigt werden, die Repräsentationen – „am richtigen Ort“ (Winter,

¹³¹ Vgl. Branford (1913, S. 107), Jahnke und Ufer (2015, S. 332), Winter (1975, S. 109)

1975, S. 109) – stärker in den Mittelpunkt rücken (vgl. Abschnitt 2.2.1–2.2.4).¹³² Gera de in der Schulmathematik gibt es zahlreiche Beweise, die sich auf praktische Handlungen oder Repräsentationen im Realkontext stützen (Kirsch, 1979, S. 261; Winter, 1983, S. 70). Sie können Denkprozesse initiieren, die grundlegend sind, um beweisend aktiv zu werden (Holland, 2007, S. 132, 134, 196–198; M. Meyer & Voigt, 2009, S. 14).

Branford (1913¹³³, S. 100) bringt erstmals das Beweisen mit Anschauungen in Zusammenhang und unterscheidet drei Beweisstufen: die experimentelle, die intuitiv-anschauliche¹³⁴ und die rein logische. Er konkretisiert beispielgebunden, dass aus einem Experiment nur eine hinreichende, jedoch keine absolute Genauigkeit resultiert, und warnt vor fehlerhaften Schlussfolgerungen bezüglich des zugrunde liegenden Wahrheitsgehaltes (Branford, 1913, S. 102; vgl. Abschnitt 2.2.3). Experimentelles Beweisen grenzt er bewusst von den beiden anderen Beweistypen ab, da es sich streng genommen nicht um *echte* Beweise handelt (Branford, 1913, S. 100–108; vgl. Blum & Kirsch, 1991, S. 184). Dem intuitiv-anschaulichen Beweisen spricht er für die Entwicklung mathematischen Verständnisses eine hohe Bedeutung zu, weil aus ihm „allgemeine und streng gültige Wahrheiten“ resultieren können, die, „wenn es nötig wird, auf Postulate[n] der sinnlichen Erfahrung“ beruhen (Branford, 1913, S. 103). Wissenschaftliches (rein logisches) und auf Intuition beruhendes Beweisen unterscheiden sich aus seiner Sicht, im Gegensatz zum experimentellen Beweisen, lediglich in seiner jeweils zugrunde liegenden mathematischen Strenge, denn die „gewonnenen Wahrheiten sind allgemeingültig“ (Branford, 1913, S. 108–109; vgl. E. C. Wittmann & Müller, 1988, S. 248–249).

Die operativen Beweise, welche erstmals in dieser Form bei Semadeni (1974¹³⁵, 1984) vorzufinden sind, griffen die branfordsche Typisierung auf und schärften die Abgrenzungen zwischen experimentellem und inhaltlich-anschaulichem Beweisen (E. C. Wittmann & Müller, 1988, S. 249–255). Auch wenn die dreistufige Typisierung heute

¹³² Z. B. Blum und Kirsch (1991, S. 200), Branford (1913, S. 107), Jahnke (1978, S. 211), Krauthausen (2001, S. 103), Walsch (2000, S. 5), Winter (1983, S. 67), E. C. Wittmann (2014, S. 213), E. C. Wittmann und Müller (1988, S. 254)

¹³³ Branford veröffentlichte erstmals 1908 seine Beweistypisierungen (Branford, 1913, S. 100; s. a. Blum und Kirsch, 1991, S. 202). Diese Quelle konnte nicht recherchiert werden, weshalb sich die Ausführungen auf die deutsche Übersetzung von Schimmack und Weinreich aus dem Jahre 1913 stützen.

¹³⁴ Heute wird anstelle des intuitiv-anschaulichen Beweisens vielmehr vom inhaltlich-anschaulichen Beweisen gesprochen (z. B. Wittmann & Müller, 1988), wobei sich hiermit trotzdem auf die Typisierung von Branford (1913) bezogen wird. Aus Gründen der Orientierung wird nachfolgend der heute geläufigeren und bekannteren Bezeichnung gefolgt.

¹³⁵ Die Originalquelle (1974) konnte nicht recherchiert werden. Sie wird vielfach zitiert (z. B. Kirsch, 1979, S. 261; Walther, 1984, S. 10; Winter, 1983, S. 94; E. C. Wittmann & Müller, 1988, S. 249) und von dem Autor selbst in einer späteren Publikation aufgegriffen (Semadeni, 1984, S. 32).

ehler mit Wittmann und Müller als mit Branford oder Semadeni (z. B. Blum & Kirsch, 1991, S. 189; E. C. Wittmann & Müller, 1988, S. 248–249) assoziiert wird, wird sie immer noch verfolgt: 1) experimentelles Beweisen, 2) inhaltlich-anschauliches Beweisen und 3) formal-deduktives Beweisen. Die einzelnen Typen werden nicht als isoliert nebeneinanderstehend, sondern vielmehr als Stufen betrachtet, welche es ermöglichen, die Entwicklung des Wissens abzubilden (Branford, 1913, S. 240, 270; Brunner, 2014a, S. 20).

Aufgrund ihrer Relevanz für die zugrunde liegende Untersuchung werden im Folgenden das experimentelle Beweisen und das inhaltlich-anschauliche Beweisen theoretisch fundiert aufgezeigt und vom formal-deduktiven Beweisen abgegrenzt. Allerdings zielt diese theoretische Einbettung nicht darauf, ausschließlich an den Beweistypen nach Branford (1913) bzw. Wittmann und Müller (1988) festzuhalten. Es lassen sich in der Literatur weitere Ansätze klassifizieren, die im Kern, abgesehen von ihrem zugrunde liegenden Präzisierungsgrad, nicht von den nachfolgend vorgestellten Beweistypen abweichen (Brunner, 2014a, S. 20–21). Aufgrund dessen und aufgrund der Tatsache, dass die Beweistypen heute noch Relevanz haben, sollen sie mögliche Abstraktionsstufen exemplarisch aufzeigen, die neben dem formal-deduktiven Beweisen existieren und für die Grundschule von essenzieller Bedeutung sein können. Angesichts der Tatsache, dass formal-deduktive Beweise in der Grundschule nicht zu erwarten sind, werden sie im Folgenden nicht eigenständig theoretisch aufgezeigt.

2.2.3 Argumentieren als experimentelles Beweisen

Basiert das Beweisen im Unterricht nicht auf logischen Schlussfolgerungen, sondern vielmehr auf der „Verifikation oder Falsifikation einer endlichen Zahl von Beispielen“, so handelt es sich um experimentelles Beweisen (Brunner, 2014a, S. 17; E. C. Wittmann & Müller, 1988, S. 249; vgl. Krauthausen, 2001, S. 107). Es gilt per se nicht als unmathematisch, wenn gegebene Sachverhalte experimentell erschlossen werden (Freudenthal, 1973a, S. 141; Jahnke & Ufer, 2015, S. 341; Philipp, 2013, S. 2), da die Lösenden an Beispielen zugrunde liegende Muster entdecken und Hypothesen formulieren, die an weiteren Beispielen zu überprüfen sind (Krauthausen, 2001, S. 107; Thies, 2002, S. 100). Pólya unterscheidet diesbezüglich zwischen „suggestiven Beobachtungen“ (1962, S. 22), wenn anhand von Beispielen Hypothesen generiert werden, und „stützenden Beobachtungen“ (1962, S. 24), wenn zusätzliche Beispiele herangezogen werden, um die Hypothesen zu überprüfen. Ersteres fasst Peirce (1932, S. 153) unter *Abduktion*, Letzteres unter *Induktion* zusammen. Werden erste Muster entdeckt bzw. vermutet, so ist zu erwarten, dass weitere Beispiele systematisch ausgewählt werden und kein naives Ausprobieren erfolgt (Brunner, 2014a, S. 18; Krauthau-

sen, 2001, S. 107). Eine gezielte, systematische Auswahl von Beispielen basiert größtenteils auf deduktiven Schlüssen (Philipp, 2013, S. 93).

Wenn das Generieren und Prüfen von Hypothesen auf der Basis von Beispielen erfolgt und im peirceschen Sinne (1932, S. 375) den Prozess der Abduktion und Induktion vereint, dann spricht Philipp (2013, S. 2) von innermathematischem Experimentieren, das neues mathematisches Wissen entstehen lässt. Bezug nehmend auf Peirce sieht Philipp (2013, S. 10) den „Anfang jeder Erkenntnis“ in der Abduktion, die im Gegensatz zur Induktion und Deduktion erkenntnisserzeugend ist. Beispiele fungieren in diesem Zusammenhang als „empirische Daten“, „Träger mathematischer Phänomene“ und somit als Basis des Experimentierens (Philipp, 2013, S. 2, 89; vgl. Bezold, 2009, S. 38–39; M. Meyer, 2007a, S. 28–30). Sie sind grundlegend für den Erwerb von Fähigkeiten der bewussten Auseinandersetzung, wenn sie vom Lernenden selbst konstruiert und nicht von der Lehrkraft vorgegeben werden.¹³⁶ Erstere können Erkundungsprozesse in zweierlei Richtungen unterstützen und lenken: Fehlt den Lernenden ein erster, spontaner Lösungsgedanke, so können sie zunächst mithilfe eigener Beispiele bzw. Repräsentationen den Problemraum erkunden. Liegt ihnen bereits eine Lösungsidee vor, dann können sie mithilfe eigener, konkreter Beispiele den Problemraum eingrenzen oder ihre Vermutungen ersten Prüfungen unterziehen (siehe Abschnitt 1.3.3.2). Gemein haben beide Richtungen, dass die Lernenden in der explorativen, experimentellen Phase ihre (Denk-) Handlungen auf eigene Konstruktionen stützen können, welche im nächsten Schritt strukturiert und geordnet werden müssen (Freudenthal, 1973b, S. 558–565; Jahnke & Ufer, 2015, S. 341; Philipp, 2013, S. 89–90). Ideal-erweise identifizieren sie Zusammenhänge oder Beziehungen, die sie in Hypothesen umwandeln und erneut überprüfen.¹³⁷

Es ist zu erwarten, dass eigene, experimentell gewonnene Entdeckungen und das Aufstellen von Hypothesen erste Argumentationen und Beweisideen wecken und das Beweisen somit weniger einer „aufgesetzten Tätigkeit“ gleicht (M. Meyer & Voigt, 2009, S. 14).¹³⁸ Wenngleich experimentellen Zugängen keine allgemeine Gültigkeit zugrunde liegt, was zur Folge hat, dass sie nicht als mathematische Beweise akzeptiert werden (z. B. Brunner, 2014a, S. 17–18; Holland, 2007, S. 132; E. C. Wittmann & Müller, 1988, S. 249), bleibt jedoch anzumerken, dass sie jüngeren und vor allem lernschwächeren Kindern, dank der nicht im Vordergrund stehenden formalen Kriterien, die Möglichkeit bieten, bestimmte Aufgabengebiete explorativ zu erschließen (Brunner,

¹³⁶ Freudenthal (1973b, S. 389, 569), Berk (2011, S. 310), A. Meyer (2015, S. 16), Steinheider (2014, S. 159), Steinweg (2001, S. 24)

¹³⁷ Für einen vertiefenden Einblick in innermathematisches Experimentieren und dessen Verbindung zum experimentellen Beweisen wird auf Philipp (2013) verwiesen.

¹³⁸ Vgl. Holland (2007, S. 132–134), Krauthausen (2001, S. 104), Krumsdorf (2009, S. 9), Schwarzkopf (2000, S. 14), Winter (1983, S. 78)

2014a, S. 17–18; Holland, 2007, S. 132). Branford (1913, S. 239) bezeichnet die Intuition und das Experiment als „Werkzeuge der Entdeckung par excellence“. Mithilfe beider Werkzeuge gelingt es, Entdeckungen aufzuspüren, welche eher verborgen blieben, wenn ausschließlich der streng deduktive Weg fokussiert würde (Branford, 1913, S. 240).

Ein weiterer Vorteil besteht darin, dass aus anschaulich-empirischen Aktivitäten, wie das beispielbasierte Begründen und Schließen, im nächsten Schritt sprachlich-symbolische Verallgemeinerungen bis hin zu deduktiven Argumentationen resultieren können.¹³⁹ Experimentelle „Beweise“ können demzufolge in Beweise höherer Stufe münden, wenn es gelingt, sich von den herangezogenen Beispielen zu lösen sowie zugrunde liegende Beziehungen aufzuspüren und zu identifizieren. Lernumgebungen, die das Entdecken mathematischer Phänomene begünstigen, stellen natürliche Differenzierungen dar, die Argumentationen unterschiedlicher Niveaustufen zulassen.¹⁴⁰ Jeder Lernende kann entsprechend seiner Voraussetzungen seine eigenen Funde nutzen, um argumentativ oder beweisend tätig zu werden. Mathematisches Argumentieren kann hier als „ergebnisoffen“ bezeichnet werden, wenn es ein mathematisches Phänomen zu entdecken, zu prüfen, abzusichern und zu begründen gilt (Reiss & Ufer, 2009, S. 157).

2.2.4 Argumentieren als inhaltlich-anschauliches Beweisen¹⁴¹

Das inhaltlich-anschauliche Beweisen unterscheidet sich vom experimentellen Beweisen dahin gehend, dass die algebraische Struktur in der gewählten Repräsentation adäquat repräsentiert sein muss (Brunner, 2014a, S. 18). „Inhaltlich anschauliche, operative Beweise stützen sich dagegen auf Konstruktionen und Operationen, von denen intuitiv erkennbar ist, daß sie sich auf eine ganze Klasse von Beispielen anwenden lassen und bestimmte Folgerungen nach sich ziehen“ (E. C. Wittmann & Müller, 1988, S. 249). Daraus lässt sich ableiten, dass inhaltlich-anschauliche Beweise nicht ausschließlich beispielgebunden erfolgen können. Vielmehr sollen Verallgemeinerungen „ables-

¹³⁹ Vgl. Brunner (2014b, S. 233), Holland (2007, S. 133), Krauthausen (2001, S. 107), Mac Lane (1981, S. 465), Pickert (1957, S. 49), Rehm (1990, S. 109), Winter (1983, S. 66)

¹⁴⁰ Freudenthal (1974, S. 167), Hirt und Wälti (2010, S. 14–15), Hußmann und Leuders (2006, S. 70), E. C. Wittmann (2003, S. 29)

¹⁴¹ In der Literatur wird begrifflich das inhaltlich-anschauliche Beweisen nicht vom operativen Beweisen abgegrenzt. Vielmehr werden die Bezeichnungen synonym verwendet, da sich inhaltlich-anschauliches Beweisen immer auf Operationen zurückführen lässt (Brunner, 2014a, S. 18). Weil der Fokus der Untersuchung auf der Veranschaulichung von Problemsituationen liegt und somit das adäquate Repräsentieren des Sachverhalts erfolgreich bewältigt werden muss, damit das Operieren mit der Darstellung gelingen kann, wird im Folgenden der Bezeichnung *inhaltlich-anschauliches Beweisen* gefolgt. Diese Bezeichnung spiegelt den Kern der vorliegenden Untersuchung aussagekräftiger wider, als es mit der Bezeichnung operativer Beweis gelingt.

bar“ sein, auch wenn sie streng genommen nicht dem Anspruch eines formal-deduktiven Beweises genügen (Brunner, 2014a, S. 18; Duncker, 1935, S. 66).

Charakteristisch für inhaltlich-anschauliche Beweise ist, dass sie in einer „schlichten, symbolarmen Sprache“ erfolgen und sie häufig Anwendung finden, wenn es mathematische Problemstellungen zu lösen gilt (E. C. Wittmann, 2014, S. 226). Greifen Lernende während des Erforschungsprozesses auf ihnen vertraute nichtsymbolische Darstellungen zurück, so kann dies das Klären von Sachverhalten unterstützen (siehe Abschnitt 1.5.5.1). Dadurch wird eine „leicht zugängliche Quasi–Realität“ geschaffen, die Strukturen und Beziehungen offenlegt sowie Hypothesen aufzustellen und zu überprüfen erlaubt (E. C. Wittmann, 2014, S. 227, 229). Das Agieren mit Materialien wie Plättchen und Rechenstäben oder mit externen Repräsentationen, wie geometrischen Figuren, Zahlengeraden, Stellentafeln oder Rechnungen, lassen vorliegende Muster als Beziehungen, die auf Wirkungen von Operationen beruhen, transparent werden.

Inhaltlich-anschauliches Beweisen bringt das Verstehen von Gesetzmäßigkeiten voran, indem die Lernenden ihre Fähigkeiten und Fertigkeiten im Umgang mit Darstellungsmittern ausbauen (Bezold, 2009, S. 78; E. C. Wittmann, 2014, S. 230). Ferner unterstützt es explizit auch den Austausch untereinander, da es hohes *Kommunikationspotenzial* aufweist (Bezold, 2009, S. 78; vgl. Lakatos, 1979, S. XII; E. C. Wittmann & Müller, 1988, S. 254). Während formal-deduktive Beweise eher eine Kommunikationsgrundlage für Kinder darstellen, die selbst die formal-symbolische Sprache ausüben und verstehen können, bieten inhaltlich-anschauliche Beweise allen Lernenden einen Zugang (Bezold, 2009, S. 78; Hanna, 1989a, S. 22–23). Letztere erleichtern zusätzlich das Nachvollziehen und Mitteilen von Beweisideen (Bezold, 2009, S. 78).

An dieser Stelle kommt (wie bei experimentellen Beweisen auch) die Frage danach auf, ob inhaltlich-anschauliche Beweise die nötige mathematische Strenge des Beweises erfüllen. Wittmann (2014, S. 228–229) setzt die Bedeutung der Wirkungen der Operationen (bei inhaltlich-anschaulichen Beweisen) mit der Bedeutung der Axiome (in der formalen Mathematik) gleich, weshalb er sie zur Gruppe der stichhaltigen Beweise zählt. Weiterhin spricht dafür, dass sie Verallgemeinerungen und eine Transformation in die formal-symbolische Notation zulassen (Thies, 2002, S. 83, in Anlehnung an Blum & Kirsch, 1989, 1991).

Formale Beweise stützen sich auf *symbolische Beschreibungen* mathematischer Objekte und symbolische Operationen im Rahmen systematisch-deduktiver Theorien, operative Beweise direkt auf *Darstellungen* dieser Objekte und *Operationen* an ihnen. Die Stichhaltigkeit operativer Beweise ergibt sich daraus, dass diese Operationen allgemein ausführbar sind, unabhängig von speziellen Beispielen, an denen sie demonstriert werden. (E. C. Wittmann, 2014, S. 226, Hervorhebung im Original)

Zusammenfassend bleibt zum Ausdruck zu bringen, dass sich das inhaltlich-anschauliche und formale Beweisen ausschließlich in den „eingesetzten Mitteln“ un-

terscheidet (E. C. Wittmann, 2014, S. 226): Ersteres stützt sich auf die ikonische Ebene, wohingegen sich Letzteres auf die symbolische Ebene stützt. Die enge Verknüpfung inhaltlich-anschaulichen und formal-deduktiven Beweisens hebt auch Winter (1983) hervor. Er weist „anschaulich-empirischem Tun“, dem er heuristische Tätigkeiten wie „Wahrnehmen, Beobachten, Messen, Analogisieren, Induzieren“ zuordnet, eine essenzielle, wissensentdeckende Funktion zu (Winter, 1983, S. 65). Er warnt davor, die Anschauung streng vom Beweisen abzugrenzen und positioniert sich sogar dergestalt, dass eine Trennung beider Bereiche zumindest nicht in der Sekundarstufe I – und somit erst recht nicht in der Grundschule – möglich sei. Vielmehr würden sich Anschaulichkeit und Deduktion wechselseitig verstärken:

Durch deduktives Ordnen, und dies besteht im wesentlichen in sprachlich gefassten (symbolhaften) Verallgemeinerungen, wird das anschaulich-intuitive Handeln nicht abgeschafft oder überflüssig gemacht, sondern im Gegenteil verfeinert, erhöht, sublimiert, das sinnliche Wahrnehmen wird strukturiert, vergeistigt, theoretisiert. Und umgekehrt erhalten deduktive Ableitungen erst Sinn und Bedeutung durch das intuitiv gegebene Material, an dem es arbeitet. Insoweit bedeutet Beweisen nicht Abkehr von der Empirie (Beobachten, Messen), sondern geradezu eine verstärkte Zuwendung. (Winter, 1983, S. 67; vgl. Jahnke, 1978, S. 211)

Es besteht unangefochtene Einigkeit, dass formales Beweisen und deren inkludierte Forderung der „lückenlosen Argumentation“ von Grundschulkindern nicht zu bewältigen ist und dieser Beweistyp für die Grundschule keinesfalls geeignet ist.¹⁴² Im Gegensatz zu formal-deduktiven Beweisen sind inhaltlich-anschauliche Beweise in der Grundschule zwar möglich (vgl. Peterßen, 2012, S. 20), aber im Vergleich zu experimentellen Beweisen eher im Einzelfall zu erwarten. Sich von Beispielen zu lösen, also die Wirkungen von Operationen, wie es das inhaltlich-anschauliche Beweisen vor sieht, verallgemeinernd darzustellen, konfrontiert Grundschulkinder mit einem Problem, das sie spontan und ohne Unterstützung nicht bewältigen können (Neumann et al., 2014, S. 118).

Der Nachweis der Gültigkeit unterscheidet inhaltlich-anschauliches Beweisen maßgeblich von mathematischem Begründen (Peterßen, 2012, S. 20). Während Ersteres auf Verallgemeinerung abzielt, reicht es beim Begründen aus, wenn Entdeckungen beispielbezogen begründet werden und dadurch ihre Gültigkeit erfahren. Nachfolgend soll die Einordnung des Begründens die gegenseitige Abgrenzung der drei Begriffe komplementieren und abrunden.

¹⁴² Bezold (2009, S. 76), Jahnke und Ufer (2015, S. 333), Krauthausen (2001, S. 100–101), Krummheuer (2003b, S. 125), Peterßen (2012, S. 12, 20), Reiss und Ufer (2009, S. 156)

2.2.5 Begründen als übergeordnete Kompetenz

Mathematischem Begründen und Beweisen wird eine gemeinsame Funktion zugesprochen: das Verifizieren und Falsifizieren von Aussagen (Neumann et al., 2014, S. 114, vgl. Abschnitt 2.2.3). Sie gelten daher als „hochgradig sprachlich verankert“ und werden terminologisch nicht strikt voneinander getrennt.¹⁴³

Der Begriff des Begründens ist jedoch weiter gefasst als der des Beweisens und impliziert stärker auch außermathematische Tätigkeiten, die im allgemeinbildenden Mathematikunterricht vermittelt werden sollen (M. Meyer & Prediger, 2009, S. 5). Er hat mit dem Argumentieren gemein, dass die Begriffe nicht an ein deduktives Vorgehen gebunden sind und ihnen deshalb ein größeres Bedeutungsspektrum zugrunde gelegt werden kann, als es bei dem Beweisen der Fall ist (vgl. Walsch, 2000, S. 7). Da das Beweisen mit dem formal-deduktiven Charakter konnotiert ist, bildet es lediglich die abstrakte Facette des Begründens ab (Brunner, 2014a, S. 29–31; Walsch, 2000, S. 7).

Hieraus lässt sich ableiten, dass das Begründen sowohl mit dem Argumentieren als auch mit dem Beweisen Schnittmengen besitzt und sich nicht explizit abgrenzen lässt. Aus diesem Grund zieht Malle (2002, S. 4) meist „das sanftere Wort ‚Begründen‘ dem Wort ‚Beweisen‘ vor, weil das letztere manchmal abschreckend wirkt“. Auch in der Grundschule wird bevorzugt vom Begründen gesprochen (z. B. Fetzer, 2011, S. 28; Meissner, 1979, S. 307; Neumann et al., 2014, S. 114). Nach Meissner (1979, S. 307) sollen Begründungen im Grundschulkontext konkret Aufschluss geben, warum

- dieser mathematische Sachverhalt zutrifft,
- diese Aussage gilt,
- dies die Lösung der Aufgabe ist,
- diese Technik hier sinnvoll ist,
- gerade diese Regel zum Erfolg führt,...

„Verstehen und Verständnis mitteilen“ sollen dabei als wechselseitige Schritte verstanden werden (Meissner, 1979, S. 307; vgl. Bauersfeld, 2002, S. 11–12; Hußmann & Leuders, 2006, S. 84). Das Begründen im Sinne des *eigenen Verstehens* erfolgt monologisch und bringt in seiner epistemischen Funktion den Erkenntnisgewinn des Einzelnen voran (Neumann et al., 2014, S. 115). Wenn dagegen das *Verständnis anderen mitgeteilt* wird, dann dominiert die kommunikative Funktion. Das Begründen ist so dann dialogisch ausgelegt und von den Kommunikationspartnern beeinflusst (vgl. Abschnitt 1.5.5). Berücksichtigt man darüber hinaus die deutschdidaktische Perspektive mit der Annahme, dass sich das Begründen aus der „situativen Alltagssprachlichkeit“ entwickelt, so sollen die Lernenden in der Lage sein, vorangegangene Begründungen

¹⁴³ Dörfler (1979, S. 9), M. Meyer und Prediger (2009, S. 5), Neumann et al. (2014, S. 114), Rehm (1990, S. 95), Walsch (2000, S. 7)

aufzugreifen und sich zu Nutzen zu machen, wenn es darum geht, neue Sachverhalte zu begründen (Neumann et al., 2014, S. 115). Die Autorinnen leiten hieraus in Anlehnung an Neumeister und Voigt (2012) und Spreckels (2009) ab, dass mathematisches Begründen als ein Spektrum verstanden werden kann, das sich „in den sprachlichen Mustern von der Deskription über die Explanation hin zur begründeten Argumentation mit steigenden Anforderungen bewegt“, wobei es gilt, sowohl die epistemische als auch die kommunikative Funktion zu berücksichtigen. Analog fasst Brunner (2014a, S. 30, 110) in Bezug auf Duval (1991) das Begründen als Kontinuum auf, wobei das Begründungsspektrum von einem niedrigen bis hin zu einem maximalen Maß an Strenge und formaler Notation reicht. Das Argumentieren und das Beweisen stellen folglich zwei spezifische Formen des Begründens dar, „die sich auf unterschiedliche Kontexte beziehen und damit auch teilweise unterschiedlichen Regeln folgen und andere Mittel verwenden“ (Brunner, 2014a, S. 30).

Darüber hinaus wird in der Literatur auch die Meinung geteilt, dass das Begründen eine „argumentative Tätigkeit unter anderen“ ist (Peterßen, 2012, S. 22).¹⁴⁴ Hier gilt das Argumentieren als allgemeinster, das Beweisen als speziellster und das Begründen eher als unschärfster der drei Begriffe (vgl. Walsch, 2000, S. 7).

2.3 Argumentieren im Rahmen der Untersuchung

Obwohl die Begriffe Argumentieren, Begründen und Beweisen in der mathematischen Community kontrovers diskutiert werden und keine übereinstimmende Terminologie besteht, lässt sich eine Gemeinsamkeit identifizieren, die als Ausgangspunkt aller Tätigkeiten verstanden werden kann: mathematische Entdeckungen.¹⁴⁵ Generieren Lernende beim Lösen von Problemen eigene Denk- und Erkenntniswerkzeuge, die ihnen durch das Aufdecken von Zusammenhängen zu mathematischen Entdeckungen verhelfen, so gilt es mit deren Hilfe, Gefundenes zu beschreiben, zu formulieren und zu begründen (vgl. Bezzold, 2009, S. 38; M. Meyer, 2007b, S. 300). Es eignet sich, individuelle Lösungsprozesse in einem sozialen und kommunikativen Kontext argumentativ darzulegen.¹⁴⁶

Dieser Konsens wird sich in der Untersuchung zu eigen gemacht und davon ausgegangen, dass mathematische Entdeckungen die inhaltliche Basis kommunikativer und argumentativer Tätigkeiten darstellen (vgl. Bezzold, 2009, S. 37; Neumann et al., 2014,

¹⁴⁴ Reiss und Ufer (2009, S. 158), Walsch (2000, S. 7), KMK (2005b, S. 8)

¹⁴⁵ Z. B. Bezzold (2009, S. 38–39), Freudenthal (1973b, S. 558–565), Jahnke und Ufer (2015, S. 341), Philipp (2013, S. 2), Ruwisch (2014, S. 43), Vollrath (1980, S. 34–35)

¹⁴⁶ Bezzold (2009, S. 60), Brunner (2014a, S. 14, 28), Gallin (2010, S. 6–8), Krummheuer und Brandt (2001, S. 18), Selter (1994, S. 66–67), Winter (1983, S. 81)

S. 114; Ruwisch, 2014, S. 43). Es gilt, zugrunde liegende mathematische Phänomene aufzuspüren, zu beschreiben, zu hinterfragen, zu begründen oder auch auf Korrektheit zu prüfen (Bezold, 2009, S. 37).¹⁴⁷

Die Verflochtenheit des Argumentierens und Problemlösens spiegelt sich schwerpunktmäßig in der Phase des Aufspürens mathematischer Phänomene wider (vgl. Roppelt & Reiss, 2012, S. 37). Diese kreative Phase des Lösens kann insbesondere durch externe Repräsentationen Unterstützung finden (z. B. Franke & Ruwisch, 2010, S. 66). Darüber hinaus können die konstruierten Denk- und Erkenntniswerkzeuge im gesamten argumentativen Prozess unterstützend herangezogen werden, wenn es darum geht, das eigene Vorgehen (*was*, *wie* und *warum*) offenzulegen (z. B. E. C. Wittmann, 1998, S. 158). Dabei stehen das *Was* als Beschreibung mathematischer Entdeckungen, das *Wie* als Beschreibung getätigter Handlungsabläufe und das *Warum* als Beschreibung des Ursache-Wirkungs-Zusammenhangs im Vordergrund.¹⁴⁸

Wird davon ausgegangen, dass Argumentieren heißt „Vermutungen über mathematische Eigenschaften und Zusammenhänge zu äußern (zu formulieren), diese zu hinterfragen sowie zu begründen bzw. hierfür eine Begründungsidee zu liefern“ (Bezold, 2009, S. 38), so wird implizit die Nähe zum Kommunizieren deutlich. Entdeckungen stellen daher einen Ausgangspunkt dar, das Problemlösen, die externe Repräsentationen wie auch das mathematische Kommunizieren und das Argumentieren zu vernetzen.

Die soziale und die kommunikative Dimension des Argumentierens (M. Meyer & Prediger, 2009, S. 5; Ruwisch, 2014, S. 43; vgl. Selter, 1994, S. 26) veranlassen, Kommunikationsprozesse im Mathematikunterricht näher in den Blick zu nehmen sowie deren Verknüpfung zum Argumentieren herauszustellen. Nachfolgend werden unter Berücksichtigung empirischer Ergebnisse die Vorteile intersubjektiven Austauschs herausgestellt. Die Ausführungen beschränken sich hierbei auf Kommunikationen, die sich in Zusammenhang mit dem Argumentieren ergeben.

2.4 Potenzial intersubjektiven Austauschs

Intersubjektiver Austausch ermöglicht ein Voneinander- und Miteinander-Lernen (Abschnitt 2.4.1), wobei sich die Kinder gegenseitig helfen oder auch ihre Lösungsgedanken wechselseitig und gleichberechtigt vorstellen und erklären können (Abschnitt

¹⁴⁷ Vgl. Franke und Ruwisch (2010, S. 137), M. Meyer (2007a, S. 28–30), Neumann et al. (2014, S. 114)

¹⁴⁸ Neumann et al. (2014, S. 114), Neumann und Ruwisch (2015, S. 320), Neumeister und Vogt (2012, S. 565–568, 575–581)

2.4.3). Inwiefern die metakognitiven Aktivitäten diese Prozesse beeinflussen, wird in Abschnitt 2.4.2 aufgezeigt.

2.4.1 *Voneinander- und Miteinander-Lernen*

Wenn man das *Lernen unter Gleichaltrigen* fokussiert, dann ist Jean Piaget (1972, S. 60) die Erkenntnis zu verdanken, dass Interaktionen unter Gleichaltrigen einen kognitiven Gewinn darstellen.¹⁴⁹ Kinder lernen nicht nur aus Sozialerfahrungen, sondern auch in kognitiven Bereichen Wichtiges nur von Gleichaltrigen.¹⁵⁰ Lernen Kinder „auf gleicher Stufe“ voneinander, dann sind die Lösungen des Kommunikationspartners erst einmal nicht mehr oder weniger wert als die eigenen (Forman, 1989, S. 59; Oswald, 1994, S. 11; R. Rasch, 2001b, S. 72). „Kinder scheinen sich mit ihresgleichen erheblich leichter und ungezwungener über z. T. sehr anspruchsvolle mathematische Themen unterhalten zu können als mit ihren Lehrern“ (Selter, 1994, S. 27; vgl. Webb, 1989, S. 25; Piaget, 1972, S. 72). Sie machen sich ihre eigene Sprache zugänglich und zum Ausgangspunkt der Argumentationen und des Verständnisses (Slavin, 1993, S. 167).

Belehrungen durch die Lehrkraft geschehen dagegen nicht auf dem „Boden von Gleichheit“ (Oswald, 1994, S. 11), erfahren daher weniger Anstrengungsbereitschaft seitens der Lernenden, da keine aktive Auseinandersetzung mit dem Sachverhalt notwendig ist. Werden Definitionen, mathematische Sätze oder Regeln einfach reproduziert, so fehlt den Kindern die „begriffliche Grundlage und Einsicht in sinnstiftende Zusammenhänge“ (Maier, 2000, S. 10). Schüler-Schüler-Interaktionen gelten als „wichtiger Motor der kognitiven Entwicklung“ (Selter, 1994, S. 27; vgl. Damon, 1984, S. 335; Slavin, 1993, S. 159). Insbesondere dann, wenn das Wissen der Lernenden im gemeinsamen Austausch „erschüttert“ wird, weil es nicht auf Anhieb erklärt bzw. das Geäußerte des Gegenübers nicht direkt verstanden werden kann, kann altes Wissen umstrukturiert werden (Nührenbörger & Schwarzkopf, 2010, S. 14). Durch das Transparentwerden der unterschiedlichen Sichtweisen weist der intersubjektive Austausch enormes Potenzial für argumentative Tätigkeiten auf und regt Lernprozesse an (Damon, 1984, S. 335; Nührenbörger & Schwarzkopf, 2010, S. 14).

Beim *Lernen von Gleichaltrigen* können ausgewählte Eigenproduktionen zum Gesprächs- und Diskussionsgegenstand der Gruppe werden und Impulse für die Weiter-

¹⁴⁹ Vgl. Cobb, Yackel und Wood (1992, S. 25), Götze (2008, S. 71), Slavin (1993, S. 159), Webb (1989, S. 22)

¹⁵⁰ Nührenbörger und Schwarzkopf (2010, S. 14), Oswald (1994, S. 11), Piaget (1972, S. 72), Steinbring und Nührenbörger (2010, S. 168)

arbeit geben (siehe Abschnitt 1.5.5.3).¹⁵¹ Es liegt daher der Leitgedanke zugrunde, aus subjektiven Auffassungen und Thesen intersubjektives Wissen zu generieren.¹⁵² Das gemeinsame Interagieren ermöglicht eine reziproke Verknüpfung von Produkt und Prozess. Austauschphasen, die auf Eigenproduktionen der Kinder (den Produkten) beruhen, können wiederum neue Arbeitsphasen (Prozesse) anstoßen, bei denen neue Produkte zu erzeugen sind (Lengnink, 2013, S. 218; vgl. Hefendehl-Hebeker & Hußmann, 2010, S. 105–106). Hieraus resultiert die Forderung an den Unterricht, dass individuelle Lösungsgedanken nicht ausschließlich verbal, sondern zuvor auch schriftlich versprachlicht werden müssen, um Argumentationskompetenzen fördern zu können (Fetzer, 2009, S. 23; Neumann et al., 2014, S. 115; Ruwisch, 2014, S. 43). Erst wenn beide Prozesse im Unterricht verfolgt werden, kann das Argumentieren die geforderte epistemische und kommunikative Funktion erfüllen (Neumann et al., 2014, S. 115, vgl. Abschnitt 1.5.5.1–1.5.5.3).

2.4.2 Metakognitive Aktivitäten

Nachfolgend werden sowohl die metakognitiven Aktivitäten im Verstehensprozess als auch die metakognitiven Aktivitäten im Interaktionsprozess differenziert aufgezeigt.

Allein die Tatsache, zu wissen, dass der aktiv erarbeitete Lösungsprozess öffentlich gemacht und mit Gleichaltrigen geteilt werden soll, kann im *Verstehensprozess* eine Klärung der eigenen Gedanken bewirken (Rathgeb-Schnierer, 2005, S. 166). Lernende, die fähig sind, situationsbedingt zu entscheiden, wann sie welches Wissen aktivieren müssen, wie dieses zu koordinieren und zu kontrollieren ist, sind metakognitiv aktiv (Mayer, 1998, S. 53). Jeder muss sich seiner individuellen Entdeckungen, mentalen Bilder und Denkvorgänge bewusst werden, um sie kommunizieren und erklären zu können (Lorenz, 2005, S. 161; Rathgeb-Schnierer, 2005, S. 166). Die Auseinandersetzung mit der Entdeckung kann intensiviert werden, indem die Lösungsgedanken schriftlich fixiert und dadurch geordnet werden. (Bezold, 2009, S. 87–88, vgl. Abschnitt 1.5.5). Das Einnehmen einer Metaperspektive ist erforderlich, um die eigene Vorgehensweise von außen rezipieren und reflektieren zu können.¹⁵³ Angefertigte Notizen können unterstützend hinzugezogen werden und helfen, eine Begründung anzubahnen (Bezold, 2009, S. 87). Gelingt es dem Lösenden, seine Antworten und Schlussfolgerungen zu reflektieren und nachvollziehbar zu machen, dann setzt er sich intensiv

¹⁵¹ Eichler (2015, S. 57), Lengnink (2013, S. 218), Selter (1994, S. 66–67, 283–284), E. C. Wittmann (2003, S. 30–33)

¹⁵² Gallin und Ruf (1993, S. 5–9), M. Meyer (2007a, S. 19, 102, 152), M. Meyer und Voigt (2009, S. 20), Selter (1994, S. 65–67)

¹⁵³ Bezold (2009, S. 87–88), Lorenz (2005, S. 154), Mayer (1998, S. 61), Schoenfeld (1985, S. 139–140, 1992, S. 349), Schütte (2002, S. 17)

ver mit dem vorliegenden Sachverhalt auseinander, erhält dadurch eine tiefere Einsicht in die zugrunde liegenden mathematischen Strukturen, was infolgedessen in einer intensiveren Verarbeitung mathematischen Wissens mündet.¹⁵⁴ Letztlich kann dies jeden Einzelnen befähigen, seine eigenen Entdeckungen und Schlussfolgerungen den Klassenkameraden begründend mitzuteilen. Auch wenn erste Beschreibungs- und Begründungsversuche vom Gegenüber unverstanden bleiben, ist ein reflektierter Lerner eher als ein unreflektierter in der Lage, sie neu oder auf eine andere Weise zu formulieren, neue, ergänzende Beispiele zu generieren oder auch weitere Repräsentationen heranzuziehen (Webb, 1989, S. 29).

Im *Interaktionsprozess* steht der Lernende nicht nur vor der Aufgabe, seinen eigenen Lösungsprozess vorzustellen und zu erklären, sondern er wird auch mit alternativen Sicht- und Vorgehensweisen seiner Klassenkameraden konfrontiert. Das Kommunizieren fördert das Lernen voneinander, wenn sich die Beteiligten auf die fremden, noch unbekannten Lösungsprozesse der anderen einlassen, sich in diese hineindenken, diese nachzuvollziehen, zu verstehen und zu hinterfragen versuchen, sie aber auch als Alternativen annehmen, wenn sie überzeugen.¹⁵⁵ Mit dem Ziel, ein geteiltes, sich deckendes Verständnis zu erzeugen und Wissen gemeinsam zu elaborieren, steht der Vergleich der Lösungsprozesse im Vordergrund (Fehse, 2001, S. 72; Naujoks, 2002, S. 67; Roschelle, 1992, S. 255, 269). Auch metakognitive Aktivitäten, die im Gespräch erwachsen, können das Überdenken und Reflektieren der eigenen Lösung voranbringen und eine kognitive Entwicklung anregen.¹⁵⁶ Die Erweiterung des Wissens durch den Austausch von Begründungen – sei es aktiv durch Lernen beim Begründen oder passiv durch Lernen aus Begründungen – gilt als zentraler Vorteil kommunikativer Settings (vgl. Fehse, 2001, S. 67–72).

2.4.3 Prozesse des Helfens und Kollaborierens

Beim Problemlösen kann nicht davon ausgegangen werden, dass Lernende direkt die richtige Lösung produzieren.¹⁵⁷ Anstelle einer vorschnellen Belehrung durch die Lehrkraft kann eine gemeinsame Auseinandersetzung unter Gleichaltrigen das Aufspüren und Identifizieren von Stolpersteinen, Widersprüchen oder Fehlern anstreben (Gallin,

¹⁵⁴ Fehse (2001, S. 67), Kuntze und Prediger (2005, S. 1), Lorenz (2005, S. 160), Maier (2000, S. 13), Muir et al. (2008, S. 231), Rathgeb-Schnierer (2005, S. 166), Slavin (1993, S. 160), Webb (1989, S. 29)

¹⁵⁵ Fehse (2001, S. 64), Hengartner (1992, S. 25–27), Rathgeb-Schnierer (2005, S. 167), Sternberg (1998, S. 128)

¹⁵⁶ Bezold (2009, S. 86), Fehse (2001, S. 67), Lorenz (2005, S. 154), Rathgeb-Schnierer (2005, S. 205), Selter (1994, S. 27)

¹⁵⁷ Groß (2013, S. 114–133), Hohn (2012, S. 106), R. Rasch (2001b, S. 89), Stahl (1975, S. 279), Thevenot und Oakhill (2005, S. 1317)

2010, S. 6–8; M. Miller, 1986, S. 10; Steinbring & Nührenbörger, 2010, S. 169). Oswald (1994, S. 12) konnte zeigen, dass Kinder hierzu fähig sind und kognitiv voneinander profitieren können. Basiert der Austausch auf abweichenden Lösungen, so steht das Team vor der Herausforderung, mögliche Fehlstrategien zu identifizieren und aufzulösen sowie erklärend zu agieren, um neue Einsichten entwickeln zu können (M. Miller, 1986, S. 10; Naujoks, 2002, S. 67–69; Rathgeb-Schnierer, 2005, S. 166). Hierzu müssen die Lernenden kollektiv argumentieren und diskutieren, aber auch Empfehlungen aussprechen und annehmen, indem sie ihre unterschiedlichen Ansätze miteinander vergleichen (Lorenz, 2004, S. 48; Naujoks, 2002, S. 67; Rathgeb-Schnierer, 2005, S. 167).

In diesem Zusammenhang werden zwei wesentliche Kooperationstypen unterschieden: das Helfen einerseits und das Kollaborieren andererseits (Naujoks, 2002, S. 68). Beide lassen Interaktionen mit hoher Intensität erkennbar werden. Wenn sich die Interaktionspartner ihre individuellen Gedanken wechselseitig und gleichberechtigt vorstellen und erklären, dann handelt es sich um die Tätigkeit des Kollaborierens, die ein hohes Maß an Lernförderlichkeit aufweist (Naujoks, 2002, S. 68). Beim Begründen wird über das eigene Vorgehen reflektiert, wodurch die Lernenden an Klarheit gewinnen (Goos, Galbraith, & Renshaw, 1996, S. 243; Naujoks, 2002, S. 69).

Ferner initiieren Interaktionen Situationen, in denen sich die Lernenden gegenseitigen Helfen. Der Helfende nimmt hierbei die Lehrerrolle ein und beschreibt und begründet seinem Partner seine Erfahrungen und/oder mathematischen Entdeckungen (Naujoks, 2002, S. 68; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003, S. 242). Zudem stehen die Kinder vor der Aufgabe, auf die Fragen ihres Gegenübers einzugehen und diese zu beantworten. Die Lernenden bilden ein didaktisches Dreieck (Lehrer, Lernender) mit dem Lerninhalt (Oswald, 1994, S. 11; Schnotz, 2009, S. 48). Dabei stehen die Helfer vor der Herausforderung, ihren Denk- und Lösungsprozess reflektiert und strukturiert abzubilden und gegebenenfalls Umstrukturierungen vorzunehmen, sodass er für den Interaktionspartner nachvollziehbar wird (Webb, 1989, S. 29). Es sind hauptsächlich die fortgeschrittenen Mitglieder, die helfen, die Zone der nächsten Entwicklung zu erreichen (Goos et al., 1996, S. 243–244; Piaget, 1972, S. 32).

Mitunter konnte beobachtet werden, dass Prozesse des gegenseitigen Helfens schwierig sind und erst beschwerlich erlernt werden müssen (Oswald, 1994, S. 11–12). Ein Grund hierfür könnte sein, dass gerade in Situationen des Helfens die Gleichrangigkeit verletzt ist (Forman, 1989, S. 59; Oswald, 1994, S. 11). Ein Kind verfügt über Wissen, das dem anderen fehlt. Ersteres ist jedoch nicht verpflichtet, sein Wissen mitzuteilen, Letzteres vielleicht nicht bereit, Wissen von diesem Helfenden anzunehmen. Ob Interaktionspartner befreundet sind oder nicht, kann daher durchaus Einfluss auf deren Kooperationsverhalten haben (Oswald, 1994, S. 12).

Webb (1989, S. 35) konnte beim Vergleich von 19 Studien zeigen, dass es einen positiven Zusammenhang zwischen beobachteter Interaktion und erzielter Leistung bei Lernenden gibt, die ihre Gesprächspartner im Austausch unterstützten (Slavin, 1993, S. 160–161). Insbesondere dann, wenn sie schrittweise beschrieben und begründeten, wie das Problem gelöst werden kann bzw. Teile des Problems gelöst werden können (Webb, 1995, S. 110). Erhielten Lernende aufgrund eines Fehlers oder auf Nachfrage hilfreiche Tipps, war ein positiver Zusammenhang nicht kontinuierlich festzustellen (Webb, 1989, S. 35). Der Zusammenhang war sogar negativ, wenn ihnen ausschließlich lösungsbetreffende Informationen wie die Lösungszahl als solche, Ja-Nein-Antworten oder Hinweise, wie beispielsweise dazu, welches Problem zuerst zu bearbeiten ist, mitgeteilt wurden.

Den Ergebnissen nach Webb (1989, S. 35) zufolge ist die reine Integration von Schülernkommunikationen in den Unterricht kein Garant für Erfolg. Der Zusammenhang wird von einer komplexen Interaktion möglicher Faktoren bestimmt. Es ist daher wenig überraschend, dass Individuen nicht automatisch erfolgreich waren, als sie eine hilfreiche Argumentation erhielten. Vielmehr war entscheidend, ob im argumentativen Austausch auf die Bedürfnisse des Fragenden eingegangen wurde und ob diese auch erfüllt wurden, ob das Gesagte für den Fragenden tatsächlich hilfreich war, ob er es verstanden hatte und ob er es sich zu eigen machen konnte oder noch machen wird. Lernende, die auf ihr Hilfegesuch lediglich lösungsbetreffende Informationen ohne Erklärung und Begründung erhielten, schafften es in den seltensten Fällen, von alleine Problembarrieren zu überwinden und erfolgreich zu sein. Die Qualität von Kommunikationen sowie die argumentativen, sozialen und metakognitiven Fähigkeiten der einzelnen Teilnehmer entscheiden, ob Kommunikationen Erfolg bewirken können.

Weitere Studien¹⁵⁸ konnten zeigen, dass kommunikative Settings unter Gleichaltrigen als Lernunterstützung gelten und einen Lernzuwachs bewirken können: Unterricht, der auf Kommunikation und Kooperation beruhte, führte langfristig gesehen zu besseren Mathematikleistungen als Unterricht, der die gleichen Inhalte, jedoch nicht den kommunikativen Fokus verfolgte. Allerdings ist bisher noch unerforscht, ob der Austausch unter Gleichaltrigen die Entwicklung von Repräsentationskompetenzen fördert.

¹⁵⁸ Exemplarisch wird hier auf Boaler und Staples (2008), Boaler (2008), Oswald (1994), Rohrbeck et al. (2003) und Tarim (2009) verwiesen.

Empirischer Teil

Der zweite Teil der Arbeit widmet sich der vorliegenden empirischen Untersuchung. In Kapitel 3, das den empirischen Teil der Arbeit eröffnet, werden aus den theoretischen Ausführungen und dem Stand der Forschung die Forschungsfragen und Hypothesen der Untersuchung abgeleitet. Zu welchen methodischen Überlegungen und Entscheidungen dies geführt hat, wird in Kapitel 4 dargestellt. Erkenntnisse und Konsequenzen, die aus der Pilotierung für die Hauptuntersuchung abgeleitet wurden, werden ausführlich geschildert und diskutiert. Darüber hinaus werden das zugrunde liegende Design der experimentellen Interventionsstudie sowie das Training, der Kern der Untersuchung und die Durchführung der Untersuchung vorgestellt. In diesem Zusammenhang werden auch die Messinstrumente vorgestellt und diskutiert, sowie die Operationalisierungsgedanken offen gelegt. Im Ergebniskapitel 5 werden Bezug nehmend auf die Forschungsfragen die Hypothesen beantwortet und im anschließenden Diskussionskapitel 6 ausführlich diskutiert. Welche Implikationen für die Praxis erwartet werden und welche Anknüpfungspunkte für weitere Forschungen sich ableiten lassen, runden neben der Schlusszusammenfassung den empirischen Teil und somit die vorliegende Arbeit ab.



3 Förderung der Generierung und Nutzung externer Repräsentationen

In den vorangegangenen Kapiteln wurde das Problemlösen, die Facetten externer Repräsentationen und deren Bedeutung für das Problemlösen sowie der intersubjektive Austausch über die individuellen mathematischen Entdeckungen diskutiert. Deren Verflochtenheit wird neben den sich hieraus ergebenden Forschungsdesideraten in einer kurzen Zusammenfassung des jeweiligen theoretischen Hintergrunds (Abschnitt 3.1) sowie des aktuellen Forschungsstandes (Abschnitt 3.2) akzentuiert und begründet. In Abschnitt 3.3 werden die Zielsetzungen der Untersuchung sowie die Forschungsfragen mit den jeweiligen Hypothesen aufgezeigt.

3.1 Schlussfolgerungen aus den theoretischen Ausführungen

Wenn man zunächst auf Prozesse des Problemlösens zurückblickt, konkret auf das Lösen problemhaltiger Textaufgaben, dann liegt der *Representational Change Theory* (siehe Abschnitt 1.3.3.2 und 1.4) die Annahme zugrunde, dass Lernende auf Basis ihrer individuellen Ressourcen Repräsentationen konstruieren, die als Gedächtnis-Trigger fungieren und Wissen für die Lösungsfindung aktivieren. Dieses aktivierte Wissen kann für den weiteren Lösungsprozess einerseits erkenntnis- und lösungsunterstützend, andererseits aber auch hemmend und störend sein. Letztgenanntes verhindert das Abrufen erforderlicher Operatoren und lässt den Lösungsprozess stocken. Dies hat zur Folge, dass Problemlösende in ihrem Denk- und Verstehensprozess auf Barrieren stoßen, die es zu überwinden gilt. Umstrukturierungen der Problemsituation sind umgänglich. Es wurde argumentiert, dass repräsentationale Veränderungen des Problemraums, sei es durch Erweiterung oder Eingrenzung, als Schlüssel zum Erfolg gelten (vgl. Abschnitt 1.3.3.2). Tätigkeiten wie das Elaborieren, Re-Interpretieren, Lockern von Bedingungen sowie Zerlegen von Gruppierungen können Lernenden helfen, die erforderlichen Veränderungsprozesse zu vollziehen. Gelingt es dem Einzelnen, die noch zu vollziehenden Schritte zu reduzieren und den eigenen Wissensstand dem Zielzustand anzunähern sowie den passenden Weg durch den neu konstruierten Problemraum zu finden, können daraus Einsicht und Erfolg resultieren (vgl. Abschnitt 1.3.3.2). Das Fundament erfolgreichen Problemlösens ist die Generierung adäquater Repräsentationen (vgl. Abschnitt 1.5.2), weshalb die Frage danach aufkommt, wie man Lernende in diesem Prozess unterstützen kann (vgl. Abschnitt 1.5.4 und 2.3).

Das Externalisieren eigener Denk- und Lösungsprozesse kann aus mehreren Gründen – wie in Abschnitt 1.5.5 ausführlich dargestellt – die Generierung adäquater Repräsentationen unterstützen sowie Schwierigkeiten im Verstehensprozess entgegenwirken.

Mentale Modelle nehmen durch das Externalisieren äußere Gestalt an. Mit den externen Repräsentationen kann dann sichtbar und nicht nur vor dem geistigen Auge handelt werden, was insbesondere dann von Vorteil ist, wenn die Informationen und zugrunde liegende Sachzusammenhänge noch nicht bzw. noch nicht vollständig erschlossen wurden. Hinzu kommt, dass sich Prozesse des Konkretisierens, Präzisierens, Manipulierens und Weiterentwickelns experimentell an externen Repräsentationen leichter realisieren lassen als an internen (vgl. Abschnitt 2.2.1, 2.2.3 und 2.2.4). Lernende, die ihre Gedankengänge zu Papier bringen, gewinnen Ressourcen, die ihnen für das Ausführen notwendiger Umstrukturierungen und die Lösungsfindung zur Verfügung stehen. Da diese Abläufe nicht ausschließlich im Kopf vollzogen werden müssen, kann der Problemraum flexibler und beweglicher verändert und spezifisch angepasst werden. Mehrere Schritte oder auch Irrwege sind dabei weder hinderlich noch störend, sondern können sogar in manchen Situationen lernförderlich und klarend sein. Externe Repräsentationen unterstützen daher die Strukturierung der Denkprozesse sowie die individuelle Wissensaneignung und werden zu Denk- und Erkenntniswerkzeugen, die Lernenden in Arbeitsphasen, Kommunikations- oder Argumentationssituatonen Sicherheit und Rückhalt geben können (vgl. Abschnitt 1.5.5.1, 2.2.1–2.2.4 und 2.3).

Nicht nur der Produzent selbst, sondern auch Dritte können von ihnen profitieren (Abschnitt 1.5.5.2.–1.5.5.3). An dieser Stelle muss betont werden, dass externe Repräsentationen Produzenten und Rezipienten stützen und stärken können. Damit sie eine stützende Funktion übernehmen, muss der Unterricht Situationen schaffen, die das Externalisieren anregen, den Lernenden das Sammeln vielfältiger Erfahrungen ermöglichen und in denen die Eigenschaften externer Repräsentationen, ihre Vor- und Nachteile, ihre Einsatzmöglichkeiten und ihre Vernetzung thematisiert, reflektiert und diskutiert werden (vgl. Abschnitt 1.5.5.4). Analoges kann auf einen kommunikativen Austausch auf Basis externer Repräsentationen übertragen werden (vgl. Kapitel 2): Er kann die Interaktionspartner in ihrem Denk- und Lösungsprozess voranbringen, wenn sie sich auf reflexive, kommunikative und argumentative Tätigkeiten einlassen.

Zusammenfassend bleibt festzuhalten, dass das Problemlösen (nicht nur) für ungeübte Löser enorme Schwierigkeiten und Herausforderungen birgt, welchen aus psychologischer Sicht durch die Generierung und Nutzung externer Repräsentationen, aber auch durch intersubjektiven Austausch entgegengewirkt werden kann.

3.2 Schlussfolgerungen aus dem Stand der Forschung

Bisher fokussierten viele Studien Ist-Stand-Analysen, um vorhandene Problemlösekompetenzen der Lernenden zu erfassen und zu diagnostizieren. Den empirischen Er-

gebnissen zufolge ist der Lösungserfolg bei problemhaltigen Textaufgaben nicht zufriedenstellend. Es besteht Konsens, dass es ungeübte Löser in den seltensten Fällen mit eigenen Kräften schaffen, ProblembARRIEREN zu überwinden (vgl. Abschnitt 1.3.3.2 sowie 1.4.3). Dies geht mit der Tatsache einher, dass sie nur vereinzelt von sich aus ihren Denk- und Lösungsweg externalisieren, obwohl es empirisch gestützt als lernförderlich gilt.

Ungeübte Problemlösende tendieren dazu, die Lösung problemhaltiger Textaufgaben im Kopf zu produzieren. Greifen sie doch auf externe Repräsentationen zurück, so zeigen sie sich beim Konstruieren unsicher und häufig nicht imstande, die Problemstruktur adäquat abzubilden. Gehäuft wird ein konzeptloses, unflexibles und unsystematisches Vorgehen erkennbar, das der geschilderten Problemsituation nicht gerecht wird. Es fehlt ihnen ein Repertoire an Handlungsweisen, auf das sie in der jeweiligen Problembewältigung zurückgreifen können.

Es zeichnet sich ab, dass der Mathematikunterricht der Grundschule diese Förderung versäumt bzw. ihr nicht im erforderlichen Maße gerecht wird (z. B. Brenner et al., 1997, S. 663; Zhe, 2012, S. 63, 68). Es lassen sich Forderungen ableiten, Kinder zur Konstruktion externer Repräsentationen und zu einem explorativen Agieren mit diesen anzuregen. Das Hauptaugenmerk muss darauf gerichtet sein, die Problemstruktur adäquat wiederzugeben, um das für die Lösungsfindung benötigte Wissen aktivieren zu können. Ungeübte Problemlösende müssen erst erlernen, wie sie ihr mentales Modell abbilden können, d. h., wie sie geeignete Denk- und Erkenntniswerkzeuge konstruieren können, um der Forderung adäquater Strukturabbildung gerecht zu werden (siehe Abschnitt 1.5.5.5 und 1.5.6). Dabei können Prozesse repräsentationaler Veränderungen stärkend fungieren und erforderliche Umstrukturierungen voranbringen. Anhand vielfältiger Bearbeitungen galt und gilt es zu erfahren, was adäquate Repräsentationen auszeichnet und was mit ihnen erreicht werden kann. Entscheidend und nicht zu unterschätzen ist hierbei, dass Lösende ihre externen Repräsentationen als wichtige Denk- und Erkenntniswerkzeuge, konkret ihre Sinnhaftigkeit, erleben müssen, um sie aktiv als Gedächtnis-, Kommunikations- oder auch Argumentationsstütze in ihren Lösungsprozess integrieren zu können. Der Unterricht muss dahin gehend unterstützen und sensibilisieren, mit dem Bewusstsein, dass externe Repräsentationen nur Gebrauch finden, wenn sie nicht als Mehraufwand empfunden werden. In diesem Zusammenhang sollen mögliche Darstellungswechsel herausgearbeitet und langfristig forciert werden (KMK 2005b, S. 8; Prediger, 2013a, S. 172). Noch nicht empirisch untersucht wurde, ob sich ein derartiges Unterrichtskonzept positiv auf die Leistungen von Grundschulkindern auswirkt.

Einige auch interdisziplinäre Studien zum Problemlösen konnten bereits belegen, dass der kompetente Umgang und die Konstruktion externer Repräsentationen erlernbar

sind. Es wurde herausgestellt, dass sich mit zunehmender Erfahrung sowohl die Problemlösekompetenzen als auch die Konstruktion externer Repräsentationen in gegenseitiger Abhängigkeit entwickeln, und dass sich die kognitive Belastung nachweislich reduzieren ließ. Gerade hinsichtlich der Konstruktion und Nutzung depiktionaler Repräsentationen zielten viele empirische Befunde in diese Richtung (vgl. Abschnitt 1.5.6.1). Es wurde in diesem Zusammenhang explizit die Bedeutsamkeit der adäquaten Strukturabbildung sowie des notwendigen Vor- und Erfahrungswissens hervorgehoben. Es wird die Auffassung vertreten, dass Problemlösende langfristig ein facettenreiches Repräsentationsrepertoire aufbauen müssen, „um in Abhängigkeit von dem gerade zu lösenden Problem eine hierfür geeignete externe Repräsentation auswählen zu können“ (Fehse, 2001, S. 53; vgl. Cox, 1999, S. 360; Cox & Brna, 1995, S. 282; Schreiber, 2014, S. 9; Steinau, 2014, S. 34; Stern, 1998, S. 105). Es gilt zu überprüfen, ob Grundschulkinder dazu fähig sind, ein Repertoire aufzubauen und ihre Vorgehensweisen zu variieren.

Explizite Forschung zum mathematischen Problemlösen fand nach Heinrich, Bruder und Bauer (2015, S. 289) bisher schwerpunktmäßig im Bereich der Förderung von Heurismen statt (z. B. Bruder & Collet, 2011a, 2011b; Collet, 2009; König, 1992). Hierbei wurden insbesondere die Entwicklung und Nutzung heuristischer Vorgehensweisen fokussiert und erforscht (Bruder & Collet, 2011a, S. 30, 2011b, S. 37). Externe Repräsentationen ordnen Bruder und Collet (2011b, S. 45) den heuristischen Hilfsmitteln zu und somit heuristischen Vorgehensweisen unter. Aus der Problemlöseforschung der Darmstädter Arbeitsgruppe geht hervor, dass Sekundarstufenschülerinnen und -schüler, die mit dem Fokus auf heuristische Vorgehensweisen über ein Schuljahr hinweg von geschulten Lehrkräften unterrichtet wurden, ihre Problemlösekompetenzen erheblich ausbauen konnten (Collet, 2009, S. 186–190, 230–233). Auch deren Einsatz heuristischer Vorgehensweisen nahm im Prä-Post-Vergleich zu und hat sich im Durchschnitt nahezu verdreifacht (Collet, 2009, S. 190–191, 230). Dabei wurde stets betont, dass es sich um ein langfristiges Fördern von Problemlösekompetenzen handeln muss und Lernumgebungen geschaffen werden müssen, die Leistungsschwächeren wie Leistungsstärkeren ansprechen (z. B. Bruder & Collet, 2011b, S. 143–145; Lester & Kroll, 1990, S. 56). Die Heterogenität der Lerngruppe, auch hervorgerufen durch Migration, fordern Lernangebote, die binnendifferenziert ausgelegt sind (vgl. Collet, 2009, S. 57) und sowohl epistemische als auch kommunikative Funktionen erfüllen. Im Vergleich zur Nutzung heuristischer Strategien sei die Nutzung heuristischer Hilfsmittel mühseliger zu vermitteln, da sie geistige Handlungen leichter transparent und Deutungen nachvollziehbar werden lassen (König, 1992, S. 26). Nach König (1992, S. 27) lassen sich Lernende „noch ziemlich einfach“ zu einer bewussten Verwendung solcher Werkzeuge aktivieren, da sie gerade in der kreativen Phase des Problemlösens helfen,

die Aufgabe zu verstehen und mathematische Phänomene aufzuspüren. In Bezug auf die Grundschule ist es daher lohnenswert, zunächst einen Baustein heuristischer Vorgehensweisen, beispielsweise externe Repräsentationen, zu fokussieren und damit die Lernenden an die Heurismen heranzuführen bzw. an die Heurismen zu gewöhnen (Bruder, 2003, S. 31; Bruner, 1973, S. 72). Ob hierbei auch Prozesse des Helfens und Kollaborierens unterstützend fungieren können, gilt es zu überprüfen (vgl. Abschnitt 2.4.3).

3.3 Forschungsfragen und Hypothesen

Grundlage der vorliegenden Untersuchung stellt ein Trainingsexperiment dar, das sich an den individuellen Ressourcen der Lernenden orientiert, sie aufgreift und zum Trainingsgegenstand innerhalb des regulären Mathematikunterrichts werden lässt. Für eine detaillierte Darstellung des Trainings sei auf Abschnitt 4.4 verwiesen. Nachfolgend werden die Ziele, die mit dem Trainingsexperiment verbunden sind erläutert (Abschnitt 3.3.1). Im Anschluss daran werden Forschungsfragen nebst den Hypothesen vorgestellt (Abschnitt 3.3.2).

3.3.1 Ziele des Trainingsexperiments

Vor dem Hintergrund aktueller empirischer Ergebnisse und des ausbleibenden Erfolgs zahlreicher Lernender beim Lösen problemhaltiger Textaufgaben ist es Ziel des Trainingsexperiments, sich das Potenzial externer Repräsentationen zunutze zu machen. Es wird erwartet, mehr über die vielfältigen Herangehensweisen der Lernenden zu erfahren, indem sie zum Externalisieren ihrer mentalen Problemlöseprozesse angeregt werden. Die selbstgenerierten Denk- und Erkenntniswerkzeuge werden eingesetzt, um gezielt die individuellen Ressourcen der Lerngruppe aufzugreifen und daran anzuknüpfen. Dabei werden deren Unterschiedlichkeit, Individualität und Kreativität genutzt und im Rahmen des regulären Mathematikunterrichts zum Gegenstand des Trainings gemacht.

Das Hauptaugenmerk der Untersuchung ist darauf gerichtet, herauszufinden, ob sich Drittklässler im Verlauf eines Repräsentationstrainings auf der Basis selbstgenerierter Repräsentationen Techniken sowie Vor- und Herangehensweisen aneignen, die sie zur Steigerung ihres Lösungserfolgs und ihrer Problemlösekompetenzen befähigen (vgl. hypothesenprüfendes Vorgehen bei Interventionsmaßnahmen; Hager & Hasselhorn, 2000, S. 42). Darüber hinaus gilt zu untersuchen, ob ein Austausch unter Gleichaltrigen oder ein Austausch im Klassenverband die Lernenden unterstützen kann. Es wird überprüft, inwiefern Lernende von den drei *Potenzialen* externer Repräsentationen

profitieren können: Welche Bedeutung haben sie beim Lösen problemhaltiger Textaufgaben für

1. jeden Einzelnen (siehe Abschnitt 1.5.5.1 sowie 2.2, 2.4.2),
2. die Interaktion unter Gleichaltrigen (siehe Abschnitt 1.5.5.2 sowie 2.2, 2.4),
3. für Kommunikations- und Reflexionsphasen im Mathematikunterricht (siehe Abschnitt 1.5.5.3 sowie 2.2, 2.4)?

Ersteres wird überprüft, indem untersucht wird, ob sich die Lernenden auf die Generierung externer Repräsentationen einlassen und infolgedessen häufiger in Problemlösungsprozessen an und mit ihnen arbeiten. Es ist dabei von Interesse, zu untersuchen, inwiefern eine steigende Generierung und Nutzung externer Repräsentationen den Lösungserfolg und die Problemlösekompetenzen beeinflussen.

In Bezug auf die Interaktion unter Gleichaltrigen ist es das Ziel, herauszufinden, ob Austauschphasen über die individuellen Lösungsprozesse (vgl. Abschnitt 2.3), das Hinterfragen sowie das gegenseitige Beschreiben und Argumentieren die Lernenden voranbringen, ihr Denken strukturieren und sie letztlich in ihren fortlaufenden Problemlösungen stärken. Die Reflexions- und Kommunikationsphase hat als eigentliche Trainingsphase die Intention, den Lernenden ein breites Lösungsspektrum in ihrer Sprache, der *Sprache des Verstehens*, zu offerieren und ins Bewusstsein zu rücken. Das Training zielt auf die Anbahnung eines Repräsentationsrepertoires, das die Kinder befähigt, je nach Aufgabe entscheiden zu können, welche externen Repräsentationen ihnen am ehesten zum Erfolg verhelfen und warum.

Um Aussagen diesbezüglich treffen zu können, müssen die individuellen Ressourcen der Lernenden im Vorfeld im Rahmen einer Pilotierung erfasst und ausgewertet werden. Es stellten sich daher die Fragen danach, mit welchen externen Repräsentationen Grundschulkinder vertraut sind, welche sie zum Lösen problemhaltiger Textaufgaben heranziehen und über welche Merkmale diese verfügen. Darüber hinaus ist von großem Interesse, herauszufinden, welchen Einfluss die individuellen Eigenproduktionen auf den Lösungserfolg bzw. die Problemlösekompetenzen haben.

3.3.2 Forschungsfragen

Vor dem Hintergrund des Hauptziels der Untersuchung, das darin besteht, herauszufinden, inwiefern ein Repräsentationstraining auf Basis selbstgenerierter Repräsentationen den Lösungserfolg und die Problemlösekompetenz von Drittklässlern bei der Bearbeitung problemhaltiger Textaufgaben steigern kann, ergeben sich die nachfolgenden Forschungsfragen. Die Differenzierung zwischen Lösungserfolg und Problemlösekompetenz beruht auf der Tatsache, dass mit Ersterem das erzielte Produkt und Letzterem der erzielte Prozess im Fokus der Betrachtung stehen sollen. Durch das Erfassen

beider Variablen können ganzheitlichere Aussagen getroffen werden (z. B. Adibnia & Putt, 1998, S. 45; Charles et al., 1987, S. 11; Lester & Kroll, 1990, S. 56; Muir et al., 2008, S. 233; s. a. Abschnitt 4.5.1.1).

Im Fokus der Fragestellungen stehen die durch die Drittklässler konstruierten Denk- und Erkenntniswerkzeuge.

Der Untersuchung liegt folgende Ausgangsfragestellung zugrunde:

Regt die Interventionsmaßnahme Viertklässler an, externe Repräsentationen beim Lösen problemhaltiger Textaufgaben häufiger zu generieren und zu nutzen?

Zahlreiche Studien deckten auf, dass Problemlösenovizen bei der Konfrontation mit einer ProblembARRIERE nur vereinzelt auf externe Repräsentationen zurückgreifen und zum Bestandteil ihres Lösungsprozesses machen konnten (siehe Abschnitt 1.4.2). Dies trifft insbesondere für junge, ungeübte Lernende zu. Den Ergebnissen von Hohn (2012, S. 113–114, 117) zufolge greifen untrainierte Drittklässler zum einen signifikant seltener auf externe Repräsentationen zurück und zeigen sich zum anderen unflexibler als Jahrgangsaltäre. Kindfield (1993, S. 10) und van Essen und Hamaker (1990, S. 308) konnten explizit zeigen, dass sich die Anzahl selbstgenerierter externer Repräsentationen durch eine Intervention signifikant steigern lässt. Im Rahmen heuristischer Strategien konnte Collet (2009, S. 190–191, 230) ebenso einen Trainingserfolg verzeichnen. Darüber hinaus konnte in unterschiedlichen Kontexten bestätigt werden, dass Lernende in Problemlöseprozessen durchaus in der Lage sind, externe Repräsentationen zu generieren (vgl. Cox, 1999, S. 360; Dörfler, 2006, S. 210; Fehse, 2001, S. 50; Prediger, 2013a, S. 172; R. Rasch, 2001b, S. 47). Es kann somit davon ausgegangen werden, dass die Anzahl selbstgenerierter externer Repräsentationen mit zunehmender Erfahrung zunimmt. Es wurde jedoch herausgestellt, dass die Konstruktion adäquater, externer Repräsentationen erst erlernt und trainiert werden muss, um sie als Denk- und Erkenntniswerkzeuge nutzen zu können (z. B. Cox, 1999, S. 360; s. Abschnitt 1.5.5.4). Die Lernenden benötigen Unterstützung beim Externalisieren ihrer mentalen Modelle.

Die Untersuchung basiert daher auf der Annahme, dass sich Grundschulkinder dazu animieren lassen, externe Repräsentationen zu konstruieren und aktiv in ihren Lösungsprozess einzubeziehen. Ob dies für die vorliegende Altersgruppe sowie für die in der Untersuchung herangezogenen problemhaltigen Textaufgaben zutrifft, muss zunächst bestätigt werden, um später überprüfen zu können, ob das Repräsentationstraining eine Erhöhung des Lösungserfolgs und der Problemlösekompetenzen bewirken kann. Diese Überprüfung erfolgte im Rahmen der Pilotierung. Es ist zu erwarten, dass

bisherige empirische Befunde auf das Lösen problemhaltiger Textaufgaben übertragbar sind.

Es lässt sich folgende Wirksamkeitshypothese ableiten:

Das Repräsentationstraining befähigt Viertklässler, externe Repräsentationen beim Lösen problemhaltiger Textaufgaben zu generieren und zu nutzen.

Diese Wirksamkeitshypothesentestung, dass ein Vergleich *innerhalb* der Gruppen vorhersagekonforme Verbesserungen aufdeckt, muss sichergestellt werden, bevor getestet wird, ob sich die einzelnen Gruppen nach der Intervention vorhersagekonform unterscheiden und somit ein Vergleich *zwischen* den Gruppen erfolgen kann (Wirksamkeitsunterschiedshypothesentestung) (Hager, 2000a, S. 212, 2008, S. 728).

Forschungsfrage 1: Effekt der Interventionsmaßnahme

Können Drittklässler durch die Interventionsmaßnahme ihren Lösungserfolg, ihre Problemlösekompetenzen und die Anzahl selbstgenerierter externer Repräsentationen erhöhen?

Die Studie von Geier (2012, S. 286) zeigte, dass die Implementierung problemhaltiger Textaufgaben – wie sie auch in vorliegender Studie verwendet werden – in den Mathematikunterricht sowohl für Leistungsschwächere als auch Leistungsstärkere mit einer positiven Leistungsentwicklung einherging, jedoch nur für Letztere signifikant wurde. In der Sekundarstufenuntersuchung von Collet (2009, S. 230) konnte auch eine Leistungssteigerung der Kontrollgruppe verzeichnet werden. Hieraus lässt sich schließen, dass bereits die Integration problemhaltiger Textaufgaben über die Gruppen hinweg eine positive Leistungsentwicklung bedingen kann.

Folgende Hypothesen lassen sich ableiten:

- 1.1 *Über alle Gruppen hinweg lösen Drittklässler problemhafte Textaufgaben nach der Interventionsmaßnahme erfolgreicher als davor.*
- 1.2 *Über alle Gruppen hinweg verfügen Drittklässler nach der Interventionsmaßnahme über höhere Problemlösekompetenzen als davor.*

Der Grund dafür, dass ungeübte Problemlösende ihre Problemlösung selten auf externe Repräsentationen stützen, wird u. a. in ihrem fehlenden Vor- und Erfahrungswissen gesehen (vgl. Abschnitt 1.5.3.4, 1.5.4 sowie 3.2). Folglich kann ihnen (noch) kein breit gefächertes Repräsentationsrepertoire zur Verfügung stehen, auf das sie je nach Situation zurückgreifen können. Vor dem Hintergrund der Befunde, dass die Generierung externer Repräsentationen trainiert werden kann, gilt es zu überprüfen, ob diese Er-

gebnisse – auch die Ergebnisse der Pilotierung – replizierbar sind (siehe Abschnitt 1.5.5.4). Folgende Hypothese ergibt sich hieraus:

- 1.3 *Über alle Gruppen hinweg generieren Drittklässler nach der Interventionsmaßnahme mehr externe Repräsentationen als davor.*

Forschungsfrage 2: Effekt des Repräsentationstrainings

Unterscheiden sich trainierte und untrainierte Drittklässler nach der Interventionsmaßnahme in ihren Leistungen hinsichtlich des Lösungserfolgs, des Problemlösekompetenzzuwachses und der Anzahl selbstgenerierter externer Repräsentationen?

Es wird davon ausgegangen, dass Lernenden, die ihre Lösungsgedanken und -ideen zu Papier bringen, mehr Kapazitäten für den Lösungsprozess zur Verfügung stehen und es ihnen eher gelingt, die Lösung des Problems an und mit ihren externen Repräsentationen zu erarbeiten, als wenn diese Denk- und Erkenntniswerkzeuge nicht explizit gefördert werden (vgl. Abschnitt 1.5.5 und 3.2). Dementsprechend höhere Problemlösekompetenzen werden erwartet. Es wird vermutet, dass die Leistungsentwicklung und letztlich der Kompetenzzuwachs stärker in der Erfassung des Prozesses als in der Erfassung des Produktes erkennbar wird. Dies beruht auf der Annahme, dass der Lösungserfolg ausschließlich die Richtigkeit der erzielten Lösungszahl berücksichtigt. Schritte, die dem Erreichen der Lösung dienlich, aber noch nicht hinreichend sind, können erworbene Teilkompetenzen nur mit Blick auf den Prozess, nicht mit Blick auf das Ergebnis, nachweisen (vgl. Abschnitt 4.5.1.1). Weil sich Problemlösekompetenzen langfristig entwickeln müssen (siehe Abschnitt 1.5.5.4), wird davon ausgegangen, dass Trainingseffekte sowohl in der Betrachtung des Lösungsprozesses als auch in der Betrachtung des Endproduktes erzielt werden, stärker jedoch in Ersterem.

Es lassen sich folgende Wirksamkeitshypothesen ableiten (siehe Tabelle 3.1):

- 2.1 *Drittklässler lösen problemhaltige Textaufgaben erfolgreicher, wenn sie trainiert werden.*
- 2.2 *Drittklässler erzielen einen höheren Problemlösekompetenzzuwachs, wenn sie trainiert werden.*

Ferner ist davon auszugehen, dass sich die Konstruktion externer Repräsentationen trainieren lässt (vgl. Abschnitt 1.5.6 sowie 3.2). Aufgrund dessen wird vermutet, dass Kinder, die trainiert wurden, externe Repräsentationen zu generieren, diese auch explizit häufiger in ihren Lösungsprozess integrieren als Kinder, die nicht trainiert wurden. Demzufolge ist ein Unterschied zwischen Trainings- und Nichttrainingsklassen zugunsten erstgenannter zu erwarten. Es resultiert folgende Hypothese:

2.3 Drittklässler generieren nach der Interventionsmaßnahme mehr externe Repräsentationen, wenn sie trainiert werden.

Die vermuteten Entwicklungstendenzen können Tabelle 3.1 entnommen werden.

Tabelle 3.1. Hypothesen zu dem Lösungserfolg, den Problemlösekompetenzen und der Anzahl selbstgenerierter externer Repräsentationen trainierter und untrainierter Lernender

	Erwartete Richtung des Effekts	
	T+	T-
Lösungserfolg	++	+
Problemlösekompetenzen	+++	+
Anzahl selbstgenerierter Repräsentationen	+++	+

Anmerkung. T+ = Training; T- = kein Training. Erwartete Zuwächse: 0: es wird kein Zuwachs erwartet; +: es wird ein geringer Zuwachs erwartet; ++: es wird ein Zuwachs erwartet; +++: es wird ein deutlicher Zuwachs erwartet.

Forschungsfrage 3: Effekt kommunikativer Zweiersettings¹⁵⁹

Unterscheiden sich Drittklässler, die sich unter Gleichaltrigen austauschen durften, in ihren Leistungen hinsichtlich des Lösungserfolgs, des Problemlösekompetenzzuwachses und der Anzahl selbstgenerierter externer Repräsentationen von Drittklässlern, die sich nicht unter Gleichaltrigen austauschen durften?

Empirischen Studien zufolge kann ein intersubjektiver Austausch für den kognitiven Gewinn lernförderlich sein (vgl. Abschnitt 2.4). An dieser Stelle wird auf die Studie von Webb (1989, S. 35) verwiesen, die beim Vergleich von 19 Untersuchungen aufdecken konnte, dass es einen positiven Zusammenhang zwischen beobachteter Interaktion und erzielter Leistung bei Lernenden gibt, die ihre Gesprächspartner im Austausch unterstützen. Dieses Bild spiegelte sich jedoch nicht konsistent wider. Situationen, in denen kein oder nur ein Interaktionspartner profitiert, blieben nicht aus. Kramarski et al. (2010, S. 185) stellen die Lernunterstützung des angeleiteten kommunikativen Austausches für Lernende heraus. Ihren Ergebnissen zufolge gehören Drittklässler, die sich explizit mit dem Verstehen der Aufgabe („What is the problem about?“), dem

¹⁵⁹ In Anlehnung an die prozessbezogene Kompetenz des Kommunizierens wird anstelle intersubjektiven Austauschs der Bezeichnung *kommunikative Settings* (K) gefolgt (vgl. KMK 2005b, S. 7–8). Ziel war es, eine eindeutige, abkürzende Schreibweise festzulegen (A könnte in diesem Zusammenhang sowohl für Argumentation als auch für Austausch verstanden werden). Weshalb spezifisch Zweiersettings fokussiert werden, wird in Abschnitt 4.2.1.2 erläutert.

Herstellen von Verknüpfungen („What is the similarity or the difference between the two problems/explanations?“, „How do you justify your conclusion?“), der Planung und Auswahl adäquater Strategien („What strategy/representation can be used and how in order to solve the problem/task?“, „Why?“) und dem Reflektieren und Kontrollieren ihres Lernprozesses („Does the solution make sense?“, „Is the solution reasonable?“) auseinandersetzen und darüber kommunizieren, zu den besseren Problemlösern als Drittklässler, die diese Unterstützungsmaßnahme nicht erhalten (Kramarski et al., 2010, S. 185). Darüber hinaus konnten beispielsweise Söbbeke und Steenpaß (2010, S. 222–223) bestätigen, dass Lernende in der Lage sind, externe Repräsentationen auf unterschiedliche Weise zu deuten und sich darüber auszutauschen.

Es gilt daher, zu prüfen, ob ein Austausch unter Gleichaltrigen die Entwicklung von Repräsentations- und Problemlösekompetenzen fördern kann. Folglich könnte der Austausch als zusätzliches Training bzw. als eine kommunikative Unterstützungsmaßnahme interpretiert werden – ein Alternativtraining, das auf der Ebene der Gleichheit und nicht der Belehrung erfolgt.

Nachfolgende Wirksamkeitsunterschiedshypothesen lassen sich ableiten (siehe Tabelle 3.2):

- 3.1 *Drittklässler lösen problemhaltige Textaufgaben nach der Interventionsmaßnahme erfolgreicher, wenn sie sich in kommunikativen Settings austauschen.*
- 3.2 *Drittklässler erzielen einen höheren Problemlösekompetenzzuwachs, wenn sie sich in kommunikativen Settings austauschen.*

Dass die Implementierung kommunikativer Settings einen bedeutsamen Einfluss auf die Anzahl selbstgenerierter externer Repräsentationen ausübt und die Kinder dazu bewegt, mehr externe Repräsentationen zu generieren, wird nicht erwartet. Dies beruht auf der Tatsache, dass sich die Kinder im kommunikativen Austausch unter Gleichaltrigen zentral auf ihre eigene(-n) externe(-n) Repräsentation(-en) stützen. Sie nutzen sie, um ihre individuellen mathematischen Entdeckungen zu beschreiben, zu formulieren und auch zu begründen, sodass ihr Denken schwerpunktmäßig auf den eigenen Lösungsprozess gerichtet ist (vgl. Abschnitt 2.3 sowie 1.5.5.1). Jeder Einzelne steht vor der Aufgabe, sein Verständnis seinem Partner mitzuteilen (z. B. Meissner, 1979, S. 307 s. a. Abschnitt 2.2.5). Durch das Hineindenken in die Lösung des Gegenübers wie auch das Nachvollziehen dessen Konstruktions- und Nutzungsgedankens ist nicht auszuschließen, dass Denk- und Erkenntniswerkzeuge des Partners in anschließenden Problemlösungen erinnert, aufgegriffen und ausprobiert werden. Im Gegenteil, wie in Abschnitt 2.4.1 aufgezeigt, kann das Voneinander- und Miteinander-Lernen durchaus hohes Lernpotenzial aufweisen und Impulse für die Weiterarbeit geben. Dass Kinder kognitiv voneinander profitieren können, wird – wie in Hypothese 3.1 und 3.2 darge-

legt – angenommen. Jedoch wird keine bzw. nur eine geringfügige Beeinflussung des intersubjektiven Austauschs auf die Anzahl selbstgenerierter externer Repräsentationen erwartet. Die Vorhersage, dass intersubjektiver Austausch mit hoher Wahrscheinlichkeit zu einer größeren Anzahl von Repräsentationskonstruktionen führt, kann daher nicht getroffen werden, weshalb sich folgende Hypothese ergibt:

3.3 *Drittklässler, die sich in kommunikativen Settings austauschen, unterscheiden sich in der Anzahl selbstgenerierter externer Repräsentationen nicht von Drittklässlern, die sich nicht in kommunikativen Settings austauschen.*

Die vermuteten Entwicklungstendenzen können Tabelle 3.2 entnommen werden.

Tabelle 3.2. Hypothesen zu dem Lösungserfolg, den Problemlösekompetenzen und der Zahl selbstgenerierter externer Repräsentationen der Gruppe mit und ohne Kommunikation

	Erwartete Richtung des Effekts	
	K+	K-
Lösungserfolg	++	+
Problemlösekompetenzen	+++	+
Anzahl selbstgenerierter Repräsentationen	0	0

Anmerkung. K+ = kommunikative Zweiersettings; K- = keine kommunikativen Zweiersettings. Erwartete Zuwächse: 0: es wird kein Zuwachs erwartet; +: es wird ein geringer Zuwachs erwartet; ++: es wird ein Zuwachs erwartet; +++: es wird ein deutlicher Zuwachs erwartet.

Forschungsfrage 4: Effekt unterschiedlicher Gruppen

Welchen Einfluss haben die experimentellen Bedingungen aus den Faktoren Repräsentationstraining und kommunikative Zweiersettings auf die Leistungen (Lösungserfolg, Problemlösekompetenzen, Anzahl selbstgenerierter externer Repräsentationen) von Drittklässlern?

Vor dem Hintergrund, dass sowohl ein Repräsentationstraining als auch ein angeleiteter kommunikativer Austausch für Lernende eine Unterstützungsmaßnahme darstellen kann, ist davon auszugehen, dass ein Mathematikunterricht, der die Lernenden trainiert, adäquate Repräsentationen zu konstruieren, und der einen kommunikativen Austausch über die individuellen Lernprozesse (*Was, Wie, Wann und Warum*) berücksichtigt, in seiner Wirksamkeit einem Mathematikunterricht, der nur einen Faktor (Training oder Kommunikation) bzw. keinen der beiden Faktoren fokussiert, überlegen ist.

Hieraus lassen sich folgende Wirksamkeitsunterschiedshypothesen ableiten (siehe Tabelle 3.3):

- 4.1 *Trainierte Drittklässler, die sich zusätzlich unter Gleichaltrigen austauschen, lösen problemhafte Textaufgaben nach der Interventionsmaßnahme am erfolgreichsten.*
- 4.2 *Drittklässler, die weder trainiert wurden noch sich unter Gleichaltrigen austauschen, steigern nach der Interventionsmaßnahme ihren Lösungserfolg am geringsten.*
- 4.3 *Trainierte Drittklässler, die sich zusätzlich unter Gleichaltrigen austauschen, erzielen nach der Interventionsmaßnahme den größten Problemlösekompetenzzuwachs.*
- 4.4 *Drittklässler, die weder trainiert werden noch sich unter Gleichaltrigen austauschen, erzielen nach der Interventionsmaßnahme den geringsten Problemlösekompetenzzuwachs.*

Tabelle 3.3. Hypothesen zu dem Lösungserfolg, den Problemlösekompetenzen und der Anzahl selbstgenerierter externer Repräsentationen der vier experimentellen Bedingungen

	Erwartete Richtung des Effekts			
	T+K+	T+K-	T-K+	T-K-
Lösungserfolg	+++	++	++	+
Problemlösekompetenzen	+++	++	++	+
Anzahl selbstgenerierter Repräsentationen	+++	+++	+	+

Anmerkung, T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings. Erwartete Zuwächse: 0: es wird kein Zuwachs erwartet; +: es wird ein geringer Zuwachs erwartet; ++: es wird ein Zuwachs erwartet; +++: es wird ein deutlicher Zuwachs erwartet.

Bezüglich der Anzahl selbstgenerierter externer Repräsentationen wird kein Unterschied zwischen den trainierten Klassen mit und ohne kommunikativen Settings bzw. den nicht-trainierten Klassen mit und ohne kommunikativem Setting erwartet. Als Konsequenz ergeben sich folgende Hypothesen:

- 4.5 *Trainierte Drittklässler, die sich zusätzlich unter Gleichaltrigen austauschen, unterschieden sich beim Lösen problemhafte Textaufgaben nach der Interventionsmaßnahme nicht in der Anzahl selbstgenerierter externer Repräsentationen von trainierten Drittklässlern, die sich nicht austauschen durften.*

4.6 *Drittklässler, die weder trainiert werden noch sich unter Gleichaltrigen austauschen, generieren gleich viele externe Repräsentationen wie untrainierte Drittklässler, die sich austauschen.*

Die vermuteten Entwicklungstendenzen können der Tabelle 3.3 entnommen werden.

Fragestellung 5: Effekt externer Repräsentationen

Welchen Einfluss kann selbstgenerierten externen Repräsentationen auf den Lösungserfolg bzw. die Problemlösekompetenzen zugesprochen werden und welche Repräsentationen lassen sich als gute Prädiktoren zur Vorhersage des Lösungserfolgs bzw. der Problemlösekompetenzen identifizieren?

In der Fachdidaktik lässt sich eine Vielzahl an Repräsentationsformen identifizieren, die Lernende im Rahmen des Problemlösens konstruieren können. In Abschnitt 1.5.6 wurden diese Formen im Hinblick auf charakteristische Eigenschaften und Merkmale präzisiert. In diesem Zusammenhang wurde herausgestellt, dass es nicht *die* Repräsentationsform gibt, die Einsicht und Erfolg garantiert (vgl. Cox, 1999, S. 344).

In der Pilotierung stand im Fokus der Betrachtung, welche externen Repräsentationen Grundschulkinder beim Lösen problemhaltiger Textaufgaben konstruieren und ob diese konform gehen mit den Repräsentationsformen, die die Fachdidaktik vorsieht (vgl. Abschnitt 4.2.3). Von Interesse war dabei, herauszufinden, welche inter-individuellen Unterschiede sich bei der Bearbeitung der gleichen problemhaltigen Textaufgaben durch verschiedene Kinder ergeben. Darüber hinaus galt es in der Vor-Analyse zu entschlüsseln, welche Eigenschaften die Schülerrepräsentationen verkörpern und charakterisieren, um in der Hauptuntersuchung aufzuklären zu können, welchen Einfluss externe Repräsentationen auf den Lösungserfolg und die Problemlösekompetenzen nehmen.

Zu untersuchen ist, welche externen Repräsentationen beim Lösen problemhaltiger Textaufgaben gute Prädiktoren für die Vorhersage des Erfolgs bzw. der Problemlösekompetenzen sind. Dabei werden sowohl der Lösungsprozess als auch das Endprodukt vergleichend in den Blick genommen. Würde ausschließlich der Lösungserfolg einbezogen, blieben erreichte Teilkompetenzen außerhalb des Blickfeldes. Erzielt ein Kind ein falsches Ergebnis, sein dokumentierter Lösungsprozess lässt aber eine Kompetenzsteigerung durch die Generierung einer externen Repräsentation erkennbar werden, so kann dieser Einfluss der externen Repräsentationen nur erfasst werden, wenn die Problemlösekompetenzen operationalisiert werden, nicht jedoch der Lösungserfolg. Da sich Problemlösekompetenzen langsam und schrittweise entwickeln, ist zu erwarten, dass nicht jede konstruierte externe Repräsentation direkt zum Lösungserfolg führt, wes-

halb ein größerer Einfluss der externen Repräsentationen auf den Kompetenzzuwachs prognostiziert wird als auf den Lösungserfolg.

Im Zentrum des Interesses stehen folgende Wirksamkeitsunterschiedshypothesen:

- 5.1 *Das Generieren externer Repräsentationen beeinflusst den Lösungserfolg zum Zeitpunkt des Posttests positiv.*
- 5.2 *Das Generieren externer Repräsentationen beeinflusst die Problemlösekompetenzen zum Zeitpunkt des Posttests positiv.*
- 5.3 *Der Einfluss selbstgenerierter externer Repräsentationen auf die Problemlösekompetenzen ist zum Zeitpunkt des Posttests größer als auf den Lösungserfolg.*

Wird berücksichtigt, dass jede Repräsentationsform Vor- und Nachteile aufweist, aber gleichzeitig zur Aufgabe wie auch zum Lösen selbst passen muss (vgl. Abschnitt 1.5.5.4), ist davon auszugehen, dass keine alleinige Repräsentationsform in der Vorhersage dominiert. Während in der Literatur insbesondere die depiktionalen Repräsentationen als lernförderlich akzentuiert werden (vgl. Abschnitt 1.5.6.1), spiegelte sich in der Untersuchung von Hohn (2012, S. 110–111) bei den Räuber-, Dampfer-, Märchen- sowie Schneckenaufgaben eine Dominanz deskriptionaler Repräsentationen wider, während lediglich bei der Eislaufgabe depiktionale Repräsentationen dominierten. Diese Ergebnisse beruhten auf einer Ist-Stand-Analyse, bei der die Kinder weder trainiert noch explizit aufgefordert wurden, externe Repräsentationen zu generieren. Es gilt herauszufinden, ob sich diese Ergebnisse replizieren lassen oder ob ein Training, das sowohl die Generierung depiktionaler als auch deskriptionaler Repräsentationen fört, die Kinder befähigt, unterschiedliche Repräsentationen in Abhängigkeit von der Aufgabe oder ihren Präferenzen einzusetzen.



4 Methode

In diesem Kapitel werden die Methoden erläutert, mit denen die in Abschnitt 3.3.2 formulierten Forschungsfragen untersucht werden sollen. Zu Beginn werden im Rahmen der methodischen Überlegungen und Entscheidungen (Abschnitt 4.1) das Unterrichtskonzept, dem die Untersuchung folgt, und die Aufgabenauswahl begründet aufgezeigt. Sowohl das Unterrichtskonzept als auch die Aufgabenauswahl wurden im Rahmen der Pilotierung erprobt und evaluiert. Die hieraus resultierenden Erkenntnisse und Ergebnisse sowie deren Konsequenzen für die Hauptuntersuchung werden in Abschnitt 4.2 dargestellt. Im Anschluss daran stehen das Design (Abschnitt 4.3) sowie der Kern der Studie, das Training (Abschnitt 4.4), im Vordergrund. Die Variablen und ihre Operationalisierung werden in Abschnitt 4.5 ausführlich erläutert und erörtert. Es folgt ein Überblick über den zeitlichen Verlauf und die Durchführung der Studie (Abschnitt 4.6) sowie die Einordnung der Stichprobe in das Design (Abschnitt 4.7). Die Erläuterung und Begründung, der zur Auswertung verwendeten statistischen Methoden, runden den Methodenteil ab (Abschnitt 4.8).

4.1 Methodische Überlegungen und Entscheidungen

Die theoretisch fundierten Entscheidungen, die sich sowohl für das Unterrichtskonzept, an dem sich die Untersuchung orientiert (Abschnitt 4.1.1), als auch für die Wahl der Aufgaben ergeben (Abschnitt 4.1.2), werden nachfolgend aufgezeigt.

4.1.1 Unterrichtskonzept

Unterricht, der die in den vorangegangenen Abschnitten 3.1 und 3.2 genannten Schlussfolgerungen berücksichtigt, muss sich an dem Vor- und Erfahrungswissen der Lernenden orientieren (Stern, 2004, S. 11). „Doppelte Aufgabe der Instruktion ist es deshalb, dem Lernenden hinreichend Gelegenheit zu explorativem Verhalten, zu produktiven Fehlern und zu intuitiven Schlussfolgerungen zu geben, gleichzeitig aber vorwissensaktivierend, informierend und ermutigend zu wirken“ (Weinert, 1996, S. 28). Der Mathematikunterricht hat dabei die Pflicht, eine gezielte Auswahl zentraler, unterschiedlicher Denk- und Lösungsprozesse, die bei herausfordernden Problemaufgaben zu erwarten sind (Winter, 1983, S. 81), mit all ihrer Kreativität, Individualität und Dynamik aufzugreifen und transparent werden zu lassen, um Flexibilität im Denken und Handeln anzuregen und zu fördern (Lorenz, 2005, S. 152; Selter, 1994, S. 26). Hinsichtlich der Tatsache, dass intuitives Denken selten stringent und auf direktem Wege, ohne Umwege und Fehler zum Ergebnis führt, ist dies von besonderer Bedeutung (F. Link, 2011, S. 77). Es wird ein Voneinander-Lernen auf gleicher Stufe ermöglicht, wenn

individuelle Vorgehensweisen und deren Eignung in der konkret vorliegenden Situation thematisiert und reflektiert werden, und über sie nachgedacht und diskutiert wird (z. B. Lorenz, 2005, S. 154; Rathgeb-Schnierer, 2005, S. 166; Selter, 1994, S. 26–28; Winter, 1983, S. 81–82; vgl. Abschnitt 1.5.5.3). Lernende sollen sich das Potenzial argumentativen Austauschs für ihre eigenen Lösungsprozesse und persönliche Weiterentwicklung zunutze machen (vgl. Abschnitt 1.5.5 sowie 2.2). Ein regelmäßiger Austausch kann Lernende langsam an argumentative Tätigkeiten heranführen und ihnen den Zugang zum mathematischen Argumentieren erleichtern.

Demnach wurde sich für ein Unterrichtskonzept ausgesprochen, das den Lernenden ausreichend Zeit für individuelle Entdeckungen und den Austausch darüber entsprechend ihren Möglichkeiten einräumt. Gleichzeitig bietet es ihnen eine feste organisatorische Struktur, die ihnen fortwährend Kontinuität garantiert, und die notwendige Sicherheit bietet (vgl. R. Rasch, 2001b, S. 77). In jeder Problemlösestunde wurde eine neue, herausfordernde problemhaltige Textaufgabe in das Unterrichtskonzept eingebettet. In Anlehnung an das dialogische Lernen wird zwischen folgenden Unterrichtsphasen differenziert (vgl. Hahn & Janott, 2011, S. 15; Lester, Garofalo, & Kroll, 1989b, S. 26; R. Rasch, 2001b, S. 79–82, 2009, S. 89; Ruf & Gallin, 1998, S. 11):

- I. Reflexionsphase (Training),
- II. Vorstellen der Textaufgabe,
- III. explorative Phase,
- IV. kommunikative Settings.

Nachfolgend wird nicht die erste Problemlösestunde der Interventionsmaßnahme vorgestellt, sondern eine repräsentative Folgestunde. Es muss davon ausgegangen werden, dass die Lernenden bereits in einer vorhergehenden Mathematikstunde eine Problemstellung bearbeitet haben. Der trainierenden Lehrkraft lag dadurch ein Fundus selbstgenerierter Schülerrepräsentationen vor. Hiervon wurden unterschiedliche Vorgehensweisen ausgewählt und in der Folgestunde zum Gegenstand der Reflexionsphase gemacht (vgl. Abschnitt 4.4.2).

I. Reflexionsphase

Das Training findet in der Reflexionsphase statt. An dieser Stelle erfolgt lediglich ein kurzer Anriß dessen, was in der Phase des Stundeneinstiegs im Fokus stand. Für einen ausführlichen und detaillierten Einblick in das Training wird auf den Abschnitt 4.4 verwiesen, in dem die Ausgangslage, die zentralen Ziele und der Aufbau des Trainings im Fokus der Betrachtung stehen.

Die Reflexionsphase verfolgte das Ziel, eine bewusste Auswahl individueller Eigenproduktionen in den Unterricht zu implementieren, um ein breites Spektrum möglicher

Vorgehensweisen zur Verfügung zu stellen und ins Bewusstsein zu rücken sowie Problemlösenovizen Anreiz und Orientierung für den eigenen Lösungsprozess zu bieten (z. B. Dörfler, 2006, S. 212–213; Schreiber, 2014, S. 9; Söbbeke & Steenpaß, 2014, S. 13; s. a. Abschnitt 1.5.5.3). Der Vorteil wird darin gesehen, dass es sich bei der exemplarischen Auswahl nicht um von der Lehrkraft vorgegebene, sondern um selbstgenerierte Lösungsbeispiele der Lernenden handelt, und dass daher Lernen auf gleicher Stufe ermöglicht wird. Um am Denken der Kinder anzuknüpfen und eine gut überlegte Auswahl an Lösungsbeispielen auszuwählen, bezieht sich die Reflexionsphase immer auf die in der vorangegangenen Stunde von den Kindern konstruierten Denk- und Erkenntniswerkzeuge (z. B. Schütte, 2008a, S. 137). Jede Problemlösestunde wurde demzufolge mit einem direkten Bezug zur zuletzt vollzogenen Problembehandlung eröffnet.

Bei der Auswahl alternativer Vorgehensweisen standen unterschiedliche, aber zur Problemstellung passende Schülerrepräsentationen im Vordergrund (vgl. Abschnitt 1.5.6 und 4.4.4). Die Lernenden standen vor der Aufgabe, die einzelnen Repräsentationen zu beschreiben und die Lösungsprozesse zu rekonstruieren. Darüber hinaus sollte deren Eignung sowie Vor- und Nachteile in Abhängigkeit von dem gerade zu lösenden Problem thematisiert und diskutiert werden, um die Lernenden für adäquate Konstruktionen zu sensibilisieren. Im Vordergrund standen einerseits das Kennen-, Nachvollziehen- und Verstehenlernen lösungsunterstützender Repräsentationsformen, aber andererseits auch das Herausarbeiten von Vernetzungen sowie Darstellungswechsel (vgl. Abschnitt 4.4.4). Im Klassengespräch blieben die Produzenten der jeweiligen Eigenproduktionen anonym, sofern sie nicht selbst offenlegten, dass es sich um ihre Eigenproduktion handelte. Der Verlauf wurde von der Versuchsleiterin diskret gesteuert, indem sie moderierend und unterstützend tätig war und darauf achtete, dass alle Schritte des mathematischen Argumentierens vollzogen wurden (vgl. Abschnitt 2.3 und 4.2.1.2). Für diese Phase wurden in der Regel 15 Minuten benötigt.

II. Vorstellen der Textaufgabe

Nachdem gemeinsam im Sitzkreis verschiedene Vorgehensweisen an einer konkreten Problemstellung durchdacht, reflektiert und ausgewertet wurden, las die Versuchsleiterin den Lernenden die neue Textaufgabe zweimal vor.¹⁶⁰ Von Interesse war es, unterschiedlichen Ausgangsvoraussetzungen entgegenzuwirken und eine nahezu gleiche Wissens- und Verstehensbasis zu schaffen (vgl. Abschnitt 1.5.4). Die Lernenden sollten zunächst ihre volle Konzentration dem Zuhören widmen können, bevor im Anschluss daran die Aufgabe auf einem vorgefertigten Papierstreifen, der in das Knobel-

¹⁶⁰ Bransford und Johnson (1972, S. 720) konnten in ihrer Untersuchung zeigen, dass zweimaliges Vorlesen einmaligem Vorlesen überlegen ist.

heft einzukleben war, ausgeteilt wurde. Es wurde erwartet, dass durch das Hören der Textaufgabe der Sachverhalt nicht nur visuell, sondern auch auditiv verarbeitet wird und die Lernenden dadurch eine zusätzliche Unterstützung erfahren. Die Kinder standen dann vor der Aufgabe, ihr Vor- und Erfahrungswissen auf die neue Situation zu übertragen, adaptiv anzuwenden und sich in der Lösungsfindung zu erproben.

Eine erste kurze Austauschphase nach dem Vorlesen verfolgte das Ziel, das Verständnis des Problems zu sichern, indem erste spontane Assoziationen (keine Lösungsansätze) geäußert wurden und die Sachsituation in eigenen Worten wiedergegeben wurde (Lester et al., 1989b, S. 26; vgl. Collet, 2009, S. 40). Kinder, die unmittelbar eine Lösungsidee im Kopf hatten, konnten sich nach dem Hören der Aufgabe und der kurzen Austauschphase direkt auf ihren Platz begeben und mit der Problembewältigung starten (z. B. R. Rasch, 2001b, S. 79). Bei Verständnisschwierigkeiten auf Wort-, Satz- oder Textebene oder sonstigen Barrieren, gab es die Möglichkeit, im Sitzkreis zu verbleiben und Fragen zu stellen und eigene Unsicherheiten, Unklarheiten oder Probleme anzusprechen und zu thematisieren (vgl. Abschnitt 1.5.4). Nachdem Verständnisbarrieren überwunden wurden, begaben sich auch die restlichen im Sitzkreis verbliebenen Lernenden in die individuelle Lösungsphase. Für diese Unterrichtsphase wurden ca. 5 Minuten benötigt.

III. Explorative Phase

Die Lernenden wurden in den eigenständigen Arbeitsphasen animiert und ermutigt, externe Repräsentationen zu generieren, um das aktuell zu lösende Problem zu bewältigen (siehe Abschnitt 4.4.4). Hierzu stand ihnen ein Blanko-Knobelheft zur Verfügung, in dem sie ihre Gedanken, Lösungsideen und -wege sowie ihre Selbstkonstruktionen skizzierten und protokollierten (Selter, 1994, S. 61). Die jeweils zu bearbeitende Aufgabe wurde von der Lehrkraft ausgeteilt und von den Kindern eingeklebt. Das Knobelheft kann als individuelle sowie kollektive Sammlung von Eigenproduktionen beim Lösen problemhaltiger Textaufgaben verstanden werden, das dem Lernenden „Anlaß zur Selbstbeobachtung und zur Selbstkontrolle“ gibt (Selter, 1994, S. 39).

Einschränkende Vorgaben durch die Lehrkraft gab es in der eigenständigen Lösungsphase keine. Vielmehr glich sie einem experimentellen Erproben und Erforschen des zugrunde liegenden Zusammenhangs (Fricke, 1970, S. 97). Die Lernenden widmeten sich ganz individuell der Problemstellung, indem sie den Problemraum mit geeigneten Repräsentationen zunächst zu explorieren und dann durch eine Erweiterung oder Eingrenzung zu verändern versuchten (vgl. Abschnitt 1.3.3.2). Sie standen vor der Aufgabe, den passenden Weg durch den neu konstruierten Problemraum zu finden. Es war denkbar, dass sie Heran- und Vorgehensweisen aus der Reflexionsphase aufgriffen, erprobten oder adaptiert anwandten, um eigene Erfahrungen damit zu sammeln (vgl.

Abschnitt 1.5.4). Das Externalisieren gegebener Informationen und das Arbeiten mit ihnen sollte sie dazu befähigen, Beziehungen aufzuspüren und zu verstehen sowie zu grunde liegende mathematische Phänomene zu entdecken (vgl. Abschnitt 2.3). Repräsentationale Veränderungen unterstützten diese Prozesse im Falle nicht adäquater Repräsentationen. Ziel war es, in dieser Phase ein Erfahrungswissen aufzubauen, das die Lernenden langfristig zu flexiblem Lösen befähigt.

Ferner schufen die Lernenden indirekt die Voraussetzung für argumentative Tätigkeiten, traten durch das Nutzen ihrer eigenen Denk- und Erkenntniswerkzeuge in Interaktion mit sich selbst und schärften eigene Argumentationen (vgl. Abschnitt 1.5.5.1). Weil Grundschulkinder selten von sich aus erzielte Lösungen hinterfragen und reflektieren, wurden argumentative Tätigkeiten durch einen zusätzlichen Arbeitsauftrag initiiert (Franke & Ruwisch, 2010, S. 76; Jahnke & Ufer, 2015, S. 341; Steinweg, 2001, S. 262, siehe Kapitel 1.5.3.2). Nach jeder Bearbeitung wurden sie explizit aufgefordert, ihre Entdeckungen zu beschreiben: „Stelle dir vor, du müsstest deinem Freund erklären, wie du zu deiner Lösung gekommen bist. Du hast dir die Sachaufgabe durchgelesen. Was hast du dir dann gedacht?“¹⁶¹ (Fetzer, 2011, S. 46; Rehm, 1990, S. 95, vgl. Abschnitt 1.5.6.5 und 4.2.1.2). Auf Basis ihrer individuellen Denk- und Erkenntniswerkzeuge sollten sie erlernen, ihre mathematischen Entdeckungen und aufgespürten Zusammenhänge „zu äußern (zu formulieren), diese zu hinterfragen sowie zu begründen bzw. hierfür eine Begründungsidee zu liefern“ (Bezold, 2009, S. 38, siehe Abschnitt 2.3). Um diesem Bereich sichtlich einen hohen Stellenwert einzuräumen, fand die Vorgangsbeschreibung der Kinder im Post- und Follow-up-Test immer auf einer gesonderten Seite statt (siehe Anhang B.2). Die Kinder hatten auf der Doppelseite ausreichend Platz, ihre Entdeckungen explorativ zu entwickeln, zu verschriftlichen und zu versprachlichen. Im Idealfall enthielten die Beschreibungen der Kinder bereits Begründungen oder Begründungsansätze. Da die Kinder im Prätest erwartungsgemäß weniger externalisieren, wurde aus ökonomischen Gründen nur eine Seite pro Aufgabe zur Verfügung gestellt, was sich als ausreichend erwies (siehe Anhang B.1).

IV. Kommunikative Settings

In den Austauschphasen wurden Lernende in einer kleinen, ihnen vertrauten Gruppe argumentativ tätig, indem sie sich gegenseitig ihre Entdeckungen präsentierten und Begründungen für ihr Vorgehen verbal artikulierten (vgl. Fehse, 2001, S. 68; Hahn & Janott, 2011, S. 15). Die selbstgenerierten Denk- und Erkenntniswerkzeuge sollten in der Austauschphase sowohl für den Produzenten als auch für den Rezipienten zur

¹⁶¹ Während der Arbeitsauftrag in der Pilotierung noch „Was hat dir bei der Lösungssuche geholfen? Beschreibe möglichst genau“ lautete, wurde er in der Hauptuntersuchung adressatenbezogen präzisiert (siehe differenzierte Ausführungen in Abschnitt 4.2.1.2).

Kommunikations-, Argumentations- und Gedächtnissstütze werden. Während der Produzent seine Beschreibung und Begründung an der externen Repräsentation vollziehen konnte, hatte der Rezipient die Möglichkeit, dessen Ausführungen leichter nachzuvollziehen, wenn ihm dessen Denk- und Erkenntniswerkzeug visuell vorlag und er nicht nur akustisch folgen musste (vgl. Abschnitt 1.5.5).

Im Austausch sollten nicht nur die eigenen, sondern auch die *fremden* Vorgehensweisen des Partners beschrieben, hinterfragt, diskutiert sowie verglichen und auf Richtigkeit geprüft werden. Insbesondere dann, wenn der Austausch auf unterschiedlichen Lösungen basierte, galt es, den Fehler aufzuspüren und zu korrigieren. Sowohl Prozesse des Helfens als auch Prozesse des Kollaborierens zeichneten sich in dieser Phase ab (vgl. Abschnitt 2.4.3). Als Zeitaufwand wurden 5–7 Minuten angesetzt.

Die Implementierung des Unterrichtkonzeptes forderte eine Lernkultur, die sowohl an die zu unterrichtende Lehrkraft (Abschnitt 4.1.1.1) und die Schülerinnen und Schüler (Abschnitt 4.1.1.2) als auch an die Auswahl der Problemstellungen (Abschnitt 4.1.2) konkrete Anforderungen stellte. Diesbezüglich getroffene Entscheidungen werden nachfolgend vorgestellt. In Abschnitt 4.2 erfolgt eine erste Evaluierung im Rahmen der Pilotierung.

4.1.1.1 Rolle der Lehrkraft

Die Lehrkraft steht im Unterricht, der das entdeckende Lernen fokussiert, vor der Aufgabe, ein ausgewogenes Verhältnis von individueller Wertschätzung und Intersubjektivität herzustellen (M. Meyer & Voigt, 2009, S. 20). Konkret bedeutet dies, dass sie sowohl der Individualität all ihrer Lernenden gerecht werden, als auch diese wohl durchdacht zusammenführen muss, sodass es allen Lernenden ermöglicht wird, den konkreten Unterrichtsgegenstand gleichermaßen nachzuvollziehen und zu verstehen (M. Meyer & Voigt, 2009, S. 20; Krauthausen & Scherer, 2010, S. 8; Schütte, 2008b, S. 136). In diesem Spannungsfeld und im Hinblick auf die Förderung von Problemlöse- und Argumentationskompetenzen ist darauf zu achten, dass die Lehrperson im Unterricht den gemeinsamen Austausch fokussiert und zum Nachfragen, Widersprechen, Hinterfragen und vor allem Diskutieren alternativer Vorgehensweisen anregt (Fetzer, 2009, S. 24; Hengartner, 1992, S. 22; A. Meyer, 2015, S. 16). Dabei sind dem individuellen Wissen der Schülerinnen und Schüler eine hohe Bedeutung und Wertschätzung beizumessen und mit dem eigenen Professionswissen zu vereinen (Gallin & Ruf, 1993, S. 5; Ruf & Gallin, 2005, S. 33–42; vgl. Lindmeier, 2011, S. 100). Nach Lindmeier (2011, S. 105–109) benötigt die Lehrperson neben dem Basiswissen, zu dem sie das Fachwissen einerseits und das fachdidaktische Wissen andererseits zählt, reflexive und aktionsbezogene Kompetenzen. Nachfolgend werden die beiden letzten Kompe-

tenzen aufgegriffen und erläutert. Erstere wird an dieser Stelle vorausgesetzt und nicht explizit erwähnt.

Auf die Frage danach, wie Lehrende in diesem Zusammenhang die Selbsttätigkeit und Eigenständigkeit ihrer Lernenden fordern und fördern können, antworten M. Meyer und Voigt (2009, S. 20) mit der Implementierung einer Unterrichtskultur, die auf Entdeckeraufgaben und deren Bearbeitungsfreiheiten basiert (vgl. Boaler, 2003, S. 8; Schütte, 2008b, S. 136). Dies hat zur Folge, dass unterschiedliche Vorgehensweisen, die aus Bearbeitungsfreiheiten erwachsen, zum Ausgangspunkt der Reflexion werden müssen (vgl. Abschnitt 1.5.5.3). Die Herausforderung auf Lehrerseite besteht darin, die verschiedenen Lösungsansätze der Kinder zu moderieren und mit wohlüberlegten Frage- und Impulstechniken Begründungen zu wecken (Krauthausen & Scherer, 2010, S. 8; Scherer & Steinbring, 2001, S. 199–200; Selter, 1994, S. 33). Die Lehrkraft „stellt derweil das Gerüst (engl. ‚scaffold‘) für das Gespräch“ (Brunner, 2014b, S. 235). Der Lehrer wird als Vermittler zwischen „inventions“, sämtlichen Vorgehensweisen der Lernenden, und „conventions“, mathematischen Konventionen, gesehen (Lampert, 1990, S. 254): „I deliberately acted as a mediator between students‘ thinking and mathematical conventions as I moderated discussions and taught students about the meaning of symbols and expressions“.

Gerade in der Reflexionsphase ist die *aktionsbezogene* Kompetenz der Lehrkraft ausschlaggebend gefordert (Vollstedt, Ufer, Heinze, & Reiss, 2015, S. 578; vgl. Lindmeier, 2011, S. 108). Der Lehrkraft muss es unmittelbar im Unterrichtsgespräch gelingen, adäquat auf die individuellen Schülerbeiträge zu reagieren, „um Lernprozesse zu initiieren, aufrecht zu erhalten oder auf relevante Fragen zu fokussieren“ (Vollstedt et al., 2015, S. 578; vgl. Lampert, 1990, S. 264). Die Anforderungen sind sowohl aus fachlicher als auch aus didaktischer Sicht herausfordernd, da für die Reflexion der eigenen Handlung kaum Zeit bleibt und die Reaktionen daher „einen spontanen und unmittelbaren Charakter aufweisen“ (Vollstedt et al., 2015, S. 578).

Darüber hinaus muss die Lehrkraft im Vorfeld entscheiden, welche Eigenproduktionen sich eignen, um im Plenum besprochen und diskutiert zu werden, aber auch, wie viele sich aufzugreifen lohnen (vgl. Scherer, 2014, S. 54; Scherer & Steinbring, 2001, S. 189; Thies, 2002, S. 86; Vollstedt et al., 2015, S. 578; E. C. Wittmann, 1993, S. 395–396, 1998, S. 160). Eine kompetente Analyse und die Diagnose der entstandenen Eigenproduktionen werden unerlässlich (Hirt et al., 2010, S. 12) und erfordern demnach *reflexive* Kompetenzen seitens der Lehrperson (Lindmeier, 2011, S. 107; Vollstedt et al., 2015, S. 578). Sie muss stets den Lernstand und Lernzuwachs der Kinder berücksichtigen, um zu entscheiden, welche Repräsentationen und Aufgaben sie auswählt und nutzt. Reflexive Kompetenzen werden benötigt, um die Vor- und Nachbereitung des Unterrichts professionell zu managen (Lindmeier, 2011, S. 106–107; Vollstedt et

al., 2015, S. 578). Das Ziel soll immer das Erzeugen und Anreichern intersubjektiven Wissens sein (Fetzer, 2007, S. 253; M. Meyer & Voigt, 2009, S. 20). Die Lehrkraft muss situationsabhängig entscheiden, wie „ausgehend von subjektiven, individuellen Vermutungen ein intersubjektives Wissen entstehen kann“ (M. Meyer & Voigt, 2009, S. 20). Je nach Situation kann es ausreichen, die Begründung auf eine Schülerlösung zu stützen. Das Heranziehen mehrerer Begründungen bringt jedoch den Vorteil mit sich, dass aufgrund vielfältiger Ansätze ein vernetztes Wissen bzw. Verständnis erzeugt werden kann (Meyer & Voigt, 2009, S. 20; Söbbeke & Steinbring, 2007, S. 63, vgl. Abschnitt 1.5.5.3–1.5.5.4). Im Allgemeinen sollen die Selektion und die Restriktion sorgfältig durchdacht sein, um ein Weiterdenken und Reflektieren realisieren zu können (Fetzer, 2007, S. 247; E. C. Wittmann, 1998, S. 160; vgl. Scherer & Steinbring, 2001, S. 189; Krauthausen, 1994, S. 13).

Ferner ist von großer Bedeutung, wie sich die Lehrkraft in der Arbeitsphase der Lernenden verhält. „Der Lehrer steht nicht *im* Weg, sondern *am* Weg – er *steht bei*“ (Krauthausen, 1994, S. 13, Hervorhebung im Original). Ein vorschnelles, lenkendes Eingreifen seitens der Lehrkraft kann die Lernenden in ihrer Kreativität und Eigentätigkeit bremsen, stören oder gar zu einer Unterbrechung ihrer Aktion(-en) führen (M. Meyer & Voigt, 2009, S. 20). Stattdessen sollte die Lehrkraft anstreben, ihre eigene Person zurückzunehmen, den Lernenden ausreichend Zeit zur Aufgabenbearbeitung und zum Austausch einzuräumen sowie sich auf die Ideen, Ansätze und Lösungen ihrer Schülerinnen und Schüler einzulassen und diese zum Unterrichtsgegenstand zu machen (Hengartner, 1992, S. 22; Krumsdorf, 2009, S. 11; Naujoks, 2002, S. 61; Selter, 1994, S. 12, 28). Sie hat die Funktion des Begleiters und Wegbereiters inne (Seidel, Blomberg, & Stürmer, 2010, S. 297; Steinbring & Nührenbörger, 2010, S. 169).

4.1.1.2 Rolle der Lernenden

Im konstruktiv, aktiv-entdeckenden Unterricht sind die Lernenden die „Konstrukteure ihres eigenen Wissens“, indem sie sich eigenständig mathematisches Wissen aneignen und sich aktiv damit auseinandersetzen (Lorenz, 2005, S. 151).

Um über ihr Lösungsvorgehen und ihre Gedanken kommunizieren zu können, benötigen Grundschulkinder eine Kommunikationsgrundlage (Krauthausen, 2001, S. 104; Lorenz, 2005, S. 153, 155 vgl. Abschnitt 1.5.5.2). Es darf nicht unterschätzt werden, dass ihnen einerseits noch die sprachlichen Voraussetzungen fehlen und die AlltagsSprache nicht genügt, um sich auf rein sprachlicher Ebene mit Mitschülern austauschen zu können (Krummheuer & Fetzer, 2005, S. 29; Lorenz, 2005, S. 153, 156; Ruwisch, 2014, S. 41). Andererseits wird von ihnen gefordert, zugrunde liegende Denkwege zu versprachlichen, obwohl es sich hierbei um einen Prozess handelt, der bei Kindern nach Lorenz und Radatz (1993, S. 41) auf der bildlichen und nicht auf der sprachlichen

Ebene erfolgt (vgl. Lorenz, 2004, S. 53, 56). Die Sprache stellt somit für viele Grundschulkinder eine Hürde dar. Diese kann mithilfe geeigneter Kommunikationsmittel überwunden werden (Lorenz, 2005, S. 153 vgl. Abschnitt 1.5.5 und 2.4). Insbesondere das schriftliche Verbalisieren der Lösungsgedanken hilft den Lernenden, den eigenen Lösungsprozess nachzuvollziehen und zu optimieren. Dabei können sie in Kommunikation mit sich selbst treten, indem sie sich auf ihre eigene Konstruktion stützen. Darüber hinaus besteht die Möglichkeit, die Eigenproduktionen im Austausch mit Gleichaltrigen oder dem Plenum als Unterstützung heranzuziehen (z. B. Fetzer, 2009, S. 21–23; Maier, 2000, S. 13; Rathgeb-Schnierer, 2005, S. 166–175; E. C. Wittmann, 1993, S. 395). Die Denk- und Erkenntniswerkzeuge der Kinder werden zu solchen Kommunikationsmitteln (Lorenz, 2005, S. 155; Selter, 1994, S. 66). Ein lernförderlicher Austausch mit dem Ziel, sich über den Lösungsprozess auszutauschen und voneinander zu profitieren, setzt voraus, dass sich die Kinder darauf einlassen, sich gegenseitig sowohl ihre individuellen Prozesse als auch Produkte vorzustellen und zu begründen (Rathgeb-Schnierer, 2005, S. 166–169). Essenziell ist an dieser Stelle auch, dass sie den Vergleich ihrer Lösung mit der des Partners (bzw. der Partner) wagen und Bezüge herzustellen bereit sind. Damit dies erfolgreich gelingt, müssen die Lernenden aufeinander eingehen, sich im richtigen Moment zurücknehmen, dem Gegenüber zuhören und offen sein für dessen Lösungsideen, -ansätze und -wege sowie dessen konstruierte Repräsentationen oder aufgetretenen Probleme. Sowohl das Vorstellen eigener Vorgehensweisen und Begründungen als auch das Einlassen auf die Lösung des Gleichaltrigen, das aktives Zuhören voraussetzt, erfordern Kompetenzen, die sich Grundschulkinder erst aneignen müssen (Rathgeb-Schnierer, 2005, S. 171). An dieser Stelle sei auf die Abschnitte 1.5.3–1.5.5 verwiesen, in denen einerseits ausführlich aufgezeigt wird, welche Vorzüge Grundschulkinder durch das Externalisieren ihrer Denkwege erfahren, aber andererseits auch welche Schwierigkeiten sie dabei begleiten.

4.1.2 Kriterien zur Aufgabenauswahl

Um dem Grundgedanken des Trainings zu entsprechen, unterschiedliche Vorgehensweisen der Kinder aufzugreifen und zum Gegenstand des Unterrichts zu machen, mussten die ausgewählten Problemstellungen gewisse Kriterien erfüllen. Sie sind jedoch nicht als *selbstwirksam* zu verstehen, weshalb es einer Lernkultur bedarf, die das Potenzial der Aufgaben ausschöpft und die individuellen Ressourcen der Einzelnen berücksichtigt (Krauthausen & Scherer, 2010, S. 7). Das in Abschnitt 4.1 vorgestellte Unterrichtskonzept kann als solche Lernkultur verstanden werden. Mit dem Ziel, ein Repräsentationsrepertoire aufzubauen, das die Kinder befähigt, je nach Aufgabe entscheiden zu können, welche externen Repräsentationen ihnen am ehesten zum Erfolg

verhelfen, wurden die problemhaften Textaufgaben entsprechend folgenden Kriterien ausgewählt:¹⁶²

- *Begründungsvielfalt*: Die Wahl der Aufgaben sollte nach Brunner (2014a, S. 125; vgl. Vollrath, 1980, S. 32) dadurch beeinflusst sein, dass nicht alle nach dem gleichen Muster begründet werden können. Vielmehr sollten sie unterschiedliche Begründungsstrukturen initiieren, um eine Begründungskultur zu konstituieren (vgl. Ruwisch, 2003, S. 6).
- *Klarheit (semantisch und sprachlich-syntaktisch)*: Es wird vielfach herausgestellt, dass die semantische und sprachlich-syntaktische Struktur von Textaufgaben die Aufgabenschwierigkeit mitbestimmt (vgl. Deseniss, 2015, S. 98–100; Duarte et al., 2011, S. 44–45; Lange, 2014, S. 178; Rösch & Paetsch, 2011, S. 71–73). Daher galt es, die Aufgaben auf das sprachliche Niveau der Grundschulkinder abzustimmen und auf unnötige, verkomplizierende Textlänge, Wortwahl oder Satzbau zu verzichten (Eichler, 2015, S. 54). Es wird empfohlen, Problemstellungen, die Barrieren auf semantischer und sprachlich-syntaktischer Ebene hervorrufen, dosiert auszuwählen und sich möglicher Problematiken bewusst zu sein (vgl. Lange, 2014, S. 178; Rösch & Paetsch, 2011, S. 71–73). Ein sprachsensibler Mathematikunterricht wirkt (nicht nur für mehrsprachige Lernende) unbeabsichtigten Verständnisschwierigkeiten entgegen, welche „eine mathematikbezogene Auseinandersetzung mit der Aufgabe verhindern würden“ (Deseniss, 2015, S. 347; vgl. Duarte et al., 2011, S. 49; Rösch & Paetsch, 2011, S. 73; s. Abschnitt 1.5.4).
- *Lösbarkeit*: Herausfordernde, aber lösbare Aufgaben bewirken eine größere Anstrengungsbereitschaft und ermöglichen dadurch das Erreichen der Zone der nächsten Entwicklung (Lange, 2014, S. 177; Vygotskij, 2002, S. 331; Wegerich, 2015, S. 266). Ferner motivieren sie Lernende zum Kommunizieren und Argumentieren (Goos et al., 1996, S. 244; Hußmann & Leuders, 2006, S. 80; Lorenz, 2004, S. 48). Konträr zu dem von Stein (1999, S. 4) aufgezeigten Prinzip der Unlösbarkeit, sollte das Ziel für Grundschulkinder erreichbar sein.
- *Mathematische Vielfalt*: Insgesamt sollte darauf geachtet werden, dass nicht alle ausgewählten Textaufgaben mit dem gleichen Grundmodell des Rechnens gelöst werden können. In der Summe sollten die geschilderten Zusammenhänge alle vier Grundmodelle des Rechens und ihre Varianten einbeziehen sowie das domi-

¹⁶² Die Reihenfolge der Kriterien ist nicht als Hierarchie oder Gewichtung zu interpretieren. Eine alphabetische Ordnung soll diesem Gedanken entgegenwirken. Um die zentralen Aspekte ins Blickfeld zu rücken, wurde sich trotz der Fülle an Informationen für eine auflistende Darstellung ausgesprochen.

nierende additive Denken der Grundschulkinder aufbrechen (R. Rasch, 2009, S. 77).

- *Motivation:* Die geschilderten Sachsituationen sollten die Kinder geistig anregen und Neugier und Interesse auslösen (Eichler, 2015, S. 54).
- *Natürliche Differenzierung/Offenheit:* Problemaufgaben, die eine gewisse Offenheit innehaben, ermöglichen Lernen auf eigenen Wegen (Gallin & Ruf, 1993, S. 5, 18; Greefrath, 2004, S. 16–17; Lange, 2014, S. 177; Ulm, 2008, S. 8). Sie regen zum Nachdenken an, lassen individuelle Entdeckungen zu und fördern unterschiedliche Heran- und Vorgehensweisen. Jeder Lernende, gleich welchen Wissenstandes, gleich welcher Lernvoraussetzungen oder gleich welchen Vorwissens, hat die Möglichkeit, entsprechend seinen Ressourcen eigene Lösungen zu entwickeln (Freudenthal, 1974, S. 167; Gallin & Ruf, 1993, S. 18; Hirt & Wälti, 2010, S. 14–15; Hußmann & Leuders, 2006, S. 70; Krauthausen & Scherer, 2011, S. 4; E. C. Wittmann, 2003, S. 29; E. C. Wittmann & Müller, 2013a, S. 6). Auch im Rahmen des Argumentierens rufen solche Aufgaben Begründungen unterschiedlichen Niveaus hervor (Vollrath, 1980, S. 32, s. Abschnitt 2.2.1 sowie 2.2.2). „*Offene Aufgaben bieten Raum für die ganz spezifische Leistungsfähigkeit des Einzelnen*. Dies ist die beste Grundlage für den weiteren Wissensaufbau“ (R. Rasch, 2008a, S. 145, Hervorhebung im Original).
- *Problemhaftigkeit:* Die Aufgaben zeichnen sich durch ihren Problemcharakter aus, sie erfordern ein Lösen „jenseits einfacher Routine“ (Ruwisch, 2003, S. 6, vgl. Abschnitt 1.4.1). Ihnen liegt eine anspruchsvolle mathematische Struktur und Komplexität zugrunde (R. Rasch, 2001b, S. 26–27, 2005, S. 94, 96, 2008a, S. 5–6, 2015a, S. 204; Ulm, 2008, S. 8), die es erschwert, die Lösung im Kopf zu produzieren.
- *Problemlösungsspektrum:* Ferner lag dem Trainingsgedanken zugrunde, ein breites Spektrum an Problemaufgaben abzubilden. Das Repräsentationstraining war nicht auf die Bearbeitung einzelner, ausgewählter Aufgabentypen ausgerichtet. Stattdessen sollten sich die Lernenden an diversen Aufgabentypen erproben, vielfältige Entdeckungen wagen sowie Erfahrungen sammeln, modifizieren und auf andere Situationen übertragen (vgl. Eichler, 2015, S. 54). Sie sollten bewusst erleben, dass externe Repräsentationen beim Entdecken unterschiedlicher, mathematischer Strukturen lösungsunterstützend sein können und das „*Weiter-Denken*“ unterstützen (vgl. Ruwisch, 2003, S. 5–6). Vielfältige Erfahrungen sollten sie in Abhängigkeit von der zu lösenden Aufgabe zu einer gezielten Repräsentationsauswahl befähigen (z. B. Cox, 1999, S. 360; s. Abschnitt 1.5.5.3).
- *Repräsentationsvielfalt:* Lernende sollten mit Problembarrieren konfrontiert werden, die es auf vielfältigen Wegen zu explorieren und zu bewältigen gilt (vgl. Deseniss, 2015, S. 99). Deshalb blieben Aufgaben unberücksichtigt, wenn be-

kannt war, dass bei der Problemlösung bestimmte Repräsentationsformen fokussiert werden und aufgrund dessen die Lösungsvielfalt eingeschränkt war (vgl. Ruwisch, 2003, S. 6; Winter, 1983, S. 82). Eine aktiv-entdeckende Auseinandersetzung mit der Problemstellung soll über eine „spielerisch-experimentierende Annäherung“ bzw. „heuristisch-forschende Erarbeitung“ möglich sein (Bauer, 1993, S. 77).¹⁶³

- *Voraussetzungslosigkeit:* Die Bearbeitung der Aufgaben sollte mit elementaren Rechenfertigkeiten gemäß der Altersstufe möglich sein und keine inhaltsspezifischen Kenntnisse erfordern (vgl. Stein, 1999, S. 4).

4.2 Erkenntnisse und Konsequenzen aus der Pilotstudie

Die Pilotstudie wurde in zwei vierten Jahrgangsstufen im zweiten Drittel des Schuljahres 2012/2013 durchgeführt. Insgesamt nahmen 39 Kinder ($N_1 = 19$, $N_2 = 20$) an der Untersuchung teil, 23 Mädchen ($N_1 = 10$, $N_2 = 13$) und 16 Jungen ($N_1 = 9$, $N_2 = 7$). Über einen Zeitraum von fünf Wochen bearbeiteten die Kinder in einer festen Mathematikstunde pro Woche eine problemhafte Textaufgabe. Um mehr Aufgaben berücksichtigen zu können, wurden vor dem Training, direkt nach dem Training sowie vier Wochen nach dem letzten Erhebungszeitpunkt zwei weitere problemhafte Textaufgaben von den Viertklässlern bearbeitet.

Der Ablauf der Unterrichtsstunden folgte dem in Abschnitt 4.1.1 vorgestellten Unterrichtskonzept. Beide Klassen wurden von der Versuchsleiterin trainiert. Der Unterschied bestand jedoch darin, dass in einer der beiden Klassen die *kommunikativen Settings* (Phase IV) entfiel. Anstelle des Austauschs hatten die Kinder dementsprechend mehr Zeit, um an der jeweiligen, problemhaften Textaufgabe zu verweilen.¹⁶⁴

Im Fokus der Pilotierung stand, möglichst viele verschiedene Problemaufgaben zu testen, das Unterrichtskonzept zu erproben und die organisatorische Struktur zu überprüfen, um noch erforderliche Feinabstimmungen und Optimierungen im Hinblick auf die Hauptuntersuchung vornehmen zu können. Für den Einsatz verschiedener Problemaufgaben sprach, dass die externen Repräsentationen der Kinder kategorisiert werden sollten. Eine begrenzte Auswahl an Aufgabentypen hätte zur Folge gehabt, dass die

¹⁶³ Diesem Kriterium kann auf den ersten Blick vorgeworfen werden, das Gleiche wie das *natürliche Differenzierungs-Kriterium* zu erfassen. Die beiden Kriterien lassen sich jedoch voneinander abgrenzen: Lernen auf unterschiedlichen Wegen muss nicht explizit auf unterschiedlichen Repräsentationen basieren, d. h., Kinder können beispielsweise auf die gleiche Repräsentationsform zurückgreifen, und trotzdem ganz unterschiedliche Denk- und Lösungswege absolvieren bzw. unterschiedliche Strategien anwenden, wenn die Aufgabe offen formuliert ist. Aus diesem Grund wird zwischen diesen beiden Kriterien differenziert.

¹⁶⁴ Für diesen Wegfall sprach das Design der Hauptstudie, da sowohl Trainingsklassen mit als auch ohne kommunikative Settings berücksichtigt wurden (vgl. Abschnitt 4.3).

Kategorisierung mit hoher Wahrscheinlichkeit vom untersuchten Aufgabentyp abhängig wäre. Um jedoch ein möglichst breites Spektrum an externen Repräsentationen erfassen und klassifizieren zu können, wurden Aufgaben aus verschiedenen Aufgabentypen herangezogen. Darüber hinaus galt es herauszufinden, ob das geplante Repräsentationstraining wirksam ist, konkret, ob eine mehrwöchige Interventionsmaßnahme Grundschulkinder anregt beim Lösen problemhaltiger Textaufgaben externe Repräsentationen zu generieren und häufiger zu nutzen (siehe Ausgangsfragestellung, Abschnitt 3.3.2).

Mit diesen Zielsetzungen ließ sich im Rahmen der Pilotierung die Prüfung der Effektivität des Trainings, und explizit dessen Auswirkungen auf die Leistungen und Kompetenzen, in Form eines Prätest-Posttest-Vergleichs nicht vereinen. Ein solcher Vergleich setzt voraus, dass den ausgewählten Testaufgaben gleiche Schwierigkeit und Vergleichbarkeit sowie den Klassen gleiche Ausgangsvoraussetzungen zugrunde liegen. Alle drei Anforderungen konnten weder sichergestellt noch kontrolliert werden, sodass die Ergebnisse hinsichtlich eines Leistungsvergleichs nicht interpretiert werden durften. Beobachtete Veränderungen werden an den entsprechenden Stellen berichtet, werden jedoch maximal als Tendenz interpretiert.

Nachfolgend werden zunächst die Umsetzung des Unterrichtskonzepts und aus der Pilotierung resultierende Schlussfolgerungen aufgezeigt und diskutiert (Abschnitt 4.2.1). Im Anschluss werden die Aufgaben vorgestellt und deren Wahl begründet (Abschnitt 4.2.2). Welche externen Repräsentationen Grundschulkinder bei der Problemlösung entwickelten und inwiefern sich diese kategorisieren ließen, wird in Abschnitt 4.2.3 vorgestellt. Darüber hinaus wurde die Pilotierung genutzt, um den organisatorischen Rahmen der Untersuchung festzulegen (Abschnitt 4.2.4). Hierbei standen sowohl die Festsetzung der Jahrgangsstufe als auch der Trainingsdauer im Mittelpunkt. Die Ausgangsfragestellung wird im Rahmen der Forschungsfragen in Abschnitt 5.2 beantwortet, und die damit einhergehende Wirksamkeitsprüfung erbracht.

4.2.1 Monitoring des Unterrichtskonzepts

Die Pilotierung ermöglichte eine Erprobung des in Abschnitt 4.1.1 aufgezeigten Unterrichtskonzepts und dessen zeitliche und organisatorische Umsetzung. Im Fokus stehen dabei sowohl die Reflexionsphase, sprich das *Training* (Abschnitt 4.2.1.1) als auch die *komunikativen Settings* (Abschnitt 4.2.1.2).

4.2.1.1 Trainingskomponenten

Wie einleitend in Abschnitt 4.2 bereits aufgezeigt, strebte die Pilotierung keinen Prätest-Posttest-Vergleich der erzielten Leistungen an. Vielmehr galt es, die Trainingskomponenten aufeinander abzustimmen sowie sicherzustellen, dass das Training die

Lernenden anregt und motiviert, individuelle Repräsentationen zu generieren und für ihren eigenen Lösungsprozess zu nutzen. Ob sich die Lernenden auf unterschiedliche Denk- und Erkenntniswerkzeuge einliessen und diese aufzugreifen bereit und fähig waren, galt es aufzudecken. Bezuglich der Unterrichtsorganisation wurde kontrolliert, ob das angesetzte Zeitfenster für die Trainingsphase angebracht erschien, um die Leistungen der Lernenden entsprechend zu würdigen, aber gleichzeitig die Konzentration und das Arbeitsgedächtnis nicht zu überlasten.

Vor dem Hintergrund, dass das Aufgreifen individueller Eigenproduktionen im Mathematikunterricht vielfach unterstützt, sogar gefordert wird (siehe Abschnitt 1.5.5.3), lag in der Pilotstudie ein weiterer Fokus auf die Klärung der Frage, wie viele exemplarische Schülerrepräsentationen im Training aufgegriffen werden können bzw. sollen. Hiermit sollte der prognostizierten Gefährdung, dass deren Berücksichtigung einer Reizüberflutung oder lediglich einem Abarbeiten von Lösungsbeispielen gleicht, ausreichend Beachtung geschenkt werden (siehe Abschnitt 1.5.5.4).

Beobachtungen zufolge ließen sich die Probanden motiviert auf die Problemstellungen ein und sammelten rege Möglichkeiten, diese zu lösen. Sie äußerten sehr schnell differenzierte Ideen, sei es rechnerisch, zeichnerisch oder auch tabellarisch, wie an die Problemstellung herangegangen werden könnte. Vereinzelt wurde auch die Möglichkeit, Stichpunkte in Form von Rechennotizen aufzuschreiben, genannt. Die von den Lernenden konstruierten Denk- und Erkenntniswerkzeuge eigneten sich, sie zum Gegenstand der Reflexion zu machen. Vor diesem Hintergrund musste für das Training eine Auswahl an Schülerrepräsentationen getroffen werden. Unter Berücksichtigung der einzelnen Kategorien, denen sich die Eigenproduktionen zuordnen ließen (vgl. Abschnitt 4.2.3.2), wurde zunächst der Empfehlung, wenige, dafür zentrale Eigenproduktionen, zu implementieren, entsprochen (vgl. Bauer, 1993, S. 79; Thies, 2002, S. 86; Verboom, 2014a, S. 43).

Zeitlich standen für das Training 15 Minuten zur Verfügung, was sich konzentrations- und motivationsbedingt als günstig erwies. Durch den hohen Anteil an Eigenaktivität seitens der Lernenden kann diese Phase nicht als lehrerzentriert interpretiert werden. Sie lebte vielmehr von den Aussagen, Impulsen und Vorschlägen sowie Nachfragen der Lernenden, die in erster Linie an die gesamte Gruppe gerichtet waren. Aufgrund der anfänglichen Tendenz, wenige Vorgehensweisen zu fokussieren, wurden nur drei unterschiedliche Eigenproduktionen in der Reflexionsphase aufgegriffen (Verboom, 2014a, S. 43). Für jede Thematisierung standen demnach fünf Minuten zur Verfügung.

Bei der Auswahl der externen Repräsentationen wurde bevorzugt darauf geachtet, dass die Rekonstruktion der zugrunde liegenden Struktur für die Grundschulkinder leicht möglich und vor allem nachvollziehbar war. Darüber hinaus wurde das Augenmerk auf ein möglichst breites Spektrum an Denk- und Erkenntniswerkzeugen gerichtet. In

Situationen, in denen sich die Lernenden einfach in die Vor- und Herangehensweisen hineindenken und ihnen folgen konnten, blieb Zeit, eine weitere Repräsentation zu analysieren. Im Verlaufe des Trainings zeigte sich, dass auch vier Eigenproduktionen in dem Zeitfenster gut zu bewältigen waren, fünf jedoch die Konzentration und Leistungsbereitschaft der Lernenden überanstrengte. Eigenproduktionen, die den Kern sehr deutlich und nachvollziehbar herausstellten und von den Kindern u. a. als *einfach* und *genial* bezeichnet wurden, konnten in weniger als fünf Minuten analysiert werden. Darüber hinaus konnte stets beobachtet werden, dass die Lernenden externe Repräsentationen als Hilfe wahrnahmen und mit der Zeit von sich aus in den Lösungsprozess integrierten.

4.2.1.2 *Kommunikative Settings*

Zum einen wurde die Pilotierung genutzt, um die Größe der Settings zu überprüfen, zum anderen, um zu untersuchen, ob die Kinder in diesen Settings argumentativ tätig werden. Beide Aspekte werden nachfolgend diskutiert.

Größe der Settings

Weil Grundschulkinder als Novizen im Bereich des Problemlösens, des mathematischen Kommunizierens und des Argumentierens gelten und es ihnen an Erfahrung und Sicherheit, auf die sie zurückgreifen können, mangelt, wurden Tandems Settings mit mehreren Kommunikationspartnern aus folgenden Gründen vorgezogen.

Novizen sollen ihre ersten argumentativen Tätigkeiten in einem kleinen, geschützten Rahmen sammeln, in dem sie sich wohlfühlen und in dem gegenseitige Wertschätzung dominiert (Rathgeb-Schnierer, 2005, S. 169; vgl. Bruder, 2000, S. 3; Seidel et al., 2010, S. 297). Wenn sie ihrem Kommunikationspartner vertrauen und sich respektiert fühlen, wagen sie eher, ihre Entdeckungen oder erste Lösungsideen vorzustellen, Fehler zu machen oder auch Fragen zu stellen (Oswald, 1994, S. 12).

Darüber hinaus besteht bei mehr als zwei Interaktionspartnern die Gefahr, dass sich nicht alle Lernenden aktiv an der Kommunikation beteiligen und sich mit zunehmender Anzahl von Interaktionspartnern die zur Verfügung stehende Sprechzeit jedes Einzelnen reduziert (Kauffeld & Grote, 2007, S. 311). Die Möglichkeit, dass Gedachtes nicht in die Gruppendiskussion eingebracht wird, ist in größeren Gruppen wahrscheinlicher als in kleinen. Sinkt letztlich die Chance, eigene Entdeckungen einbringen zu können, so sinkt auch die Motivation des Einzelnen zum Nachdenken. Dies führt dazu, dass prinzipiell aktivierbare Ressourcen, sprich die selbstgenerierten Repräsentationen aller Gruppenmitglieder, nicht ausreichend gewürdigt und genutzt werden (z. B. Eichler, 2015, S. 57; Hahn & Janott, 2011, S. 18; Lorenz & Radatz, 1993, S. 58) oder auch Fragen unbeantwortet und Unstimmigkeiten ungeklärt bleiben (Webb, 1989, S. 32).

Kurze, oberflächliche Thematisierungen der Eigenproduktionen wären die Folge und als wenig bis gar nicht gewinnbringend einzuschätzen. Webb (1989, S. 32) konnte in ihrer Vergleichsanalyse zeigen, dass Interaktionspartner intensiv aufeinander eingingen, sich gegenseitig ihre Fragen beantworteten und sich unterstützten, wenn sie in Partnerarbeit arbeiteten. Wenn die Gruppengröße zunahm, dann blieben diese Beobachtungen aus.

Als weiterer Grund für die Implementierung von Zweiersettings sprach die Unterrichtsorganisation. Aufgrund des Unterrichtskonzeptes und des Zeitrahmens einer Unterrichtsstunde blieb für die Austauschphase ein Zeitfenster von zehn Minuten. Für eine Gruppe von vier Lernenden würde dies bedeuten, dass jeder sein Lösungsvorgehen in zwei Minuten argumentativ darlegen müsste. Dies hätte zur Konsequenz, dass die restlichen Gruppenmitglieder die für sie neuen, unbekannten Denk- und Lösungsprozesse binnen zwei Minuten rekonstruieren müssten. Zeit, um sie zu vergleichen, zueinander in Beziehung zu setzen oder gar Fehler aufzuspüren und zu diskutieren, bliebe aus. Rückfragen seitens der Mitschülerinnen und Mitschüler wären nur begrenzt bis gar nicht möglich. Vor diesem Hintergrund wird bezweifelt, dass die individuellen Eigenproduktionen in dieser Zeit verstanden werden können (vgl. Lorenz, 2005, S. 157, 2013, S. 9 sowie Abschnitt 1.5.5.4). Folglich wurden in der Pilotierung vorrangig Zweiergruppen und in Ausnahmesituationen Dreiergruppen zugelassen.¹⁶⁵

Es zeigte sich in der Pilotierung, dass die oben beschriebenen Schwierigkeiten bereits beim Zusammenschluss von drei Kindern eintraten. Zum einen war zu beobachten, dass sich in den Dreiersettings nur zwei der drei Kommunikationspartner aktiv beteiligten und somit der größte Redeanteil auf zwei Lernende fiel. Schwächeren und zurückhaltenderen Lernenden gelang es, sich zurückzunehmen und die Beobachterrolle einzunehmen. Brachten sich alle Gesprächspartner aktiv in die Kommunikation ein und diskutierten rege über ihre Lösungen und unterschiedlichen Lösungswege, so war das Zeitfenster von zehn Minuten zumeist nicht ausreichend. Um optimale Voraussetzungen für einen aktiven, konstruktiven Austausch zu schaffen und jedem Kind ausreichend Sprechzeit einzuräumen, wurden in der Hauptstudie ausschließlich Zweiersettings implementiert, aber es wurde zugleich versucht, die lernförderliche, auf Freundschaft basierende Teambildung aufrechtzuerhalten.

¹⁶⁵ Die Lehrkraft der Klasse sprach sich für die gewohnte Gruppenbildung aus, die aus der Sitzordnung resultierte. Dadurch ergaben sich neben den Zweiergruppen vier Dreiergruppen. Der Empfehlung der Lehrkraft wurde vor dem Hintergrund, dass Kooperationen, die auf Freundschaften basieren, lernförderlich sein können, entsprochen (Oswald, 1994, S. 12).

Argumentative Tätigkeiten in kommunikativen Settings

Neben gruppendifamischen Entscheidungen wurde in der Pilotstudie zum einen überprüft, ob das Lösen problemhaltiger Textaufgaben sowie die explizit vorhandene schriftliche Begründungsaufforderung argumentative Tätigkeiten hervorriefen (*epistemische Funktion*). Zum anderen wurde untersucht, ob die Probanden in den auf Kommunikation ausgerichteten Unterrichtsphasen tatsächlich argumentativ aktiv wurden (*kommunikative Funktion*). Dabei wurde stets der Vierschritt des mathematischen Argumentierens berücksichtigt: Entdecken, Beschreiben, Hinterfragen und Begründen mathematischer Gesetzmäßigkeiten und Zusammenhänge (Bezold, 2009, S. 37; vgl. Franke & Ruwisch, 2010, S. 137; Neumann et al., 2014, S. 114, siehe Abschnitt 2.2 und 2.3).

Im Hinblick auf die *epistemische Funktion* argumentativer Tätigkeiten deckte die Pilotierung auf, dass es den Lernenden trotz des begründungsinitierenden Arbeitsauftrages schwerfiel, ihre Beobachtungen zu verbalisieren. Knappe Äußerungen wie *nichts*, *meine Rechnung*, *meine Zeichnung* oder auch *mein Kopf* dominierten neben fehlenden Notizen als Antworten auf die Frage „Was hat dir bei der Lösungssuche geholfen? Beschreibe möglichst genau“. In den seltensten Fällen wurden Entdeckungen in vollständigen Sätzen beschrieben. Einige der Lernenden notierten, sie wüssten nicht, wie sie es erklären sollten, strichen die Seite durch oder ließen sie leer. Es war nicht auszuschließen, dass der Arbeitsauftrag für die Kinder einerseits nicht eindeutig und präzise genug formuliert oder andererseits neuartig war. Aufgrund dessen wurde er in Kooperation mit zwei Grundschullehrkräften überarbeitet. Hierbei wurden zwei wesentliche Aspekte fokussiert: (1) Der Aufforderungscharakter sollte klarer aus dem Arbeitsauftrag hervorgehen und den Lernenden Impulse für den Begründungseinstieg geben. (2) Um eine stärkere adressatenbezogene Aktivierung zu erzeugen, wurde die Interaktion mit einem (fiktiven) Freund simuliert (Hefendehl-Hebeker & Hußmann, 2010, S. 95; Maier, 2000, S. 11). Den Kindern wurde mitgeteilt, dass ein Freund mithilfe der Notizen das Lösungsvorgehen rekonstruieren können muss. Beide Aspekte mündeten in folgendem Arbeitsauftrag: „Stelle dir vor, du müsstest deinem Freund erklären, wie du zu deiner Lösung gekommen bist. Du hast dir die Sachaufgabe durchgelesen. Was hast du dir dann gedacht?“ Es wurde erwartet, dass dieser Arbeitsauftrag argumentative und begründende Tätigkeiten initiiert.

In den ersten Stunden der Pilotierung fiel hinsichtlich der *kommunikativen Funktion* argumentativer Tätigkeiten auf, dass einige Zweierteams bereits nach wenigen Minuten ihren Austausch beendeten. Dies war trotz der Diskrepanz der vorliegenden Ergebnisse zu beobachten. Die unterschiedlichen Ergebnisse wurden selten von Lernerseite hinterfragt, teilweise auch beide als *richtig* angenommen. Nachfragen ließen darauf deuten, dass die Lernenden beim ersten Schritt, der Beschreibung, innehielten und kei-

ne Begründung wagten bzw. für notwendig erachteten. Gegensätzliche Behauptungen wurden ohne gegenseitige Zugeständnisse und Begründungen ausgetauscht (Piaget, 1972, S. 34). In diesen Fällen blieb das bewusste Reflektieren sowohl des eigenen Lösungsweges als auch des Ergebnisses aus, sodass beides ohne Rückbezug auf die Sachsituation als richtig anerkannt wurde. Gleiches galt für die Lösung des Interaktionspartners, sie wurde ohne anzuzweifeln akzeptiert. Diesen Teams gelang es weder, sich auf die externen Repräsentationen des Kommunikationspartners einzulassen, noch sie zu hinterfragen oder gar zum Anstellen von Vergleichen zu nutzen. Das kollektive Aushandeln und Diskutieren der Lösungen blieb aus, sodass von einem reinen Informationsaustausch gesprochen werden kann. Beide nahmen eine rezipierende Rolle ein. Die Argumentation endete dabei bereits nach dem *1. Schritt*.

Die Unterrichtsbeobachtungen sowie die empirischen Erkenntnisse veranlassten die Entwicklung eines Gesprächsleitfadens, der die Kinder in den kommunikativen Settings unterstützen sollte. Es wurden vier W-Fragen (*Was? Wie? Wann? Warum?*) aufgenommen, die sich an die „self-questioning strategy of *WWWH* (*what, when, why, and how*)“ von Kramarski et al. (2010, S. 181, Hervorhebung im Original) sowie an die Unterscheidung „Erklären-was“, „Erklären-wie“ und „Erklären-warum“ von Neumeister und Vogt (2012, S. 565–568, 575–581) anlehnten.

Zunächst sollte konkret benannt und beschrieben werden, *was* entdeckt bzw. konstruiert wurde, um die Lösung zu erarbeiten. Anschließend sollten die Lernenden beschreiben, *wie* sie beim Darstellen vorgegangen sind und *wann* welcher Schritt folgte. Begründungen bzw. Begründungsideen für ihre Entdeckungen, *warum* sie so vorgegangen sind, sollten Anlass für eine rege Diskussion sein. Um den Gesprächsablauf zu festigen, wurde das Training derart abgestimmt, dass die vier W-Fragen zentraler Bestandteil der Austausch- und Reflexionsphase waren und immer wieder aufgegriffen wurden (Kramarski et al., 2010, S. 183–184; Neumeister & Vogt, 2012, S. 565–568, 575–581; vgl. Neumann et al., 2014, S. 114; Neumann & Ruwisch, 2015, S. 320; Winter, 1975, S. 109). Die Leitfrage „*Wann?*“ wurde ergänzt, um die Interaktionspartner beim Beschreiben und Begründen an die Reihenfolge der einzelnen Schritte zu erinnern (vgl. Rathgeb-Schnierer, 2005, S. 169). Wird diese eingehalten, so ist zu erwarten, dass es zum einen dem Produzenten leichter fällt, seine Begründung zu gliedern und zu strukturieren, und dass es zum anderen dem Rezipienten leichter fällt, das Vorgestellte zu rekonstruieren.

Argumentationsprozesse lassen sich im Unterrichtskonzept sowohl in der eigenständigen Lösungsphase als auch in der Austausch- und Reflexionsphase wiederfinden. In der Austauschphase werden die Kinder erstmalig zu argumentativen Tätigkeiten aufgefordert, indem sie ihr Vorgehen einem fiktiven Freund beschreiben und begründen sollen. Ob sie tatsächlich argumentativ aktiv werden, kann in dieser Unterrichtsphase,

in der die Eigenaktivität dominiert, nicht sichergestellt werden. Die Wahrscheinlichkeit erhöht sich jedoch in den Austauschphasen spätestens dann, wenn der Gesprächspartner Begründungen für das jeweilige Vorgehen einfordert. Der Gesprächsleitfaden soll das Einfordern von Begründungen begünstigen.

Letztlich können argumentative Tätigkeiten auch in dieser Phase nur impulsgebend angestoßen, jedoch nicht kontrolliert oder zielgerichtet gefördert werden. Das Klassengespräch hat die Aufgabe, die Argumentationskompetenzen der Lernenden aufzugeben, zu fördern und zu stärken. Sollten die Argumentationen in den Reflexionsphasen auf dem Niveau von Beschreibungen verharren, so hat die Lehrkraft die Möglichkeit, zu intervenieren, indem sie die Lernenden zum Begründen ermutigt und dabei unterstützend agiert.

4.2.2 Wahl der Aufgaben

Bei der Wahl der Aufgaben muss zwischen den Problemaufgaben der Pilotierung und der Problemaufgaben der Hauptuntersuchung unterschieden werden. Die ausgewählten Pilotierungsaufgaben werden in Abschnitt 4.2.2.1 vorgestellt. Welche Textaufgaben aus testtheoretischer Sicht und aus Sicht der gewählten Kriterien in der Hauptuntersuchung zu selektieren bzw. beizubehalten sind, wird begründet erläutert. Die sich hieraus für die Hauptuntersuchung ergebenden Konsequenzen werden in Abschnitt 4.2.2.2 aufgezeigt sowie die Trainingsaufgaben aufgelistet.

4.2.2.1 Aufgaben der Pilotstudie

Auf Basis der in Abschnitt 4.1.2 aufgeführten Kriterien wurden die Problemstellungen für die Pilotierung ausgewählt. Die Nummerierung gibt Aufschluss über die tatsächliche Bearbeitungsreihenfolge, auch wenn die Gruppierung gemäß dem Aufgabentyp erfolgt.

Aufgaben mit kombinatorischem Hintergrund (R. Rasch, 2008a, S. 48–49, 89; vgl. Hohn, 2012, S. 95)

- 1) Handschlagaufgabe: Murks, Quicki, Streblinde und zwei Freunde der Kinder gehen in die Ferien. Jedes Kind verabschiedet sich von jedem mit Handschlag. Wie viele Handschläge sind es?
- 10) Eisaufgabe: Emma, Tim und Paul möchten sich ein Eis kaufen. Der Eisverkäufer bietet 3 Sorten an: Schoko, Vanille und Himbeereis. Emma kauft sich 2 Kugeln. Wie viele verschiedene Möglichkeiten hat sie?

Aufgaben mit unbekanntem Anfangszustand (R. Rasch, 2008a, S. 72)

2) Teufelaufgabe: Der Teufel sagte zu einem armen Manne: Wenn du über diese Brücke gehst, will ich dein Geld verdoppeln, doch musst du jedes Mal, wenn du zurückkommst, 8 Taler für mich ins Wasser werfen. Als der Mann das dritte Mal zurückkehrte, hatte er keinen blanken Heller mehr. Wie viel hatte er anfangs?

Aufgaben mit komplexen Informationen (R. Rasch, 2008a, S. 92; E. C. Wittmann & Müller, 2013b, S. 88)

3) Pferde-Fliegen-Aufgabe: An einem Herbsttag werden in einem Stall 15 Tiere gezählt. Es sind Pferde und Fliegen. Zusammen haben sie 72 Beine. Wie viele Pferde und wie viele Fliegen sind es?

Bewegungsaufgaben einschließlich Vergleichsproblematik (R. Rasch, 2008a, S. 80, 83)

4) Zugaufgabe: Auf einem Bahnhof fahren zur gleichen Zeit zwei Züge ab. Sie fahren in entgegengesetzte Richtungen. Der eine fährt pro Stunde 80 km, der andere fährt pro Stunde 60 km. Wie weit sind die beiden Züge nach $1\frac{1}{2}$ Stunde voneinander entfernt?

9) Wettkaufaufgabe: Quicki und Murks laufen um die Wette. In der Zeit, in der Quicki 100 m läuft, schafft Murks 75 m. Quicki gibt Murks 100 m Vorsprung. Wann hat sie ihn eingeholt?

Bewegungsaufgaben (R. Rasch, 2008a, S. 85, 65)

5) Schneckenaufgabe: Eine Schnecke in einem 20 m tiefen Brunnen will nach oben auf die Wiese. Sie kriecht am Tag immer 5 m hoch und rutscht nachts im Schlaf wieder 2 m nach unten. Am wievielten Tag erreicht sie den Brunnenrand?

11) Ameisenaufgabe: Der Weg der kleinen Ameise auf dem Quadrat. Die Seite des Quadrats ist 200 m lang. Tagsüber legt die Ameise genau 200 m zurück. Aber während der Nacht bläst sie ein starker Wind die halbe Strecke, die sie während des Tages zurückgelegt hat, wieder zurück. Am Montagmorgen geht sie los. Sie läuft von A aus über B, C und D und wieder zurück zu A. Wann wird sie wieder in A ankommen?

Vergleichsaufgaben (R. Rasch, 2016, S. 59, 101)

- 6) YU-GI-OH!®-Kartenaufgabe: Lukas und Jonas haben zusammen 30 YU-GI-OH!®-Karten. Lukas hat 6 mehr als Jonas. Wie viele Karten hat Lukas? Wie viele Karten hat Jonas?
- 8) Schneeballweitwurfaufgabe: Die Kinder der 4. Klasse freuen sich über den ersten Schnee. Luisa und Eric werfen ihre Schneebälle besonders weit. Addiert man beide Strecken sind es 25 m. Eric wirft 1 m weiter als Luisa. Wie weit wirft Eric? Wie weit Luisa?

Problemaufgaben räumlich-statischer Situationen (R. Rasch, 2009, S. 83, vgl. 2008a, S. 32)

- 7) Pralinenaufgabe: Marc betrachtet ärgerlich seine Pralinenschachtel. In der Schachtel sind 6 Reihen mit Pralinen und in jeder Reihe sind 7 Stück. Doch seine beiden Brüder haben schon welche stibitzt. Rundherum fehlt die äußere Reihe, stellt er fest. Wie viele Pralinen bleiben für ihn noch übrig?

Ergebnisse und Konsequenzen aus der Pilotierung

Die Aufgaben wurden im Mathematikunterricht der Viertklässler erprobt und im Anschluss einer Itemanalyse unterzogen, um eine Qualitätseinschätzung zu erwirken.¹⁶⁶ Die Aufgaben 1 und 2 wurden vor dem Training, die Aufgaben 8 und 9 unmittelbar nach dem Training und die Aufgaben 10 und 11 acht Wochen nach dem Training erhoben. Die Aufgaben 3–7 waren „Trainingsaufgaben“. Die Zuweisung der Aufgaben erfolgte randomisiert.¹⁶⁷

Die im Zuge der Itemanalyse gewonnenen Ergebnisse wurden nicht wie üblich genutzt, um ein neues Messinstrument zu konstruieren. Vielmehr bildeten die der Itemanalyse unterzogenen Problemaufgaben mögliche Trainings- und Testaufgaben der Hauptuntersuchung ab, die fiktiv als Testaufgaben eines Messinstruments verstanden werden können. Ziel war es herauszufinden, ob Aufgabenselektionen vorgenommen werden müssen. Hierzu wurden Itemschwierigkeit, Itemvarianz und Itemtrennschärfe

¹⁶⁶ Aufgaben, die bereits in den Untersuchungen von Hohn (2012) und Groß (2013) zum Einsatz kamen, wurden bis auf wenige Ausnahmen trotzdem pilotiert. Die Handschlagaufgabe wurde aufgegriffen, um den Lernenden einen geeigneten Einstieg in problemlösende, darstellende und argumentative Tätigkeiten zu ermöglichen (vgl. „Eisbrecher-Aufgaben“ nach Lienert & Raatz, 1998, S. 115). Die Implementierung der Teufel-, Schnecken- und EisAufgabe verfolgte das Ziel, zu überprüfen, ob die Ergebnisse von Hohn replizierbar sind, wohingegen mit der Implementierung der EisAufgabe überprüft werden sollte, ob tendenziell eine Leistungssteigerung zu beobachten ist.

¹⁶⁷ An dieser Stelle wird auf die Zielsetzung der Pilotierung in Abschnitt 4.2 verwiesen. Beabsichtigt war nicht die Durchführung eines Prätest-Posttest-Vergleichs, sondern die Erprobung möglichst vieler verschiedener Aufgaben, des Unterrichtskonzeptes sowie des organisatorischen Rahmens.

bestimmt und simultan berücksichtigt, und es wurde die Erfüllung der Kriterien überprüft (vgl. Abschnitt 4.1.2).

In Anlehnung an Lienert und Raatz (1998, S. 73) wird unter dem Schwierigkeitsindex P_i der Aufgabe i bei Niveau-Tests¹⁶⁸ der prozentuale „Anteil der auf diese Aufgabe entfallenden richtigen Antworten in Beziehung zur Analysestichprobe von der Größe n “ verstanden. Demzufolge ist der Schwierigkeitsindex „bei schwierigen Aufgaben niedrig, bei leichten hoch“ (Lienert & Raatz, 1998, S. 73; vgl. Fisseni, 1997, S. 41). Auf eine Zusatzkorrektur wurde verzichtet, da mit hoher Wahrscheinlichkeit das Erraten des richtigen Ergebnisses ausgeschlossen werden kann (Kelava & Moosbrugger, 2012, S. 80). Die Itemvarianz $\sigma^2(x_i)$ erfasst die Differenzierungsfähigkeit einer Aufgabe i hinsichtlich der Stichprobe. Die Trennschärfe r_{it} einer Aufgabe i gibt den korrelativen Zusammenhang zwischen den Itemwerten x_{vi} und den Testwerten x_v der Lernenden an (Kelava & Moosbrugger, 2012, S. 84). Bezuglich der Selektion messgenauer Items stellt die Trennschärfe nach Fisseni (1997, S. 47) den wichtigsten Kennwert dar. Ein hoher Trennschärfekoeffizient einer Aufgabe bedeutet, dass die Aufgabe geeignet ist, zwischen leistungsstarken und leistungsschwachen Lernern zu unterscheiden (Lienert & Raatz, 1998, S. 78). Erstere würden die Aufgabe meist richtig beantworten, wohingegen Letztere die Aufgabe eher auslassen oder falsch beantworten würden. Da problemhafte Textaufgaben ohne zusätzliche Unterstützung für ungeübte Löser, gleich ob leistungsschwach oder -stark, eine Herausforderung darstellen und der Lösungserfolg losgelöst vom Leistungsniveau des Kindes eher gering ausfällt (siehe Abschnitt 1.4.2), wurden dementsprechend niedrige Trennschärfen erwartet. Mit zunehmender Trainingszeit wurde prognostiziert, dass die Aufgaben besser geeignet sind, um zwischen *guten* und *schlechten* Lernenden zu differenzieren, und dass sich dies in höheren Trennschärfen widerspiegelt. Dafür spricht u. a. auch das Kriterium der natürlichen Differenzierung, das den ausgewählten Aufgaben zugrunde liegt (vgl. Abschnitt 4.1.2). Die Lernenden haben den Freiraum, sich kreativ zu zeigen und entsprechend ihren Fähigkeiten Lösungen auszuarbeiten, weshalb vermutet wurde, dass mit zunehmender Trainingszeit und Erfahrung sowohl der Schwierigkeitsindex als auch der Trennschärfekoeffizient leicht steigen würden. Die Ergebnisse ließen dies jedoch nicht stringent erkennbar werden (siehe Tabelle 4.1). Die Trennschärfen r_{it} und Schwierigkeitsindizes P_i der Aufgaben 2 und 9 bzw. der Schwierigkeitsindex P_{11} der Aufgabe 11 fielen diesbezüglich besonders auf.

¹⁶⁸ Niveau-Tests grenzen sich von sogenannten Speed-Tests ab, da die Bearbeitungszeit nicht begrenzt ist. Es ist davon auszugehen, dass die zur Verfügung stehende Zeit mit einer Unterrichtsstunde ausreicht, um den Problemlösetest zu bewältigen, sodass von einem Niveau-Test gesprochen werden kann.

Tabelle 4.1. Ergebnisse der Itemanalyse für die 11 problemhaften Textaufgaben der Pilotstudie

Aufgabe	P_i	$\sigma^2(x_i)$	r_{it}
1 Handschlag	32.5	.22	.124
2 Teufel	00.0	.00	.000
3 Pferde – Fliegen	24.4	.18	.179
4 Zug	42.5	.24	.230
5 Schnecke	25.0	.19	.179
6 YU-GI-OH!®-Karten	39.4	.28	.435
7 Pralinen	62.5	.23	.305
8 Schneeballweitwurf	36.3	.23	.469
9 Wettkampf	07.5	.07	.199
10 Eis	65.0	.23	.385
11 Ameisen	17.5	.14	.435

Anmerkung. $N = 39$. P_i = Schwierigkeitsindex; $\sigma^2(x_i)$ = Itemvarianz; r_{it} = korrigierte Trennschärfe. Die Nummerierung der Aufgaben steht für die Bearbeitungsreihenfolge. Alle Angaben basieren auf aggregierten Werten von vier Ratern. Die Beobachterübereinstimmung wurde mit dem Koeffizienten Cohens Kappa κ bestimmt (Wirtz & Caspar, 2002, S. 120). Sie lagen zwischen .66 und 1.00, was für gute bis sehr gute Übereinstimmungen spricht (Döring & Bortz, 2016, S. 346; Wirtz & Caspar, 2002, S. 59). Die einzelnen aufgabenspezifischen κ -Koeffizienten können im Anhang E.1 eingesehen werden.

Ausschluss von Aufgaben

Aufgrund der deskriptiv statistischen Evaluation wurden die Aufgaben 2 (Teufel) und 9 (Wettkampf) ausgeschlossen. Ihre Schwierigkeitsindizes $P_1 = 0$ (Teufel) bzw. $P_9 = 7.5$ (Wettkampf) ließen für diese Stichprobe auf sehr schwere, für Grundschulkinder nahezu unlösbare Aufgaben schließen (vgl. Tabelle 4.1). Analog zu den Untersuchungsergebnissen von Hohn (2012, S. 64) wurde die Teufelaufgabe von keinem Kind richtig gelöst. Die Trennschärfe des Items ist demzufolge gleich 0, weil es keine Varianz gab und daher keine Korrelation mit anderen Aufgaben geben konnte. Die Aufgabe gilt als hochgradig komplex, da zum einen das empfehlenswerte Rückwärtsarbeiten für viele Lernende ungewohnt und beschwerlich ist (R. Rasch, 2008a, S. 74), zum anderen aber häufig auch die Komplexität der Aufgabe übersehen und die Lösung vermeintlich im Ergebnis der Rechenaufgaben 3–8 gesehen wird.

Die Ergebnisse der Wettkaufaufgabe, insbesondere ihr geringer Schwierigkeitsindex, waren auf den ersten Blick überraschend, da es sich um die neunte Problembewältigung handelte und eine Aufgabe des gleichen Aufgabentyps bereits vorausging (Aufgabe 4, Zug). Diese Problematik zeigte ganz deutlich, dass Aufgaben gleicher Struktur und gleichen Aufgabentyps durchaus unterschiedlich schwer sein können (z. B. Stern, 1998, S. 107–110, 2003, S. 119). Gründe können darin gesehen werden, dass beiden Aufgaben unterschiedliche Handlungsmöglichkeiten und Einschränkungen zugrunde liegen (Greeno, Moore, & Smith, 1993, S. 160–161; Stern, 2001, S. 178, 182). Während bei der Zugaufgabe die Bewegungen in entgegengesetzte Richtungen erfolgen und im zweiten Schritt die beiden zurückgelegten Distanzen addiert werden müssen, ist bei der Wettkaufaufgabe ein anderes Lösungsvorgehen notwendig. Eine Wettkaufsituation impliziert, dass die Bewegungen in die gleiche Richtung erfolgen. Das Besondere dieser Vergleichsproblematik ist, dass die Läufer weder am gleichen Punkt starten noch gleich schnell laufen. Murks, der 100 m Vorsprung erhält, startet dadurch *weiter vorne* als Quicki. Um festzustellen, dass Murks nach 400 m eingeholt wird, ist es von Vorteil, schrittweise die jeweiligen Laufintervalle (100 m für Quicki und 75 m für Murks) zu vollziehen sowie die Positionen immer wieder aufs Neue zu vergleichen. Für Kinder, die fälschlicherweise davon ausgehen, dass Quicki und Murks am gleichen Punkt starten und dann lediglich die zurückgelegten Wegstrecken der beiden direkt vergleichen (100 m zu 75 m), ist es zum einen schwer zu vereinen, dass Quicki, die doch sowieso die schnellere Läuferin ist, Murks irgendwann einholen muss, und/oder zum anderen unklar, wie mit dem 100-m-Vorsprung umzugehen ist. Unter diesen Gesichtspunkten ist nachvollziehbar, warum die Wettkaufaufgabe höhere Anforderungen an die Lernenden stellt als die Zugaufgabe. Bei diesen Schwierigkeiten ist nur geringfügig ausschlaggebend, dass den Kindern Wettkaufsituationen aus ihrem Alltag und Sportunterricht vertrauter sind, als dies bei der Bestimmung der Distanz zweier entgegengesetzt fahrender Züge der Fall ist. Hinzu kommt, dass sie sich leichter ein Bild von der Situation machen können. Mit „Das geht doch nicht“, „Geht das überhaupt, dass die auch Gleichstand haben können?“ (R. Rasch, 2008a, S. 84) oder vergleichbaren Aussagen brachten auch die Viertklässler der Pilotstudie ihre Zweifel sowie mangelnde Erfahrungen und Vertrautheit mit der Situation zum Ausdruck. Die erzielten Ergebnisse lassen sich demnach auf die semantische Struktur der Aufgabe zurückführen (vgl. Abschnitt 1.5.3.1). Die bei der Wettkaufaufgabe von den Kindern generierten Zeichnungen und Tabellen (53 % konstruierten eine Zeichnung, 10 % eine Tabelle), verhalfen jedoch selten zum Erfolg, auch wenn die gewählten externen Repräsentationen aus aufgabenspezifischer Sicht als Überwindung der Barriere geeignet scheinen.

Kelava und Moosbrugger (2012, S. 87) empfehlen, Items mit mittlerer Schwierigkeit zu berücksichtigen, da sie am ehesten deutliche Differenzierungen zwischen den Lernenden hoher Merkmalsausprägung und Lernenden niedriger Merkmalsausprägung erzeugen (vgl. Weise, 1975, S. 219). Aufgaben, deren Schwierigkeitsindex zwischen $5 \leq P_i \leq 20$ bzw. $80 \leq P_i \leq 95$ liegt, sollten nur berücksichtigt werden, wenn eine Differenzierung zwischen Kindern mit extremen Merkmalsausprägungen angestrebt wird. Beide Problemaufgaben (Teufel und Wettkauf) blieben im Training unberücksichtigt, da herausfordernde, aber lösbare Aufgaben im Vordergrund standen (Verletzung des *Lösbarkeits-Kriteriums*, Abschnitt 4.1.2). Zu schwere Trainingsaufgaben gehen mit dem Risiko einher, die Drittklässler zu demotivieren und zu entmutigen. Immerwährende Misserfolge wirken sich nachweislich negativ auf die Motivation, das Selbstwertgefühl und letztlich die Leistung aus (Eichler, 2015, S. 55; Winter, 1983, S. 83). Es ist daher zu erwarten, dass frustrierende Erlebnisse auch für den Erwerb von Repräsentationskompetenzen kontraproduktiv sind (vgl. Lange, 2014, S. 177).

Die Analyse der Zugaufgabe (4) deckte auf, dass fast alle Viertklässler, welche die Aufgabe falsch lösten (58 %), Schwierigkeiten hatten, die geschilderte Bewegung zu verstehen (vgl. Deseniss, 2015, S. 297). Sowohl die Eigenproduktionen der Kinder als auch deren Begründungen in der Reflexionsphase spiegelten Fehlvorstellungen wider. R. Rasch (2008a, S. 82) berichtete analoge Beobachtungen. Auch in ihrer Stichprobe gelang es den meisten Kindern nicht, die Situationsbeschreibung „Sie fahren in entgegengesetzte Richtungen“ in ein passendes Situationsmodell zu überführen (vgl. Deseniss, 2015, S. 297). Dass die Züge am gleichen Punkt starteten und unterschiedlich schnell fuhren, externalisierten sie sicher. Jedoch fuhren beide Züge in die gleiche Richtung, quasi nebeneinander her. Folglich lösten die Kinder die Vergleichsproblematisik, indem sie die absolvierte Strecke des langsameren Zuges von der des schnelleren Zuges subtrahierten. Würden die Züge in die gleiche Richtung fahren, wäre dies der richtige Schluss. Es ist zu erwarten, dass der Problemcharakter der Aufgabe abnimmt und der Lösungserfolg steigt, wenn erörtert werden würde, was *in entgegengesetzte Richtung fahren* bedeutet (vgl. R. Rasch, 2008a, S. 81). Aufgrund sprachlicher Barrieren wurde die Aufgabe für das Training nicht berücksichtigt (Verletzung des *Klarheit-Kriteriums*, Abschnitt 4.1.2).

Die Pralinenaufgabe (7) zählt mit einem Schwierigkeitsindex von $P_7 = 62.5$ noch zu den Aufgaben mittlerer Schwierigkeit (vgl. Weise, 1975, S. 219). Im Vergleich zu den anderen Problemaufgaben der Pilotierung erwies sie sich jedoch als die einfachste. Die Ergebnisse gehen konform mit den Erfahrungen und Einschätzungen von R. Rasch (2008a, S. 33), welche die Aufgabe der Klassenstufe 2 zuordnet. Es kann davon ausgegangen werden, dass die Textaufgabe für den Großteil der Viertklässler keinen Problemcharakter aufweist, weshalb sie ausgeschlossen wurde (Verletzung des *Prob-*

lemhaftigkeit-Kriteriums, Abschnitt 4.1.2). Sie kam außerdem nicht als Trainingsbestandteil in Betracht, weil die Kinder bei der Bearbeitung der Lösung fast ausnahmslos eine bestimmte externe Repräsentation fokussierten (Verletzung des *Repräsentationsvielfalt-Kriteriums*, Abschnitt 4.1.2). 90 % aller Viertklässler erstellten eine Zeichnung, um das Problem zu lösen, 35 % davon kombinierten die Zeichnung mit einer Rechnung. Eine Tabelle wurde nur von 3 %, ein begründender Text von keinem Kind gewählt. R. Rasch (2008a, S. 33) zog für solche Beobachtungen als Erklärung heran, dass den Kindern gerade im Hinblick der Multiplikation die rechteckige Felddarstellung vertraut ist. Unter diesem Gesichtspunkt war es nicht verwunderlich, dass sie bei der Lösung verstärkt auf die vertraute Darstellung der Zeichnung zurückgriffen.

Kriterienerfüllende Aufgaben

Die Handschlagaufgabe (1) kann mit einem Schwierigkeitsindex von $P_1 = 32.5$ noch zu den Problemaufgaben mittlerer Schwierigkeit gezählt werden. Als häufige Fehlerquelle zeigte sich, dass die von den Kindern ermittelte Anzahl an Handschlägen dem Doppelten der tatsächlichen Anzahl entsprach. In diesen Fällen brachten die Lernenden die Fehlvorstellung mit, dass sich bei einem Handschlag zwei Personen die Hände schütteln, weshalb zwei Handschläge gezählt wurden (vgl. Deseniss, 2015, S. 303). Nach diesem Verständnis wurde die Anzahl der bei einem Handschlag beteiligten Hände, hier zwei, gezählt. Dies geht einher mit der Tatsache, dass Komposita, wie hier *Handschläge* als Zusammensetzung der Substantive *Hand* und *Schläge* für junge Lernende, insbesondere für Lernende mit migrationsbedingter Mehrsprachigkeit, nicht immer leicht zu verstehen sind (vgl. Abschnitt 1.5.3.1). Eine weitere Fehlinterpretation ergab sich im Zusammenhang mit der Aussage „Jedes Kind verabschiedet sich von jedem mit Handschlag“. Einige gingen davon aus, dass sich beispielsweise Anke von Birgit, aber auch Birgit von Anke verabschiedet, woraufhin dieses Prozedere auf alle weiteren Paare übertragen wurde (vgl. R. Rasch, 2001b, S. 248). Auch in diesem Zusammenhang lag fälschlicherweise die tatsächliche Anzahl an Handschlägen verdoppelt vor. Es zeigte sich jedoch, dass Kinder, die ihre Lösungsprozesse auf Darstellungen stützten, häufiger in der Lage waren, die richtige Anzahl zu ermitteln, als Kinder, die keine Darstellungen anfertigten. Die Aufgabenkriterien sind somit erfüllt, weshalb das Problem Bestandteil der Untersuchung blieb.

Die Eisaufgabe (10) gehört, wie die Handschlagaufgabe (1), den Aufgaben mit kombinatorischem Hintergrund an. Es zeigte sich, dass die Viertklässler nach dem fünfwochigen Kurztraining und einer vierwöchigen Pause die Aufgabe 10 (Eis, $P_{10} = 65.0$) häufiger richtig lösten als die Aufgabe 1 (Handsclag, $P_1 = 32.5$) zu Beginn des Trainings, (siehe Tabelle 4.1). Dieser Anstieg war jedoch nicht signifikant, $z = -1.72$,

$p = .086$, $r = .20$.¹⁶⁹ Es sei erneut darauf hingewiesen, dass die Leistungssteigerung nur als Tendenz zu interpretieren ist und keinesfalls ausschließlich auf das Training zurückgeführt werden darf. Höhere Leistungen können durch die kurze Interventionszeit auf Erinnerungseffekte oder auch auf unterschiedliche Aufgabenschwierigkeiten zurückgeführt werden. Bezuglich der Trennschärfen konnte ein beachtlicher Anstieg von $r_{it} = .124$ (Handschlag) auf $r_{it} = .385$ (Eis) beobachtet werden. Dies kann dergestalt interpretiert werden, dass sich Aufgaben mit kombinatorischem Hintergrund eignen, um nach einem Training zwischen leistungsstarken und leistungsschwachen Lernenden zu unterscheiden (vgl. Lienert & Raatz, 1998, S. 78). Insgesamt gelang es den Grundschulkindern gut, sich in die kombinatorischen Zusammenhänge einzudenken, sodass sie in der Lage waren, vielfältige geeignete Repräsentationen zu generieren (vgl. R. Rasch, 2001a, S. 89). Aus inhaltlichen Gründen wurde die EisAufgabe im Rahmen der Intervention berücksichtigt (Bühner, 2011, S. 183).

Die Aufgaben 6 (YU-GI-OH!®-Karten) und 8 (Schneeballweitwurf) lassen sich den Vergleichsaufgaben zuordnen. Sie wiesen für Problemaufgaben zufriedenstellende, mittlere Schwierigkeiten auf, die nahezu vergleichbar sind ($P_6 = 39.4$ und $P_8 = 36.3$). Wenn verschiedene Untersuchungen zu Vergleichsaufgaben herangezogen werden, dann stimmen die in der Pilotierung beobachteten Lösungsraten nicht immer mit dortigen Beobachtungen überein. Während nur 15 % der von Stern (1998, S. 16) untersuchten Viertklässler Aufgaben gleicher Struktur richtig lösten, waren in der Untersuchung von R. Rasch bereits 33 % Erstklässler bzw. 50 % Zweitklässler mit Materialunterstützung in der Lage, die Aufgaben erfolgreich zu lösen (R. Rasch, 2009, S. 76, 79). Es muss jedoch unterstrichen werden, dass es sich bei den Aufgabenbewältigungen der Schulanfänger (Erst- und Zweitklässler) um einen sehr kleinen Zahlenraum handelte. Die vorliegenden Pilotierungsergebnisse sind nahezu übereinstimmend mit den Beobachtungen R. Raschs (2009, S. 82) bei gleicher Altersgruppe: 38 % der Dritt- und Viertklässler, die sich in ihrem Lösungserfolg nicht mehr signifikant voneinander unterschieden, lösten die Vergleichsaufgabe richtig. Den Rückgang in höheren Grundschuljahrgängen führte sie sowohl auf kaum verwendete unterstützende Maßnahmen als auch auf das dominierende Bemühen, das Problem im Kopf bewältigen zu wollen, zurück (R. Rasch, 2009, S. 82).

Ursachen für das schlechte Abschneiden der Lernenden bei Vergleichsaufgaben werden in deren semantischer Struktur gesehen (vgl. Abschnitt 4.2.2.4). Die Bezeichnungen *addiert* und *haben zusammen wecken* bei Lernenden die Assoziation des Plusrechnens. Problematisch ist jedoch, dass die beiden Besitz- bzw. Vergleichsmengen, die zu addieren sind, gesucht sind. Es müssen zeitgleich zwei Bedingungen erfüllt sein. Dem-

¹⁶⁹ Die Analyse basiert auf dem nichtparametrischen Wilcoxon-Test für gepaarte Stichproben, da die Normverteilungsannahme verletzt war (Field, 2013, S. 235).

zufolge ist es auch nicht ausreichend, wenn die Gesamtmenge lediglich halbiert wird. Die gegebene Differenz muss berücksichtigt werden, um die beiden Vergleichsmengen bestimmen zu können (vgl. Moser Opitz, 2013, S. 204; R. Rasch, 2009, S. 76; Stern, 1998, S. 116). Die Schneeballweitwurfaufgabe birgt die Gefahr in sich, dass sie eine Situation darstellt, die den Kindern aus dem Alltag eher fremd ist und daher konstruiert wirkt. Die Tatsache, dass zwei Wurfweiten addiert werden, ist eher ungewöhnlich. Dagegen sind den Kindern Situationen, in denen z. B. Gegenstände in unterschiedlichen Besitzmengen vorliegen, aufgrund der Lebensnähe durchaus vertraut. Dies könnte ein Grund dafür sein, dass sich die Aufgaben in ihrer Schwierigkeiten unterscheiden und keine Leistungssteigerung zu beobachten ist, $z = -0.93$, $p = .352$, $r = .11$.¹⁷⁰

Den Ergebnissen der Itemanalyse zufolge zählen die Bewegungsaufgaben 5 (Schnecke) und 11 (Ameise) mit zu den schwierigsten Aufgaben ($P_5 = 25.0$, $P_{11} = 17.5$), wobei auch hier die Ameisenaufgabe schwieriger zu sein scheint, obwohl sie zeitlich später behandelt wurde. Der beobachtete Leistungsabfall konnte inferenzstatistisch bestätigt werden, $z = -2.47$, $p = .014$, $r = .28$ ¹⁷¹, wird aber aufgrund mangelnder Vergleichbarkeit nicht interpretiert. Die Reflexionsphase sowie die Analyse der selbstdurchsetzten Denk- und Erkenntniswerkzeuge deckten auf, dass die Kinder große Unsicherheiten in der Konstruktion und Beschriftung eines Quadrats hatten. Sprachlich und mathematisch bedingte Schwierigkeiten (vgl. Abschnitt 1.5.3.1) erschwerten es hier den Lernenden, die Hürde zu nehmen und die Situation adäquat zu repräsentieren. Diese Verständnisschwierigkeit brachte in beiden Klassen niemand zum Ausdruck, als die Textaufgabe in der 2. Phase des Unterrichts (siehe Abschnitt 4.1.1) vorgestellt wurde, sodass die mathematische Bedeutung des Begriffes *Quadrat* als allen geläufig vorausgesetzt wurde. Im Nachhinein bestätigten die Mathematiklehrkräfte, dass Beschriftung und Konstruktion eines Quadrates bisher nicht im Fokus ihres Mathematikunterrichts standen. Mit diesem Erfahrungswissen wurde trotz der Verletzung des *Voraussetzungslosigkeits-Kriteriums* (siehe Abschnitt 4.1.2) die Stärke der Aufgabe in den Kriterien *Repräsentationsvielfalt* und *Begründungsvielfalt* gesehen. Im Gegensatz zu der Zugaufgabe verliert die Aufgabe nicht ihre Problemhaftigkeit, wenn der Begriff *Quadrat* in der Vorstellungsphase der Textaufgabe geklärt wird.

Abschließend gilt es noch, die Aufgabe (3) mit den Pferden und Fliegen aufzugreifen. Sie wurde in der ersten Trainingsstunde bearbeitet, sodass die Lernenden hauptsächlich auf ihre eigenen Ressourcen und bisherigen Explorationserfahrungen zurückgreifen konnten. Unter diesem Gesichtspunkt wurde der geringe Schwierigkeitsindex P_3

¹⁷⁰ Analog zur Analyse der Aufgaben mit kombinatorischem Hintergrund (1 und 10) basiert die Analyse der Vergleichsaufgaben (6 und 8) auf dem nichtparametrischen Wilcoxon-Test für gepaarte Stichproben, da die Normverteilungsannahme verletzt war (Field, 2013, S. 235).

¹⁷¹ Siehe Fußnote 170.

= 24.4, der die Aufgabe noch zu den Aufgaben mittlerer Schwierigkeit zählen lässt, nicht überbewertet. Die Vorgehensweisen der Viertklässler waren facettenreich und eigneten sich, in der Reflexionsphase aufgegriffen zu werden. Darüber hinaus zeigten sich die Lernenden überdurchschnittlich motiviert, die Pferde-Fliegen-Aufgabe zu lösen, sodass insgesamt das *Motivations-Kriterium* und das *Repräsentationsvielfalts-Kriterium* erfüllt waren.

Fazit

Die Ergebnisse spiegeln den mathematisch anspruchsvollen Charakter der problemhaltigen Textaufgaben wider. Es zeichnete sich eindeutig ab, dass Aufgaben gleicher, mathematischer Struktur keinesfalls gleich schwer sein müssen. Diesen Anspruch konnte und wollte das Training zu keinem Zeitpunkt erfüllen. Im Gegenteil galt es, ein Training zu implementieren, das beim Lösen einer Bandbreite problemhaltiger Textaufgaben unterstützend fungieren kann und die Kinder dazu befähigt, ihnen nicht vertraute Zusammenhänge losgelöst von dominierendem additiven Denken aufzuspüren und zu erschließen. Die Priorität der Untersuchung lag auf der Konstruktion externer Repräsentationen als Maßnahme, Lösungsbarrieren zu überwinden und den Lösungserfolg zu erhöhen. Daher haben schwierige Aufgaben, die nicht unmittelbar auf Anhieb vom Großteil der Kinder gelöst werden können, hohes Potenzial, individuelle Lösungsprozesse und Veränderungen abzubilden. Aus den Erkenntnissen der Pilotierung ließen sich nachfolgende Trainingsaufgaben ableiten.

4.2.2.2 Trainingsaufgaben

In der Hauptuntersuchung wurden schließlich 12 problemhafte Textaufgaben aus sechs verschiedenen Aufgabentypen berücksichtigt. Die Anzahl der zu bearbeitenden Problemaufgaben ergab sich aus dem zur Verfügung stehenden Untersuchungszeitraum (vgl. Abschnitt 4.2.4). Um jeden Aufgabentyp adäquat abbilden zu können, wurde sie durch zwei repräsentative Probleme repräsentiert. Eine höhere Zahl wurde ausgeschlossen, um die Wahrscheinlichkeit des Eintretens von Erinnerungseffekten nicht zusätzlich zu erhöhen.

Sechs Aufgaben gingen aus der Pilotstudie ein: Handschlag, Eis, Pferde-Fliegen, Schnecke, Ameisen, YU-GI-OH!®-Karten, Schneeballweitwurf (siehe Abschnitt 4.2.2.1). Diese repräsentierten vier der sechs Aufgabentypen:

- Aufgaben mit kombinatorischem Hintergrund,
- Aufgaben mit komplexen Informationen,
- Bewegungsaufgaben,
- Vergleichsaufgaben.

Mit dem Ziel, das Problemlösespektrum nach dem Wegfall zweier Aufgabentypen¹⁷² wieder zu erweitern, wurden zum einen *Ausgleichsaufgaben* (R. Rasch, 2001b, S. 314, 2008a, S. 104) und zum anderen *Aufgaben zum Verhältnis von Zwischenraum und Begrenzung* (R. Rasch, 2008a, S. 41, 2001b, S. 294) ergänzt. Die Ausgleichsaufgaben waren bereits durch die Räuberaufgabe¹⁷³, die Aufgaben zum Verhältnis von Zwischenraum und Begrenzung durch die Märchenaufgabe¹⁷⁴ in vorangegangenen Untersuchungen von Groß (2013, S. 89–90), Hohn (2012, S. 95–96) und R. Rasch (2001b, S. 96, 269) vertreten. Beide Aufgaben erwiesen sich als herausfordernd, aber lösbar (Groß, 2013, S. 114, 129; Hohn, 2012, S. 64; R. Rasch, 2001b, S. 96). Wenn auch die Märchenaufgabe zu den schwierigeren zählte, zeichnete sich ein hoher Nutzen für externe Repräsentationen ab (Groß, 2013, S. 145). Beide Aufgaben schienen die zugrunde gelegten Kriterien zu erfüllen (vgl. Abschnitt 4.1.2). Exemplarisch sei auf die Begründungsvielfalt, die Problemhaftigkeit, das Problemlösespektrum und die Repräsentationsvielfalt hingewiesen. Deren Integration komplettierte die zu berücksichtigenden sechs Aufgabentypen, indem mit ihnen zwei neue hinzukamen.

Jede problemhaltige Textaufgabe wurde in zwei Schwierigkeitsstufen im Training angeboten (R. Rasch, 2008b, S. 144, vgl. 2009, S. 85, 2012a, S. 7). Die Variation der Anforderungen wird durch eine Veränderung der Zahlenwerte bewirkt, um sowohl stärkeren als auch schwächeren Problemlösenden die Aufgabenbewältigung zu ermöglichen und Schwierigkeiten im Verstehensprozess entgegenzuwirken (vgl. Abschnitt 1.5.3). Nach R. Rasch (2009, S. 85–86) dürfen gerade beim Textaufgabenlösen die Lernvoraussetzungen der Kinder nicht unberücksichtigt bleiben. Differenzierte Aufgabenstellungen ermöglichen es, unterschiedlichen Entwicklungsstufen gerecht zu werden (R. Rasch, 2008b, S. 144–145). Während leistungsstärkere Lernende weniger Kapazitäten für das Kopfrechnen benötigen und somit ausreichend Kapazitäten für Denkprozesse zur Verfügung haben, so kann die gleiche Aufgabenstellung die Kapazitäten leistungsschwächerer Lernender überschreiten. „Das Operieren mit kleinen Zahlen bedeutet im günstigen Fall für schwächere Kinder, dass das Kopfrechnen zu bewältigen ist und dass Kapazitäten bleiben, um über die situativen Zusammenhänge zielfüh-

¹⁷² Durch den Wegfall der Teufelaufgabe schieden die *Aufgaben mit unbekanntem Anfangszustand*, durch den Wegfall der Zug- und Wettkampfaufgabe die *Bewegungsaufgaben mit eingeschlossener Vergleichsproblematik* aus.

¹⁷³ Zwei Räuber entdecken einen vergrabenen Schatz, 2 Beutel Goldmünzen. Sie zählen die Münzen. In einem Beutel sind 34 Münzen, in dem anderen sind 52 Münzen. Sie wollen die Beute unter sich gerecht teilen. Wie viele Münzen müssen sie aus dem volleren Beutel herausnehmen und in den anderen füllen, damit in beiden Beuteln gleich viele Münzen sind (R. Rasch, 2001b, S. 314, 2008a, S. 34)?

¹⁷⁴ Emma liest Paul und Tim aus einem Märchenbuch vor. Am Abend beendet sie das Vorlesen auf Seite 69. Das ist eine rechte Buchseite. Das Märchen endet auf Seite 130. Wie oft muss Emma noch umblättern (Hohn, 2012, S. 96; vgl. R. Rasch, 2008a, S. 75)?

rend nachzudenken und einen entsprechenden Zahlenaussatz zu notieren“ (R. Rasch, 2009, S. 86). Dem Argument, dass Leistungen, die auf differenzierten Aufgabenstellungen beruhen, nicht vergleichbar sind, kann entgegengesetzt werden, dass beiden Aufgabenstellungen die gleiche mathematische Struktur zugrunde liegt, die vom Problemlöser erkannt werden muss. Die Verwendung kleinerer Zahlen heißt nicht, dass dadurch die Lösung direkt abrufbar wird. Die Kinder stehen gleichermaßen vor der Aufgabe, die Sachsituation zu veranschaulichen und ihre Gedankengänge und Lösungsideen zu verbalisieren. Gelingt es leistungsschwächeren Grundschulkindern, ihre Kopfrechen- und Denkprozesse anhand eines einfachen Zahlenaussatzes abzubilden, so wird es ihnen im Laufe der Zeit auch bei schwierigeren Zahlenaussatzes gelingen. Für das Aneignen von Repräsentationskompetenzen, das Erkennen und Repräsentieren von Zusammenhängen und das Verstehen der Aufgabe kann der Größe der Zahlenwerte eine eher untergeordnete Rolle zugesprochen werden, solange der Problemcharakter der Aufgabe erhalten bleibt.

Nachfolgend werden die Trainingsaufgaben entsprechend ihrer Gruppenzuordnung vorgestellt. Aus ökonomischen Gründen wird in der Übersicht ausschließlich eine Schwierigkeitsstufe dargeboten. Für das vereinfachte Zahlenaussatz wird auf den Anhang C.1 verwiesen. Teilweise weichen die Formulierungen der Trainingsaufgaben von den Originalformulierungen der genannten Quellen ab. Dies ist auf eine Expertenbefragung zweier erfahrener Grundschullehrkräfte zurückzuführen. Die Änderungen werden in den Fußnoten aufgezeigt und begründet.

Vergleichsaufgaben (R. Rasch, 2016, S. 59, vgl. Schneeballweitwurfaufgabe 2016, S. 101)

- 1) YU-GI-OH!®-Kartenaufgabe: Lukas und Jonas haben zusammen 30 YU-GI-OH!®-Karten. Lukas hat 6 mehr als Jonas. Wie viele Karten hat Lukas? Wie viele Karten hat Jonas?
- 2) Luisa und Eric werfen an den Bundesjugendspielen ihre Bälle besonders weit. Rechnet man beide Strecken zusammen, sind es 27 m. Eric wirft 3 m weiter als Luisa. Wie weit wirft Eric? Wie weit Luisa?¹⁷⁵

Ausgleichsaufgaben (R. Rasch, 2001b, S. 314, 2008a, S. 104)

- 1) Ein Esel und ein Maultier sind mit schweren Säcken beladen. Sie trotten nebeneinander her. Dabei stöhnt der Esel fürchterlich unter der großen Last. Das Maultier spricht zu ihm: „Warum stöhnst du so? Ich trage doch mehr Säcke als

¹⁷⁵ In Anlehnung an die Jahreszeit wurde der Kontext ausgetauscht. Der Schneeballweitwurf wird durch den Ballweitwurf an Bundesjugendspielen ersetzt. Mit den Bundesjugendspielen assoziiieren die Kinder eine authentische Situation.

du. Würdest du mir einen Sack abnehmen, dann hätten wir gleich viele Säcke auf dem Rücken. Würdest du mir aber einen Sack abgeben, so hätte ich dreimal so viele Säcke wie du.“ Wie viele Säcke trägt das Maultier, wie viele der Esel?¹⁷⁶

- 2) Zwei Räuber entdecken einen vergrabenen Schatz, 2 Beutel Goldmünzen. Sie zählen die Münzen. In einem Beutel sind 34 Münzen, in dem anderen sind 52 Münzen. Sie wollen die Beute unter sich gerecht teilen.
Wie viele Münzen müssen sie aus dem volleren Beutel herausnehmen und in den anderen füllen, damit in beiden Beuteln gleich viele Münzen sind?

Bewegungsaufgaben (R. Rasch, 2008a, S. 85, 65)

- 1) Eine Schnecke in einem 20 m tiefen Brunnen will nach oben auf die Wiese. Sie kriecht am Tag immer 5 m hoch und rutscht nachts im Schlaf wieder 2 m nach unten. Am wievielten Tag erreicht sie den Brunnenrand?
- 2) *Der Weg der kleinen Ameise auf dem Quadrat.* Die Seite des Quadrats ist 200 m lang. Tagsüber legt die Ameise genau 200 m zurück. Aber während der Nacht bläst sie ein starker Wind die halbe Strecke, die sie während des Tages zurückgelegt hat, wieder zurück. Am Montagmorgen geht sie los. Sie läuft von A aus über B, C und D und wieder zurück zu A. Wann wird sie wieder in A ankommen?

Aufgaben mit kombinatorischem Hintergrund (R. Rasch, 2008a, S. 48–49, 89; vgl. Hohn, 2012, S. 95)

- 1) Anke, Birgit, Christian und zwei Freunde der Kinder gehen in die Ferien. Jedes Kind verabschiedet sich von jedem mit Handschlag. Wie oft werden Hände geschüttelt?¹⁷⁷

¹⁷⁶ Der Originalformulierung nach R. Rasch (2008a, S. 104) liegt der Konjunktiv II zugrunde: „Nähmst du mir einen Sack ab, dann hätten wir gleich viele Säcke auf dem Rücken. Gäbest du mir aber einen Sack ab, so hätte ich doppelt so viele Säcke wie du.“ Um im Hinblick des *Klarheit-Kriteriums* Missverständnisse zu verhindern, wird auf die Ersatzform des Konjunktivs II mit „würde“ ausgewichen.

¹⁷⁷ Die Fantasienamen Quicki, Murks und Streblinde werden in Anlehnung an die 2. Geschichte der Denk- und Sachaufgaben für Klassenstufe 3 (R. Rasch, 2008a, S. 49) durch die Namen Anke, Birgit, Christian ersetzt. Darüber hinaus regten die Grundschullehrkräfte an, die Fragestellung „Wie viele Handschläge sind es?“ durch „Wie oft werden Hände geschüttelt?“ zu ersetzen. Letztere wäre für die Kinder greifbarer. Kinder könnten sich unter der Handlung „Hände schütteln“ mehr vorstellen als unter einem Handschlag. Diese Formulierung wurde von R. Rasch (2001a, S. 144) bereits in früheren Veröffentlichungen verfolgt.

2) Emma möchte sich ein Eis kaufen. Der Eisverkäufer hat 3 Sorten Eis: Schoko, Vanille und Himbeereis. Sie kauft sich 2 Kugeln. Wie viele verschiedene Möglichkeiten hat sie?¹⁷⁸

Aufgaben zum Verhältnis von Zwischenraum und Begrenzung (R. Rasch, 2008a, S. 41, 75; vgl. Hohn, 2012, S. 96)

- 1) Tom findet eine Kiste. Er kramt. Es ist eine Kiste mit lauter alten Ketten, mit einer Krone, mit einem Kleid und mit Diamanten. Er entdeckt eine Karte aus einem fernen Land. Ein Schatz ist eingezeichnet. Um den Schatz zu finden, muss man an 11 großen Bäumen vorbeigehen. Zwischen den Bäumen ist ein Abstand von 10 Metern. Am 11. Baum ist der Schatz vergraben. Wie weit ist es vom 1. bis zum 11. Baum?¹⁷⁹
- 2) Emma liest Paul und Tim aus einem Märchenbuch vor. Am Abend beendet sie das Vorlesen auf Seite 69. Das ist eine rechte Buchseite. Das Märchen endet auf Seite 130. Wie oft muss Emma noch umblättern?¹⁸⁰

Aufgaben mit komplexen Informationen (R. Rasch, 2001b, S. 91, 2008a, S. 92; E. C. Wittmann & Müller, 2013b, S. 88)

- 1) An einem Herbsttag werden in einem Stall 15 Tiere gezählt. Es sind Pferde und Fliegen. Zusammen haben sie 72 Beine. Wie viele Pferde und wie viele Fliegen sind es?
- 2) Der Opa von Hannes ist Kaninchenzüchter. Er besitzt Ställe für ein und für zwei Kaninchen. Insgesamt sind es 25 Ställe. Er kann darin 40 Tiere unterbringen. Wie viele Ställe sind es für ein Kaninchen? Wie viele Ställe sind es für zwei Kaninchen?

¹⁷⁸ Die Aufgabenstellung nach R. Rasch (2008a, S. 13) sah drei Kinder vor: „Streblinde, Quicki und Murks möchten sich ein Eis kaufen.“ Hohn (2012, S. 95) griff diese Aufgabe auf und ersetzte die Fantasienamen durch die echten Namen Emma, Tim und Paul. In Anlehnung an Hohn wurden die Namensänderungen übernommen. Unter dem Gesichtspunkt, dass sich ausschließlich Emma ein Eis kauft, empfahlen die Experten das Streichen der beiden Namen Tim und Paul. Darüber hinaus sollte deren Ansicht nach die irrelevanten Informationen, „jedes Kind hat Geld für 2 Kugeln Eis“, unberücksichtigt bleiben, um Trugschlüsse zu vermeiden.

¹⁷⁹ Da die Aufgabenstellung sehr lang war, rieten die Experten zur Textreduktion. Vereinzelte, entbehrliche Ausschmückungen wurden von ihnen gestrichen (Originalaufgabe aus R. Rasch, 2008a, S. 41).

¹⁸⁰ Die Originalaufgabe nach R. Rasch (2008a, S. 75) weist sehr große Zahlen auf, die enorme Kapazitäten binden: „Am Abend beendete sie das Vorlesen auf der Seite 169. Das ist eine rechte Buchseite. Das Märchen endete auf der Seite 230“. In Anlehnung an Hohn (2012, S. 96) wurden vereinfachte Zahlenwerte eingesetzt.

4.2.3 Kategorisierung externer Schülerrepräsentationen

Um herauszufinden, welche selbstgenerierten Schülerrepräsentationen den Erfolg beim Lösen problemhaltiger Textaufgaben am besten vorhersagen und welche Merkmale diesen zugesprochen werden können, wurden die empirisch gewonnenen Daten der Pilotierung inhaltsanalytisch ausgewertet. Im Fokus der Betrachtung stand, *wie* die Lernenden die einzelnen Problemstellungen angehen und in ihrer eigenen Sprache entfalten, und *welche* Merkmale die externen Repräsentationen charakterisieren und sie zur Lösungsunterstützung werden lassen. Von der Analyse der individuellen Eigenproduktionen wurde sich versprochen, mehr über die Heran- und Vorgehensweisen der Lernenden sowie deren Sicht- und Denkweisen zu erfahren (Kelle & Kluge, 2010, S. 70; Mayring, 2002, S. 19). Neben dem Bestreben, mit quantitativen Methoden herauszufinden, ob sich die Lernenden nach der Interventionsmaßnahme verbessert bzw. einen Kompetenzzuwachs erzielt haben, sollten die qualitativen Methoden helfen, Gründe für die beobachteten Leistungsentwicklungen in der Hauptuntersuchung zu identifizieren. Dabei ist theoriegeleitet denkbar, dass die konstruierten Denk- und Erkenntniswerkzeuge maßgeblich über Erfolg bzw. Misserfolg bestimmen können. In Abschnitt 1.5.3.2 wurde bereits ausführlich geschildert, dass inadäquate Repräsentationen die Überwindung von Problembarrieren verwehren können und repräsentationale Veränderungen unumgänglich werden. Es galt daher, aufzuspüren, welche Repräsentationsmerkmale erkenntnisfördernd und lösungsunterstützend sind.

Werden in einem ersten Schritt die Eigenproduktionen der Lernenden qualitativ ausgewertet, so kann der Einfluss dieser Ergebnisse auf die Leistungsentwicklung in einem nächsten Schritt untersucht werden. „Die qualitativen und quantitativen Methoden helfen in diesem Fall, Licht auf unterschiedliche Aspekte des untersuchten Phänomens zu werfen. Die qualitativen und quantitativen Ergebnisse können sich daher nicht wechselseitig ersetzen und sind damit auch nicht zur wechselseitigen Validierung brauchbar“ (Lüken, 2012, S. 149; vgl. Neumann & Ruwisch, 2015, S. 327).

4.2.3.1 Entwicklung des Kategoriensystems

Um die von den Lernenden generierten externen Repräsentationen näher zu beschreiben, zu spezifizieren und zu analysieren, wurde in Anlehnung an die qualitative Inhaltsanalyse nach Mayring (2015) ein Kategoriensystem entwickelt. Es gilt als zentrales Instrument qualitativer wie auch quantitativer Analysen (Mayring, 2015, S. 51). Die 351 individuellen Eigenproduktionen der Lernenden, welche in der Pilotierung von 39 Viertklässlern zu neun verschiedenen Problembewältigungen im Mathematik-

unterricht entstanden sind, fungierten als Ausgangsmaterial¹⁸¹. Sie wurden im Klassenverband und demnach in einem natürlichen, alltäglichen Umfeld erhoben (Mayring, 2002, S. 22). Im Gegensatz zu transkribierten Interviews handelte es sich dabei nicht ausschließlich um niedergeschriebene Texte. Vorrangig lag Bildmaterial zugrunde, das analog zu Texten, Audio- oder auch Videomaterial analysiert werden kann (Kuckartz, 2014a, S. 111). Die Kinder externalisierten ihren Lösungsprozess sowohl grafisch, tabellarisch und rechnerisch als auch in Textform oder in kombinierten Vorgehensweisen. Die Bandbreite an Vor- und Herangehensweisen sollte ganzheitlich mithilfe der qualitativen Inhaltsanalyse erfasst und kategorisiert werden. Mayring (2014, S. 64–65, 2015, S. 67–68) differenziert zwischen zwei Interpretationsformen, der induktiven Kategorienbildung und der deduktiven Kategorienanwendung, und zieht Mischformen beider Formen in Betracht.¹⁸² Dabei werden drei voneinander unabhängige Analysetechniken unterschieden: Zusammenfassung, Explikation und Strukturierung (Mayring, 2015, S. 67, 2016, S. 115).

In der vorliegenden Untersuchung wurde sich für die *inhaltlich strukturierende Inhaltsanalyse* ausgesprochen, eine Mischform beider Interpretationsformen, die eher einer deduktiven Logik folgt (Gläser-Zirkuda, 2015, S. 125; Mayring, 2014, S. 104, 2015, S. 68). Gegen eine rein induktive Kategorienbildung sprach, dass die Reduktion und Zusammenfassung des Datenmaterials möglicherweise nicht fähig ist, die Gesamtheit der Daten wie gefordert abzubilden. Hinzu kommt, dass in der Mathematikdidaktik bereits vielfältige externe Repräsentationen und Vorgehensweisen unterschieden werden, weshalb eine induktive Vorgehensweise künstlich und konstruiert wäre. Es sind Rechnungen/Terme/Gleichungen, Zeichnungen/Skizzen, Tabellen, Listen oder auch Graphen und Diagramme, die in der Mathematik der Gruppe externer Repräsentationen zugeordnet werden (siehe Abschnitt 1.5.6). Vor diesem Hintergrund gilt es vielmehr, zu überprüfen, welche dieser Repräsentationen von Grundschulkindern im Rahmen des Problemlösens generiert werden. Eine ausschließlich induktive Vorgehensweise, ohne Berücksichtigung der Theorie und des bisherigen Forschungsstandes, wurde daher nicht in Betracht gezogen.

Im Rahmen einer *explizierenden Inhaltsanalyse* wird die postulierte Erweiterung des Datenmaterials (Mayring, 2015, S. 90) in der vorliegenden Untersuchung lediglich in Form von Interviews mit den Lernenden als zusätzlicher Gewinn angesehen. Durch sie würde man zusätzliche Information über die Entstehung der Eigenproduktion und die vollzogenen Denkwege gewinnen, die nicht unmittelbar sichtbar sind. Ein weiterer

¹⁸¹ Da alle 39 Kinder 9 problemhafte Textaufgaben bearbeiteten, entstanden 351 (9 mal 39) Eigenproduktionen, die der Analyse zugrunde lagen.

¹⁸² Für detaillierte Ausführungen der einzelnen Analyseform sei an dieser Stelle auf die Abschnitte 6.3, 6.5, 6.6 in Mayring (2014) sowie Abschnitt 5.5.2 und 5.5.4 in Mayring (2015) verwiesen.

Vorteil wird darin gesehen, dass sie Aufschluss über Fehlvorstellungen geben könnten. Die Interviews würden das Verständnis erweitern, indem die Produzenten selbst ihre individuellen Vorgehensweisen beschreiben, begründen und erklären (Mayring, 2015, S. 67). Deren Durchführung war jedoch aus ökonomischen Gründen nicht möglich (siehe Abschnitt 4.5.1.3 sowie 6.3 und 6.4). Darüber hinaus wäre das Heranziehen zusätzlicher Materialien, beispielsweise aus den vorherigen Untersuchungen von Groß (2013) und Hohn (2012), denkbar, jedoch als wenig gewinnbringend eingeschätzt. Hierbei handelte es sich um Ist-Stand-Analysen, die zeigten, dass die Kinder selten ihre Denkwege externalisieren. Die Anreicherung des Datenmaterials wäre dadurch kein Zugewinn. Eine weitere Möglichkeit der Explikation wäre, Schulhefte der Kinder als zusätzliche Materialien einzubeziehen. Deren Auswertung ist wenig erkenntnisfördernd, wenn es darum geht, die interpretationsbedürftigen Eigenproduktionen zu explizieren und für sich sprechen zu lassen. Dafür sind die individuell angefertigten Produkte zu speziell, sodass sich kaum hierzu passende Hefteinträge finden und heranziehen lassen.

Gegen ein rein deduktives Vorgehen sprach, dass eine theoriegeleitete Kategorienanwendung ebenso wenig wie ein rein induktives Vorgehen fähig ist, das zugrunde liegende Lösungsspektrum der Individualität und Kreativität vollständig zu erfassen. Gründe hierfür sind, dass bisherige Analysen von Problemlöse- und Modellierungsprozessen verstärkt verwendete Lösungsstrategien in den Blickpunkt nahmen, weniger externe Repräsentationen. Dadurch liegen für heuristische Strategien fundiertere empirische Erkenntnisse vor, als dies für externe Repräsentationen der Fall ist. Viele Studien, die ihren Fokus auf externe Repräsentationen richten, gleich ob auf selbstgeneierte oder auf vorgefertigte, spezialisieren sich vorwiegend auf *eine* externe Repräsentationsform: Zeichnungen bzw. Skizzen (Ott, 2014, 2015; Pantziara, Gagatsis, & Elia, 2009; Rellensmann, Schukajlow, & Leopold, 2015; Uesaka, Manalo, & Ichikawa, 2007; van Essen & Hamaker, 1990, sowie Abschnitt 1.5.6.1).

Hieraus ließen sich Merkmale und Kriterien guter Zeichnungen ableiten, die ein theoriegeleitetes Vorgehen ermöglichen. Dies trifft jedoch nur bedingt für Rechnungen, Tabellen und Begründungen zu. Die vorhandene empirische Beweislage machte es diesbezüglich erforderlich, einen Teil der Kategorien aus dem Material abzuleiten, d. h. induktiv heranzugehen, um eine Passung des Kategoriensystems an das Datenmaterial sicherstellen zu können (Kuckartz, 2014a, S. 112; Mayring, 2014, S. 104; Schreier, 2014). Aus diesen Überlegungen folgt die Entscheidung für eine Mischform beider Interpretationsformen. Dabei wurde die Entwicklung eines niedrig inferenten Kodiersystems angestrebt, das präzise und eindeutig formulierte Kategorien enthält, die wenig Interpretationsspielraum zulassen, objektive Entscheidungen ermöglichen und „wenig schlussfolgernde Kognition beim Beobachter“ erfordern (Petko, Waldis,

Pauli, & Reusser, 2003, S. 275; vgl. Waldis, Gautschi, Hodel, & Reusser, 2006, S. 171). Ein Ratertraining soll die anzustrebenden hohen Beobachterübereinstimmungen sichern (Petko et al., 2003, S. 276).

Inhaltlich strukturierende qualitative Inhaltsanalyse

Die Kategorisierung der externen Schülerrepräsentationen erfolgte mithilfe einer inhaltlich strukturierenden Inhaltsanalyse, um die zugrunde liegende Struktur aus dem Material herauszufiltern (Mayring, 2014, S. 95, 2015, S. 97). Ihr Kernstück ist die Generierung eines Kategoriensystems mit „Definitionen, Ankerbeispielen, Kodierregeln und Fundstellenbezeichnungen“ (Gläser-Zirkuda, 2015, S. 125).

Der inhaltliche Fokus macht deutlich, dass es die im Datenmaterial zu bestimmenden Inhalte, hier Merkmale externer Repräsentationen, zu extrahieren und zusammenzufassen galt (Mayring, 2015, S. 99). Die Entwicklung des Kategoriensystems erfolgte in zehn Schritten (siehe Abbildung 4.1). Zu Beginn der Inhaltsanalyse galt es, die Analyseeinheit zu bestimmen (*Schritt 1*, siehe Abbildung 4.1) (Mayring, 2015, S. 61; Mayring & Fenzl, 2014, S. 553). Den kleinsten auszuwertenden Materialbestand, die Kodiereinheit, stellte ein *Bildzeichen* bzw. ein *Schriftzeichen*¹⁸³ dar, das mit einer bestimmten Kategorie bzw. einem bestimmten Merkmal externer Repräsentationen in Verbindung stand (vgl. Kuckartz, 2014b, S. 47). Da der Auswertung nicht nur Text-, sondern vorrangig auch Bildmaterial zugrunde lag, bezog sich Ersteres in diesem Zusammenhang auf den minimalsten graphemischen Textteil, Letzteres auf den minimalsten grafischen Bildteil. Bildzeichen lassen – wie in Abschnitt 1.5.1 bzw. 1.5.3 bereits aufgezeigt – durch ihre Strukturmerkmale und räumliche Position konkrete Rückschlüsse auf das Repräsentierte zu. Schriftzeichen liegen demgegenüber Konventionen zugrunde, die deren Beziehung zum Repräsentierten bestimmen. Dabei kann es sich sowohl um einzelne Wörter als auch um Wortkombinationen oder Sätze handeln. Die gesamte Eigenproduktion eines Kindes wurde als größter Bestandteil, der unter eine Kategorie fallen kann, definiert. Mayring (2015, S. 61) spricht hier von der Kontexteinheit. Als Auswertungseinheit, sprich welche Teile nacheinander ausgewertet werden, fungierten alle vorliegenden 360 Knobelhefte.

Die Entwicklung des Kategoriensystems für die inhaltliche Strukturierung der Eigenproduktionen basierte auf der fachdidaktischen Unterscheidung externer Repräsentationen (siehe Abschnitt 1.5.6). Sie bildeten die inhaltlichen Hauptkategorien, die theoriegeleitet festgelegt wurden: *Tabellen*, *Rechnungen*, *Zeichnungen*, *Listen* und *schriftsprachliche Beschreibungen* (*Schritt 2*).

¹⁸³ Hier wird der Bezeichnung von Schnotz und Dutke gefolgt (2004, S. 63) und verbale (graphemische) und piktoriale (grafische) Informationen unterschieden.

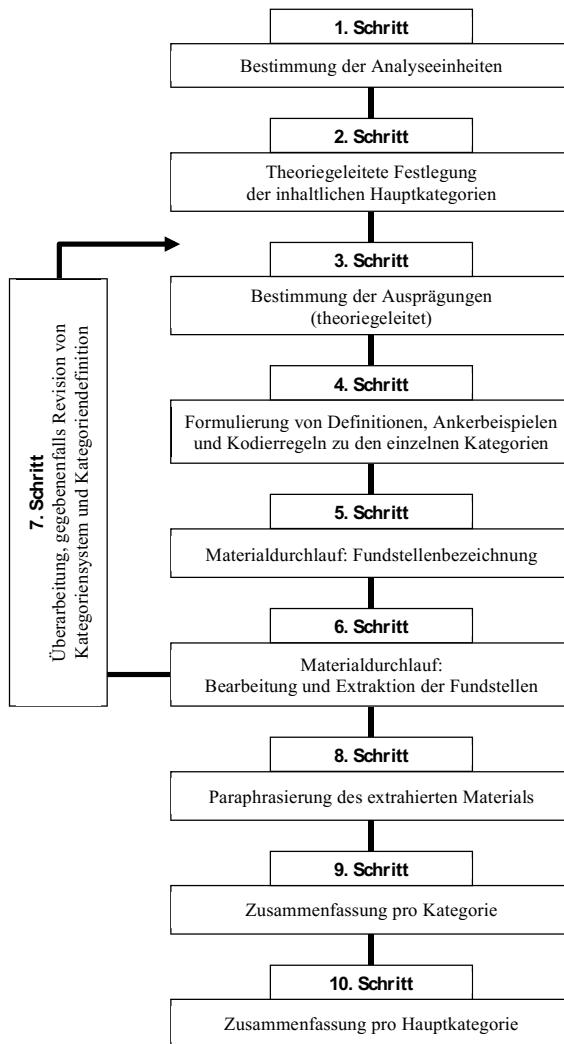


Abbildung 4.1. Ablaufmodell der inhaltsanalytischen Strukturierung nach Mayring(2014, S. 95, 2015, S. 97), entnommen aus Steigleder (2008, S. 33).

Die Hauptkategorien übernahmen die Aufgabe eines „Suchrasters“, mittels dessen das Material auf die entsprechenden Merkmale hin durchsucht und grob kategorisiert wur-

de (Kuckartz, 2014b, S. 69). Insofern wurden in einem ersten Materialdurchgang alle Eigenproduktionen den thematischen Hauptkategorien zugeordnet, welche erst im nächsten Schritt ausdifferenziert, verfeinert und präzisiert wurden (Kuckartz, 2014a, S. 112; Mayring, 2015, S. 97). Wurde mehr als eine externe Repräsentation konstruiert, so konnten auch mehrere Hauptkategorien zutreffen.

Das konsensuelle Kodieren¹⁸⁴ von vier unabhängigen Ratern¹⁸⁵ ergab lediglich Unstimmigkeiten bei der Zuordnung des Datenmaterials bezüglich der Hauptkategorie *Listen*. Diese ließ sich von der Kategorie *Tabellen* einerseits und von der Kategorie *Zeichnungen* andererseits nicht trennscharf abgrenzen. Da Tabellen ohne Tabellenkreuz oder auch Tabellen mit lediglich einer Spalte ebenfalls als Liste identifiziert bzw. Strichlisten auch als Zeichnungen gedeutet wurden, war eine eindeutige Kategorisierung nicht mehr möglich und eine Konsensfindung erforderlich. In einer Kodierkonferenz (Mayring, 2014, S. 114) aller Rater wurde dieser Schwierigkeit begegnet und ein Konsens gefunden, indem die Kategorie *Listen* aufgegeben wurde, was eine Überarbeitung der Kategoriedefinition *Tabellen* nach sich zog.

Dagegen bestätigte sich die Vermutung, dass externe Repräsentationen, wie Diagramme und Graphen, von den Grundschulkindern nicht generiert wurden, was deren Nichtberücksichtigung als Hauptkategorien legitimierte. Nach Kelle und Kluge (2010, S. 72) stellt ein Wegfall einer Kategorie oder eine Ergänzung auch bei theoriegeleiteten Kategorienanwendungen keine Seltenheit dar. Alle als *Liste* identifizierten externen Repräsentationen konnten dann eindeutig entweder der Kategorie *Tabellen* oder der Kategorie *Zeichnungen* zugeordnet werden. Der Forderung, die „Kategorien sollen disjunkt, erschöpfend und präzise sein“, wurde somit entsprochen (Diekmann, 2007, S. 589; vgl. Kuckartz, 2014b, S. 61).

Nach dieser ersten Zuordnung galt es, die noch relativ allgemein formulierten Hauptkategorien auszudifferenzieren (Schritt 3, Kuckartz, 2014b, S. 83). Dabei wurde eine Mischform induktiver und deduktiver Kategorien- und Subkategorienbildung favorisiert (Bortz & Döring, 2006, S. 330; Kuckartz, 2014a, S. 112, 2014b, S. 77; Mayring, 2014, S. 65, 104; Schreier, 2014). Zum einen wurden aus der Theorie Kategorien

¹⁸⁴ Unter konsensuellem Kodieren versteht Kuckartz (2014b, S. 82) eine Technik, bei der in der Regel zwei Kodierer unabhängig voneinander das Datenmaterial kodieren. Voraussetzung hierfür ist das Vorhandensein eines „Kategoriensystems mit hinreichend präzisen definierten Kategorien“ (Kuckartz, 2014b, S. 83).

¹⁸⁵ Aufgrund der Tatsache, dass neben dem theoriegeleiteten auch das empiriegeleitete Vorgehen angestrebt wurde, wurden für die Entwicklung des Kategoriensystems vier Rater eingesetzt. Zu ihnen zählten zwei Studierende des Bachelorstudiums, ein Studierender des Masterstudiums sowie die Versuchsleiterin. Zum späteren Zeitpunkt konnte zusätzlich zu der Versuchsleiterin nur noch ein Rater, ein Studierender des Masterstudiums, weiterbeschäftigt werden. Hierbei handelte es sich um den Rater, der die höchste Übereinstimmung mit der Versuchsleiterin aufwies.

gewonnen, die sich den einzelnen Hauptkategorien unterordnen ließen (vgl. Abschnitt 1.5.6). Zum anderen wurde direkt am Material nach möglichen Ausprägungen dieser theoriegeleiteten Kategorien gesucht. Dies diente „dazu, das empirische Spektrum zu erschließen, das von den anfangs festgelegten Kodierkategorien aufgespannt wird und diese damit zu konkretisieren bzw. empirisch anzureichern“ (Kelle & Kluge, 2010, S. 73). Insbesondere Ähnlichkeiten und Unterschiede wurden durch die Erweiterung der Subkategorien klar herausgestellt, um eine zuverlässige Abgrenzung der Kategorien sicherzustellen (Kelle & Kluge, 2010, S. 73; Kuckartz, 2014b, S. 61).

Um entscheiden zu können, wann ein Merkmal der Eigenproduktion unter eine Kategorie fiel, wurde im nächsten Schritt genau definiert und festgelegt (*Schritt 4*, Mayring, 2014, S. 96, 2015, S. 97–98, s. Abschnitt 4.2.3.2). Eine repräsentative Auswahl konkreter Ankerbeispiele aus dem Datenmaterial sollte die Kategorien explizieren sowie Transparenz und Orientierung schaffen. Zusätzliche Regeln wurden in Form von Kodierkommentaren verfasst, um Abgrenzungen zwischen einzelnen Kategorien gewährleisten zu können. Unter diesem Gesichtspunkt wurde dann mit einem ersten „Materialdurchlauf“ kontrolliert, „ob die Kategorien überhaupt greifen, ob die Definitionen, Ankerbeispiele und Kodierregeln eine eindeutige Zuordnung ermöglichen“ (*Schritt 5*, Mayring, 2015, S. 97). Den vier Mitarbeitern der Projektgruppe stand mit dem Kodierleitfaden eine Handanweisung zur Verfügung (Mayring, 2002, S. 119). Sie markierten in den Knobelheften die Teile, die unter eine Kategorie fielen, die sogenannten „Fundstellen“ (z. B. Mayring, 2002, S. 120), und filterten daraufhin die im Material gekennzeichneten Merkmale heraus (*Schritt 6*).

Dieser Probendurchlauf zog „eine Überarbeitung, eine teilweise Neufassung vom Kategoriensystem und seinen Definitionen“ nach sich (*Schritt 7*, Mayring, 2015, S. 98–99). Bei schwierigen Kodierungen wurden zusätzliche Kodierregeln ergänzt und neue Ankerbeispiele aufgenommen oder verwendete gegen eindeutigere ausgetauscht. Einzelne Ausprägungen wurden an Stellen gestrichen, wo sie zu feingliedrig waren, aber auch ergänzt, wo sie fehlten. Des Weiteren mussten neben Definitionen auch Kodierregeln präzisierend überarbeitet werden. Gemeinsame Diskussionen in Kodierkonferenzen dienten der Klärung adäquater Einordnungen, schärften ungenaue Abgrenzungen und mündeten in einem überarbeiteten Kodierleitfaden. In einem iterativen Prozess wurde die Passung des Kategoriensystems an das Datenmaterial angestrebt und verfolgt (Wiederholung der *Schritte 3–6*). In den letzten *Schritten 8–10* wurden die Kernmerkmale externer Repräsentationen extrahiert und pro Kategorie und Hauptkategorie zusammengefasst.

Das vorliegende Kategoriensystem enthält Kategorien mit nominalem und ordinalem Skalenniveau (Mayring & Fenzl, 2014, S. 548). Es wird nachfolgend gegliedert nach den Hauptkategorien vorgestellt. Was die jeweilige Kategorie (das Merkmal) aufzuklä-

ren fähig ist, wird in der Tabelle mithilfe einer Fragestellung notiert. Aus Gründen der Nachvollziehbarkeit und des besseren Verständnisses werden direkt deren Ausprägungen aufgezählt. Auf die Präsentation des Kodierleitfadens und der Ankerbeispiele wird an dieser Stelle verzichtet und es wird auf Anhang A.1 und A.2 verwiesen.

Ferner wurde für alle vier Kategorien die Güte der Beurteilung überprüft und ausgewiesen. Die Reliabilitäten kennzeichnen, inwieweit die beiden unabhängigen Rater mit gleichem Wissensstand in ihrem Urteil übereinstimmten (Wirtz & Caspar, 2002, S. 15). Die Übereinstimmungsberechnungen basierten auf der Tatsache, dass jeder Rater alle Eigenproduktionen der Kinder von allen drei Messzeitpunkten beurteilte. Es galt dann zu entscheiden, ob das jeweilige Repräsentationsmerkmal erkennbar war oder nicht. In Abhängigkeit der Skalenniveaus wurden die entsprechenden Koeffizienten bestimmt. Weil alle Repräsentationsmerkmale aggregiert vorlagen und somit als intervallskaliert angenommen werden dürfen, wurde die unjustierte Intraklassenkorrelation (ICC) als Übereinstimmungsmaß gewählt.

4.2.3.2 Kategoriensystem zur Kodierung des Datenmaterials

Die aus fachdidaktischer Sicht für die Grundschule repräsentativen externen Repräsentationen *Zeichnungen*, *Rechnungen*, *Tabellen*, *Listen* und *schriftsprachliche Beschreibungen* waren die Basis für die Entwicklung des Kategoriensystems. Die inhaltlich strukturierende qualitative Inhaltsanalyse identifizierte 4 Hauptkategorien, weil die Kategorie *Listen* inhaltsanalytisch nicht bestätigt werden konnte. Die in den Abschnitten 1.5.6.1–1.5.6.5 erläuterten Repräsentationsmerkmale bestimmten theoriegeleitet die Ausprägungen der Hauptkategorien und wurden empiriegeleitet durch das Datenmaterial angereichert. Im vorangegangen Abschnitt 4.2.3.1 wurde der Prozess der Kategorienfindung ausführlich geschildert. Nachfolgend werden lediglich die Ergebnisse der Inhaltsanalyse, sprich die Subkategorien vorgestellt und erklärt.

Zeichnungen

Für die Hauptkategorie *Zeichnungen* ergaben sich fünf Subkategorien, die mit ihren Ausprägungen in Tabelle 4.2 dargestellt sind. Sie leiten sich aus den im Theorieteil aufgezeigten Merkmalen ab und lehnen sich an deren Reihenfolge an (vgl. Aufzählung in Abschnitt 1.5.6.1). Die Reihenfolge wird auch im Kodierleitfaden aufgegriffen (siehe Anhang A.1).

Ist der Fokus auf den *dargestellten Zustand* gerichtet, so lassen sich verschiedene Merkmalsausprägungen unterscheiden. Eine vollständige Rekonstruktion des Lösungsprozesses ist insbesondere dann möglich, wenn neben dem Anfangs- und Zielzustand auch der Lösungsweg transparent wird. Daher lassen sich folgende dargestellten Zustände ableiten: Anfangszustand, Zielzustand sowie die Manipulation der Situation

(Lösungsweg). Es wurde bei jeder Zeichnung entschieden, ob 1) der Anfangszustand sichtbar oder im Kopf konstruiert wurde, 2) der Zielzustand sichtbar oder im Kopf konstruiert wurde und 3) die Transformation der Situation sichtbar oder im Kopf vollzogen wurde. Berücksichtigt man die Tatsachen, dass die Lernenden nicht immer explizit ihre Lösungsgedanken notieren (Fetzer, 2011, S. 42) und ihnen das begleitende wie auch das nachträgliche Dokumentieren schwerfällt (R. Rasch, 2009, S. 84; Schütte, 2004, S. 139), ist damit zu rechnen, dass die Zeichnungen selten alle Zustände und erfolgten Transformationen abbilden. Es stellt sich die Frage danach, ob und wie sich der tatsächlich dargestellte Zustand und die denkbaren Kombinationen auf den Lösungserfolg auswirken.

Tabelle 4.2. Kategorien von Zeichnungen auf Basis einer inhaltlich strukturierenden Inhaltsanalyse

Kategorien	Informationsgehalt	Ausprägung
Dargestellter Zustand	Was ist in der Skizze dargestellt?	Anfangszustand; Zielzustand; Lösungsweg; Anfangszustand und Lösungsweg; Anfangs- und Zielzustand; Zielzustand und Lösungsweg, Anfangs-, Zielzustand und Lösungsweg, keine eindeutige Zuordnung möglich
Grad der Strukturabbildung	Werden alle strukturelevanten Informationen adäquat und vollständig abgebildet?	Aufgabenstruktur falsch abgebildet, nur Oberflächenmerkmale (dekorativ); Aufgabenstruktur falsch abgebildet, aber mit Bild-Text-Bezug; Aufgabenstruktur unvollständig abgebildet, aber mit Bild-Text-Bezug; Aufgabenstruktur richtig und vollständig abgebildet
Abstraktionsgrad	Wie originalgetreu, ausgeschmückt bzw. dekorativ werden die Informationen aus dem Text dargestellt?	originalgetreue Darstellung (Kontextbezug), abstrakte Darstellung (kein direkter Kontextbezug)
Kontrolle	Werden die Zusammenhänge zwischen Gegebenem und Gesuchtem in der Zeichnung kontrolliert? Wird das in der Skizze deutlich?	Aufgabenbedingungen werden nicht sichtbar kontrolliert; Aufgabenbedingungen werden sichtbar kontrolliert
Systematisches Vorgehen	Wird sich systematisch probierend oder unsystematisch planlos der Lösung genähert?	unsystematisches Vorgehen; systematisches Vorgehen

Beim *Grad der Strukturabbildung* gilt es, zu berücksichtigen, ob die Struktur vollständig richtig, teilweise richtig oder falsch in der Zeichnung abgebildet wird und dem-

nach eindeutige Bild-Text-Bezüge herausgestellt wurden oder nicht (vgl. Pantziara et al., 2009, S. 40; Uesaka et al., 2007, S. 328). Beim Kodieren zeigte sich, dass diese drei Abstufungen nicht genügten. Vielmehr wurde erforderlich, die Ausprägung *falsche Strukturabbildung* differenzierter zu erfassen, weshalb eine zusätzliche Abstufung vorgenommen wurde. Die Unterscheidung, ob lediglich Oberflächenmerkmale abgebildet wurden oder aber ein konkreter Text-Bild-Bezug bestand, bei dem lediglich die Aufgabenbedingungen falsch repräsentiert wurden, ergab sich induktiv aus dem Material. Im Materialdurchlauf zeichnete sich für die Rater ein klares Gefälle zwischen diesen beiden Ausprägungen ab, sodass sie als unterschiedliche Abstufungen der Strukturabbildung in die Analyse eingingen.

Darüber hinaus sollte untersucht werden, ob sich die jeweilig stattgefundenen Abstraktionsprozesse auf den Lösungserfolg auswirken (*Abstraktionsgrad*). Da sie helfen, die Problemstruktur aufzuspüren und die Lösung auszuarbeiten (z. B. Bruder, 2000, S. 6; de Bock et al., 2003, S. 446; König, 1992, S. 35; M. Meyer & Prediger, 2012, S. 4, 6; van Essen & Hamaker, 1990, S. 302–303), ist von Interesse, ob sich detailgetreue und nah am Kontext orientierte Zeichnungen in ihrer Vorhersagekraft des Lösungserfolgs von abstrakten unterscheiden.

Letztlich sollte überprüft werden, ob sich ein *kontrollierendes* bzw. *systematisches Vorgehen* in den Zeichnungen widerspiegelt und die Lernenden wie prognostiziert häufiger zum richtigen Ergebnis führt (siehe Abschnitt 1.5.6).

Die Übereinstimmungen beider Rater lagen bei allen fünf Subkategorien zwischen $ICC_{min} = .98$ und $ICC_{max} = 1.00$, sodass von nahezu identischen Urteilen beider Rater ausgegangen werden kann (Döring & Bortz, 2016, S. 347; Wirtz & Caspar, 2002, S. 160).

Rechnungen

Für die Kategorie *Rechnungen* ergaben sich acht Subkategorien, die mit ihren Ausprägungen in Tabelle 4.3 dargestellt sind. Sie leiten sich analog zu den Zeichnungen aus den im Theorieteil aufgezeigten Merkmalen ab und lehnen sich an deren Reihenfolge an (vgl. Aufzählung in Abschnitt 1.5.6.2). Diese Reihenfolge wird auch im Kodierleitfaden verfolgt (siehe Anhang A.1).

Die Kategorie *Nutzungsmöglichkeiten* verfolgte das Ziel, verschiedene Anlässe des Rechnens zu ermitteln. Wenn man an dieser Stelle aufgreift, dass (ungeübte) Grundschulkinder tendenziell dazu neigen, die in der Textaufgabe geschilderte Sachsituation erst einmal in eine Rechnung zu überführen (R. Rasch, 2009, S. 69), dann lassen sich in der Regel zwei verschiedene Verhaltensweisen beobachten: Entweder sie verharren in „Darstellungen mit Zahlen“ (Schreiber, 2014, S. 8) und nutzen die Rechnung zur Erarbeitung, oder aber sie kombinieren und wechseln zwischen Repräsentationsformen

(Fuchs, 2006, S. 92; Käpnick, 1998, S. 254–255; Lorenz, 2013, S. 13; R. Rasch & Schütte, 2008, S. 74). Letzteres schließt ein, dass die erzeugte Rechnung zum einen als Einstieg fungiert, welche in einer weiteren Repräsentation fortgeführt wird, oder zum anderen ein auf eine andere Art und Weise gewonnenes Ergebnis kontrolliert (Eichler, 2015, S. 60–61; Gerster & Schultz, 2004, S. 388). Eine Kombination aus Einstieg und Kontrolle geht ebenfalls aus dem Material hervor, weshalb sie als weitere Nutzungs möglichkeit aufgenommen wurde.

Tabelle 4.3. Kategorien von Rechnungen auf Basis einer inhaltlich strukturierenden Inhaltsanalyse

Kategorien	Informationsgehalt	Ausprägung
Nutzungs möglichkeiten	Wozu wird die Rechnung genutzt? Welche Funktion kann ihr zugesprochen werden?	Erarbeitungsfunktion; Kontrollfunktion; Einstiegsfunktion; Einstiegs- und Kontrollfunktion
Anzahl der Rechnungen	Werden Zwischenschritte und daher mehr als eine Rechnung notiert?	eine Rechnung; mehr als eine Rechnung
Abstraktions grad	Wird sich in der Rechnung auf den Kontext bezogen?	Kontextbezug; kein direkter Kontext be zug
Geschriebene mathematische Sprache	Wird das Gleichheitszeichen als Symbol einer Beziehung oder einer Rechenoperation verwendet?	Symbolverwendung als Rechenoperationen; Symbolverwendung als Beziehung
Konstanz eines Faktors	Wird eine Größe oder das Ergebnis konstantgehalten?	eine Größe wird konstantgehalten; keine Größe wird konstantgehalten
Variation der Rechenoperationen	Wie viele verschiedene Rechenoperationen werden verwendet?	Fokussierung einer Rechenoperation; Fokussierung von mehr als einer Rechenoperation
Kontrolle	Werden die Zusammenhänge zwischen Gegebenem und Gesuchtem in der Rechnung kontrolliert? Wird das in der Rechnung deutlich?	Aufgabenbedingungen werden nicht sichtbar kontrolliert; Aufgabenbedingungen werden sichtbar kontrolliert
Systematisches Vorgehen	Wird sich systematisch probierend oder unsystematisch planlos der Lösung genähert?	unsystematisches Vorgehen; systematisches Vorgehen

Wenn auch in der Literatur nur vereinzelt genannt (z. B. R. Rasch & Schütte, 2008, S. 72–73), fiel insbesondere bei der Materialdurchsicht auf, dass die *Anzahl der Rechnungen* variierte. Entweder „zwängten“ die Lernenden gegebene Informationen in eine Rechnung, brachen ihren Lösungsprozess nach einer Rechnung ab oder nutzten mehrere

re Rechnungen, um sich schrittweise der Lösung anzunähern. Es handelt sich hierbei um keine aus bisherigen Theorien hervorgegangene Kategorie. Sie resultiert vielmehr aus der induktiven Kategorienfindung.

Rechnungen gelten im Allgemeinen als kontextunabhängig und bringen den Vorteil mit sich, kurz und präzise aufzuzeigen, wie die Lösung gefunden wurde (vgl. Bruder & Collet, 2011b, S. 68; Steinweg, 2013, S. 190). Inwiefern sich bei Rechnungen eine Kontextnähe oder -ferne abzeichnet, sollte mit dem *Abstraktionsgrad* erfasst werden.

Fokussiert man in Lösungsprozessen ausschließlich die *geschriebene mathematische Sprache*, so fällt auf, dass die Lösenden beim Niederschreiben ihrer Rechnung weniger auf die korrekte Notation achten. Einzelne Rechenschritte werden hintereinanderhängend gemäß der stattgefundenen Denkprozesse aufgeschrieben und mit einem Gleichheitszeichen verbunden, sodass dieses nicht mehr als Symbol einer Beziehung, sondern als Rechenoperation Einsatz findet (vgl. Gerster & Schultz, 2004, S. 361; R. Rasch, 2001b, S. 241, 2009, S. 86; Steinweg, 2013, S. 97). Interessant ist es, zu überprüfen, inwiefern sich eine solche „Notationsschwäche“ auf die Zielerreichung auswirkt.

Systematische Veränderungen zählen in der Fachdidaktik zu den raffinierten, erfolgsversprechenden Vorgehensweisen (vgl. M. Link, 2012, S. 125, 274; R. Rasch & Schütte, 2008, S. 71; Steinweg, 2005, S. 16, 2013, S. 153–162). Das Konstanthalten einer Größe oder des Ergebnisses (*Konstanz eines Faktors*) kann als eine solche Veränderung gedeutet werden.

Während Grundschulkinder in Textaufgaben geschilderte Zusammenhänge bevorzugt additiv erschließen und selten die Multiplikation und Division heranziehen (R. Rasch, 2009, S. 75–81; vgl. Selter, 1994, S. 117, 121, 125–128, 132–133; Steinweg, 2013, S. 102), rückte jedoch im Material verstärkt in den Blick, dass die Lernenden entweder bei einer *Rechenoperation* innehielten oder mehrere kombinierten. Aufgrund dessen wurde diese induktiv gebildete Kategorie aufgegriffen und verfolgt.

Ferner wurde analog zu den Zeichnungen überprüft, inwiefern die Rechnung(-en) *kontrollierendes* bzw. *systematisches Vorgehen* transparent werden lassen (siehe Abschnitt 1.5.6).

Insgesamt bleibt festzuhalten, dass die vorhandenen empirischen Befunde zwar die Schwierigkeit aufseiten der Kinder, Lösungsgedanken und -ideen zu mathematisieren, hervorheben, aber weniger die Rechnungen als solche oder deren charakteristische Merkmale in den Vordergrund stellen. Theoriegeleitet konnten daher nur einzelne Aspekte aufgegriffen werden, die mithilfe des Materials ausdifferenziert und angereichert wurden. Die Kategorien- sowie die Subkategorienbildung erfolgte aufgrund dessen bei der Kategorie *Rechnungen* schwerpunktmäßig induktiv.

Die Übereinstimmung beider Rater lag bei den Rechnungsmerkmalen im Mittel zwischen

$ICC_{min} = .88$ und $ICC_{max} = .99$. Auch in diesem Fall ist das Urteil der unterschiedlichen Beobachter zufriedenstellend (Döring & Bortz, 2016, S. 347; Wirtz & Caspar, 2002, S. 160).

Tabellen

Sechs Subkategorien, die mit ihren Ausprägungen in Tabelle 4.4 dargestellt sind, repräsentieren die Kategorie *Tabellen*. Sie leiten sich aus den im Theorieteil aufgezeigten Merkmalen der Tabellen und Listen ab (vgl. Abschnitt 1.5.6.3 sowie 1.5.6.4). Der Wegfall der Kategorie *Listen*, der sich, wie in Abschnitt 4.2.3.1 aufgezeigt, durch die inhaltlich strukturierende Inhaltsanalyse ergab, erforderte die Hinzufügung der ersten Kategorie *Dimensionen*. Demnach stehen Tabellen mit nur einer Dimension für Listen, während die Ausprägungen zwei Dimensionen und mehr als zwei Dimensionen Tabellen von Listen abgrenzen. Die Reihenfolge der weiteren Kategorien orientiert sich sowohl an der Aufzählung der Merkmale in Abschnitt 1.5.6.3 sowie der Reihenfolge im Kodierleitfaden (siehe Anhang A.1).

Tabelle 4.4. Kategorien von Tabellen auf Basis einer inhaltlich strukturierenden Inhaltsanalyse

Kategorien	Informationsgehalt	Ausprägung
Dimensionen	Wie viele Dimensionen werden in der Tabelle erfasst?	eine Dimension; zwei Dimensionen; mehr als zwei Dimensionen
Abstraktionsgrad	Sind die Spalten mit den gegebenen Größen (Kontext) beschriftet?	Beschriftung der Spalten (Kontextbezug); keine Beschriftung der Spalten (kein direkter Kontextbezug)
Nutzungsmöglichkeiten	Wozu wird die Tabelle genutzt? Welche Funktion kann ihr zugesprochen werden?	Ergebnisprotokoll; Trial-and-Error-Protokoll; Protokoll des Erarbeitungsprozesses
Kontrolle	Werden die Bedingungen aus der Aufgabenstellung in der Tabelle sichtbar kontrolliert?	Aufgabenbedingungen werden nicht sichtbar kontrolliert; Aufgabenbedingungen werden sichtbar kontrolliert
Konstanz eines Faktors	Wird eine Größe oder das Ergebnis konstantgehalten?	eine Größe wird konstantgehalten; keine Größe wird konstantgehalten
Systematisches Vorgehen	Wird sich systematisch probierend oder unsystematisch planlos der Lösung genähert?	unsystematisches Vorgehen; systematisches Vorgehen

Die Beschriftung der Spalten wird als anspruchsvolle Denkleistung angesehen, die von den Lernenden nicht ohne Anstrengungsbereitschaft zu bewältigen ist (Bruder & Collet, 2011b, S. 59). Es ist daher zu erwarten, dass Lernende, denen es gelingt, die Spalten adäquat zu beschriften, in der Lage sind, zwischen Wesentlichem und Unwesentlichem zu abstrahieren und die zugrunde liegende Struktur zu entschlüsseln (Bruder & Collet, 2011b, S. 59; vgl. Düll, 2009b, S. 54). Hieraus resultiert die Annahme, dass der Lösungserfolg höher ist, wenn den Lösenden die Beschriftung der Spalten gelingt.

In der Theorie werden verschiedene *Nutzungsmöglichkeiten* von Tabellen unterschieden: (1) Annäherung an die Lösung über eine Auflistung absolvierten Suchprozesse (Bruder & Collet, 2011b, S. 57; Dedekind, 2012, S. 18; Franke & Ruwisch, 2010, S. 153–154; Fricke, 1987, S. 76, 79–84) und (2) schrittweise Erarbeitung der Lösung über ein Wegeprotokoll, das den Einbezug vorangegangener Schritte erforderlich macht (Düll, 2009a, S. 28–29, 2009b, S. 120–121; Franke & Ruwisch, 2010, S. 150; Fricke, 1987, S. 86–87; Krauthausen & Scherer, 2014, S. 117). Da teilweise nur das Ergebnis tabellarisch dokumentiert vorlag, aber nicht mit der Tabelle gearbeitet wurde, wurde das Ergebnisprotokoll als weitere Nutzungsmöglichkeit aufgenommen.

Die Kategorien *Kontrolle*, *Konstanz eines Faktors* und *Systematisches Vorgehen* zählten nicht nur zu typischen Merkmalen von Rechnungen, sondern konnten sowohl theoriegeleitet als auch induktiv am Material erschlossen ebenfalls den Tabellen zugesprochen werden. Die Ausprägungen dieser Kategorien unterschieden sich nicht zwischen Rechnungen und Tabellen.

Bei den Tabellenmerkmalen lagen die Interklassenkorrelationen beider Rater im Mittel zwischen $ICC_{min} = .62$ und $ICC_{max} = 1.00$. Der geringste Wert liegt nicht mehr im guten Bereich. Es handelte sich hierbei um das Repräsentationsmerkmal *Abstraktionsgrad*.

Schriftsprachliche Beschreibungen

In Anlehnung an Neumann, Beier und Ruwisch (2014, S. 118) wurden in der Hauptkategorie *schriftsprachliche Beschreibungen* drei Subkategorien unterschieden: *Mathematische Begründungsstruktur*, *sprachliche Begründungsstruktur* und *mathematische Zusammenhänge* (siehe Tabelle 4.5).

Die in den Versprachlichungen der Lernenden sichtbar gewordenen *Begründungsstrukturen* wurden in dem interdisziplinären Forschungsprojekt der Autorinnen nach zwei Gesichtspunkten analysiert: Zum einen standen mathematische, zum anderen sprachliche Indikatoren im Vordergrund.

Tabelle 4.5. Kategorien zur Erfassung schriftsprachlicher Beschreibungen

Kategorien	Informationsgehalt	Ausprägung
Mathematische Begründungsstruktur	Wird im Text ausschließlich die mathematische Entdeckung beschrieben oder sind Begründungselemente enthalten?	mathematische Entdeckungen beschrieben; mathematische Entdeckungen ansatzweise begründet; mathematische Entdeckungen beispielbezogen begründet; mathematische Entdeckungen verallgemeinernd begründet
Schriftsprachliche Begründungsstruktur	Welche sprachlichen Indikatoren sind im Text zu identifizieren, die auf eine Beschreibung, Begründung oder Verallgemeinerung hindeuten?	Indikatoren ohne Begründungsstruktur; Grund-Folge-Beziehung ohne Aufgabenbezug; Grund-Folge-Beziehung mit Aufgabenbezug; Vollständigkeit und Widerspruchsfreiheit
Mathematische Zusammenhänge	Werden die Aufgabenbedingungen im Text explizit versprachlicht?	keine Bedingung versprachlicht; eine Bedingung versprachlicht; alle Bedingungen versprachlicht

Anmerkung. Adaptiert von den Ratingskalen zur *Erfassung schriftlichen mathematischen Begründens* von Neumann et al. (2014, S. 118).

Mit der ersten Kategorie wurde erfasst, welcher Grad *mathematischer Begründungsstrukturen* den verfassten Texten zugesprochen werden kann. Es bestätigte sich, dass eine Differenzierung zwischen reinen Beschreibungen, Beschreibungen, die begründende Ansätze enthalten, und vollständigen, widerspruchsfreien Begründungen sinnvoll scheint (Neumann et al., 2014, S. 118, vgl. Abschnitt 1.5.6.5). Vollständig abstrakte Begründungen, die Verallgemeinerungen und formal-symbolische Notationen inkludieren, stellen das höchste Abstraktionsniveau dieser Kategorie dar (Neumann et al., 2014, S. 118). Da diese Anforderungen fern von dem von Grundschulkindern leistbaren Niveau liegen, wurde die Ausprägung gestrichen.

Stellt man die *schriftsprachlichen Begründungsstrukturen* ins Zentrum des Interesses, so gilt es entsprechende Indikatoren – auch als „sprachliche Marker“ bezeichnet (Neumann & Ruwisch, 2015, S. 321) – zu identifizieren, die für eine Begründung repräsentativ sind. Die Autorinnen bestimmten theoriegeleitet fünf Ausprägungen, „wo bei textlinguistische Kategorien wie Kohärenz, Textverknüpfung durch Konnektoren und explizite Bezüge innerhalb der Aufgabenlösung eine große Rolle spielten“ (Neumann et al., 2014, S. 118). Diese konnten eins zu eins auf den Kontext Textaufgaben übertragen werden: Elemente wie z. B. „und“, „aber“, „dass“, „immer“, „danach“ und „dann“ weisen in einem Text auf eine Beschreibung hin. Eine Begründungsstruktur ist nicht vorhanden. Lassen sich sprachliche Marker wie „wenn ... dann“, „obwohl“, „dadurch“, „weil“, „da“, „also“, „so wusste ich“ identifizieren, kommen Grund-Folge-Beziehungen zum Ausdruck, die entweder mit Aufgabenbezug oder ohne Aufgaben-

bezug realisiert werden. Vollständige und widerspruchsfreie Begründungen mit thematischem Aufgabenbezug stellen die höchste Niveaustufe dar, die für Grundschulkinder zu erreichen ist (vgl. Neumann & Ruwisch, 2015, S. 324, 326).

Ob die *mathematischen Zusammenhänge* richtig und vollständig im Niedergeschriebenen wiedergegeben werden, wurde mit einer dreistufigen Skala erfasst (Neumann et al., 2014, S. 120). Da der ursprünglichen Skalenkonstruktion arithmetische Begründungsaufgaben und keine Textaufgaben zugrunde lagen, wurde diese Kategorie modifiziert und adaptiert angewandt. Es wurde sich zunutze gemacht, dass sich mathematische Entdeckungen im Kontext des Textaufgabenlösens begründen lassen, indem explizit auf die Aufgabenbedingungen eingegangen wird. Da in den ausgewählten Aufgaben immer mindestens zwei einzuhaltende Bedingungen vorhanden waren, wurde zwischen der expliziten Versprachlichung keiner Bedingung, einer Bedingung oder aller Bedingungen differenziert.

Die Autorinnen konnten zeigen, dass alle drei Skalen die mathematische Begründungskompetenz als eindimensionales Konstrukt abbilden und es ermöglichen, stabile Schülerfähigkeitswerte vorherzusagen (Neumann et al., 2014, S. 120; Neumann & Ruwisch, 2015, S. 325–326; vgl. Ansatz von Roppelt & Reiss, 2012, S. 42–44). Für die internen Konsistenzen (Cronbachs Alpha) wurden in den einzelnen Skalen hinreichend zufriedenstellende Ergebnisse erzielt (Entdeckerskala: $\alpha = .86$; mathematische Begründungsskala: $\alpha = .81$; sprachliche Begründungsskala: $\alpha = .71$) (Neumann et al., 2014, S. 120; Ruwisch, 2014, S. 46).

Die Interklassenkorrelationen lagen bei diesen drei Repräsentationsmerkmalen im Mittel bei $ICC = .78$, $ICC = .88$ und $ICC = .89$, sodass auch hier sehr zufriedenstellende Überstimmungen beider Rater vorliegen (Döring & Bortz, 2016, S. 347; Wirtz & Caspar, 2002, S. 160).

4.2.4 Ableitungen des organisatorischen Rahmens

Die Beobachtungen und Erfahrungen aus der Pilotierung trugen neben der Optimierung und Feinabstimmung des Unterrichtskonzepts, der kriterienorientierten Auswahl der Aufgaben und der Kategorisierung der externen Schülerrepräsentationen dazu bei, die Altersgruppe (Abschnitt 4.2.4.1) sowie die Trainingsdauer (Abschnitt 4.2.4.2) festzusetzen. Die Wahl beider Faktoren wird nachfolgend diskutiert und festgelegt.

4.2.4.1 Wahl der Altersgruppe

Das Trainingsexperiment wurde in der Grundschule verortet, um bereits in den ersten Schuljahren der Förderung problemlösender, aber auch darstellender, kommunikativer und argumentativer Kompetenzen zu entsprechen. Grundschulkindern fehlt es in der

Regel an Vor- und Erfahrungswissen im Problemlösen, was in enger Verbindung mit nur gering ausgeprägten Problemlöse- und Repräsentationskompetenzen steht. Sie verfügen insofern über einen „Mangelzustand, also ein Defizit“, der eine Interventionsmaßnahme rechtfertigt (Hager, 2008, S. 723).

Für die Primarstufe sprach auch das Vorhandensein von Interesse und Selbstvertrauen, das besonders bei Grundschulkindern nachgewiesen wurde (Spiel, Lüftenegger, Wagner, Schober, & Finsterwald, 2011, S. 308–309). Die Autoren zeigten, dass neben dem Interesse und der höheren Einschätzung der Selbstwirksamkeit das Vertrauen in die eigenen Fähigkeiten und das Streben nach Kompetenzerweiterung bei Viertklässlern ausgeprägter vorliegt als bei Sekundarstufenschülerinnen und -schülern (Spiel et al., 2011, S. 308). Diese Ergebnisse stützten die Wahl der Zielgruppe insofern, als Einflussfaktoren wie Attitudes und Beliefs¹⁸⁶ bei Grundschulkindern nahezu als kontrolliert angesehen werden können.

Die Festsetzung der dritten Jahrgangsstufe, in der die Intervention durchgeführt wurde, wird nachfolgend empirisch gestützt diskutiert und begründet.

Die ersten beiden Grundschuljahrgänge wurden für die Interventionsmaßnahme aus mehreren Gründen ausgeschlossen. Ausschlaggebend war der Implementierungszeitpunkt der Untersuchung zu Schuljahresbeginn, der sich auf den Wissensstand der Kinder auswirkt. Für Erstklässler hätte dieser Zeitpunkt zur Folge gehabt, dass sie zu Beginn des Schuljahres mit dem Wissenstand eines Vorschulkindes in die Maßnahme gestartet wären. Analog hätte dies für Zweitklässler bedeutet, mit dem Wissensstand eines Erstklässlers zu starten. Dass es problemhafte Textaufgaben für die beiden ersten Klassenstufen gibt, wird vielfach aufgezeigt (z. B. R. Rasch, 2008a, S. 11–46; Riley et al., 1983, S. 163; Stern, 1998, S. 89, 102–103, 122–123, 130–131). Die Entscheidung des Ausschlusses fußte stattdessen auf der Tatsache, dass Schulanfänger noch nicht zwingend über ein ausgeprägtes, mathematisches Verständnis verfügen, um diese Aufgaben erfolgreich lösen zu können. Häufig können sie beim Lösen handelnd auf Zählprozeduren zurückgreifen, ohne Beziehungen zwischen den beschriebenen Mengen herstellen zu müssen (Stern, 1998, S. 99). Wenn das Denken der Lernenden noch sehr stark an Wahrnehmungen gebunden ist, kann es nicht ohne Weiteres verallgemeinert werden (Lack, 2009, S. 13). Es fehlt insbesondere noch jungen Lernenden

¹⁸⁶ Unter *Attitudes* (Einstellungen) wird in Anlehnung an Grigutsch, Raatz und Törner (1998, S. 5 in Bezug auf Triandis, 1975, S. 3) verstanden, „daß (1) eine Einstellung eine Bereitschaft zur Reaktion auf eine Situation ist und daß (2) eine Einstellung durch Konsistenz der Reaktionen gekennzeichnet ist.“ In Anlehnung an Pehkonen (1995, S. 12) werden *Beliefs* verstanden „as one's stable subjective knowledge (which also includes his feelings) of a certain object or concern to which tenable grounds may not always be found in objective considerations. The reason why a belief is adopted are defined by the individual self – usually unconsciously.“

die Flexibilität und die Fähigkeit, Beziehungen zwischen zwei Mengen näher zu spezifizieren (Halford, 1992, S. 208; Stern, 1998, S. 41).

In der Grundschule können nahezu alle mathematischen Sachverhalte mit den Kardinalprinzipien, den sogenannten Zählprinzipien, erschlossen werden (Stern, 1998, S. 79). Ein reines Kardinalzahlverständnis reicht jedoch nicht aus, um mathematisches Verständnis aufzubauen. Vielmehr ist für dessen Entwicklung der Erwerb des Relationsverständnisses im Sinne des Teil-Ganzes-Schemas erforderlich (Stern, 1998, S. 78, 108). Kinder müssen über ein „arithmetisches Netzwerk“ verfügen, das es ihnen ermöglicht, Beziehungen zwischen zwei Mengen herzustellen und zu spezifizieren (Stern, 1998, S. 84, 100). Gerade hinsichtlich der Zielsetzung des Trainingsexperiments sollten die Grundschulkinder ihrer Entwicklung entsprechend in der Lage sein, Beziehungen zwischen Mengen mithilfe externer Repräsentationen darzustellen (vgl. Abschnitt 4.5.1.3). Darüber hinaus müssen sie bei der Bearbeitung auf ein sprachliches und mathematisches Wissen zurückgreifen können, welches in der dritten und vierten Jahrgangsstufe stärker ausgeprägt ist als in den Jahren zuvor.

Ferner bestätigen die Ergebnisse von van Essen und Hamaker (1990, S. 306, siehe Abschnitt 1.5.6.1) sowie Gagatsis und Elia (2004, S. 453) die Fokussierung höherer Grundschulklassen hinsichtlich Problemlöseaufgaben. Beide Studien konnten zeigen, dass Erst- und Zweitklässler noch nicht über die Fähigkeit verfügen, unterschiedliche Repräsentationen selbstständig und flexibel zu nutzen, während ältere Lernende hierzu bereits in der Lage waren. Insbesondere Erstklässler müssen die hierzu erforderlichen mentalen Fähigkeiten erst noch entwickeln.

Empirischen Untersuchungen zufolge ist das Lösen problemhaltiger Textaufgaben auch für Grundschulkinder der Klassenstufen 3 und 4 keinesfalls trivial (Groß, 2013, S. 111–133; Hohn, 2012, S. 106–108; R. Rasch, 2001b, S. 88–89, 96–97, 104–105). Die vorliegende Untersuchung bezieht teilweise dieselben Problemaufgaben ein, wie sie in jüngeren Untersuchungen von Groß (2013) und Hohn (2012), aber auch in weiter zurückliegenden von R. Rasch (2001b) Anwendung fanden. Erstgenannte konnten im Rahmen ihrer Ist-Stand-Analyse zeigen, dass Drittklässler ohne Unterstützungsmaßnahme im Mittel $M = 0.45$ ($SD = .69$) von 5 problemhaften Textaufgaben richtig lösten, dagegen Viertklässler im Mittel $M = 0.83$ ($SD = 1.11$) Aufgaben lösten (Hohn, 2012, S. 106). Tendenziell fiel es den Drittklässlern schwerer, die vorgelegten problemhaften Textaufgaben zu lösen, als den älteren Grundschulkindern. Die Leistungen der beiden Jahrgänge unterschieden sich jedoch nicht signifikant voneinander, woraus geschlossen werden kann, dass sich die Dritt- und Viertklässler dieser Stichprobe in ihren Problemlösefähigkeiten nur geringfügig voneinander unterschieden (Hohn, 2012, S. 106).

Basierend auf den fachdidaktischen, entwicklungspsychologischen und empirischen Erkenntnissen wurde die Pilotierung für die vierte Jahrgangsstufe angesetzt. Es wurde sich davon versprochen, dass Kinder dieses Alters in der Regel über ein mathematisches Verständnis verfügen, das sie befähigt, externe Repräsentationen zu konstruieren und als Lösungsunterstützung wahrzunehmen.

Erfahrungen und Konsequenzen aus der Pilotierung

Die Eigenproduktionen brachten zum Ausdruck, dass sich die Viertklässler mit der Generierung und Nutzung externer Repräsentationen, wie beispielsweise dem Erstellen von Tabellen oder Zeichnungen, befassten und ihnen zuwandten. Sie entwickelten sehr motiviert kreative und vielfältige Lösungsideen. Es zeigte sich aber auch, dass den Viertklässlern bereits vor dem Training die Konstruktion externer Repräsentationen als Denk- und Erkenntniswerkzeuge vertraut war (siehe Ausgangsfragestellung, Abschnitt 5.2). Ihre Eigenproduktionen ließen insbesondere bei den ersten Problembewältigungen Rückschlüsse auf den bisherigen Mathematikunterricht, dessen Prägung und favorisierte Lösungswege zu. Während die Lernenden einer Klasse fast ausschließlich die ikonische Ebene wie das Erstellen von Zeichnungen fokussierten, dominierten in der Parallelklasse verstärkt arithmetische Lösungsansätze (siehe Abbildung 5.3, Abschnitt 5.2). Ein Teil der Viertklässler konnte nach und nach seine eingewöhnlichen und gewohnten Vorgehensweisen aufbrechen und sein Repertoire durch neue ergänzen, auch wenn einige Lernende an ihren Automatisierungen festhielten. Insgesamt bleibt jedoch festzuhalten, dass es mit der Zielsetzung dieser Studie nachteilig zu sein scheint, wenn die Lernenden bereits externe Repräsentationen favorisieren und demnach deren Generierung nicht neutral und unbefangen gegenüberstehen sowie in ihrem Lösungsverhalten wenig vergleichbar sind.

Dies soll nicht verstanden werden, als sei es per se nachteilig, ein solches Training in höheren Klassen zu implementieren. Für die Untersuchung war vielmehr der Grundgedanke ausschlaggebend, dass die Kinder offen, kreativ und von ihrem bisherigen Mathematikunterricht möglichst unbeeinflusst an die Lösung von Textaufgaben herangehen sollten, um erste, unvoreingenommene Erfahrungen mit externen Repräsentationen sammeln zu können. Der Unterricht sollte gerade diese ursprünglichen und unverfälschten Herangehensweisen der Lernenden zulassen, aufgreifen und weiterentwickeln, um deren Potenzial auszuschöpfen (R. Rasch, 2001b, S. 76, 2009, S. 69). Die vorgefundenen Gegebenheiten in den vierten Klassen schränkten diesen Grundgedanken ein, weshalb die Untersuchung bereits in der dritten Klassenstufe durchgeführt wurde.

4.2.4.2 Wahl der Trainingsdauer

Die Festsetzung des zeitlichen Trainingsumfangs erfolgte unter Berücksichtigung zweier wesentlicher Faktoren. Einerseits wurden aus der Theorie und Empirie abgeleitete Empfehlungen herangezogen, andererseits wurde das Handlungsfeld *Schule* berücksichtigt. Nachstehend werden beide Faktoren zusammenfassend dargestellt.

In den Abschnitten 1.5.5.3, 1.5.5.4 und 2.4 wurde betont, dass die Implementierung unterschiedlicher Vor- und Herangehensweisen und deren Reflexionen kontinuierlich in neuen Zusammenhängen zum Gegenstand des Unterrichts gemacht werden müssen (z. B. Söbbeke & Steenpaß, 2014, S. 11, 13). Hintergrund hierfür ist, dass Wissen nur das Verständnis voranbringen und sicher, flexibel genutzt werden kann, wenn die Lernenden immer wieder aufs Neue vor der Herausforderung stehen, ihr Vor- und Erfahrungswissen zu aktivieren und adaptiert anzuwenden (Guberman & Leikin, 2012, S. 37; E. C. Wittmann, 1993, S. 394, 1998, S. 157). Mit zunehmenden Erfahrungen wird ein gezielter Rückgriff auf das eigene Repräsentationsrepertoire prognostiziert, sodass in Abhängigkeit von dem gerade zu lösenden Problem flexibler entschieden und ausgewählt werden kann, welche externe Repräsentation sich eignet. Es geht eindeutig hervor, dass dies langfristige, in mehreren Schritten ablaufende Prozesse sind, welche immerfort mit Reflexionsphasen einhergehen müssen (vgl. Bromm, 2009, S. 42; Freesemann & Breucker, 2014, S. 11; König, 1992, S. 25).

Eingeschlossen ist, dass sich die Fähigkeit, zwischen Darstellungen wechseln zu können, langfristig entwickelt (Freesemann & Breucker, 2014, S. 11; Prediger, 2013a, S. 172) sowie die Konstruktion externer Repräsentationen erlernt werden muss (Cox, 1999, S. 360; Dedekind, 2012, S. 8; Dörfler, 2006, S. 210; Fehse, 2001, S. 50; Prediger, 2013a, S. 172; R. Rasch, 2001b, S. 47).

„The ability to solve mathematics problems develops slowly over a very long period of time because success depends on much more than mathematical content knowledge“ (Lester & Kroll, 1990, S. 56; vgl. de Bock et al., 2003, S. 460). Nur durch kontinuierliches Anwenden und Üben können Lernprozesse optimiert werden, weshalb deren Förderung ein unabdingbares Ziel der Intervention darstellt.

Einerseits soll mit dem Fokus, Problemlösekompetenzen zu fördern – auch unter dem Gesichtspunkt einer aufgabenunabhängigen Förderung –, ausreichend Zeit eingeplant werden. Andererseits galt es auch zu berücksichtigen, dass Lehrkräfte gewonnen werden mussten, die ihre Unterrichtszeit zur Verfügung stellten bzw. sich einverstanden zeigten, in diesem Zeitraum die vorgegebenen Textaufgaben in ihrem Unterricht bearbeiten zu lassen und den Verlauf zu protokollieren. Letzteres bedeutete eine nicht zu unterschätzende Belastung und Mehrarbeit für die Lehrkräfte, was ein entscheidender Grund dafür war, eine für die Lehrkräfte noch zumutbare Trainingsdauer festzusetzen.

Erfahrungen und Konsequenzen aus der Pilotierung

Rückblickend schienen sechs Trainingswochen in der Pilotierung zu kurz zu sein, weshalb sich in der Hauptuntersuchung für eine längere Zeitspanne ausgesprochen wurde. Ein Interventionszeitraum von 12 Unterrichtswochen brachte mit dem direkten Start nach den Sommerferien den Vorteil mit sich, dass lediglich eine zweiwöchige Unterbrechung durch die Herbstferien nach sechs Unterrichtswochen stattfand. Somit konnten eine zusätzliche, zweite ferienbedingte Unterbrechung verhindert und eine abschließende Testung vor den Weihnachtsferien ermöglicht werden. Die Rekrutierung der Schulen ergab, dass dies für die einen einen Zeitraum maximaler Länge, für die anderen wiederum einen zu langen Zeitraum darstellte. Schulen mit dem erstgenannten Argument konnten im Gegensatz zu Schulen mit dem zuletzt genannten Argument für die Teilnahme an der Untersuchung gewonnen werden.

4.3 Design der experimentellen Interventionsstudie

Ziel der Untersuchung ist es, die Wirksamkeit des Trainings bzw. der unterschiedlichen Interventionsmaßnahmen zu überprüfen. Sie gelten im Allgemeinen als wirksam oder effektiv, wenn beobachtbare Leistungen durch eine Interventionsmaßnahme hypothesenkonform beeinflusst werden (Hager, 2008, S. 721; Hager & Hasselhorn, 2000, S. 44). Demnach musste empirisch nachgewiesen werden, ob es aufgrund der Intervention zu einer Veränderung bzw. (gerichteten) Verbesserung der Leistung kam, die auch über die Zeit erkennbar wird (Hager & Hasselhorn, 2000, S. 65; Seidel & Shavelson, 2007, S. 456; Klauer, 2001, S. 22). Aus diesem Grund wurde in der Untersuchung ein Prätest-Posttest-Kontrollgruppen-Design verfolgt (Hager, 2008, S. 727; Morris, 2008, S. 365). Durch die Datenerhebung über drei Messzeitpunkte hinweg liegt der Untersuchung ein 2 (Training) \times 2 (Kommunikation) \times 3 (Zeit) –faktorielles Design mit Messwiederholung auf dem letzten Faktor zugrunde. Um aufgabenspezifische Unterschiede zu prüfen, wurde der Faktor Aufgabentyp mit den Abstufungen Vergleichsaufgaben, Aufgaben mit komplexen Informationen und Aufgaben zur Verhältnisverteilung aufgenommen.

Der erste Faktor *Repräsentationstraining* (T) unterscheidet sich darin, dass die Drittklässler beim Bearbeiten problemhaltiger Textaufgaben durch ein Repräsentationstraining unterstützt bzw. nicht unterstützt wurden. Der Faktor trägt demnach die Ausprägungen *Training* (T+) sowie *kein Training* (T-). Klassen, die nicht trainiert wurden, bearbeiteten die Aufgaben mit ihrer Lehrkraft im regulären Mathematikunterricht. Den Lehrkräften war die Intention und Zielsetzung der Studie nicht bekannt (vgl. Hager & Hasselhorn, 2000, S. 48). Der zweite Faktor *Kommunikation* (K) hat die Ausprägung *komunikative Zweiersettings* (K+), wenn sich die Kinder in Paaren über ihre indivi-

duellen Entdeckungen austauschten, und die Ausprägung *keine kommunikativen Zweiersettings* (K-), wenn kein Austausch stattfand.

Inwiefern die unabhängigen Variablen *Repräsentationstraining* (T) und *Kommunikation* (K) die abhängigen Variablen¹⁸⁷ beeinflussen, wurde anhand eines Quasi-Experiments¹⁸⁸ erprobt. Ziel war es, die Effizienz unterschiedlicher Unterrichtsarrangements zu messen sowie die Wirkung des Repräsentationstrainings zu erfassen (vgl. Seidel & Shavelson, 2007, S. 456). Um mögliche Wirkungen vergleichen zu können, wurden die zu evaluierenden Interventionsmaßnahmen durch den „Vergleich verschiedener Kombinationen“ beider Komponenten (T und K) variiert (Rost, 2000, S. 137). Hieraus gingen vier Kombinationen und somit vier experimentelle Bedingungen hervor: T+K+, T+K-, T-K+ und T-K- (siehe Abbildung 4.2). Es galt zu untersuchen, ob Drittklässler von einem Repräsentationstraining (T+), einem Kommunikationstraining (K+) oder einer Kombination beider Trainingsformen (T+K+) profitieren können und ob sie sich nach der Intervention von Gleichaltrigen, welche die gleichen Aufgaben in ihrem regulären Mathematikunterricht, der Kontrollgruppe T-K-, bearbeiteten, unterscheiden (Hager, 2008, S. 724).

		Training (T)	
		Training (T+)	kein Training (T-)
Kommunikation (K)	kommunikative Zweiersettings (K+)	T+K+	T-K+
	keine kommunikativen Zweiersettings (K-)	T+K-	T-K- (Kontrollgruppe)

Abbildung 4.2. Experimentelle Bedingungen des Trainingsexperimentes.

Der Untersuchung lag nach Hager (2008, S. 724–725) eine Kombination aus isolierter und vergleichender Evaluation zugrunde, da die Wirksamkeit der drei Trainingsarten noch nicht abgesichert war. Um allerdings einen möglichen Leistungsanstieg zwischen den Messungen auf die beiden Faktoren Training und Kommunikation zurückführen

¹⁸⁷ Für eine ausführliche Darstellung der abhängigen Variablen und ihrer Operationalisierung wird auf den nachfolgende Abschnitt 4.5.1 verwiesen.

¹⁸⁸ Es handelt sich um ein quasi-experimentelles Design, weil eine vollständig randomisierte Zuweisung der Klassen zu den einzelnen experimentellen Bedingungen aufgrund des Parallelisierungsverfahrens nicht realisiert werden konnte (Rost, 2000, S. 137).

zu können, und um sagen zu können, dass die Interventionsmaßnahmen wirksam sind, wurde die Kontrollgruppe T-K- implementiert (Hager & Hasselhorn, 2000, S. 48–49; Schumann, 2012, S. 117).¹⁸⁹ Erst die Tatsache, dass eine Leistungssteigerung in den experimentellen Bedingungen, nicht aber vergleichbar in der Kontrollgruppe zu beobachten ist, gilt als Indiz für die Effektivität eines Trainings (Hager, 2008, S. 731; Schumann, 2012, S. 117).

Die Kontrollgruppe T-K- wurde weder trainiert noch waren kommunikative Zweiersettings Bestandteil des Unterrichts. Die Wahrscheinlichkeit, dass die als wirksam erachteten Maßnahmen (Training und Kommunikation) in dieser Interventionssituation nicht realisiert wurden, konnte mit der Tatsache, dass den Lehrenden der Fokus der Untersuchung nicht bekannt war, erhöht werden (Hager & Hasselhorn, 2000, S. 49; Rost, 2000, S. 137). Die Lehrkräfte standen vor der Aufgabe, die problemhaften Textaufgaben ohne Ablaufplan und ohne vorgegebene Strategieanweisung in ihren regulären Mathematikunterricht zu integrieren (vgl. Heinze, 2007, S. 17). Um der Forderung unterschiedlicher Fokussierungen innerhalb der Kontrollgruppe nachzukommen (Hager, 2008, S. 724), wurden die Lehrkräfte angehalten, ihre Problemlösestunden zu dokumentieren. Darüber hinaus notierte jedes Kind uneingeschränkt alles, was es üblicherweise in sein Mathematikheft oder auf einen Notizblock schreibt, in ein ihm zur Verfügung gestelltes Blanko-Knobelheft. Hierzu zählten alle erstellten Notizen aus Einzel- oder Gruppenarbeitsphasen sowie der allgemeine Unterrichtsmitschrieb, das Tafelbild.

Die Unterrichtsprotokolle der Lehrkräfte (Lehrerperspektive) und die Knobelhefte der Kinder (Lernerperspektive) ermöglichen rückblickend Einblicke in den jeweils erfolgten Unterricht. Somit konnte der stattgefundenen Unterricht aus Lehrer- und aus Lernerperspektive retrospektiv nachvollzogen werden. Wenn in den Knobelheften aller Kinder einer Klasse der gleiche Unterrichtsmitschrieb vorzufinden war und keine Hinweise auf individuelle Lösungen oder Teamlösungen gegeben waren, dann waren dies Indizien für einen lehrerzentrierten Unterricht. Die Lehrerprotokolle und Knobelhefte der Kinder wurden darüber hinaus genutzt, um Informationen über die Verwendung zusätzlicher Materialien in den einzelnen Klassen zu gewinnen. Zeichneten die Kinder bzw. benannten die Lehrkräfte den Einsatz von Cuisenaire-Stäben, Rechenplättchen, Rechenketten oder dem Zahlenstrahl, um nur eine Auswahl möglicher Anschauungsmittel zu nennen, wurde davon ausgegangen, dass diese im Unterricht Verwendung fand.

¹⁸⁹ Es gilt, die Wirksamkeitshypothese zu prüfen, die besagt, dass das Training im Vergleich zu einem Kontrolltraining effektiv ist (Hager, 2008, S. 724).

Die Durchführung des Kontrolltrainings oblag dem gleichen Rahmen (1 Mathematikstunde pro Woche, max. 45 Minuten) und den gleichen Randbedingungen (1 Aufgabe pro Stunde, identische Aufgabenreihenfolge, Dokumentation des Lösungsprozesses im Knobelheft) wie der Durchführung des Repräsentationstrainings (Hager, 2008, S. 724). Hätte die Kontrollgruppe im Interventionszeitraum keine problemhaltigen Textaufgaben bearbeitet, sondern den konventionellen Unterricht absolviert, so könnte ein möglicher Leistungsanstieg bzw. Kompetenzzuwachs der Trainingsgruppen nicht eindeutig auf die Faktoren Training, Kommunikation oder deren Kombination zurückgeführt werden (vgl. Grenzen einer Wartegruppe als Vergleichsgruppe, Hasselhorn & Hager, 2006, S. 344). Eine zusätzliche Variation auf Aufgabenebene mit den Ausprägungen *problemhaltige Textaufgaben* und *keine problemhaltigen Textaufgaben* hätte Wirkungen zur Folge, die sich auf die abhängigen Variablen auswirken, obwohl sie mit den interessierenden Faktoren nicht in Zusammenhang stehen (Eid, Gollwitzer, & Schmitt, 2011, S. 57). Allgemeiner formuliert würde dies auch auf Zuwendungsebene zutreffen, wenn zusätzlich zwischen den Ausprägungen *Intervention* und *regulärer Unterricht* differiert worden wäre. Konfundierungen im Sinne von Scheineffekten blieben unkontrolliert, wenn sich die Versuchsbedingungen in mehr als den zu untersuchenden Faktoren unterschieden hätten. Aus diesem Grund werden die experimentellen Manipulationen als konstruktvalide angesehen, wenn alle Probanden problemhaltige Textaufgaben bearbeiten und keine inhaltlichen Variationen vorgenommen werden (Eid et al., 2011, S. 60). Würde lediglich ein Teil der Probanden an der Interventionsmaßnahme teilnehmen und der andere Teil einer Wartegruppe gleichen, welche sich nicht dem Lösen problemhaltiger Textaufgaben widmet, so wäre die interventionsgebundene Wirkung, welche aus der verstärkten Unterstützung und Zuwendung beim Lösen problemhaltiger Textaufgaben hervorgeht, unerwünscht, wenn die beiden Gruppen verglichen werden sollen (Hager & Hasselhorn, 2000, S. 46; Hasselhorn & Hager, 2006, S. 344).

Die T-K+-Gruppe, die das Alternativtraining absolvierte, unterschied sich dahin gehend von der Kontrollgruppe, dass sich die Kinder in kommunikativen Zweiersettings (K+) über ihre individuellen Entdeckungen austauschten.

Die Gruppen T+K+ und T+K- wurden beim Lösen problemhaltiger Textaufgaben durch ein Repräsentationstraining (T+) unterstützt. Sie unterschieden sich darin, dass sich Lernende der Gruppe T+K+ zusätzlich über ihre Entdeckungen austauschten (K+). Auch wenn der zeitliche Umfang des Experimentaltrainings nicht zwingend mit dem des Alternativtrainings übereinstimmen muss (Hasselhorn & Hager, 2006, S. 724), wurde in den experimentellen Bedingungen T+K+ und T+K- darauf geachtet, dass die ersten drei Unterrichtsphasen übereinstimmend konzipiert wurden und sie sich lediglich in der letzten Unterrichtsphase unterschieden. In der Zeit, in der sich die

Kinder der Gruppe mit Kommunikation (T+K+) gegenseitig über ihre Entdeckungen austauschten, konnten die Kinder der Gruppe ohne Kommunikation (T+K-) länger an der eigenständigen Problembearbeitung verweilen oder Kopfrechenaufgaben lösen. Für eine detaillierte Beschreibung des Trainings, dessen Organisation und Umsetzung sowie die Variation der einzelnen Bedingungen wird auf Abschnitt 4.4 verwiesen.

Neben der Ausprägung *Training* wurde der Ausprägung *kein Training* des ersten Faktors ein hoher Stellenwert eingeräumt. Der Fokus war darauf gerichtet, möglichst den *regulären* Problemlöseunterricht abzubilden. Weil die in den jeweiligen Klassen unterrichtenden Mathematiklehrkräfte hierfür als Experten angesehen wurden, führten sie den Problemlöseunterricht durch. Ihnen wurde Freiheit in der Unterrichtsgestaltung zugesprochen.

In diesem Paradigma [der vergleichenden Evaluation] müssen die Trainings von Personen durchgeführt werden, die Experten für das von ihnen durchzuführende Training sind. Dadurch ergeben sich – im Gegensatz zur isolierten Evaluation – pro Training andere durchführende Personen, und diese Kombination von Programm und Durchführenden stellt sozusagen ein „Paket“ dar, das nicht aufgeschnürt werden darf. (Hasselhorn & Hager, 2006, S. 724, Ergänzung durch die Autorin)

Die Versuchsleiterin, die das Trainingskonzept entwickelt hat, kann als Expertin für dessen Umsetzung und Durchführung angesehen werden (Pressley & Harris, 1994, S. 196). Im Rahmen des Alternativtrainings (T-K+) wäre sie jedoch aufgrund ihres Wissens zugunsten des Experimentaltrainings voreingenommen. Pressley und Harris (1994, S. 196–198) akzentuieren, wie wichtig es ist, dass gerade Alternativtrainings von einem Experten und nicht der Versuchsleitung durchgeführt werden. Daher sind ausgebildete Grundschullehrkräfte am ehesten als Experten für die Durchführung des Alternativtrainings anzusehen. Hinzukommt, dass sowohl die Knobelhefte aller Lernenden als auch die Unterrichtsprotokolle der Lehrkräfte der Auswertung zugrunde lagen, sodass Rückschlüsse auf den jeweiligen Unterricht möglich waren. Dem Einwand, dass der Unterricht einer nicht-trainierten Klasse nicht rekonstruierbar war und keine generalisierbaren Schlüsse zuließ, kann somit begegnet werden. Darüber hinaus wurden keine Analysen auf Klassen- und Individualebene, sondern nur auf Gruppenebene angestrebt, sodass davon ausgegangen werden kann, dass sich Effekte, die auf den Unterricht einzelner Klassen zurückzuführen sind, herausmitteln und die Verletzung der Durchführungsobjektivität vernachlässigbar ist.

4.4 Das Training

Nachfolgend wird der Kern der Untersuchung, das Training, vorgestellt. In diesem Zusammenhang werden neben der Trainingsrationale (Abschnitt 4.4.1), die Aus-

gangslage (Abschnitt 4.4.2), die zentralen Ziele (Abschnitt 4.4.3) und der Aufbau des Trainings (Abschnitt 4.4.4) dargelegt.

4.4.1 *Trainingsrationale*

Damit ein Training der Forderung nach einem hohen Grad an Standardisierung sowie Validität und Schlüssigkeit gerecht werden kann, müssen in der Konzipierung der Intervention bestimmte Anforderungen berücksichtigt werden. Die einzelnen Trainingsbedingungen müssen vergleichbar sein, individuelles Lernen ermöglichen sowie anschlussfähiges Wissen fördern (Felbrich, 2005, S. 129).

Um die Vergleichbarkeit der Trainingsbedingungen sicherstellen zu können, wurde lediglich das Unterrichtssetting variiert und sonstige lernförderliche Einflussmöglichkeiten und Lernbedingungen wurden konstant gehalten (siehe Abschnitt 4.3). Der organisatorische Rahmen, die zu bearbeitenden Aufgaben und die zeitliche Struktur waren standardisiert und demnach für alle Treatmentbedingungen gleich. Folglich widmeten sich Klassen, gleich welcher experimentellen Bedingungen, innerhalb einer festen Mathematikstunde pro Woche dem Problemlösen. In jeder Stunde wurde immer nur eine problemhafte Textaufgabe bearbeitet. Insgesamt berücksichtigte die Intervention sechs verschiedene Aufgabentypen. Jede war mit zwei repräsentativen Aufgaben vertreten, sodass sich eine Interventionszeit von zwölf Wochen ergab. Die Bearbeitungsreihenfolge der Aufgaben wurde randomisiert ermittelt: Zuerst wurden alle Textaufgaben entsprechend ihrem Aufgabentyp durchnummeriert. Anschließend wurde eine Zufallsstichprobe ohne Zurücklegen mit Berücksichtigung der Reihenfolge gezogen (Eid et al., 2011, S. 148). Die Zufallsziehung unterlag einer Bedingung: Zwei der zwölf Trainingsaufgaben, die YU-GI-OH!®-Kartenaufgabe und die Kaninchenaufgabe (vgl. Abschnitt 4.2.2.2), mussten unter den ersten sechs zu bearbeitenden Aufgaben liegen. Sie zählen zu den Aufgabentypen, die sowohl im Training als auch in den Testungen berücksichtigt wurden. Um möglichst voneinander unabhängige Messungen zu erwirken und Übungs- und Erinnerungseffekte zu reduzieren, wurde die Zufallsziehung so lange wiederholt, bis diese Bedingung erfüllt war. Die Bearbeitungsreihenfolge war dann für alle Gruppen identisch und bindend. Eine Permutation der Aufgabenfolge wurde verworfen, da insgesamt $12!$ mögliche Reihungen existierten. Da sich jede Treatmentgruppe aus leistungsstärkeren und leistungsschwächeren Klassen zusammensetzte, wäre eine zusätzliche Variation der Aufgabenreihenfolge innerhalb jeder Gruppe prekär. Eine Konfundierung zwischen Leistungsniveau der Klassen und der Bearbeitungsreihenfolge wäre nicht auszuschließen. Darüber hinaus hätte eine Permutation zur Folge gehabt, dass die Testleistungen der Klassen, die unmittelbar zuvor die strukturgeleichen Trainingsaufgaben bearbeiteten, überschätzt worden wären. Letztlich hätten keine voneinander unabhängigen Messungen vorgelegen.

Wie in Abschnitt 4.1.2 bereits aufgezeigt, wurden die Problemaufgaben so ausgewählt, dass sie allen Lernenden einen Zugang ermöglichen und individuelle Entdeckungen zuließen. Lediglich die Reflexion und der Austausch über die unterschiedlichen Heran- und Vorgehensweisen wurden gemeinsam absolviert. Dabei wurde bewusst die Verantwortung an die Lernenden abgegeben, die sich im Begründen und Argumentieren erproben und positionieren sollten und die auf Reaktionen ihrer Mitschüler reagieren sollten. Im Hinblick auf die Vergleichbarkeit wurde in allen experimentellen Trainingsbedingungen die gleiche Anzahl an Repräsentationen reflektiert und es wurde darauf geachtet, dass jede Repräsentationsform (eine Zeichnung, eine Tabelle, eine Rechnung und eine Begründung) vertreten war. Die Aufgabentypen wurden bewusst variiert (vgl. Abschnitt 4.2.2.2), um anschlussfähiges und nicht kontextspezifisches Wissen zu fördern. Die Probanden sollten erfahren, dass sich die Konstruktion von Denk- und Erkenntniswerkzeugen in unterschiedlichen Kontexten eignen und deren Nutzung mit Vor- und Nachteilen einhergehen kann. Trainingsoptimierungen, die sich aufgrund der Pilotierung ergaben, wurden bereits in Abschnitt 4.2.1 vorgestellt, weshalb nachfolgend im Detail die Ausgangslage (Abschnitt 4.4.2), die verfolgten Ziele (Abschnitt 4.4.3) und der Aufbau des Trainings (Abschnitt 4.4.4) geschildert werden.

4.4.2 Ausgangslage des Trainings

Im Rahmen der Interventionsmaßnahme wurden die Lernenden motiviert, kreative Denk- und Erkenntniswerkzeuge zu konstruieren, die ihnen helfen, die ProblembARRIERE zu überwinden. Die Grundidee des Trainings bestand darin, diese individuellen Lösungen der Kinder wertschätzend und anerkennend zum Ausgangspunkt des Trainings zu machen und vorgefertigten Musterlösungen vorzuziehen (vgl. Bruder, 2000, S. 6; Hirt & Wälti, 2010, S. 13, 21; Sundermann, 2014, S. 24). Musterlösungen hätten zwar den Vorteil gehabt, dass sich alle zehn Trainingsklassen in der Reflexionsphase auf die gleiche Ausgangslage stützen würden, auf der trainiert worden wäre. Ein nicht zu unterschätzender Nachteil wäre jedoch gewesen, dass sich die Lernenden weniger mit den Lösungen hätten identifizieren können und ihre Motivation dadurch geringer gewesen wäre, als wenn sie einen Bezug zu den ausgewählten Lösungsbeispielen gehabt hätten. Ihre Anstrengungsbereitschaft, eigene Denk- und Erkenntniswerkzeuge zu konstruieren, würde sinken, wenn die Lehrkraft ausschließlich vorgefertigte (Muster-)Lösungen, an denen gelernt werden soll, bereitstellt und die Einzelleistungen der Kinder nicht anerkennt und würdigt. Darüber hinaus ist es für die Unterrichtspraxis naheliegend und von Vorteil, die klasseneigenen Repräsentationen zu nutzen und zum Reflexionsgegenstand zu machen.

Dass die Konstruktion externer Repräsentationen nicht auf Anhieb funktionieren kann, bleibt unbestritten. Gerade an diesem Punkt setzte das Training an und griff auch un-

vollständige, fehlerhafte Lösungen oder Teilleistungen auf, wenn es sich lohnte, die Heran- und Vorgehensweise öffentlich zu machen und gemeinsam weiterzuentwickeln oder zu vervollständigen. Demnach wurden nie ausschließlich vollständige und richtige Lösungen im Sinne von „Musterlösungen“ ausgewählt und eingesetzt. Sicherlich war diese Vorgehensweise für die meisten Lernenden eher ungewohnt, da häufig nur „Musterlösungen“ in Sicherungsphasen des Unterrichts Beachtung finden. Das Potenzial von fehlerhaften Schülerlösungen wurde insbesondere darin gesehen, dass sie Lernmöglichkeiten schaffen, Fehlvorstellungen aufdecken und repräsentationale Veränderungen anregen können. „Gerade im Zusammenhang mit der Bearbeitung problemhaltiger Textaufgaben relativieren sich die Fehler, die Kinder machen. Die Denkwege der Kinder sind sehr oft beeindruckend, auch die fehlerhaften“ (R. Rasch, 2001b, S. 300). Die Richtigkeit des erzielten Ergebnisses bzw. der Lösungszahl war aus diesem Grund nicht ausschlaggebend für die Wahl der externen Repräsentationen. Vielmehr sollten die ausgewählten Eigenproduktionen so strukturiert sein, dass sie die Achtsamkeit der Kinder auf relevante Eigenschaften richteten und das „Hineinsehen“ bzw. „Rekonstruieren von Beziehungen erleichter[te]n“ (Thies, 2002, S. 86). Die Denk- und Erkenntniswerkzeuge nahmen *Modellcharakter* an und forderten die Kinder heraus, sie zu prüfen, zu verstehen und gegebenenfalls als eigene Problembewältigung in Betracht zu ziehen: „Richtig verstanden haben sie Aufforderungscharakter, einen eigenen Rechenweg zu finden“ (Schütte, 2008a, S. 138). Die Reflexion erfolgte in der nächsten Stunde, damit der Versuchsleiterin ausreichend Zeit zur Verfügung stand, die einzelnen Denk- und Erkenntniswerkzeuge der Lernenden zu analysieren und auszuwerten, und um eine gezielte, durchdachte, keinesfalls spontane Auswahl zu treffen, anders als dies wohl unmittelbar nach der eigenständigen Lernphase der Fall gewesen wäre (z. B. Sundermann, 2014, S. 24).

4.4.3 Zentrale Ziele des Trainings

Das konzipierte Training kann als *kognitives Training* verstanden werden: Es stellt eine Interventions- bzw. Fördermaßnahme dar, die das Ziel verfolgt, „kognitive Fähigkeiten, Funktionen und Fertigkeiten zu verbessern“ (Hasselhorn & Hager, 2006, S. 341). Es werden Transfereffekte im Sinne erzielter Leistungsverbesserungen über die Zeit und auf andere, nicht direkt trainierte Aufgabenbereiche erwartet (Hasselhorn & Hager, 1998, S. 87).

Unter diesen Gesichtspunkten wird das vorliegende Repräsentationstraining als eine Interventionsmaßnahme verstanden, die eine Erweiterung der Problemlösekompetenzen anstrebt (vgl. Stern, 1998, S. 40). Die Leistungsentwicklung soll sich sowohl über die Zeit als auch bei einem nicht-trainierten Aufgabentyp zeigen (siehe Abschnitt 4.4.3).

Die Konzeption des Trainings orientierte sich an den Komponenten erfolgreichen Trainings (Hasselhorn & Hager, 2006, S. 342–343, vgl. 1998, S. 96–98):

1. Einüben bereichsspezifischer Strategien.
2. Explizites Vermitteln von Informationen über Nutzen und Anwendungsmöglichkeiten der Strategien.
3. Einüben genereller Techniken und Prinzipien der Selbstkontrolle und Lernregulation.
4. Direktes Einüben des Strategietransfers durch Variation der Aufgabenstellung.
5. Verknüpfung der Trainingsinhalte mit persönlicher Zielmotivation der Teilnehmer.

In der vorliegenden Untersuchung ist es zutreffender, anstelle von Strategien von externen Repräsentationen zu sprechen. Diese werden eingeübt, diskutiert und reflektiert sowie immer wieder an neuen Aufgabenstellungen erprobt. Im Training wird auch regelmäßig hervorgehoben, wie wichtig es ist, das Ergebnis zu überprüfen und zu kontrollieren. Die Lernenden sollen erleben, dass Fehler durch eine Ergebniskontrolle aufgespürt werden können. Schwierig zeigt sich jedoch die Komponente 5. Möglicherweise kann ein nachhaltiger Einsatz externer Repräsentationen mithilfe des Follow-up-Tests gezeigt werden. Dies geschieht allerdings nur, wenn die Lernenden die externen Repräsentationen im Training als hilfreich und nützlich erfahren sowie ihren funktionalen Mehrwert bewusst erleben (Hasselhorn & Hager, 2006, S. 343; Renkl & Nückles, 2006, S. 144). Es gilt, deren persönliche Motivation anzuregen. Der nachfolgend aufgezeigte Aufbau des Trainings orientiert sich an den fünf Erfolgskomponenten.

4.4.4 *Aufbau des Trainings*

Das Training war schwerpunktmäßig in der *Reflexionsphase* zu Beginn jeder Problemlösestunde verortet (siehe Abschnitt 4.1.1). Es basierte auf der Veröffentlichung von vier selbstgenerierten Schülerrepräsentationen aus der vorangegangenen Problemlösestunde. Um die Schülerzentriertheit dieser Phase in den Mittelpunkt zu rücken, wurde die Reflexionsphase im Sitzkreis realisiert. Die jeweils zu thematisierende Eigenproduktion wurde ausgelegt. Um sicherzustellen, dass alle Kinder die Lösung einsehen konnten, wurden neben der originalen Doppelseite des Knobelheftes noch vier bzw. fünf Vervielfältigungen (je nach Klassenstärke) ausgebreitet. Ziel war es, im Detail die einzelnen Lösungsprozesse, Heran- und Vorgehensweisen, zu rekonstruieren.

Die vier ausgewählten Repräsentationsformen wurden immer einzeln und nacheinander fokussiert, um die Konzentration auf eine Vorgehensweise zu bündeln und nicht durch wechselndes Aufeinander-Bezugnehmen Verwirrungen hervorzurufen. Dies ist jedoch nicht dergestalt zu interpretieren, dass das Training eine getrennte Abhandlung der Eigenproduktionen zu manifestieren suchte. Im Gegenteil sollte das Hauptaugenmerk zunächst verständnissichernd immer auf eine Heran- und Vorgehensweise gerichtet sein, um im nächsten Schritt darauf aufbauend Vernetzungen untereinander herstellen zu können. Die verschiedenen Repräsentationsformen zeitgleich zum Unter-

richtsgegenstand zu machen, was ein verfrühtes, gegebenenfalls oberflächliches Vergleichen von Eigenproduktionen nach sich gezogen und möglicherweise auf fehlendem Verständnis basiert hätte, wurde hier als wenig sinnvoll erachtet und deshalb nicht in Betracht gezogen. Insgesamt umfasst das Training fünf zentrale Bausteine (vgl. Cooper & Warren, 2008, S. 25, 27) (siehe Tabelle 4.6).

Tabelle 4.6 Zentrale Trainingsbausteine der Reflexionsphase

Zentrale Trainingsbausteine
1 Phase des Hineinversetzens, Nachvollziehens und Verstehens
2 Phase der Benennung genutzter Repräsentationen
3 Phase der Beschreibung und anfänglicher argumentativer Tätigkeiten
4 Phase kollektiver Aushandlung lösungsunterstützender Merkmale von Repräsentationen
5 Phase der Entwicklung eines externen Repräsentations-Repertoires

Wenn Lernende mit einer externen Repräsentation eines Gleichaltrigen konfrontiert werden, die sie nicht selbst produziert haben, stehen sie als Rezipient vor der Herausforderung, sich in die Heran- und Vorgehensweise des Produzenten hineinzuversetzen. Sie müssen sich in den vorliegenden Lösungsprozess eindenken und versuchen, ihn nachzuvollziehen und zu verstehen (Baustein 1). Es ist keinesfalls trivial, fremde Denk- und Erkenntniswerkzeuge nachvollziehen und verstehen zu können. Diese Fähigkeiten müssen langfristig forciert, trainiert und immer wieder besprochen werden, da Lernende alternative Heran- und Vorgehensweisen nur in ihr Repertoire lösungsunterstützender Hilfsmittel aufnehmen, wenn sie diese verstanden und deren Mehrwert erkannt haben (siehe Abschnitt 3.2). Versteht der Problemlösende die präsentierte Vorgehensweise nicht, so kann schlussgefolgert werden, dass er sie bei seinem nächsten Lösungsversuch auch nicht als mögliche externe Repräsentation heranzieht. Aus diesem Grund wurde sensibel darauf geachtet, dass ausreichend Zeit für die einzelnen Trainingskomponenten und mögliche Nachfragen zur Verfügung stand.

Nachdem jeder Lernende die präsentierte Repräsentation sichtete und zu deuten versuchte, galt es im nächsten Schritt, diese zu benennen, zu beschreiben und erste Begründungen anzustellen (Bausteine 2 und 3). Im Vordergrund stand dabei die Beantwortung der Fragen „Was?“, „Wie?“, „Wann?“ und „Warum?“ (vgl. Kramarski et al., 2010, S. 182–183 sowie Abschnitt 2.3 und 4.2.1.2). Die Benennung, ob es sich um eine Tabelle, um eine Rechnung, um eine Zeichnung oder um eine Begründung handle, stellte erwartungsgemäß kein großes Problem für die Lernenden dar. Der Zweck wurde darin gesehen, die einzelnen Repräsentationsformen im Gedächtnis der Lernen-

den zu verankern. Sie sollen mit den einzelnen Bezeichnungen mögliche Heran- und Vorgehensweisen, aber auch Vorteile assoziieren, welche sie im individuellen Lösungsprozess erinnern und aufgreifen können. Es wird erwartet, dass dies ein Stück weit dazu beiträgt, die Lernenden zu einer bewussten Auswahl von Repräsentationen zu führen.

Das Beschreiben und Begründen konkret vorliegender Handlungsweisen sollte Denk- und Erkenntniswege transparent werden lassen, das Verständnis schärfen und eine Identifizierung mit ihnen ermöglichen. Auf diese Weise wurden mögliche Vorgehensweisen öffentlich und zugänglich, über die der Einzelne so vielleicht noch nicht nachgedacht hatte. Kollektiv sollten konkrete Merkmale der externen Repräsentation ausgearbeitet, diskutiert und ausgehandelt werden (Baustein 4). Es ist wichtig, dass die Lernenden ermutigt werden, Vermutungen zu äußern, und dadurch nicht nur Rezipienten bleiben, sondern auch die Rolle von Produzenten einnehmen. Es ist zu erwarten, dass sich dies positiv auf deren individuelle Reflexionsfähigkeit auswirkt (vgl. 2.4.1). Das Erkennen von Merkmalen sowie deren Bedeutungen und Vorteile im konkreten Falle soll ihr Handlungsspektrum erweitern und im Idealfall zu einer Aufnahme der Repräsentationsform in ihr Repertoire führen (Baustein 5).

Ausreichend Raum für klärende Fragen – sei es an den Produzenten oder an die Klasse, wenn der Produzent anonym bleiben wollte – war in allen Trainingskomponenten vorhanden. Während am Anfang noch die Lehrkraft stützend und moderierend agierte, ging im Zuge der Trainingszeit die Organisation und Moderation dieser Phase vollständig in die Verantwortung der Lernenden über. Die Redekette, bei der sich die Lernenden gegenseitig aufriefen, wurde zu einem festen Bestandteil dieser Phase (vgl. B. Janssen, 2009, S. 33).

Implizit wurde auch in der explorativen Phase sowie in den kommunikativen Settings des Unterrichts trainiert (siehe Abschnitt 4.1.1). Die Kinder übten sich bei Erstgenanntem im Generieren adäquater Repräsentationen und somit im Verschriftlichen ihrer Denkwege und Lösungsideen. In den kommunikativen Settings standen sie in einem geschützten Rahmen vor der Aufgabe, ihre Vorgehensweise zu beschreiben und zu begründen sowie sich auf Alternativen ihres Gegenübers einzulassen. Wenn sie über einen längeren Zeitraum zu solchen Aktivitäten angeregt werden, ist davon auszugehen, dass sich dadurch eine Weiterentwicklung ihrer Kompetenzen anbahnt. Es wird hier von implizit ausgeführten Trainingskomponenten (Entwicklung und Konstruktion lösungsunterstützender Repräsentationen und Vergleich von Repräsentationen) gesprochen, weil die Versuchsleitung keinen direkten Einfluss auf diese Prozesse ausüben kann. Da die Lehrkraft weder direkten Einfluss auf den Ablauf noch auf die Ausführung der Trainingskomponenten hat sowie die Lernenden selbstverantwortlich tätig

sind, wird von einem Anwenden der Trainingskomponenten und nicht von einem Training selbst gesprochen.

4.5 Variablen und ihre Operationalisierung

Um die Wirksamkeit von Interventionsmaßnahmen, wie beispielsweise eines Trainings, zu messen und zu prüfen, müssen auf Basis theoretischer Überlegungen Kriterien des Interventionserfolgs festgelegt werden, die das Erfassen beobachtbarer Leistungen mit einer Prätest-Posttest-Erhebung gestatten (Hasselhorn & Hager, 2008, S. 382). In diesem Zusammenhang muss berücksichtigt werden, welche Kompetenzen gefördert werden sollen und wie diese zu messen bzw. zu operationalisieren sind (vgl. Abschnitt 4.5.1). Insgesamt ist darauf zu achten, dass wenige abhängige Variablen fokussiert werden (Cohen, 1990, S. 1305).

Darüber hinaus muss nachgewiesen werden, dass erzielte Leistungsveränderungen bzw. Unterschiede zwischen den experimentellen Bedingungen ($T+K+$, $T+K-$, $T-K+$ und $T-K-$) auf die Intervention und nicht auf potenzielle Alternativerklärungen zurückzuführen sind (Brandtstädtter, 1990, S. 218; Hager, 2000b, S. 155; Shadish, Cook, & Campbell, 2002, S. 56). Ersteres schließt die Analyse der inhaltsanalytisch gewonnenen Repräsentationsmerkmale ein (Abschnitt 4.5.2). Letztere müssen für das Zustandekommen der Ergebnisse ausgeschlossen werden können.

Die Verbesserungen können nur dann als hinreichend intensiv und das Ausmaß der Zielerreichung kann nur dann als hinreichend bewertet werden, wenn potentielle Störfaktoren, die Verbesserungen bzw. die die Zielerreichung vortäuschen könnten, kontrolliert werden. Auch sollten solche möglichen Störfaktoren kontrolliert werden, die als hinreichend intensiv bewertbare Verbesserungen auf die Ziele hin oder die ein als hinreichend bewertbares Ausmaß der Annäherung an die Ziele maskieren könnten. (Hager, 2000b, S. 155)

Ableitungen, die sich für die Wahl der Kontrollvariablen und deren Operationalisierung ergaben, werden im Abschnitt 4.5.3 thematisiert.

4.5.1 Abhängige Variablen

In der Untersuchung werden drei abhängige Variablen unterschieden: *Lösungserfolg*, *Problemlösekompetenzen* und *Anzahl selbstgenerierter externer Repräsentationen*. Die Wahl der abhängigen Variablen wird in Abschnitt 4.5.1.1 erläutert. Welche Operationalisierungsmöglichkeiten bestehen, werden im Anschluss daran aufgezeigt und diskutiert (Abschnitt 4.5.1.2). Das erfasste Aufgabenspektrum sowie die Konstruktion der Testaufgaben werden in diesem Zusammenhang offengelegt. Der sich hieraus ableitende Textaufgaben-Performance-Test wird in Abschnitt 4.5.1.3 vorgestellt. Ergänzend werden die zugrunde liegende Reliabilitäten und die faktorenanalytischen Struktur

berichtet sowie das letztlich gewählte Messinstrument diskutiert. Abschließend wird die Bildung der Testscores erläutert (Abschnitt 4.5.1.4).

4.5.1.1 Wahl der abhängigen Variablen

Was mit den drei abhängigen Variable zu messen beabsichtigt wurde, wird nachfolgend im Einzelnen dargelegt. Zudem wird herausgestellt, weshalb zwischen der Erfassung des *Lösungserfolgs* und der Erfassung der *Problemlösekompetenzen* unterschieden wird.

Lösungserfolg

Die Eigenproduktionen der Lernenden, die beim Lösen problemhaltiger Textaufgaben entstanden sind, geben Aufschluss über die individuellen Leistungen. Sie können genutzt werden, um sowohl den *Lösungserfolg* als auch die *Problemlösekompetenzen* des Einzelnen zu erfassen. Der Lösungserfolg blendet die Erfassung des Lösungsprozesses vollständig aus und fungiert als *strenge* Kriterium, das ausschließlich die aktuelle Performance, konkret die Korrektheit des erzielten Ergebnisses evaluiert. Dadurch kann sie problemlos direkt beobachtet und quantifiziert werden (Hasselhorn, 1995, S. 18). Es wird zwischen zwei Ausprägungen, richtig und falsch, differenziert. Wenn mit dem Lösungserfolg auch teilweise richtige Lösungen erfasst werden sollen, bedarf es Kriterien, die eine eindeutige Quantifizierung derer ermöglichen. Interpretationsspielräume führen zur Herabsetzung der Reliabilität und sind deshalb zu vermeiden. Um sichern zu können, dass im Wiederholungsfall nahezu identische Werte resultieren (Hussy, 2013, S. 41), wurde sich für eine dichotome Erfassung der abhängigen Variablen *Lösungserfolg* ausgesprochen.

Problemlösekompetenzen

Gleichwohl blieben erreichte Teilkompetenzen, die in Zwischenschritten oder Teilergebnissen sichtbar wurden, jedoch (noch) nicht zum richtigen Ergebnis führten, nicht unbeachtet. Gerade im Hinblick darauf, dass sich das Verständnis sowie die Problemlösekompetenzen langsam und schrittweise entwickeln (vgl. Abschnitt 1.5.5.4), ist es erforderlich, auch den Lösungsprozess und somit einzeln vollzogene Schritte einzubeziehen, anzuerkennen und zu würdigen. „*Yet, when evaluating progress in problem solving, it makes ultimate sense to attempt to design useful ways of looking at the written work that students produce*“ (Lester & Kroll, 1990, S. 63). Das alleinige Beurteilen mit richtig oder falsch kann dem Leistungspotenzial eines Grundschulkindes und seinen erbrachten geistigen Kräften und Bemühungen nicht vollständig gerecht werden. Kompetenzsteigerungen müssen nicht zwingend direkt in Erfolg münden, sind aber schon als positive Veränderungen und Leistungsentwicklungen zu deuten. Auf-

grund dessen wurde eine zweite, *weichere* abhängige Variable implementiert, die den Problemlöseprozess und somit das Konstrukt der Kompetenzen ins Zentrum des Interesses rückt (Charles et al., 1987, S. 29, 34; Lester & Kroll, 1990, S. 63). Auch wenn zu erwarten ist, dass die beiden abhängigen Variablen hoch miteinander korrelieren, da sie ähnliche Informationen messen, konkurrieren sie nicht, sondern ergänzen sich auf symbiotische Weise (vgl. Hasselhorn, 1995, S. 19). Würde entweder nur der Prozess oder nur das Produkt bewertet werden, wären Aussagen über die Wirkung der Interventionsmaßnahme nur eingeschränkt oder nur unzureichend möglich.

Die Einschätzung der Problemlösekompetenzen erfolgte kriteriumsbezogen, indem die Eigenproduktionen der Lernenden einer adaptierten Form der von Charles, Lester und O'Daffer (1987, S. 34–35) entwickelten Beurteilungsskala, der *Focused Holistic Scoring Point Scale*, unterzogen wurden (siehe Lester & Kroll, 1990, S. 64).

Die Skala wurde von den Autoren konzipiert, um vorhandene Problemlösestrategien in den Problemlöseprozessen der Lernenden zu identifizieren und nicht nur das Ergebnis zu bewerten. Aufgrund dessen wird sie auch als ganzheitliches Bewertungsverfahren (*holistic scoring*) bezeichnet (Charles et al., 1987, S. 34; Lester & Kroll, 1990, S. 64). Da das Hauptaugenmerk des Repräsentationstrainings nicht auf die Vermittlung von Strategien, sondern auf die Generierung externer Repräsentationen gerichtet war, musste die Skala adaptiert werden. Der ursprüngliche Strategiefokus wurde durch den Repräsentationsfokus ersetzt (siehe Tabelle 4.7).

Ein Charakteristikum der Bewertungsskala ist, dass nicht einzelnen Subkategorien, sondern dem gesamten Prozess auf Grundlage festgesetzter Kriterien ein Punktwert zugeordnet wird (*focused*, Charles et al., 1987, S. 34). „It is *focused* because the overall rating is a result of looking for particular, previously identified, characteristics in the students' work“ (Lester & Kroll, 1990, S. 64, Hervorhebung im Original). Die Bewertungsskala stellt einen Leitfaden dar, nach dem aufgabenbezogen die Eigenproduktionen kodiert und mit einem Aufgabenscore zwischen 0 und 4 versehen wird (siehe Tabelle 4.7). Es darf nicht fälschlicherweise davon ausgegangen werden, dass das erzielte Ergebnis im Fokus der Beurteilung stand. Dies wird an mehreren Stellen deutlich und an zwei exemplarischen Beispielen aufgezeigt: Ein richtiges Ergebnis wird nicht zwangsläufig mit einer vollen Punktzahl versehen (siehe z. B. Score 2, Tabelle 4.7).

Für den Fall, dass das Ergebnis falsch ist, können unter bestimmten Bedingungen trotzdem hohe Aufgabenscores erreicht werden (siehe z. B. Score 3 und 4, Tabelle 4.7). Eine mangelnde Darstellung des Problemlöseprozesses bei richtigem Ergebnis kann nachteilig, eine ausführliche Darstellung bei falschem Ergebnis, wenn dieses beispielsweise nur auf einem Abschreibefehler beruht, kann von Vorteil für die Erfassung

der Problemlösekompetenzen sein. Diese Tatsache zeigt sehr deutlich, dass es sich lohnt, sowohl den Prozess als auch den Lösungserfolg separat zu erfassen.

Tabelle 4.7. Instrument zur Analyse der individuellen Problemlöseprozesse

Score	Die Seite des Knobelheftes weist eine der folgenden Eigenschaften auf:
0	<ul style="list-style-type: none"> ▪ Sie ist leer. ▪ Die Daten/Informationen der Aufgabe wurden schlichtweg nur herausgeschrieben, aber es wurde nichts mit ihnen gemacht. ▪ Mit den Daten wird gearbeitet, aber scheinbar ohne das Problem verstanden zu haben. ▪ Es wird nur das Ergebnis genannt, und das ist falsch.
1	<ul style="list-style-type: none"> ▪ Eine Lösungssuche wird begonnen und geht über das Herausschreiben der Daten/Informationen der Aufgabe hinaus, ansatzweise spiegelt sich Verständnis wider, aber die Vorgehensweise führt nicht zur richtigen Lösung. ▪ Das Kind startet mit einer ungeeigneten Repräsentation, welche nicht weiter ausgeführt/verfolgt wird, und es gibt keinen Hinweis darauf, dass es zu einer anderen Repräsentationsform wechselt. Es scheint, dass der Problemlösende einen nicht zielführenden Ansatz auswählt und daraufhin aufgibt. ▪ Der Lernende versucht, ein Teilziel zu erreichen, aber schafft es nicht.
2	<ul style="list-style-type: none"> ▪ Das Kind verwendet eine ungeeignete Repräsentation und findet daraufhin eine falsche Lösung, aber der Hefteintrag zeigt ein gewisses Verständnis der Problematik auf. ▪ Eine geeignete Repräsentation wird genutzt, aber ... <ul style="list-style-type: none"> – sie wird nicht ausreichend ausgeführt, um eine Lösung zu finden; – sie wird falsch ausgeführt und somit wird kein oder ein falsches Ergebnis gefunden. ▪ Das Kind erreicht erfolgreich ein Teilziel, aber arbeitet nicht weiter. ▪ Das korrekte Ergebnis wurde notiert, aber ... <ul style="list-style-type: none"> – der Lösungsweg ist nicht nachvollziehbar; – der Lösungsweg wird nicht gezeigt.
3	<ul style="list-style-type: none"> ▪ Das Kind verwendet eine geeignete Repräsentation, die zum Ziel hätte führen können, wenn sie vollständig ausgeführt worden wäre. Es fehlt jedoch der letzte Schritt zur erfolgreichen Lösungsfindung. ▪ Eine zielführende und geeignete Repräsentationsform wird richtig ausgeführt, aber ... <ul style="list-style-type: none"> – der Lernende beantwortet ohne ertüchtlichen Grund das Problem falsch; – der Zahlenwert des Ergebnisses ist richtig und das Ergebnis wurde nicht benannt oder falsch benannt; – kein Ergebnis wurde benannt (Bezug zur Problemstellung fehlt). ▪ Das korrekte Ergebnis wurde notiert und es gibt Hinweise darauf, dass geeignete Repräsentationen ausgewählt wurden. Allerdings ist die Umsetzung/ Darstellungsweise der Repräsentation nicht eindeutig klar.
4	<ul style="list-style-type: none"> ▪ Dem Kind unterläuft beim Ausführen einer geeigneten Repräsentation ein Fehler. Dieser spiegelt weder fehlendes Verständnis der Problemstellung noch der Ausführung der Repräsentation wider. Es liegt ein Abschreib-, Rechen- oder Leichtsinnsfehler vor. ▪ Eine geeignete Repräsentation wird ausgewählt und ausgeführt. Das richtige Ergebnis wird in Bezug auf die Problemstellung gegeben.

Anmerkung. Adaptiert von der *Focused Holistic Scoring Point Scale* von Charles et al. (1987, S. 35).

Anzahl selbstgenerierter externer Repräsentationen

Entsprechend der Ausgangsfragestellung (vgl. Abschnitt 5.2) wurde sowohl in der Pilotierung als auch in der Hauptstudie untersucht, ob die Interventionsmaßnahme dazu führt, dass die Lernenden eine höhere *Anzahl externer Repräsentationen* konstruieren

(vgl. Kindfield, 1993, S. 10; van Essen & Hamaker, 1990, S. 308). Ob sich diesbezüglich Unterschiede zwischen den experimentellen Bedingungen nachweisen lassen, gilt es herauszufinden. Die Anzahl selbstgenerierter externer Repräsentationen wird als Maß für die Häufigkeit des Generierens von Denk- und Erkenntniswerkzeugen verstanden. Sie stellt ebenso wie der Lösungserfolg eine konkrete Variable dar, die direkt beobachtbar und quantifizierbar ist (vgl. Hussy, 2013, S. 38), und komplettiert die in der Untersuchung gemessenen abhängigen Variablen. Wenn die Ausgangsfragestellung hypothesenkonform erfüllt ist, dann kann sichergestellt werden, dass von dem Training eine Wirkung ausgeht. Ob diese wirksam ist, gilt es zu prüfen. Hierbei wurde zuerst auf Testebene und im darauffolgenden Schritt auf Aufgabenebene ausgewertet. Letzteres soll ergänzend zum Testscore aufdecken, ob der Aufgabentyp einen Einfluss auf die Anzahl selbstgenerierter externer Repräsentationen ausübt.

4.5.1.2 Grundgedanken zur Operationalisierung

Um zu prüfen, ob sich durch das Training die *Anzahl selbstgenerierter externer Repräsentationen* erhöhen lässt und ob sich die Intervention auf den *Lösungserfolg* und die erzielten *Problemlösekompetenzen* auswirkt, war ein Messinstrument erforderlich, das es erlaubte, sowohl das erzielte Produkt als auch den Denk- und Lösungsprozess differenziert zu erfassen. Um diesem Anspruch gerecht zu werden, musste den Lernenden die Möglichkeit eingeräumt werden, entsprechend ihrem Niveau und Lernstand eigenständig arbeiten zu können. Da sich individuelle Lernstände und -entwicklungen nur schwer abbilden lassen, wenn sie an vorgegebene Heran- und Vorgehensweisen gebunden sind, wurde ein offenes Antwortformat sowohl in den Trainings- als auch den Testaufgaben verfolgt. Konkrete Instruktionen, wie beispielsweise *Löse die Aufgabe zeichnerisch* oder auch Aufgaben im Multiple-Choice-Format, widersprachen dem offenen Zugang des Trainings, beeinflussten die Individualität, Spontanität und Kreativität der Lernenden und spiegelten nur bedingt die tatsächliche Lernentwicklung wider. Die Frage danach, inwiefern Lernende mit vorgegebenen Repräsentationen fähig sind, den Lösungsweg zu bestreiten, war zu keinem Zeitpunkt Fragestellung der vorliegenden Untersuchung. Stattdessen sollten die Lernenden herausgefordert werden, die Problemstellungen mit ihrem Vor- und Erfahrungswissen unter Rückgriff auf ihr Repräsentationsrepertoire zu bewältigen.

Ein weiterer Vorteil des offenen Formates bestand darin, dass außer der Überprüfung oben genannter Hypothesen an den erhobenen Daten hypothesengenerierend herausgefunden werden konnte, welche externen Repräsentationen geeignet sind, den Lösungserfolg bzw. die Problemlösekompetenzen vorherzusagen. Es galt auch diesbezüglich ein Verfahren zu wählen, das fähig ist, Informationen über individuelle Denk- und Erkenntniswerkzeuge aufzudecken. Die Operationalisierung der abhängigen Variablen

erfolgte daher mit einem eigens konstruierten Textaufgaben-Performance-Test, dem ein offenes Antwortformat zugrunde lag. Gedanken, die dessen Konstruktion beeinflussten, werden nachfolgend diskutiert.

Anzahl der Testaufgaben

Jeder Test (Prätest, Posttest, Follow-up-Test) umfasste drei problemhafte Textaufgaben mit offenem Antwortformat. Unmittelbar vor und nach dem Problemlösetest wurden die metakognitiven Fähigkeiten der Kinder mit dem Schülerfragenbogen II erhöhen (siehe Abschnitt 4.5.3.4). Folglich waren sie vor der Testung gefordert, ihren Lösungserfolg vorherzusagen, nach der Testung, ihn zu evaluieren. Bei einer durchschnittlichen Bearbeitungszeit von 10 Minuten pro Textaufgabe und der Messung der metakognitiven Fähigkeiten ergab sich die festgesetzte Anzahl von drei Aufgaben, die von den Kindern in einer Unterrichtsstunde (45 Minuten) zu bewältigen waren. Es sollte sichergestellt werden, dass die Lernenden sowohl den Fragebogen als auch jede einzelne Testaufgabe ohne Zeitdruck bearbeiten können, und dass darüber hinaus ausreichend Zeit für die Generierung externer Repräsentationen zur Verfügung steht.

Insgesamt galt es, die Unterrichtsstruktur der Grundschulen zu berücksichtigen, weshalb nicht vier oder mehr Testaufgaben berücksichtigt wurden. Nur in wenigen Ausnahmen waren die Unterrichtsstunden im 60-Minuten-Takt organisiert. Hätte man den Grundschulkindern ein Zeitfenster von 90 Minuten eingeräumt und sechs zu bearbeitende Aufgaben gestellt, so wäre zu erwarten gewesen, dass sowohl deren Konzentrationsfähigkeit als auch deren Motivation, die Rechenprozesse abzubilden, abfallen würden. Negative Auswirkungen auf die Konstruktionen externer Repräsentationen hätten ebenso die Folge sein können, weshalb die Testung in einer Doppelstunde für die vorliegende Untersuchung ausgeschlossen wurde.

Aufgabenspektrum der Testaufgaben

Ferner bildet der Textaufgaben-Performance-Test drei unterschiedliche Aufgabentypen mit jeweils einer repräsentativen problemhaften Textaufgabe ab. Es wurden zwei trainierte Aufgabentypen sowie ein untrainierter Aufgabentyp herangezogen. Für diese Berücksichtigung sprach, dass eine kognitive Intervention nicht nur Leistungssteigerungen in Bereichen anstrebt, die im Training berücksichtigt wurden. Vielmehr liegt nahe, dass das Training die Kinder befähigen kann, einen Transfer auf untrainierte Aufgabentypen zu leisten (Stahn, 2011, S. 45). Dies würde bedeuten, dass Kinder, die die Trainingsaufgaben mithilfe von Zeichnungen, Tabellen, Rechnungen oder schriftsprachliche Beschreibungen erfolgreich lösen können, ebenso in der Lage sind, unbekannte, neuartige Aufgaben mithilfe geeigneten Lösungsunterstützungen zu bewältigen. Ein Transfer kann durch Übertragung bereits erworbener Vorgehensweisen auf

unbekannte Aufgabentypen gelingen. Die Kritik, die kognitive Intervention entspräche ausschließlich dem Erwerb aufgabenspezifischer Lösungsstrategien, kann durch die Berücksichtigung eines untrainierten Aufgabentyps im Idealfall entkräftet werden.

Die Generalisierbarkeit von Trainingseffekten wird in der Literatur kontrovers diskutiert (vgl. Diskussion in Stahn, 2011, S. 43). Es bleibt jedoch festzuhalten, dass die vorliegende Arbeit keine generalisierbaren Schlüsse auf andere Bereiche, wie beispielsweise die allgemeinen mathematischen Basiskompetenzen, anstrebt. Die oben bezeichneten Transferleistungen werden immer innerhalb des Problemlösens gesehen. Darüber hinaus kann davon ausgegangen werden, dass Grundschulkinder Aufgaben gleicher Struktur, d. h. gleichen Aufgabentyps, nicht unmittelbar als solche identifizieren. Mit einem neuen Kontext verbinden die meisten Kinder eine neue, andersartige Aufgabe. Die Verwandtschaft der Aufgaben bleibt ihnen größtenteils verborgen (Lorenz, 2006b, S. 45).

Zu den trainierten Aufgabentypen, die in den Leistungstests abgebildet wurden, zählten die *Vergleichsaufgaben* und *Aufgaben mit komplexen Informationen*. Vor dem Hintergrund, dass zwei der drei Aufgabentypen bereits in den Trainingsaufgaben vertreten sein sollten, wurde ein dritter Bereich gewählt, der für die Kinder neuartig war. Die *Aufgaben zur Verhältnisverteilung* bilden den untrainierten Aufgabentypen ab. Nachfolgend wird die Auswahl der drei Aufgabentypen begründet.

Stellt man Lösungsraten von Grundschulkindern bei *Vergleichsaufgaben* mit anderen Textaufgaben gegenüber, so bleibt festzuhalten, dass Vergleichsaufgaben seltener richtig gelöst werden und daher zu den schwierigeren Problemaufgaben zählen (Gibb, 1956, S. 77; Moser Opitz, 2013, S. 204; R. Rasch, 2009, S. 68; Riley et al., 1983, S. 162–163; Stern, 1993, S. 7; Verschaffel, de Corte, & Pauwels, 1992, S. 85). Diese Ergebnisse werden in der Literatur weniger auf deren mathematische, sondern hauptsächlich auf deren semantische Struktur zurückgeführt (Greßmühl, 1995, S. 44; R. Rasch, 2009, S. 68). Die Schwierigkeit der Aufgaben steigt, weil die beiden Vergleichsmengen nicht gegeben sind.¹⁹⁰ Der zu leistende Vergleich gibt keine eindeutigen Hinweise auf die zu verwendende Rechenoperation (R. Rasch, 2009, S. 68), weshalb ein konkretes Lösungsschema nicht einfach abgerufen werden kann. Der Vergleich hat beispielsweise im Gegensatz zu Ausgleichsaufgaben statischen und keinen dynamischen Charakter (Riley et al., 1983, S. 161), wodurch ein größeres Abstraktionsvermögen erforderlich wird (Stern, 1997, S. 101). In solchen Problemstellungen werden beispielsweise mehrere Informationen wie „Lukas hat 6 mehr als Jonas“ einander gegenübergestellt (vgl. R. Rasch, 2016, S. 59). Wird das Geschilderte als kon-

¹⁹⁰ Gleiches kann übertragen werden, wenn die Differenzmenge oder die Gesamtmenge gesucht ist (siehe Tabelle 4.3–4.6 in Riley et al., 1983).

krete Menge verstanden, kann die Barriere nicht überwunden werden. Eine erfolgreiche Problemlösung kann nur gelingen, wenn das Gegebene als Relation zwischen zwei Mengen verstanden wird (z. B. Stern, 1997, S. 157; Stern, Felbrich, & Schneider, 2006, S. 462), was letztlich voraussetzt, dass die Lösenden den kardinalen und ordinalen Zahlaspekt beherrschen und über ein fortgeschrittenes Zahlverständnis verfügen (vgl. Fritz, Ricken, Gerlach, & Schmidt, 2007, S. 20; Stern, Felbrich, & Schneider, 2006, S. 462). Es gilt, die gegebene Zahl als Komposition zweier weiterer Zahlen zu verstehen. Um beim Beispiel der YU-GI-OH!®-Kartenaufgabe zu bleiben, heißt das, die Kinder müssten die 30 Karten als zusammengesetzte Zahl wie beispielsweise $15 + 15$, $16 + 14$, $17 + 13$ usw. verstehen. Ist das Teil-Ganze-Schema nicht verinnerlicht, misslingt die Konstruktion einer adäquaten Repräsentation (Renkl & Stern, 1994, S. 29). Textaufgaben, denen ein quantitativer Vergleich zugrunde liegt, sind ein „guter Indikator für ein fortgeschrittenes mathematisches Verständnis im Grundschulalter“ (Stern et al., 2006, S. 462). Ob das Generieren von Denk- und Erkenntniswerkzeugen Grundschulkinder befähigen kann, zugrunde liegende Relationen zwischen Mengen darzustellen und aufgrund dessen die beiden Vergleichsmengen zu entschlüsseln, gilt es zu überprüfen. Studien von Renkl und Stern (1994, S. 35) zeigten bereits, dass neben kognitiven Voraussetzungen auch Unterrichtsmerkmale den Lösungserfolg von Textaufgaben positiv beeinflussen können.

Aufgaben mit komplexen Informationen sind bereits bei Pólya (1966, S. 48) zu finden. Auf den ersten Blick scheint es sich aus Sicht der Lernenden um eher leichtere Problemaufgaben zu handeln (R. Rasch, 2008a, S. 93). Recht zügig wird deutlich, dass mehr Denkkapazitäten beansprucht werden, als zunächst vermutet. Lösende treffen unweigerlich auf eine Barriere, die sie zum Grübeln veranlasst und häufig zum probierenden Lösen führt. Ihr Problemcharakter ergibt sich aus der Dichte gegebener Informationen und heranziehender Operationen. Um die Aufgaben lösen zu können, müssen zeitgleich mehrere voneinander abhängige Bedingungen berücksichtigt werden. Auch wenn Grundschulkindern das Lösen mithilfe eines Gleichungssystems verwehrt bleibt, lassen sich vielfältige alternative Vorgehensweisen erschließen (vgl. Hoffmann, 2011, S. 12; R. Rasch, 2001b, S. 274–275; E. C. Wittmann & Müller, 2013a, S. 108, 2013b, S. 88). Die Aufgabenkategorie umfasst Aufgaben heterogener, mathematisch-logischer sowie semantischer Struktur (R. Rasch, 2001b, S. 35). Folglich unterscheiden sich die Aufgaben in ihrer Komplexität und somit auch in ihrer Schwierigkeit. Im Allgemeinen kann der Aufgabentyp als herausfordernd und repräsentativ angesehen werden, weshalb er im Problemlöse-Performance-Test vertreten war.

Insbesondere dann, wenn ungleichmäßige Verteilungen erfolgen wie es bei *Aufgaben zur Verhältnisverteilung* der Fall ist, nutzen Grundschulkinder stärker ihr Wissen über

Zahlbeziehungen als ältere Schüler (R. Rasch, 2001b, S. 226). Das Aufstellen adäquater Gleichungen schaffen im Vergleich zu Sekundarstufenschülern die meisten Grundschulkinder nicht. Vielmehr müssen gegebene Zahlenwerte entsprechend den Anforderungen aufgeteilt werden – jedoch nicht immer zu gleich großen Teilen. Um das Ziel zu erreichen, muss teilweise auch in zusätzlichen Schritten über die Gesamtzahl hinausgedacht werden.

Es bleibt festzuhalten, dass alle drei Aufgabentypen zu den Aufgaben schwieriger mathematischer Strukturen zählen, die bis zu dem Ende der Grundschulzeit und darüber hinaus ihren Problemcharakter bewahren (R. Rasch, 2001b, S. 28, 40; vgl. Hohn, 2012, S. 108). Darüber hinaus wäre der Einsatz weiterer Aufgabentypen für diese Altersgruppe denkbar gewesen (vgl. R. Rasch, 2008a, 2016). Exemplarisch sei hier auf die *Aufgaben mit kombinatorischem Hintergrund*, die *Ausgleichs-* und die *Bewegungsaufgaben* verwiesen. Nach dem Ausschlussprinzip soll die Auswahl der oben genannten Aufgabentypen begründet werden.

Kinder finden recht schnell einen Zugang zu *Aufgaben mit kombinatorischem Hintergrund*, auch wenn es ihnen anfänglich eher noch schwerfällt, sich von der Suche nach dem passenden Grundmodell zu lösen und tragfähige Modelle zu konstruieren (z. B. Hohn, 2012, S. 64; R. Rasch, 2001b, S. 84, 192–193). Gemeinsame Reflexionen über die Aufgabenstruktur sowie über mögliche Heran- und Vorgehensweisen wirken anfänglichen Schwierigkeiten entgegen und bleiben Lernenden in Erinnerung. Bei weiteren Problemstellungen gleichen kombinatorischen Charakters gelingt es ihnen häufig, diese Handlungsschemata zu reaktivieren und adaptiert anzuwenden. Diesen Aufgaben liegt ein hoher Wiedererkennungswert zugrunde. Vor- und Erfahrungswissen scheinen sich hierbei positiv im Lösungserfolg widerzuspiegeln (vgl. auch Ergebnisse der Pilotierung in Tabelle 4.1, Abschnitt 4.2.2.1). Es liegt nahe, dass Lernende nach der Bearbeitung dreier solcher Aufgaben¹⁹¹ ein Lösungsschema entwickelt und/oder automatisiert haben, das sie zum Lösen von Aufgaben mit kombinatorischem Hintergrund befähigt. Ein solches Resultat kann nicht zwingend auf das Training zurückgeführt werden, da ebenso denkbar ist, dass die Bearbeitung solcher Aufgaben im regulären Mathematikunterricht, insbesondere im Rahmen der Leitidee *Daten, Häufigkeiten und Wahrscheinlichkeit* (KMK 2005b, S. 11), hierzu verhilft. Die Pilotierung bestätigte, dass die *Handschlag-* und *EisAufgabe* zu den *leichteren* Problemaufgaben zählten (vgl. Tabelle 4.1, Abschnitt 4.2.2.1), weshalb sie als Testaufgaben nicht infrage kamen. Es wird erwartet, dass die Kinder nach dem Training in der Lage sind, die Struktur zu erkennen, sodass die Aufgaben ihren Problemcharakter verlieren.

¹⁹¹ Die Kinder wurden das erste Mal im Prätest mit einer Aufgabe mit kombinatorischem Hintergrund konfrontiert. Zwei weitere Aufgaben dieses Aufgabentyps folgten im Training.

Ausgleichs- und Bewegungsaufgaben bilden neben den Aufgaben zum Verhältnis von Zwischenraum und Begrenzung weitere zentrale Aufgabentypen ab. Es wird prognostiziert, dass den Kindern der geschilderte Zusammenhang aufgrund seiner Dynamik leichter in Erinnerung blieb und reproduziert werden konnte, als dies bei Vergleichsaufgaben der Fall sein würde. Erzielte Effekte könnten eventuell mit höherer Wahrscheinlichkeit auf die Erinnerung zurückgeführt werden. Im Allgemeinen muss jedoch betont werden, dass auch diese Aufgabentypen für das Testinstrument geeignet schienen, eine Entscheidung jedoch zugunsten der *Vergleichsaufgaben* und *Aufgaben mit komplexen Informationen* getroffen wurde.

Konstruktion der Testaufgaben

Mit dem Ziel, die Leistungsunterschiede der Lernenden vor und nach einer Intervention exakt erfassen zu können, sind vergleichbare Testinstrumente erforderlich. Demnach müssen den Tests für die verschiedenen Messzeitpunkte Aufgaben gleicher Schwierigkeit zugrunde liegen (vgl. Koerber, 2000, S. 123). Ist diese Voraussetzung verletzt, kann und darf eine positive Leistungsentwicklung nicht ausschließlich auf das Training zurückgeführt werden. Aufgaben mit heterogener Schwierigkeit können Effekte hervorrufen, die nicht aus dem Training resultieren. Dies spricht dafür, dass Aufgaben, deren Lösungsprozesse miteinander verglichen werden sollen, dem gleichen Aufgabentyp angehören. In diesem Zusammenhang muss jedoch zwingend Beachtung finden, dass sich auch Aufgaben gleicher Struktur, gleichen Typs, in ihrer Schwierigkeit unterscheiden können (z. B. Stern, 1998, S. 107–110, 2003, S. 119). Wenn man allerdings berücksichtigt, dass „nicht der oberflächliche Kontext den Transfer von mathematischem Wissen steuert, sondern die mit den quantitativen Konzepten verbundenen funktionalen Möglichkeiten und Einschränkungen“ (Stern, 1998, S. 109), dann können beispielsweise strukturgleiche Aufgaben mit identischem Satzbau und übereinstimmender semantischer Struktur durchaus als vergleichbar angesehen werden. Es handelt sich dann nicht um reformulierte Aufgaben, bei denen sich die Art der Fragestellung ändert (Stern, 1998, S. 103). Wenn ausschließlich einzelne Begriffe ausgetauscht und keine sprachlichen Veränderungen vorgenommen werden, liegen Änderungen der Oberflächenmerkmale vor, die wenig Einfluss auf die Aufgabenschwierigkeit ausüben. Werden Testaufgaben gelöst, die dieses Kriterium erfüllen, dann können gleiche mentale Modelle bzw. gleiche externe Repräsentationen herangezogen werden (Greeno et al., 1993, S. 157; Stern, 1998, S. 109). Es verbirgt sich „das gleiche abstrakte Muster hinter immer wieder neuen Verkleidungen“ (Devlin, 2003, S. 172). Ein Wissenstransfer von einer Testaufgabe zur anderen ist demnach gegeben, wenn der Bearbeitung von Problemstellungen gleiche Handlungsmöglichkeiten und Einschränkungen zugrunde liegen (Greeno et al., 1993, S. 160–161; Stern, 2001, S. 178, 182). Diese Annahme wäre verletzt, wenn beispielsweise in einer Testaufgabe nach Mengen

und in der zu vergleichenden nach Relationen zwischen den Mengen gefragt werden würde (z. B. Stern, 2001, S. 182). Solche Veränderungen sind unzulässig, wenn als Ziel der Vergleich angestrebt wird, und wurden deshalb nicht vollzogen.

Im Rahmen einer Bachelorarbeit konnte gezeigt werden, dass die Testaufgaben in ihrer Schwierigkeit nahezu vergleichbar waren (Kremer, 2015, S. 60–67). In der Summe kann daher sichergestellt werden, dass durch den Austausch einzelner Begriffe parallele, vergleichbare Tests entstehen, deren Aufgaben gleiche Strukturen aufweisen sowie gleiche Handlungsmöglichkeiten hervorrufen (Brenner et al., 1997, S. 672). Das Beibehalten gleicher Zahlenwerte zieht Koerber (2000, S. 123) als Möglichkeit in Betracht. Dies gewährleistet, Leistungsunterschiede zwischen zwei Messzeitpunkten auf das Training, und nicht auf die unterschiedlich verwendeten Zahlenwerte zurückzuführen. Im Follow-up-Test wurden die Zahlen minimal variiert, sodass den Aufgaben nach wie vor die gleichen Handlungsmöglichkeiten und Einschränkungen zugrunde lagen, aber eine weitere Wiederholung der Zahlenwerte verhindert und Erinnerungseffekte ausgeschlossen werden konnten.

Da Feedback und Reflexion der Testaufgaben im Prätest ausblieben, wird in Anlehnung an Chinn und Brewer (1993, S. 20–21) davon ausgegangen, dass Lernende von der Richtigkeit ihrer gefundenen Lösung überzeugt sind. Daher wird erwartet, dass sie selbstbewusst an ihren individuellen Lösungsvorstellungen festhalten (vgl. R. Rasch, 2001b, S. 257). Eine individuelle Verbesserung kann folglich nur durch ein Training und eine eingeschlossene Reflexion bewirkt werden, in dem das Lösungsvorgehen analysiert, hinterfragt und verbessert bzw. das Verständnis der Ausgangslage gewährleistet wird (vgl. Koerber, 2000, S. 123). Die sich hieraus ergebenden Testaufgaben, werden im folgenden Abschnitt vorgestellt.

4.5.1.3 Textaufgaben-Performance-Test

Jeder Textaufgaben-Performance-Test (Prä-, Post- und Follow-up-Test) umfasst eine *Vergleichsaufgabe*, eine *Aufgabe mit komplexen Informationen* und eine *Aufgabe zur Verhältnisverteilung* (vgl. Abschnitt 4.5.1.2). Zuerst werden die Aufgaben des Prätests (Prä) dargestellt, gefolgt von den Aufgaben des Posttests (Post) und den Aufgaben des Follow-up-Tests (Fo). Die Angaben in Klammern sind als Abkürzungen zu verstehen, die in der nachstehenden Auflistung aufgegriffen werden. Die jeweiligen Veränderungen auf Wortebene sind herausgestellt, wobei deren Wiederholung und Namensänderungen nicht explizit hervorgehoben werden. Die Messinstrumente können in Anhang B.1 und B.2, die Aufgaben zweiter Schwierigkeitsstufe in Anhang C.2 eingesehen werden.

Vergleichsaufgaben¹⁹²

(Prä) Max und Jan haben zusammen 30 *Star-Wars®-Sammelkarten*. Max hat 4 mehr als Jan. Wie viele Sammelkarten hat Max, wie viele hat Jan?

(Post) Paula und Hanna haben zusammen 30 *Stikeez®-Figuren*. Paula hat 4 mehr als Hanna. Wie viele Stikeez® hat Paula, wie viele hat Hanna?

(Fo) Tim und Lea haben zusammen 30 *Fußball-WM-Sticker*. Lea hat 6 mehr als Tim. Wie viele Fußball-WM-Sticker hat Lea? Wie viele hat Tim?

Aufgaben mit komplexen Informationen¹⁹³

(Prä) Die Klassen 3a und 3b fahren auf *Klassenfahrt*. Es sind in der Jugendherberge aber nur noch Einzelzimmer und Doppelzimmer frei. Insgesamt gibt es 25 Zimmer, in denen 40 Kinder untergebracht werden können. Wie viele Einzelzimmer gibt es? Wie viele Doppelzimmer gibt es?

(Post) Die Klassen 3a und 3b gehen in den *Computerraum*. Einige Kinder müssen zu zweit an einem Computer arbeiten. Insgesamt gibt es 25 Computer, aber 40 Kinder. Wie viele Kinder arbeiten alleine an einem Computer? Wie viele Kinder arbeiten zu zweit an einem Computer?

(Fo) Die Kinder der Klasse 3c basteln *Fensterbilder*. Einige Kinder müssen sich zu zweit eine Schere teilen. Die Klassenlehrerin zählt 20 Scheren, aber 28 Kinder. Wie viele Kinder arbeiten alleine mit einer Schere? Wie viele Kinder müssen sich zu zweit eine Schere teilen?

Aufgaben zur Verhältnisverteilung

(Prä) Mama, Papa und Tom fahren mit dem *Dampfer*. Für Kinder kostet es nur die Hälfte. Sie bezahlen insgesamt 30 €. Wie viel kostet die Karte für einen Erwachsenen und wie viel kostet sie für ein Kind?

(Post) Oma und Opa gehen mit ihrem *Enkel Michael in den Zoo*. Für Kinder kostet es nur die Hälfte. Sie bezahlen insgesamt 30 €. Wie viel kostet die Karte für einen Erwachsenen und wie viel kostet sie für ein Kind?¹⁹⁴

¹⁹² Im Rahmen einer Bachelorarbeit konnte gezeigt werden, dass sich die Vergleichsaufgaben nicht signifikant in ihrer Schwierigkeit unterschieden (Kremer, 2015, S. 60–61).

¹⁹³ Auch innerhalb dieses Aufgabentyps ergaben sich keine signifikanten Unterschiede in der Schwierigkeit (Kremer, 2015, S. 67). Deskriptiv erwies sich die Posttest-Aufgabe als die schwierigste.

¹⁹⁴ Im einführenden Sitzkreis konnte beobachtet werden, dass der Zusatz *Enkel* vielen Grundschulkindern Schwierigkeiten bereitete. Insbesondere Lernende mit Migrationshintergrund identifizierten den Ausdruck *Enkel Michael* als zwei Personen, den Enkel und Michael. Das Wort *Enkel* und dessen Bedeutung war ihnen aus ihrer Alltagssprache nicht geläufig.

(Fo) Mama, Papa und Tilo fahren mit dem Zug. Für Kinder kostet es nur die Hälfte. Sie bezahlen insgesamt 50 €. Wie viel kostet die Karte für einen Erwachsenen und wie viel kostet sie für ein Kind?

Reliabilitäten

Da beabsichtigt war, mit den Testinstrumenten inkonsistente Merkmale zu erfassen, wurde eine Reliabilitätsbestimmung in Form der internen Konsistenz durchgeführt (Lienert & Raatz, 1998, S. 179). Damit wird die Erfassung eines aktuellen Persönlichkeitsmerkmals anvisiert. Vor dem Hintergrund, dass es zu mehreren Messzeitpunkten den aktuellen Leistungsstand (Lösungserfolg, Problemkompetenzen, Anzahl selbstgenerierter externer Repräsentationen) zu messen galt, wurde der Konsistenzkoeffizient Cronbachs Alpha als Schätzung herangezogen. Dieser Reliabilitätskoeffizient gibt Aufschluss über die Homogenität eines Testinstruments – seine Messgenauigkeit im Sinne von Messfehlerfreiheit (Fissen, 1997, S. 86; Rammstedt, 2010, S. 256; Schermelleh-Engel & Werner, 2012, S. 120). Konkret bezeichnet die Schätzung „das Ausmaß, in dem von denselben Probanden alle Items in gleicher Weise beantwortet werden“ (Fissen, 1997, S. 86). Da sich die Testaufgaben innerhalb eines Tests in ihrer Schwierigkeit unterschieden (vgl. Abschnitt 4.2.2), wird die Voraussetzung essenziell τ-äquivalenter Items als erfüllt angenommen (Schermelleh-Engel & Werner, 2012, S. 131).¹⁹⁵

Insgesamt konnte die Reliabilitätsanalyse für die drei abhängigen Variablen (Lösungserfolg, Problemlösekompetenzen und Anzahl selbstgenerierter externer Repräsentationen) nur niedrige Reliabilitätskoeffizienten (< .80, siehe Tabelle 4.8) nachweisen (Weise, 1975, S. 219).

Tabelle 4.8. Textaufgaben-Performance-Test: Reliabilitäten der Skalen

Skalen	Prä	Post	Follow-up
Lösungserfolg	.39	.60	.60
Problemlösekompetenzen	.42	.68	.70
Anzahl selbstgenerierter externer Repräsentationen	.64	.68	.68

Anmerkung. Prä = Prätest; Post = Posttest; Follow-up = Follow-up-Test.

¹⁹⁵ Dies besagt sowohl, dass die wahren Werte der Probanden differieren, als auch dass verschiedene Fehlervarianzen zwischen den Probanden vorliegen (Schermelleh-Engel & Werner, 2012, S. 131).

Wenn an dieser Stelle berücksichtigt wird, dass die drei Tests jeweils nur aus drei Aufgaben bestanden und im Allgemeinen gilt, dass mit zunehmender Anzahl an Items der Reliabilitätskoeffizient steigt, dann lassen sich die Ergebnisse relativieren (Bühner, 2011, S. 166; Eid, Gollwitzer, & Schmitt, 2011, S. 833).

Für Prätest-Posttest-Vergleiche ist es nicht unüblich, dass die Reliabilität des Prätests im Vergleich zu denen der Posttests (Posttest und Follow-up-Test) geringer ausfällt und nicht zufriedenstellend ist (z. B. Collet, 2009, S. 127; Felbrich, 2005, S. 140; Groß, 2013, S. 101; Steigert & Schrenk, 2012, S. 203). Die Heterogenität der Aufgaben gilt als eine Ursache für die geringe Skalenreliabilität (Fisseni, 1997, S. 87; Lienert & Raatz, 1998, S. 177). Die drei Testaufgaben lassen sich drei verschiedenen Aufgabentypen zuordnen, was zur Folge hat, dass die Lernenden mit ganz unterschiedlichen Problemkontexten sowie unterschiedlichen Anforderungen konfrontiert wurden (Groß, 2013, S. 101, vgl. Abschnitt 1.4.1). Aus diesem Grund ist durchaus zu erwarten, dass die drei beobachteten Aufgabenscores eines Kindes unterschiedlich hoch ausfallen können (vgl. „Varianz innerhalb“ Fisseni, 1997, S. 86). Ein Teil der Varianz innerhalb des Probanden wird demnach durch die „Varianz zwischen den Items“ erklärt (Fisseni, 1997, S. 86).

Eine weitere Ursache kann in der Zusammensetzung der Stichprobe gesehen werden. Je homogener die Stichprobe ist, desto weniger wahre Unterschiede gibt es zwischen den Lernenden, sodass ein Messinstrument weniger reliabel sein kann als in einer heterogenen Stichprobe (Eid et al., 2011, S. 820; Schermelleh-Engel & Werner, 2012, S. 138; vgl. Danner, 2015, S. 1). Die niedrigen Reliabilitäten zum Messzeitpunkt 1 (Prätest) lassen sich demzufolge auch dadurch erklären, dass es im Allgemeinen den ungeübten Grundschulkindern noch schwerfällt, problemhafte Textaufgaben erfolgreich zu lösen. Sie bilden zu diesem Zeitpunkt eine sehr homogene Gruppe ab, die eine geringe Varianz ihrer wahren Werte vorweist. Nach dem Training werden deutliche Unterschiede zwischen den Probanden und folglich eine hohe Varianz ihrer wahren Werte erwartet, sodass von einer heterogeneren Stichprobe als vor dem Training ausgegangen werden kann. Hierunter darf nicht verstanden werden, dass der Problemlöse-Performance-Test nach dem Training genauer misst als vor dem Training (Schermelleh-Engel & Werner, 2012, S. 138). Vielmehr erklärt sich der von Messzeitpunkt zu Messzeitpunkt steigende Konsistenzkoeffizient durch den Kompetenzzuwachs, den die Lernenden durch das Training erzielten, und der damit einhergehenden sinkenden Varianzeinschränkung (siehe Tabelle 4.8) (Schermelleh-Engel & Werner, 2012, S. 138). Eine Wiederholung der Messung kann demnach aufgrund des stattgefundenen Lernprozesses nicht das gleiche Ergebnis erzielen (Brosius, Haas, & Koschel, 2012, S. 49–50; Lienert & Raatz, 1998, S. 178). Geht man davon aus, dass die Implementierung einer Interventionsmaßnahme immer auf einem Mangelzustand, einem Defizit fußt

(Hager, 2008, S. 723 s. Abschnitt 4.2.4.1) und demzufolge vor der Intervention Bodeneffekte zu erwarten sind (Bühner, 2011, S. 181; Schermelleh-Engel & Werner, 2012, S. 138), so ist mit größeren Messfehlern zu rechnen. Dass der Testscore in Bezug auf den Lösungserfolg nur einen begrenzten Wertebereich zwischen 0 und 1 annehmen kann, verschärft die Problematik der Bodeneffekte und spiegelt sich auch in dem geringsten Reliabilitätskoeffizienten wider (siehe Tabelle 4.8). Dies kann an folgendem Beispiel vergegenwärtigt werden (Schermelleh-Engel & Werner, 2012, S. 138): Ein Kind erhält 0 Punkte, wenn es die Aufgabe nicht zu lösen versucht. Ein weiteres Kind, das die Aufgabe versehentlich falsch löst, weil es sich verrechnet hat, weil es Teilschritte vergessen hat oder weil ihm irgendein anderer Fehler unterlaufen ist, erhält zwangsläufig auch 0 Punkte, was aber mit einem erhöhten Messfehler verbunden ist.

Auch wenn die internen Konsistenzen niedrig sind, können bei angestrebten Gruppenvergleichen die Gruppenmittelwerte trotz der „individuell stärker messfehlerbehafteten“ Testwerte richtig vorhergesagt werden (Schermelleh-Engel & Werner, 2012, S. 136).

Beobachterübereinstimmung

Die Beobachterübereinstimmung beider Rater wurde bei der abhängigen Variablen *Lösungserfolg* mit dem Koeffizienten Cohens Kappa κ bestimmt (Wirtz & Caspar, 2002, S. 120). Da bei der abhängigen Variable *Problemlösekompetenzen* die Annahme der Gleichabständigkeit der Skalenpunkte nicht gesichert war, wird der Empfehlung gefolgt, sowohl das ordinalskalierte als auch das intervallskalierte Maß zu berichten (Wirtz & Caspar, 2002, S. 127). Als ordinalskaliertes Maß wurde der Koeffizienten Spearman Rho ρ (Döring & Bortz, 2016, S. 347; Wirtz & Caspar, 2002, S. 133), als intervallskaliertes Maß die Intraklassenkorrelation ICC herangezogen (Döring & Bortz, 2016, S. 346; Wirtz & Caspar, 2002, S. 127).

Im Mittel wurde beim Lösungserfolg über alle Messzeitpunkte und alle Aufgaben eine Übereinstimmung von $\kappa = .97$ ($\kappa_{\min} = .93$, $\kappa_{\max} = 1.00$) erzielt. Diese Beobachterübereinstimmungen sind nach den konventionellen Standards als sehr gut einzustufen (Döring & Bortz, 2016, S. 346). Bei den Problemlösekompetenzen wurde über alle Messzeitpunkte und alle Aufgaben ein mittlerer ρ -Koeffizienten von $\rho = .98$ ($\rho_{\min} = .94$ und $\rho_{\max} = .99$) erzielt, was ebenso im sehr guten Bereich liegt. Die Interklassenkorrelationen bestätigen die sehr guten Beobachterübereinstimmungen. Die mittlere Intraklassenkorrelation liegt bei $ICC = .99$. Alle aufgabenspezifischen Beobachterübereinstimmungen sind im Anhang E.1, Tabelle E.1.2 einzusehen.

Faktorenanalyse

Der Konstruktion der Textaufgaben-Performance-Tests lag die Hypothese zugrunde, dass alle neun Testaufgaben das eindimensionale Konstrukt *mathematische Problemlösefähigkeiten* bzw. *Lösungserfolg* abbilden. Mithilfe einer konfirmatorischen Faktorenanalyse (CFA) wurde jeweils überprüft, ob die empirischen Daten hinreichend mit dem theoretischen Modell übereinstimmen oder ob das Modell verworfen werden muss (Moosbrugger & Schermelleh-Engel, 2012, S. 334).

Die Analyse wurde mit dem R package *lavaan* durchgeführt. Aufgrund der dichotomen Ausprägung der Indikatorvariablen (Erfolg/Misserfolg) wird von der Nutzung des Maximum-Likelihood-Schätzers (ML) abgeraten (Brown, 2015, S. 353). Um dem kategorialen Skalenniveau gerecht zu werden, sollte vielmehr eine Diagonally-Weighted-Least-Squares-Schätzmethode (DWLS-Schätzmethode) oder eine Unweighted-Least-Squares-Schätzmethode (ULS-Schätzmethode) verwendet werden (Brown, 2015, S. 354–355; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009, S. 626–627; Rosseel, 2013, S. 28, 30; Savalei & Rhemtulla, 2013, S. 202).

Weil mehrere Monte-Carlo-Simulationsstudien zeigen konnten, dass die ULS-Schätzmethode mit dem robusten ULSMV-Schätzer¹⁹⁶ der DWLS-Schätzmethode überlegen ist, wurde sich für diese ausgesprochen (Forero et al., 2009, S. 625, 635–639; Savalei & Rhemtulla, 2013, S. 201, 209–221). Wie die Fit-Indizes zeigen, weist das einfaktorielle Modell nach Brown (2015, S. 74–75) einen guten Modellfit auf:¹⁹⁷ $\chi^2(27) = 36.613$, $p = .103$; $\chi^2/df = 1.356$, $CFI = 0.985$, $TLI = 0.981$, $SRMR = .032$, $RMSEA = 0.033$ ($p = .854$).¹⁹⁸ Unter Berücksichtigung des nichtsignifikanten χ^2 -Werts erfüllen die Fit-Indizes die Vorgaben bei Stichproben mit mehr als 250 Probanden und weniger als 12 gemessenen Variablen: $CFI \geq .95$, $TLI \geq .95$, $RMSEA < .07$ (wenn $CFI \geq .07$) (Hair, Black, Babin, & Anderson, 2009, S. 753). Daher kann davon ausgegangen werden, dass der Textaufgaben-Performance-Test das eindimensionale Konstrukt *Lösungserfolg* misst.

Um die Eindimensionalität der abhängigen Variablen *Problemlösekompetenzen* zu überprüfen, wird äquivalent zum Lösungserfolg der ULSMV-Schätzer herangezogen, weil die Variablen ordinal skaliert vorliegen. Die Fit-Indizes des einfaktoriellen Modells deuten nach Brown (2015, S. 74–75) gerade noch auf einen akzeptablen Modellfit hin:¹⁹⁹ $\chi^2(27) = 58.330$, $p < .001$; $\chi^2/df = 2.160$, $CFI = 0.970$, $TLI = 0.960$, $RMSEA = .032$ ($p = .866$).

¹⁹⁶ ULSMV = robuster Unweighted-Least-Squares-Schätzer.

¹⁹⁷ Siehe auch Hair et al. (2009, S. 753), Moosbrugger und Schermelleh-Engel (2012, S. 338), Nussbeck, Eid und Lischetzke (2006, S. 196)

¹⁹⁸ Dies ist ebenfalls der Fall, wenn der Modellspezifikation der WLSMV-Schätzer zugrunde liegt: $\chi^2(27) = 36.076$, $p = .114$, $\chi^2/df = 1.336$, $CFI = 0.980$, $TLI = 0.973$, $RMSEA = 0.032$ ($p = .866$).

¹⁹⁹ Siehe Fußnote 197.

$A = 0.060$ ($p = .210$).²⁰⁰ Es kann auch in diesem Fall davon ausgegangen werden, dass der Textaufgaben-Performance-Test das eindimensionale Konstrukt *mathematische Problemlösekompetenzen* misst.

Diskussion

Insgesamt bleibt festzuhalten, dass es aus zeit- und durchführungsökonomischen Gründen den Grundschulkindern in einer regulären Schulstunde möglich war, drei problemhafte Textaufgaben zu bearbeiten, ihr individuelles Lösungsvorgehen zu beschreiben und zu begründen sowie ihren Erfolg oder Misserfolg vorherzusagen und zu evaluieren. Es hat sich bewährt, dass die Kinder ohne Zeitdruck die Möglichkeit hatten, zu jeder Aufgabe ihre Lösungsgedanken und -ideen zu externalisieren. Das Zeitfenster von 45 Minuten war hierfür in Anbetracht einer Gruppentestung ausgeschöpft und in den meisten Fällen auch ausreichend. Vereinzelt konnte beobachtet werden, dass zum Posttest und zum Follow-up-Test die Zeit nicht ausreichte und der Test nicht vollständig bearbeitet werden konnte. Das offene Aufgabenformat der Leistungstests hat sich bewährt, die Problemlösekompetenzen der Kinder zu erfassen. Die Drittklässler konnten unvoreingenommen und unbeeinflusst ihren individuellen Problemlöseprozesse sowie die genutzten Denk- und Erkenntniswerkzeuge externalisieren. Die Analyse der Eigenproduktionen ließ demzufolge nicht nur Rückschlüsse auf Erfolg oder Misserfolg, sondern auch auf deren Problemlösekompetenzen zu, indem die selbstgenerierten externen Repräsentationen sowie deren Merkmale identifiziert werden konnten.

Aus testtheoretischer Sicht ist die berücksichtigte Anzahl an Testaufgaben jedoch nicht zufriedenstellend. Rückblickend wäre es vonnöten gewesen die Anzahl der Testaufgaben zu erhöhen, insbesondere auch die des untrainierten Aufgabentyps. Die Tatsache, dass lediglich ein untrainierter Aufgabentyp erfasst wurde, hat sich retrospektiv als unzureichend erwiesen. Die erwarteten Transfereffekte im Sinne einer Leistungsverbesserung und Kompetenzentwicklung auf einen nicht direkt trainierten Aufgabentypen sind somit immer von dem jeweils berücksichtigten Aufgabentyp, hier der Aufgabe zur Verhältnisverteilung, abhängig. Generalisierbare Aussagen können aufgrund dessen nicht getroffen werden. Hätte man die Erhebung der metakognitiven Fähigkeiten, die aus sprachlichen Gesichtspunkten sehr zeitintensiv war, wie in Abschnitt 4.5.3.4 diskutiert, gekürzt, wäre die Bearbeitung einer vierten Testaufgabe realisierbar gewesen.

Ferner wurde der Einhaltung der Unterrichtsstruktur Priorität beigemessen und somit an der meist vorgefundenen 45-Minuten-Taktung festgehalten. Es zeigte sich jedoch,

²⁰⁰ Unter Verwendung des WLSMV-Schätzers, ist die Modellspezifikation vergleichbar: $\chi^2(27) = 55.555, p = .001, \chi^2/df = 2.058$, CFI = 0.963, TLI = 0.950, RMSEA = 0.057 ($p = .276$).

dass viele Schulen mitunter auch Unterrichtsstunden im 60-Minuten-Takt vorsahen. Diese hätten sich für die Testungen zum Prä-, Post- und Follow-up-Test als günstig erwiesen und wären weder mit einer zusätzlichen Belastung für die Lehrkräfte einhergegangen, noch hätte dies den Schulalltag und die Unterrichtsstruktur gestört. Rückblickend wären 60-minütige Testungen unabdingbar gewesen, da sie den Vorteil gehabt hätten, die Anzahl der Testaufgaben zu steigern und Füllitems, wie beispielsweise Routineaufgaben, zu integrieren. Insbesondere eine Testverlängerung durch die Hinzunahme von parallelen Testteilen hätte eine Reliabilitätssteigerung erwirken können (Moosbrugger, 2012, S. 112). Die erwarteten Schwierigkeiten erwiesen sich diesbezüglich als unbegründet.

Um mehr über die individuellen Denkprozesse der Kinder zu erfahren, ist in Folgeanalysen ergänzend über die Implementierung von Schülerinterviews nachzudenken. Die Interpretation der Eigenproduktionen erfordert das Hineindenken in die Lösung und die sichtbaren Strukturen. Lösungsschritte und -gedanken, die das Kind mental vollzieht und nicht zu Papier bringt, bleiben in der vorliegenden Untersuchung dem Beurteiler verborgen und können höchstens vermutet, aber nicht belegt werden. Schülerinterviews bieten die Möglichkeit, mehr über die individuellen Denkprozesse der Kinder zu erfahren. Das Schildern des individuellen Lösungsprozesses würde einerseits das Kind zum Reflektieren anregen, zum anderen die Lösungsfindung für den Interviewer nachvollziehbar machen, da die geschilderte Abfolge der Schritte mit hoher Wahrscheinlichkeit der tatsächlich vollzogenen Lösungsabfolge entspräche. Entscheidend käme hinzu, dass der Interviewer Rück- und Nachfragen stellen könnte, was gerade bei selbstgenerierten externen Repräsentationen von Vorteil ist. Die Lernenden könnten benennen, welche Funktion die externe Repräsentation für sie eingenommen hat: Es ließe sich aufdecken, ob die Kinder die Lösung *mithilfe* der externen Repräsentation erarbeitet haben, womit die externe Repräsentation zum Denk- und Erkenntniswerkzeuge wurde, oder ob sie die externe Repräsentation im *Nachhinein* generiert haben, womit sie lediglich zum Kontrollwerkzeug wurde. Insbesondere, wenn Kinder mehr als eine externe Repräsentation generierten, könnte man wichtige Hinweise über deren Nutzung und Funktion erfahren. Diesbezüglich würde man durch Schülerinterviews zum einen erfahren, ob die Kinder lediglich unterschiedliche Zugänge ausprobierten bis es ihnen gelungen war, den Problemraum einzugrenzen. Zum anderen würde transparent werden, ob vor allem die Vernetzung mehrerer Repräsentationen hilfreich war und demnach multiple Repräsentationen für die Lösungsfindung genutzt wurden. Die Schülerproduktionen decken diesen Aspekt nicht auf, geben maximal Hinweise darauf. Das hohe Potenzial von Schülerinterviews sollte in Folgeanalysen beim Umgang mit externen Repräsentationen zwingend Berücksichtigung finden.

4.5.1.4 Bildung der Testscores

Für jede abhängige Variable wird zwischen dem *Aufgabenscore*, dem erzielten Punktwert einer Aufgabe und dem *Testscore*, der Mittelwert aus den einzeln erzielten Punktwerten, unterschieden.

Testscore: Lösungserfolg

Mit den dichotomen Punktwerten der einzelnen Testaufgaben wurden verschiedene Scores gebildet:

- Der Aufgabenscore spiegelt den erreichten Punktwert der jeweiligen Aufgabe wider. Der Punktwert beträgt 0 (Aufgabe falsch, teilweise richtig oder nicht gelöst) oder 1 (Aufgabe richtig gelöst).
- Der Testscore bildet den Mittelwert aller drei Aufgabenscores ab. Er liegt zwischen 0 (keine Aufgabe richtig gelöst) und 1 (alle drei Aufgaben richtig gelöst). Als Zwischenstufen lassen sich 0,33 (eine von drei Aufgaben richtig gelöst) und 0,67 (zwei von drei Aufgaben richtig gelöst) identifizieren.

Testscore: Problemlösekompetenzen

Mit den fünfstufigen Punktwerten der einzelnen Testaufgaben (0–4) wurden analog zum Lösungserfolg verschiedene Scores gebildet:

- Der Aufgabenscore spiegelt den erreichten Punktwert 0, 1, 2, 3 oder 4 der jeweiligen Aufgabe wider. Die fünf Abstufungen stützen sich auf die von Charles et al. (1987, S. 34–35) entwickelte Beurteilungsskala, die *Focused Holistic Scoring Point Scale*. Sie sind Tabelle 4.7, Abschnitt 4.5.1.1, zu entnehmen und aufgrund ihrer Differenziertheit hier nicht erneut aufgeführt.
- Der Testscore bildet den Mittelwert aller drei Aufgabenscores ab, sodass er mindestens 0 und maximal 4 beträgt. Als Zwischenstufen lassen sich hierbei 13 mögliche Merkmalsausprägungen identifizieren, wobei es sich um alle Werte handelt, die sich als Mittelwerte aus den drei Aufgabenscores von 0 bis 4 ergeben können.

Testscore: Anzahl selbstgenerierter externer Repräsentationen

Bei jeder Aufgabe wurde mit dichotomen Punktwerten beurteilt, ob bei der Testaufgabe eine Tabelle (ja/nein), eine Zeichnung (ja/nein), eine Rechnung (ja/nein) oder eine schriftsprachliche Beschreibung (ja/nein) konstruiert wurde. Hieraus resultierten verschiedene Scores:

- Der Aufgabenscore spiegelt die Summe der selbstgenerierten externen Repräsentationen pro Aufgabe wider. Der Punktwert beträgt 0 (keine Repräsentation wurde genutzt), 1 (eine Repräsentation wurde genutzt), 2 (zwei verschiedene Repräsentationen wurden genutzt).

sentationen wurden genutzt), 3 (drei verschiedene Repräsentationen wurden genutzt) oder 4 (4 verschiedene Repräsentationen wurden genutzt). Streng genommen gibt es hier keine maximale Anzahl selbstgenerierter Repräsentationen, da es keine Konstruktionsvorgaben gab.

- Der Testscore bildet den Mittelwert aller drei Aufgabenscores ab, weshalb es auch hier keinen maximalen Testscore gibt. Es wurde erwartet, dass der Testscore zwischen 0 (alle Aufgaben wurden ohne externe Repräsentation gelöst) und 4 (jede Aufgabe wurde mit 4 verschiedenen Repräsentationen gelöst) liegt. Hierbei sind viele Zwischenstufen möglich, die sich als Mittelwert aus den einzelnen Aufgabenscores ergeben können.

4.5.2 Unabhängige Variablen: Repräsentationsmerkmale

In Abschnitt 4.2.3.2 wurden die identifizierten Repräsentationsformen, Zeichnungen, Tabellen, Rechnungen und schriftsprachlichen Begründungen nebst ihren Merkmalen aufgezeigt und erörtert. Mithilfe der strukturierenden Inhaltsanalyse wurden möglichst viele Merkmale gesammelt, welche es im nächsten Schritt auf eine wesentlich geringere Anzahl von Faktoren zu reduzieren galt (Eid et al., 2011, S. 893). Ziel war es dabei, die Redundanzen in den Merkmalen aufzudecken, wenn verschiedene Merkmale dasselbe Konstrukt erfasssten. Es wurde der Empfehlung Stevens (2009, S. 75) gefolgt, bei einer hohen Prädiktorenzahl eine faktorenanalytische Analyse vorzuschalten, um eine Reduktion der Variablen zu erzielen und wechselseitig voneinander unabhängige Faktoren zu identifizieren.

Die Auswahl der unabhängigen Variablen basiert vorrangig auf theoretischen Überlegungen (Cohen, Cohen, West, & Aiken, 2003, S. 144; Eid et al., 2011, S. 629; Stevens, 2009, S. 75), im vorliegenden Fall auf Überlegungen der strukturierenden Inhaltsanalyse. Demzufolge wären diejenigen Repräsentationsmerkmale, von denen sich aus inhaltsanalytischer Sicht eine Varianzerklärung versprochen wird, in die Regressionsgleichung aufzunehmen. Kategoriale, nicht dichotomisierte Prädiktoren wurden dummykodiert (Bortz & Schuster, 2010, S. 342; Ramsey & Schafer, 2013, S. 358), sodass aus 22 Merkmalen 32 zu messende Prädiktoren resultierten. Aufgrund der Vielzahl an beobachteten Merkmalen wurde ergänzend eine datengesteuerte Auswahl getroffen (Eid et al., 2011, S. 629), welche aufdeckte, dass aus der Menge der 32 theoretisch ausgewählten Prädiktoren nicht alle Prädiktoren die Kriteriumsvariable vorhersagten. Merkmale, die von weniger als 10 % der Kinder generiert wurden, wurden aufgrund mangelnder Repräsentativität und Vorhersagekraft aus der Analyse ausgeschlossen. Hierbei handelte es sich 1) um das Zeichnungsmerkmal *dargestellter Zustand*, konkret wenn der Anfangszustand und der Lösungsweg oder der Zielzustand und der Lösungsweg abgebildet waren und 2) um die Rechnungsmerkmale *Nutzung*,

wenn die Rechnung eine Kontrollfunktion oder Einstiegs- und Kontrollfunktion erfüllte (siehe Abschnitt 4.2.3.2). Somit reduzierte sich die Anzahl der Prädiktoren von 32 auf 28 (siehe Tabelle 4.9).

Mit dem Ziel, wechselseitig voneinander unabhängige Faktoren zu identifizieren, welche die Zusammenhänge zwischen den Repräsentationsmerkmalen erklären, wurde sich für eine Hauptachsenanalyse ausgesprochen, um latente Konstrukte bzw. Faktoren zu identifizieren, die fähig sind, das Beziehungsmuster zwischen den Variablen zu erklären (Field, 2013, S. 666; Tabachnick & Fidell, 2014, S. 660–661). Hierdurch sollte dem „Variablengeflecht eine Ordnung unterlegt“ werden (Bortz & Schuster, 2010, S. 387; vgl. Brown, 2015, S. 163) und es sollten unkorrelierte Prädiktoren für die Regressionsanalyse gewonnen werden (Stevens, 2009, S. 328).

Dieses prädiktorreduzierende Vorgehen war erforderlich, weil zum einen vereinzelte Prädiktoren sehr hoch ($.70 < r < .90$) miteinander korrelierten und daher Ähnliches zu messen schienen (Hutcheson & Sofroniou, 1999, S. 85). Zum anderen ergab die simultane Aufnahme aller 28 Repräsentationsmerkmale in das Regressionsmodell negative, nicht standardisierte sowie negative, standardisierte Regressionskoeffizienten. Ursache für negative Regressionsgewichte können negative Korrelationen zwischen einer unabhängigen Variablen mit einer abhängigen Variablen sein. Dies traf jedoch in der vorliegenden Untersuchung nicht zu. Alle Variablen – bis auf das Geschlecht – korrelierten positiv mit dem Kriterium. Der scheinbar negative Zusammenhang war ein Indiz für eine negative Suppression (Conger, 1974, S. 36; Smith, Ager, & Williams, 1992, S. 27). In diesem Fall trugen die Korrelation zwischen Prädiktor und Kriterium und der Regressionskoeffizient entgegengesetzte Vorzeichen. Das von Hutcheson und Sofroniou (1999, S. 84) empfohlene Aggregieren hoch korrelierender Variablen konnte diesem Phänomen nicht entgegenwirken.

Faktorenanalyse

Das Kaiser-Meyer-Olkin-Kriterium bestätigte, dass sich die Stichprobe für die Analysen eignete, $KMO = .75$ („middling“ Hutcheson & Sofroniou, 1999, S. 225). Alle KMO-Werte der Repräsentationsmerkmale lagen über .58 und somit über dem akzeptablen Limit von .50, weshalb keine Merkmale ausgeschlossen werden mussten (Field, 2013, S. 684–685). Nach dem Kaiser-Kriterium wurden sieben Faktoren extrahiert, die insgesamt 60 % der Varianz erklärten. Der Scree-Test ergab vier Faktoren. Die Parallelanalyse nach Horn wurde aufgrund ihrer Zuverlässigkeit herangezogen, um die tatsächliche Faktorenzahl zu bestimmen (Eid et al., 2011, S. 913; Hayton, Allen, & Scarpello, 2004, S. 201; Moosbrugger & Schermelleh-Engel, 2012, S. 332). Sie extrahierte ebenfalls vier Faktoren. Alle Merkmale, die auf den ersten Faktor luden, repräsentieren Zeichnungen, alle Merkmale, die auf den zweiten Faktor luden, repräsentieren Tabel-

len, alle Merkmale, die auf den dritten Faktor luden, repräsentieren Rechnungen (siehe Tabelle 4.9). Der letzte, vierte Faktor spiegelte die Merkmale schriftsprachlicher Beschreibungen wider.

Tabelle 4.9. Faktorladungen der Hauptachsenanalyse für die Repräsentationsmerkmale zum Zeitpunkt des Posttests

	Repräsentationsmerkmale	Faktor 1	Faktor 2	Faktor 3	Faktor 4
Zeichnung: Dargestellter Zustand (D)					
1	Anfangszustand	.85	.07	-.02	-.07
2	Zielzustand	.91	.04	-.06	-.08
3	Lösungsweg	.58	-.03	-.08	.26
4	Anfangs- und Zielzustand	.64	.13	.04	-.23
5	Anfangs- und Zielzustand und Lösungsweg	.61	.01	-.10	.26
6	Zeichnung: Grad der Strukturabbildung	.94	.04	-.05	-.06
7	Zeichnung: Abstraktionsgrad	.59	.02	-.02	-.06
8	Zeichnung: Kontrolle	.67	.07	.00	.01
9	Zeichnung: Systematisches Vorgehen	.76	.08	-.08	.13
Rechnungen: Nutzungsmöglichkeiten (D)					
10	Einstiegsfunktion	.16	.22	.15	.14
11	Erarbeitungsfunktion	-.18	-.14	.78	.09
12	Rechnungen: Anzahl der Rechnungen	.03	.06	.92	.04
13	Rechnungen: Abstraktionsgrad	-.02	.08	.25	-.00
14	Rechnungen: Geschriebene mathematische Sprache	-.07	-.05	.72	.21
15	Rechnungen: Konstanz eines Faktors	-.07	-.05	.33	.01
16	Rechnungen: Variation der Rechenoperationen	.01	.08	.65	-.03
17	Rechnungen: Kontrolle	.03	.08	.57	-.05
18	Rechnungen: Systematisches Vorgehen	.00	.02	.76	.14
19	Tabellen: Dimensionen	.14	.94	.10	-.07
20	Tabellen: Abstraktionsgrad	.13	.83	.06	-.09
Tabellen: Nutzungsmöglichkeiten (D)					
21	Trial-and-Error-Protokoll	-.03	.70	.03	.03
22	Erarbeitungsprotokoll	.09	.49	.00	-.11

23	Tabellen: Kontrolle	.00	.78	.02	-.04
24	Tabellen: Konstanz eines Faktors	-.02	.80	.00	-.00
25	Tabellen: Systematisches Vorgehen	.03	.78	-.04	.04
Schriftsprachliche Beschreibungen					
26	Mathematische Begründungsstruktur	-.08	-.05	.23	.81
27	Schriftsprachliche Begründungsstruktur	-.01	-.05	.10	.95
28	Mathematische Zusammenhänge	.09	-.06	-.02	.65

Anmerkung. $N = 366$. D = Dummykodierte Variablen. Faktorladungen $> .40$ sind hervorgehoben.

Alle Repräsentationsmerkmale, die $> .40$ auf einen Faktor luden, wurden aggregiert. Die drei Variablen 10, 13 und 15 blieben deshalb unberücksichtigt. Dadurch entstanden vier wechselseitig voneinander unabhängige Faktoren, die in das Regressionsmodell aufgenommen wurden: Zeichnungen, Rechnungen, Tabellen und schriftsprachliche Beschreibungen.²⁰¹

Diskussion

In der vorliegenden Untersuchung wurde sich dafür entschieden, der Auswertung der individuellen Eigenproduktionen eine theoriegeleitete Kategorienanwendung zugrunde zu legen, die jedoch eine induktive Kategorienfindung zulässt. Hierzu wurden die in der Fachdidaktik etablierten Kategorien aufgegriffen und durch das Einbeziehen des empirischen Datenmaterials angereichert, sodass dem individuellen Datenmaterial gerecht werden konnte. Darüber hinaus wurde festgesetzt, dass lediglich Merkmale erfasst werden sollen, die sichtbar waren und keinerlei Interpretationen und Mutmaßungen zuließen. Dies war mitunter neben den intensiven, mehrfach durchgeföhrten Raterainings ein Grund für die hohen Beurteilerübereinstimmungen.

Was aus dieser Perspektive positiv erscheint, bringt den Nachteil mit sich, dass viele Merkmale der externen Repräsentationen in den Daten unberücksichtigt und ungenutzt blieben. Durch die Ausgangslage dieses inhaltsanalytischen Ansatzes war die Analyse des Datenmaterials – oder in den Worten M. Meyers und Voigts (2009, S. 14) die Analyse des „kreative[n] Wildwuchs[es]“ der Kinder – fokussiert und zugleich eingeschränkt, sodass nicht das gesamte Potenzial der Eigenproduktionen ausgeschöpft und

²⁰¹ Alternativ wurden alle 28 Repräsentationsmerkmale in das Regressionsmodell aufgenommen. Der Erklärungsbetrag einzelner unabhängiger Variablen war in der multiplen Regression um vieles größer, als dies in der einfachen Regressionsanalyse zu beobachten war. Den Merkmalen würde dadurch ein bedeutsamerer Beitrag zugesprochen werden, als es tatsächlich der Fall war (Eid et al., 2011, S. 633–636). Es lag eine Suppression vor, die sich auch negativ zeigte. Die Ergebnisse dieser vorgesetzten Analyse werden im Rahmen der statistischen Methoden aufgezeigt und diskutiert (siehe Abschnitt 4.8.6).

erfasst werden konnte. Durch den deduktiven Ansatz wurden viele wichtige Hinweise in den Dokumenten außen vor gelassen. Dies lässt den Schluss zu, dass andere Merkmale zu einer anderen Kategorisierung und letztlich zu einer anderen Vorhersage kommen würden. Womöglich waren die Faktoren nicht stabil. Was eine Grenze für unsere Untersuchung darstellt, bietet enormes Potenzial für Folgeuntersuchungen. Während in der vorliegenden Untersuchung der *sicherere*, quantifizierbarere Weg mit Einbezug des mathematikdidaktischen Forschungsstandes gegangen wurde, ist im Rahmen einer Folgeuntersuchung ein offener qualitativer Ansatz denkbar. Für die Kategorienbildung könnte dann beispielsweise der Forschungsstil der *Grounded Theory* herangezogen werden (vgl. Breuer, 2010; Strauss & Corbin, 1998).

Ferner könnten die selbstgenerierten externen Repräsentationen aus den deutschsprachigen Untersuchungen von R. Rasch (2001b, 2008a, 2016), Hohn (2012), Groß (2013), Reuter (2016) sowie internationale Untersuchungen zusammengetragen werden, sodass ein Pool an Repräsentationsformen über viele verschiedenen Aufgabentypen entstehen würde, die es sich lohnt, qualitativ auszuwerten. Von Interesse wäre auch, ob sich die Repräsentationsmerkmale jüngerer und älterer Kinder unterscheiden bzw. inwiefern es Überschneidungen gibt.

4.5.3 Kontrollvariablen

Wenn zurückliegende und aktuelle Untersuchungen aufgegriffen werden, dann kann zusammenfassend festgehalten werden, dass das Lösen problemhaltiger Textaufgaben von mehreren schülerbezogenen Faktoren, sogenannten systematischen Störvariablen, beeinflusst wird. Zu diesen Faktoren zählen neben der Intelligenz (Abschnitt 4.5.3.1) vorrangig mathematische Basiskompetenzen und Fertigkeiten (Abschnitt 4.5.3.2), sprachliche Fähigkeiten und das Leseverständnis (Abschnitt 4.5.3.3), aber auch metakognitive Fähigkeiten (Abschnitt 4.5.3.4) sowie die Attitudes und Beliefs der Lernenden (Abschnitt 4.5.3.5). Dass darüber hinaus die Mehrsprachigkeit und das Geschlecht zu berücksichtigen sind, wird entsprechend in den Abschnitten 4.5.3.6 bzw. 4.5.3.7 aufgezeigt.

Systematische Störvariablen gefährden die interne Validität der Untersuchung, weshalb deren Einfluss kontrolliert und bestenfalls minimiert werden muss (Eid et al., 2011, S. 58; Myers & Hansen, 2012, S. 216–217). Nachstehend wird empirisch gestützt erörtert, inwiefern oben genannte Faktoren beim Lösen problemhaltiger Textaufgaben einen solchen Einfluss nehmen können. Im Zuge dessen wird das herangezogene Testinstrument vorgestellt und zugrunde liegende Reliabilitäten werden näher beschrieben. Im Anschluss daran werden die Möglichkeiten und Grenzen der Testinstrumente diskutiert. Die Kontrollvariablen gingen als Prädiktoren in die Regressions-

analysen ein. Ferner bildeten die Intelligenz, die sprachlichen und mathematischen Fähigkeiten die Basis für das Parallelisierungsverfahren.

4.5.3.1 Intelligenz

Es konnte empirisch belegt werden, dass die Intelligenz den Lösungserfolg komplexer Textaufgaben beeinflusst (Groß, 2013, S. 157–162; Renkl & Stern, 1994, S. 34). Die Ursache wurde darin gesehen, dass beispielsweise mental ablaufende Denkprozesse, das Anknüpfen am bisherigen Vor- und Erfahrungswissen sowie die Konstruktion adäquater Modelle komplexe Prozesse darstellen, welche eine hohe Abstraktionsfähigkeit voraussetzen, die wiederum als zentrales Merkmal von Intelligenz angesehen wird (Renkl & Stern, 1994, S. 30). Groß (2013, S. 183) konnte darüber hinaus zeigen, dass sich die Höhe der Beeinflussung in den Jahrgangsstufen 3 und 4 unterschied. Während in der dritten Jahrgangsstufe die Intelligenz vor dem mathematischen Vorwissen den größten Anteil der Varianz erklärte, stand es in der vierten Jahrgangsstufe nach dem linguistischen Vorwissen nur noch auf Rang zwei. Auch die Ergebnisse aus der SCHOLASTIK-Studie spiegelten die Tatsache wider, dass die Relevanz des Vorwissens in der Grundschule vor der Intelligenz dominiert (Renkl & Stern, 1994, S. 36; Stern, 2001, S. 174). Der hohe Anteil konfundierter Varianz (Intelligenz und Vorwissen) ließ darauf schließen, dass Intelligenz einen eher indirekten Einfluss auf das Leistungsniveau ausübt (Groß, 2013, S. 158–159, 161–162; Renkl & Stern, 1994, S. 37). Stern (2001, S. 174) sieht nicht die Intelligenz, sondern das Vorwissen als die „entscheidende Leistungsdeterminante“ an und hält fest:

Die nahe liegende und plausibel klingende Vorstellung von Intelligenz als Fähigkeit zum kompetenten Umgang mit Neuem ist zu kurz gegriffen. Ob verfügbares Wissen zur Bewältigung einer neuen Anforderung flexibel kombiniert werden kann oder nicht, scheint eher von Unterschieden in der Lerngeschichte als von Unterschieden in der Intelligenzleistung abzuhängen. (Stern, 2001, S. 175)

Auch die Ergebnisse der LOGIK-Studie identifizierten keine Vorteile für Lernende mit hoher gegenüber niedriger Intelligenz (Stern, 1997, S. 107–108). Lernende, gleich ob über- oder unterdurchschnittlicher Intelligenz, konnten Erfolge erzielen, wenn sie auf eine überdurchschnittliche mathematische Leistungsfähigkeit zurückgreifen konnten.

Instrument zur Erhebung der Intelligenz

Mit dem Coloured-Progressive-Matrices-Test (CPM-Test) wurde die Intelligenz der Lernenden operationalisiert (Becker, Schaller, & Schmidtke, 1980; Bullheller & Häcker, 2002). Es handelt sich dabei um einen Power-Test, der ohne Zeitvorgabe von Kindern im Alter zwischen 4 und 11 Jahren absolviert werden kann (Becker et al., 1980, S. 8, 91). Er verfolgt die Zielsetzung, die sprachfreie Intelligenz und das logi-

sche Schlussfolgern zu messen (Bullheller & Häcker, 2002, S. V; vgl. Stern, 2001, S. 187).

Der Test setzt sich aus den drei Subtests A („Erfassung von Identität und Wechsel in kontinuierlichen Mustern“), Ab („Erfassung diskreter Muster als räumlich aufeinander bezogene Einheiten“) und B („Erfassung analoger Änderungen in räumlich und logisch aufeinander bezogenen Figuren“) mit jeweils 12 Aufgaben zusammen (Becker et al., 1980, S. 11–13). Jede Aufgabe bildet innerhalb eines rechteckigen Feldes ein lückenhaftes geometrisches Muster ab (Becker et al., 1980, S. 7). Diese Lücke gilt es adäquat zu kompletieren, indem aus sechs vorgegebenen Musteralternativen die passende auszuwählen ist. Die Lernenden stehen vor der Herausforderung „geometrische Figuren zu rotieren, zu spiegeln, zu zerlegen, zu falten, übereinander zu legen oder zu komplexen Figuren zusammenzufassen“ (Stern, 2001, S. 187). Je vertrauter ihnen der Umgang mit geometrischen Körpern ist, desto leichter fällt es ihnen, entsprechend den Anforderungen die passende Antwort zu identifizieren.

Pro Aufgabe und Seite werden ein unvollständiges Muster und sechs mögliche Antwortalternativen präsentiert, sodass die Lösung (Möglichkeit 1, 2, 3, 4, 5 oder 6) auf einem separaten Antwortblatt festzuhalten ist. Auch wenn eine Testung sowohl auf Einzel- als auch auf Gruppenebene möglich ist, wird Letzteres erst ab dem Alter von 8 Jahren empfohlen (Becker et al., 1980, S. 91). Das Übertragen der Lösung auf das Antwortblatt bringt insbesondere bei jüngeren Lernenden häufiger Schwierigkeiten mit sich. Um trotz der möglichen Schwierigkeiten eine Gruppentestung zu ermöglichen und dieser Fehlerquelle entgegenzuwirken, wurde ein Antwortblatt konzipiert, das Übersichtlichkeit durch eine tabellarische Anordnung und gute Lesbarkeit durch eine ausreichend große Schrift in den Vordergrund rückte (Anhang B.3). Eine Nummerierung der Aufgaben im Testheft sowie auf dem Antwortblatt sollte den Lernenden zusätzliche Orientierung verschaffen.

Es wurde dichotom kodiert, sodass jede richtige Antwort mit einem Punkt bewertet wurde und bei 3 Subtests mit je 12 Aufgaben insgesamt 36 Punkte erreicht werden konnten. Ein direkter Vergleich mit der Normstichprobe war in der Untersuchung nicht möglich, da das Geburtsdatum der Kinder nicht erhoben wurde. Alleinig mithilfe der Altersangabe konnte die der Normierungsstichprobe entsprechende Altersgruppe nicht eindeutig bestimmt werden, da beispielsweise Kinder von acht Jahren drei Altersgruppen zugeordnet werden könnten (vgl. Tabelle 4.10) (z. B. Becker et al., 1980, S. 94). Aus diesem Grund wurde der mittlere Summenscore der Probanden ($N = 347^{202}$) auf Rohwert-Basis, der bei 29.56 ($min = 9$, $max = 36$, $SD = 4.52$) lag, mit

²⁰² Insgesamt fehlten im Erhebungszeitraum vor den Sommerferien 5 % der Probanden.

den infrage kommenden CPM-Normierungsstichproben (siehe Tabelle 4.10) verglichen.

Tabelle 4.10. Verteilungskennwerte der CPM-Normierungsstichprobe nach Altersgruppen auf Rohwert-Basis

Altersgruppe	Summenscore	min	max	SD
07;10–08;03	22.94	9	36	6.09
08;04–08;09	25.01	11	36	6.29
08;10–09;03	25.80	8	36	5.91
09;04–10;03	27.89	9	36	5.71
10;04–11;00	29.18	8	36	5.42

Anmerkung. $N = 347$, min = minimal erreichbarer Testscore als Mittelwert aller Aufgabenscores; max = maximal erreichbarer Testscore als Mittelwert aller Aufgabenscores. Altersgruppe: Die Angaben vor dem Semikolon stehen für das Alter in Jahren, die Angaben danach für die Monate (Becker et al., 1980, S. 65).

Es zeigte sich, dass die Gesamtstichprobe (Alter: $M = 8.10$, $SD = 0.47$) mit einem mittleren Summenscore von 29.60 leicht über den Durchschnitten aller möglichen CPM-Normierungsstichproben lag.

Reliabilitäten

Insgesamt wird der Test bezüglich der Durchführung und Auswertung als objektiv eingestuft (Becker et al., 1980, S. 8). Die Split-Half-Reliabilitäten (spearman-brown-korrigiert) streuen für die vergleichbaren Altersgruppen zwischen .90 und .95 (Becker et al., 1980, S. 71) und die Retest-Reliabilitäten zwischen .86 und .90 bei ein bis zwei Wochen Testwiederholungsabstand (Becker et al., 1980, S. 73; Pauen, Pahnke, & Valentiner, 2007, S. 43). Die internen Konsistenzen nach Kuder-Richardson liegen zwischen .89 und .93 (Becker et al., 1980, S. 71). Bei den internen Konsistenzen fällt auf, dass von Subtest zu Subtest ein Anstieg zu verzeichnen ist (Becker et al., 1980, S. 73). Die Kennwerte sind zufriedenstellend (vgl. Weise, 1975, S. 219). Darüber hinaus gilt die Validität des Tests als zufriedenstellend bis akzeptabel (Bölte, 2013, S. 351). Auf Gesamttestebene konnte die faktorielle Validität durch die Zusammenhänge mit dem HAWIK-Test bestätigt werden. Es zeigte sich, dass 5 von 11 HAWIK-Subtests und der CPM-Test auf einen gemeinsamen Faktor laden (Becker et al., 1980, S. 76–77; Pauen et al., 2007, S. 43). Folglich messen diese fünf Subtests und die CPM zu einem wesentlichen Anteil das gleiche Konstrukt, „die nonverbalen Anteile der allgemeinen Intelligenz“ (Becker et al., 1980, S. 76). Entscheidend für das Lösen der Matrizen ist die „simultane Verarbeitung“ als visuell-räumlicher Faktor (Pauen et al., 2007, S. 43).

Diskussion

Mit dem Coloured-Progressive-Matrices-Test liegt ein valides Instrument vor, mit dessen Hilfe die fluide Intelligenz ohne zeitlichen und materiellen Aufwand im Klassenverband erhoben werden konnte. Kritisch anzumerken ist, dass die Analyse der Daten auf Deckeneffekte hinwies, was ein Indiz für ein Messinstrument mit zu geringer Schwierigkeit ist. Alternativ hätte anstelle des eindimensionalen CPM-Tests ein mehrdimensionaler Test, wie beispielsweise der Hamburg-Wechsler-Intelligenztest (WISC-IV, ehemals HAWIK-IV) oder der Kognitive Fähigkeitstest (KFT 1–3), für Kinder dieser Altersgruppe eingesetzt werden können. Der WISC-IV hat den Vorteil, dass er neben der allgemeinen Intelligenz die Intelligenzdimensionen *Sprachverständnis, wahrnehmungsgebundenes logisches Denken, Arbeitsgedächtnis und Verarbeitungsgeschwindigkeit* erfasst (Petermann & Petermann, 2011, S. 13), worunter insbesondere die Operationalisierung des Arbeitsgedächtnisses für die vorliegende Untersuchung von großer Bedeutung gewesen wäre. Dagegen sprach jedoch die Durchführungsökonomie: Zum einen werden 65–90 Minuten veranschlagt, um alle 10 Subtests durchzuführen, zum anderen ist eine Gruppentestung ausgeschlossen (Gerstenberger-Ratzeburg, 2016, S. 59), was letztlich das Ausschlusskriterium war. Gegen den KFT 1–3 sprach, dass er nur in der ursprünglichen Ersterscheinung vorliegt, nicht mehr überarbeitet wurde und daher von dessen Anwendung abgeraten wird (Pauen et al., 2007, S. 11). Ferner hätte der Culture-Fair-Intelligence-Test (CFT) als Pendant zum eindimensionalen CPM-Test genutzt werden können, um die fluide Intelligenz zu messen. Er ist in der Durchführungsökonomie mit dem CPM-Test vergleichbar und liegt in zwei Versionen vor: CFT 1 (5;3–9;5 Jahre) und der CFT 20-R (8;5–19;11 Jahre). Beide wurden verworfen, weil die vorliegende Stichprobe im Grenzbereich der Altersbegrenzungen lag und folglich Deckeneffekte beim CFT 1 und Bodeneffekte beim CFT 20-R erwartet wurden. Rückblickend hätte dem CFT 1 der Vorzug gewährt werden sollen, weil er zu den Speed-Tests zählt, was die Gefahr von Deckeneffekten im Vergleich zu Power-Tests minimiert.

4.5.3.2 Mathematische Basiskompetenzen

Die Untersuchungen von R. Rasch (2009, S. 75, 85) konnten die zuvor bereits aufgezeigte Bedeutung mathematischer Kompetenzen für den Lösungserfolg bekräftigen (siehe Abschnitt 1.5.3.1). Sie konnte zeigen, dass leistungsstarke Rechner besser abschneiden und auf sicherere Rechenkompetenzen zurückgreifen als schwache Rechner. Es gelang fitten Rechnern eher, mit Zahlbeziehungen kompetent umzugehen und auf dieser Basis ProblembARRIEREN zu bewältigen, auch wenn passende Grundmodelle nicht direkt ersichtlich waren. An dieser Stelle sei erneut auf die Ergebnisse von Groß (2013, S. 157–162, 183) und Hohn (2012, S. 129–130) hingewiesen, die diese Zu-

sammenhänge bestätigen. Die Lernenden müssen über ein Operationsverständnis verfügen, um erkennen zu können, welche Operation es in der vorliegenden Sachsituation anzuwenden gilt (Kittel, 2013, S. 110; Mayer et al., 1984, S. 265–266). Algorithmisches Wissen gilt als wesentlicher Einflussfaktor für erfolgreiches mathematisches Problemlösen (z. B. Daroczy, Wolska, Meurers, & Nuerk, 2015, S. 2; Mayer et al., 1984, S. 265–269).

Instrument zur Erhebung der mathematischen Basiskompetenzen

Die mathematischen Basiskompetenzen wurden mit dem Heidelberger Rechentest erfasst (HRT 1–4, Haffner, Baro, Parzer, & Resch, 2005). Die standardisierte Testung eignet sich für Einzel- und Gruppentestungen ab dem Ende der ersten bis zum Anfang der fünften Klassenstufe. Aus ökonomischen Gründen wurde sich auch hier für eine Gruppentestung entschieden.

Der Test setzt sich aus 11 Subtests zusammen, aus denen drei Skalenwerte ermittelt werden (Haffner et al., 2005, S. 9): 1) Rechenoperationen (6 Subtests: Addition, Subtraktion, Multiplikation, Division, Ergänzungsaufgaben und Größer-Kleiner-Vergleiche), 2) numerisch-logische und räumlich-visuelle Fähigkeiten (5 Subtests: Zahlenreihen, Längenschätzen, Würfelzählen, Mengenzählen, Zahlenverbinden) und 3) die Gesamtleistung aller 11 Subtests. Der Test ist so konzipiert, dass es für die Kinder in einer vorgegebenen Zeit möglichst viele Itemaufgaben zu lösen gilt, weshalb er zu den Speed-Tests zählt. Dabei nimmt die Schwierigkeit und Komplexität der Aufgaben zu. Für jede richtige Antwort erhalten die Lernenden einen Punkt.

Der T-Wert der Gesamtstichprobe ($N = 351$) liegt bei 46.36 ($min = 28$, $max = 80$, $SD = 10.18$) und damit unter dem Durchschnitt der HRT 1–4-Normierungsstichprobe ($N = 3354$; $T = 50.0$).

Reliabilitäten

Die Retest-Reliabilität liegt für alle 11 Untertests zwischen .69 und .89 und für die drei Skalenwerte zwischen .87 und .93 (Haffner et al., 2005, S. 10, 30), sodass sie den mittleren Bedingungen der Testgütekriterien genügen (Weise, 1975, S. 219). „Da die Aufgaben als eine typische Auswahl aus dem Spektrum der Grundrechenarten anzusehen sind“, gilt der Test als inhaltlich valide (Haffner et al., 2005, S. 31). Die Kriteriumsvalidität des Tests ist erfüllt, weil der HRT zum einen zu $-.67$ mit der Mathematiknote und zum anderen zu $.72$ mit dem Mathematiktest DEMAT 4 korreliert (Haffner et al., 2005, S. 10).

Diskussion

Die Entscheidung fiel auf den Heidelberger Rechentest, weil er „unabhängig von curricularen Stoffplänen basale mathematische und kognitive Kompetenzen, die als Grundlagen und Voraussetzungen für die Entwicklung komplexerer mathematischen Wissens erforderlich sind“, prüft (Haffner et al., 2005, S. 13). Für die vorliegende Untersuchung war ein Instrument erforderlich, das die mathematischen Basiskompetenzen aller Kinder zu messen und zu vergleichen ermöglicht, was demzufolge im HRT 1–4 gesehen wurde. Die Möglichkeit, zwischen rechenschwachen und rechenstarken Lernenden zu differenzieren, war gerade im Hinblick auf das Parallelisierungsverfahren und den regressionsanalytischen Ansatz relevant. Vorrangig war jedoch von Interesse, dass es keine inhaltlichen Überschneidungen mit dem Textaufgaben-Performance-Test gibt. Wenn der DEMAT 3+ als Testinstrument herangezogen worden wäre, dann hätte dieser neben der Arithmetik und der Geometrie den Bereich Sachrechnen und Größen erfasst. Diese Operationalisierungsmöglichkeit wurde verworfen, weil inhaltliche Überschneidungen mit dem Textaufgaben-Performance-Test nicht ausgeschlossen werden konnten.

4.5.3.3 Sprachliche Fähigkeiten und Leseverständnis

Zweifelsohne können mathematische Textaufgaben nur erfolgreich gelöst werden, wenn der Text und seine (auch implizit) enthaltenen Bezüge verstanden werden (vgl. Abschnitt 1.5.3.1). Daher wird dem Leseverständnis und dem Textverständnis ein nicht zu unterschätzender Einfluss auf den Lösungserfolg zugesprochen (Duarte et al., 2011, S. 40; Hegarty et al., 1995, S. 18, 29; Prediger, Renk, Büchter, Gürsoy, & Benholz, 2013, S. 51; Wöller, 2016, S. 215). Die Studie von Boonen, van Wesel, Jolles und van der Schoot (2014, S. 21) konnte einen signifikanten Zusammenhang von Leseverständnis und Lösungserfolg darlegen. Auch Hohn (2012, S. 131) konnte zeigen, dass sich hohes Textverständnis von Grundschulkindern positiv auf die Anzahl richtig gelöster Aufgaben, die Verwendung multipler Repräsentationen und den flexiblen Umgang mit den Repräsentationen sowohl innerhalb als auch zwischen den Aufgaben auswirkte. In der Jahrgangsstufe 4 nahm deren Einfluss auf den erzielten Lösungserfolg gegenüber der Jahrgangsstufe 3 sogar zu (Groß, 2013, S. 158, 160).

Instrument zur Erhebung der sprachlichen Fähigkeiten und des Leseverständnisses

Das Konstrukt *Leseverständnis* wurde mit dem standardisierten Leseverständnistest erhoben (ELFE 1–6, Lenhard & Schneider, 2006). Der Test ist für 6–13-Jährige normiert und kann sowohl in Papier- als auch in Computerversion in Einzel- oder Gruppentestungen erfolgen. Aus ökonomischen Gründen wurde sich in der Untersuchung

sowohl für die Papierversion als auch für eine Gruppentestung ausgesprochen. Der Leseverständnistest zählt ebenso wie der HRT 1–4 zu den Speed-Tests.

Dem Test liegt die Annahme zugrunde, dass für das Verstehen eines Textes verschiedene Verarbeitungsschritte unterschiedlicher Komplexität erforderlich sind, weshalb drei Subtests unterschieden werden: Wort-, Satz- und Textverständnis (Lenhard & Schneider, 2006, S. 13). Der Wortverständnistest umfasst insgesamt 72 Items, die einem Bild mehrere Wortalternativen gegenüberstellen, die sich insbesondere graphemisch und phonemisch ähneln (Lenhard & Schneider, 2006, S. 32). Die Lernenden müssen die jeweils richtigen Bezeichnungen unterstreichen und möglichst viele dieser Zuweisungen in der vorgegebenen Zeit schaffen. Das Erkennen der Bedeutung einzelner Wörter ist gemäß den Autoren essenziell, um einen vorliegenden Text inhaltlich verstehen zu können.

Gelingt es darüber hinaus, die Wortbedeutung auf der Basis des Kontextes zu interpretieren, erfolgte eine Vernetzung auf Satzebene (Lenhard & Schneider, 2006, S. 13). Demnach wird im Satzverständnistest anhand von 28 Sätzen überprüft, ob die Kinder fähig sind, jeweils aus fünf Auswahlalternativen den passenden Satzteil auszuwählen. Es gilt auch hier, möglichst viele Sätze in der vorgegebenen Zeit zu vervollständigen.

Wird die Herausforderung gemeistert, einzelne Sätze miteinander in Beziehung zu setzen und die grammatischen Strukturen zu berücksichtigen, dann kann der Text als verstanden interpretiert werden. Im Test wird das Textverständnis anhand vorgegebener Textpassagen unterschiedlichen Niveaus und jeweils einer zugehörigen Fragestellung realisiert, bei der die Kinder aus vier alternativen Antworten die richtige auswählen müssen. Dieser Subtest umfasst ebenso wie der Satzverständnistest 28 Aufgaben.

Für jede richtige Antwort wurde ein Punkt vergeben, sodass jedem Kind ein Testrohwert zugeordnet werden konnte (Lenhard & Schneider, 2006, S. 26). Entsprechend den Vorgaben wurden die Testrohwerte in z-Äquivalenzwerte umgewandelt. Das Gesamtergebnis ergibt sich demnach aus der Kumulierung der z-Werte der einzelnen Subtests. Der T-Wert der Gesamtstichprobe ($N = 352$) liegt bei 46.90 ($min = 29.3$, $max = 72.4$, $SD = 8.98$) und damit leicht unter dem Durchschnitt der ELFE 1–6-Normierungsstichprobe ($N = 3621$; $T = 50.0$).

Reliabilitäten

Der Test wird bezüglich der Durchführung, Auswertung und Interpretation als objektiv eingestuft, wenn den Anweisungen des Manuals gefolgt wird (Lenhard & Schneider, 2006, S. 34). Für die internen Konsistenzen (Cronbachs Alpha) und Split-Half-Reliabilitäten (r) werden in den einzelnen Subtests zufriedenstellend hohe Kennwerte erzielt (Wortverständnis: $\alpha = .97$, $r = .95$; Satzverständnis: $\alpha = .93$, $r = .95$; Textverständnis: $\alpha = .92$, $r = .89$) (Lenhard & Schneider, 2006, S. 34). Die Retest-Reliabilität

beträgt für den gesamten Test .91 ($p < .05$) und liegt für die drei Subtests zwischen .80 und .86 ($p < .05$) (Lenhard & Schneider, 2006, S. 36), sodass zufriedenstellende Kennwerte vorliegen (Weise, 1975, S. 219). Der Test gilt als valide, weil das Wortverständnis, das Satzverständnis und das Textverständnis aus inhaltlicher und theoretischer Sicht als Indikatoren für das Leseverständnis aufzufassen sind (Lenhard & Schneider, 2006, S. 37–38). Darüber hinaus ist die kriteriumsbezogene Validität gesichert, da der Test mit .76 (.79) signifikant mit der Würzburger Leise Leseprobe (WLLP, Küspert & Schneider, 1998) und mit .73 (.76) signifikant mit dem Lehrerurteil „Lesen“ für die 2. (3.) Jahrgangsstufe korreliert (Lenhard & Schneider, 2006, S. 38–39).

Diskussion

ELFE 1–6 ermöglichte es, das Leseverständnis auf ökonomische Weise im Klassenverband zu erfassen. Insbesondere die Erfassung des Wort-, Satz- und Textverständnisses ist essenziell, wenn kontrolliert werden soll, ob der gelesene Text, einzelne Sätze oder auch die Bedeutung einzelner Wörter verstanden wurden (Leiss et al., 2010, S. 124; Prediger, 2013b, S. 27–28, vgl. Abschnitt 1.5.3). Der Einsatz dieses Testinstruments hat sich bewährt.

4.5.3.4 Metakognitive Fähigkeiten

Metakognitive Fähigkeiten sind, wie bereits mehrfach herausgestellt, für den Lernerfolg von hoher Bedeutsamkeit (z. B. Kramarski et al., 2010, S. 185–186; Wang, Haertel, & Walberg, 1993, S. 272, 278, s. Abschnitt 2.4.2 sowie 4.5.3.4). Sie beeinflussen, inwiefern die Lösenden fähig sind, ihren Lösungsprozess zu steuern und zu reflektieren (Schoenfeld, 1985, S. 139–140, 1992, S. 349, vgl. Abschnitt 1.5.3.2). Dies hat zur Konsequenz, dass Problemlösende, die ihr Ergebnis als wahr annehmen, ohne es zu hinterfragen, zu überprüfen oder zu reflektieren, tendenziell zu den leistungsschwächeren Problemlösern zählen (Desoete, Roeyers, & Buysse, 2001, S. 445; Hembree, 1992, S. 256; Kramarski et al., 2010, S. 185–186; Lester & Kroll, 1990, S. 57; Muir et al., 2008, S. 230; Perels, Gürtler, & Schmitz, 2005, S. 134).

Es ist daher nicht auszuschließen, dass inadäquate Repräsentationen aufgrund fehlender Rückschau nicht als solche identifiziert werden (Schoenfeld, 1985, S. 139–140, 1992, S. 349; vgl. Deseniss, 2015, S. 87). Überdenkt und reflektiert der Lösende jedoch sein eigenes Vorgehen oder das eines anderen, so kann er Fehler aufspüren und den jeweiligen Lösungsprozess optimieren (siehe Abschnitt 1.5.5.3). Auf der Metaebene gelingt es, Heran- und Vorgehensweisen von außen zu rezipieren und zu reflektieren (siehe Abschnitt 2.4.2). Kramarski et al. (2010, S. 185–186) konnten in ihrer Studie zeigen, dass Drittklässler, die beim Problemlösen durch ein Metakognitions-

training unterstützt werden, nach dem Training signifikant bessere Problemlöser sind als Drittklässler, die kein Training absolvieren. Ferner konnten Perels, Gürtler und Schmitz (2005, S. 134) nachweisen, dass die gekoppelte Implementierung eines Metakognitions- und Problemlösetrainings der Implementierung der einzelnen Trainingskomponenten (Metakognition oder Problemlösen) überlegen ist. Hieraus resultiert die Erwartung, dass metakognitiv aktive Kinder die besseren Problemlöser sind und eine Kontrolle dieser Störvariablen unabdingbar ist. Die Messung der metakognitiven Fähigkeiten der Lernenden erfolgte sowohl durch den Fremdbericht der Lehrkraft als auch durch Selbstberichte der Kinder. Die konvergente Validität ist gesichert, wenn die mit dem Fremdbericht und Selbstbericht erhobenen Merkmale zu einem ähnlichen Ergebnis kommen (Eid et al., 2011, S. 46). Nachfolgend werden zunächst alle Testinstrumente vorgestellt und diskutiert. Im Anschluss daran folgt eine Überprüfung, ob die konvergente Validität gesichert war und die metakognitiven Fähigkeiten als ein Konstrukt abgebildet werden konnte.

Instrument 1 zur Erhebung der metakognitiven Fähigkeiten: Lehrerfragebogen

Die metakognitiven Fähigkeiten wurden in Anlehnung an die Untersuchung von Desoete (2007, S. 723–724) mit einem Lehrerfragebogen erhoben (vgl. Tabelle 4.11 und Anhang B.4). Dieser ursprünglich 20 Items umfassende Fragebogen wurde explizit für die Erfassung metakognitiver Fähigkeiten in der Primarstufe entwickelt und in mehreren Untersuchungen erprobt (Desoete, 2001, S. 26; Desoete et al., 2001, S. 437). Die Autoren erfassten vier Fähigkeits-Konstrukte mit einer 7-stufigen Likert-Skala: *Vorhersagen* (7 Items), *Planen* (4 Items), *Kontrollieren* (6 Items), *Evaluieren* (3 Items). Um die Belastung für die Lehrkräfte durch den Fragebogen nicht zusätzlich zu erhöhen, wurde die gekürzte Version des Fragebogens eingesetzt (Desoete, 2001, S. 34; Desoete et al., 2001, S. 441). Diese entstand im Rahmen der Untersuchung von Groß (2013, S. 94, 191, 240) in Kooperation mit der Autorin des Fragebogens. Jede Subskala wurde auf zwei Items gekürzt, sodass der Fragebogen nur noch 8 Items enthielt (Groß, 2013, S. 95): *Vorhersagen* (Item 1, 2), *Planen* (Item 3, 4), *Kontrollieren* (Item 5, 6), *Evaluieren* (Item 7, 8) (siehe Tabelle 4.11).

Reliabilitäten

Die Test-Retest Korrelation von 0.81 ($p < 0.01$), die zwischen 0.99 und 1.00 variierende Interrater-Reliabilität ($p < .001$) wie auch die internen Konsistenzen (Gesamttest: $\alpha = .98$, Subtests: $\alpha = 0.97, 0.89, 0.91$ und 0.90) sind für den gesamten 20-Items umfassenden Fragebogen als hoch einzustufen (Desoete, 2007, S. 711, 2008, S. 195). Trotz der Itemreduktion sind die von Groß (2013, S. 93–95) erzielten Skalenreliabilitäten (3. Jahrgangsstufe: $\alpha = .94$; 4. Jahrgangsstufe: $\alpha = .95$) und Trennschärfen (3. Jahrgangsstufe: zwischen .58 und .90; 4. Jahrgangsstufe: zwischen .85 und .88) als hoch

einzuordnen (Weise, 1975, S. 219). Seine Ergebnisse konnten in der vorliegenden Studie nahezu repliziert werden (Gesamttest: $\alpha = .94$, Subtests: $\alpha = 0.87, 0.64, 0.84$ und 0.72). Tabelle 4.11 kann die Trennschärfen der einzelnen Items entnommen werden. Vor dem Hintergrund, dass jede Subskala nur zwei Items erfasste, sind die Reliabilitäten als zufriedenstellend zu beurteilen.

Im Gegensatz zu Groß (2013, S. 93) wurde das Item 3 nicht ausgeschlossen, auch wenn die interne Konsistenz dadurch von $.94$ auf $.95$ angestiegen wäre. Dessen Trennschärfe konnte mit $.62$ als hoch eingestuft werden (Kelava & Moosbrugger, 2012, S. 86; Weise, 1975, S. 219).

Tabelle 4.11. Item-Skala-Statistik zur Trennschärfe der 8 Lehrerfragebogenitems der metakognitiven Fähigkeiten der Lernenden

Item		<i>M</i>	<i>SD</i>	r_{it}
1	Die Schülerin/der Schüler beginnt mit einem Plan.	3.21	1.81	.84
2	Die Schülerin/der Schüler kann zwischen relevanten und irrelevanten Informationen in einer Aufgabe unterscheiden.	3.45	1.57	.86
3	Die Schülerin/der Schüler ist mit der erstbesten Lösung zufrieden.	2.44	1.84	.62
4	Die mündliche Antwort der Schülerin/des Schülers ist gut vorbereitet, vollständig und klar verständlich.	3.30	1.69	.82
5	Die Schülerin/der Schüler arbeitet präzise und systematisch.	3.32	1.72	.86
6	Die Schülerin/der Schüler kontrolliert nach dem Beenden einer Aufgabe die Rechnungen.	3.63	1.65	.82
7	Die Schülerin/der Schüler kann wiedergeben, was wichtig war, um die Aufgabe zu lösen, und welche Strategie gut funktioniert hat.	2.83	1.80	.72
8	Die Schülerin/der Schüler kann bei ähnlichen Problemen ihr/sein bereits erworbene Wissen anwenden.	3.49	1.74	.84

Anmerkung. $N = 275$. r_{it} = Trennschärfe. Antwortformat reicht von nie bis immer (0–6); Zuordnung der Items: Vorhersage: 1, 2; Planen: 3, 4; Kontrollieren: 5, 6; Evaluieren: 7, 8.

Faktorenanalyse

Ob die empirischen Daten und das theoretische Modell hinreichend übereinstimmen, konnte konfirmatorisch mit einer Faktorenanalyse (CFA) nicht überprüft werden. Durch die Kürzung des Lehrerfragebogens auf zwei Items pro Faktor lagen nicht mehr ausreichend empirische Informationen zugrunde, um die unbekannten Modellparameter schätzen zu können (vgl. Brown, 2015, S. 61–62; Moosbrugger & Schermelleh-

Engel, 2012, S. 334). Das Modell war unterspezifiziert. Ein weiterer Grund für die Unterspezifikation kann in der hohen Korrelation einzelner Items ($r > .85$) gesehen werden (Brown, 2015, S. 163; Bühner, 2011, S. 432).

Aufgrund dessen wurden im nächsten Schritt die acht Fragebogenitems explorativ einer Hauptachsenanalyse unterzogen, um die zugrunde liegende Struktur zu identifizieren (Brown, 2015, S. 163). Dabei ließ sich ein Faktor mit Ladungen zwischen .63 und .89 extrahieren, der 72 % der Varianz erklärt (vgl. 59 % in Desoete, 2007, S. 715). Dies ist vor dem Hintergrund, dass alle Items das gemeinsame Konstrukt der metakognitiven Fähigkeiten zu erfassen scheinen, durchaus theoriekonform. Die Gefahr, dass das invers formulierte Item 3 einen eigenen Faktor bildet und somit zusätzlich zu seiner Merkmalsvarianz auch systematische Methodenvarianz beinhaltete, die zur Unter- oder Überschätzung der Reliabilität führen würde (Schermelleh-Engel & Werner, 2012, S. 134), bestand somit nicht. Bezüglich der Stichprobengröße waren aufgrund der Nichtrückgabe der Fragebögen durch einzelne Kolleginnen und Kollegen große Einbußen von 91 Probanden zu verzeichnen.

Diskussion

Beachtet man, dass jede Lehrkraft die metakognitiven Fähigkeiten aller Kinder der Klasse mit 8 Items einschätzen musste, war der Lehrerfragebogen als umfangreiches Instrument anzusehen. Bei einer Klassenstärke von 15 Kindern, hatte die Lehrkraft 120 Einschätzungen zu tätigen. Auch wenn mit Blick auf die Reliabilitäten, die 20 Items umfassende Fragebogenversion vorzuziehen ist, ist deren Erhebung aus Gründen der Durchführungsökonomie nicht umzusetzen. Berücksichtigte man, dass bereits bei dem gekürzten Fragebogen 91 Fremdeinschätzungen fehlten, so käme unweigerlich hinzu, dass bei einem längeren Fragebogen die Mehrbelastung steigt, was wiederum noch höhere Nichtrückgaben der Fragebögen zur Folge haben könnte. Der dadurch zu erwartende Bias ist als hoch einzustufen (Groves et al., 2009, S. 59).

Die größte Schwäche des Fragebogens ist jedoch in seiner internen Validität zu sehen. Es konnte gezeigt werden, dass die metakognitiven Fähigkeiten in engem Zusammenhang mit den mathematischen Basiskompetenzen stehen. Die Produkt-Moment-Korrelation nach Pearson gab an, dass die Einschätzung der Lehrkräfte bzgl. der metakognitiven Fähigkeiten der Lernenden signifikant mit den mathematischen Basiskompetenzen korrelierte ($r = .53$, 95 % CI [.45, .61], $p < .001$). Dies spiegelt einerseits wider, dass metakognitiv aktive Kinder leistungsstärkere Rechner sind, weil sie ihre Ergebnisse überdenken und reflektieren. Andererseits konnte nicht ausgeschlossen werden, dass die beiden Variablen Ähnliches messen und das Urteil der Lehrerkräfte bezüglich der metakognitiven Fähigkeiten von den mathematischen Basiskompetenzen der Kinder beeinflusst ist oder gar diesen entspricht. Dies hatte zur Folge, dass der

Testscore der metakognitiven Fähigkeiten nicht eindeutig interpretierbar ist und die mathematischen Kompetenzen der Kinder eine plausible Alternativerklärung darstellen. Aus vorstehenden Gründen wurde die Möglichkeit, die fehlenden Daten ganzer Klassen zu imputieren, als nicht objektiv und gewinnbringend angesehen, weshalb die Kontrollvariable für weitere Analysen ausgeschlossen wurde. Für Folgeanalysen wäre interessant zu überprüfen, ob der Fragebogen weiterhin ein Konstrukt abbildet, wenn Items aufgenommen werden, die mathematische Basiskompetenzen erfassen.

*Instrument 2 zur Erhebung der metakognitiven Fähigkeiten:
prospektiver/retrospektiver Schülerfragebogen I*

Die metakognitiven Fähigkeiten der Lernenden wurden neben dem Lehrerfragebogen, auch mit einem Schülerfragebogen, dem *Prospective Assessment of Children* (PAC) bzw. *Retrospective Assessment of Children* (RAC) von Desoete (2001, S. 45) gemessen (vgl. Anhang B.5).²⁰³ Diese Befragungen greifen die vier metakognitiven Fähigkeitsbereiche (Vorhersagen, Planen, Kontrollieren, Evaluieren) aus dem Lehrerfragebogen auf. Die 25 Fragebogen-Items wurden ins Deutsche übersetzt und entsprechend vor (PAC) bzw. nach (RAC) der Interventionsmaßnahme erhoben (Desoete, 2007, S. 710–711). Den Fragebögen lagen diskret abgestufte, verbale unipolare Ratingskalen zugrunde (Jonkisz, Moosbrugger, & Brandt, 2012, S. 52). Auch wenn die Einschätzung im Original auf einer 7-stufigen Likert-Skala (1 = nie bis 7 = immer) basierte, wurde sich in der vorliegenden Untersuchung für eine 4-stufige Skala (0 = nie, 1 = selten, 2 = manchmal, 3 = immer) ausgesprochen. Diese Entscheidung berücksichtigte die Empfehlung, dass bei der fokussierten Altersstufe möglichst wenige und den Lernenden vertraute Intensitätsabstufungen zu wählen sind (Büttner, 2008, S. 289; Jonkisz et al., 2012, S. 60). Auf die zusätzliche Abstufung „oft“ wurde verzichtet, um mittlere Ausprägungen bzw. die Tendenz zur Mitte zu verhindern.

Reliabilitäten

Die internen Konsistenzen beider Fragebögen liegen nach Weise (1975, S. 219) im mittleren zufriedenstellenden Bereich (vgl. Tabelle 4.12). Eine Ausnahme stellt der Subtest Planung dar, dessen Reliabilitäten mit $\alpha = .55$ bzw. $\alpha = .52$ nicht zufriedenstellend sind (Desoete, 2007, S. 711).

²⁰³ Basis des Fragebogens ist das „Metacognitive Skills and Knowledge Assessment“ (MSA, Desoete et al., 2001, S. 438, 449), das neben metakognitiven Fähigkeiten auch metakognitives Wissen erfasst (Desoete, 2001, S. 26–27; Desoete et al., 2001, S. 438).

Tabelle 4.12. Metakognitive Fähigkeiten: Reliabilitäten der Skalen

	Gesamt	Vorhersagen	Planen	Kontrollieren	Evaluieren
PAC	.81	.74	.55	.75	.79
RAC	.89	.77	.52	.81	.64

Anmerkung. PAC = Prospective Assessment of Children (Prätest); RAC = Retrospective Assessment of Children (Posttest). Zuordnung der Items: Vorhersage: 1, 5, 9, 10, 11, 12, 13, 14, 15; Planung: 3, 7, 16, 17; Evaluation: 2, 8, 24, 25 und Kontrolle: 4, 6, 18, 19, 20, 21, 22, 23 (Desoete, 2007, S. 711).

Weil sich die vier metakognitiven Fähigkeitsbereiche faktorenanalytisch nicht bestätigen ließen, können die in der vorliegenden Untersuchung erzielten Skalenreliabilitäten nur bedingt mit denen der Originalstudie verglichen werden. Wenn beim Prätest beide durch die Hauptachsenanalyse gewonnenen Konstrukte berücksichtigt wurden, dann ergaben sich interne Konsistenzen von $\alpha = .70$ (Faktor 1) bzw. $\alpha = .71$ (Faktor 2).²⁰⁴ Beim Posttest lag die Reliabilität bei dem auf 15 Items gekürzten Fragebogen mit $\alpha = .84$ im mittleren, zufriedenstellenden Bereich (vgl. Weise, 1975, S. 219).

Tabelle 4.13 kann entnommen werden, dass die Items 1, 5, 9, 13 und 14, die alle beabsichtigen, das Konstrukt *Vorhersage* zu erfassen, aufgrund ihrer niedrigen Trennschärfekoeffizienten zum Prä- und/oder Posttest auffielen ($r_{it} = < .30$, Weise, 1975, S. 219). Dies deutet darauf hin, dass die mit den Vorhersage-Items erzielte Differenzierung nur in schwachem Zusammenhang mit der durch den Gesamttest erzielten Differenzierung steht. Es gelingt mit den Items kaum, zwischen Probanden mit hohem und niedrigem Testwert zu differenzieren. Darüber hinaus fällt auf, dass sich zum Posttest hin die Trennschärfen größtenteils erhöhten und die Items demzufolge nach dem Training Ähnliches maßen wie der Gesamttest. Ausnahmen bildeten insbesondere die Items 5, 13 und 14, weil die Trennschärfen zum Posttest hin zurückgingen (siehe Tabelle 4.13). Es ist nicht auszuschließen, dass diese Items unabhängig davon sind, was die restlichen Fragebogenitems messen.

Tabelle 4.13. Analyse der Schülerfragenbogenitems metakognitiver Fähigkeiten zum Prätest (PAC) und Posttest (RAC)

Item	PAC ^a			RAC ^b		
	M	SD	r_{it}	M	SD	r_{it}
1 Bevor ich mit einer Sachaufgabe beginne, über-	1.31	1.01	.26	1.57	1.06	.34

²⁰⁴ Der Reliabilitätsanalyse werden die Ergebnisse der explorativen Faktorenanalyse (EFA) zugrunde gelegt (siehe nachfolgende Erörterung).

	lege ich mir, ob die Aufgabe schwer oder leicht ist.					
2	Wenn ich fertig bin, überprüfe ich mein Ergebnis.	1.97	0.99	.46	1.90	1.07
3	Bevor ich loslege, überlege ich mir, wie ich zum Ziel kommen kann.	1.38	1.09	.34	1.83	1.05
4	Bei schwierigen Aufgaben arbeite ich genauer als bei leichten Aufgaben.	1.73	1.09	.34	1.78	1.09
5	Schon bevor ich loslege, kann ich einschätzen, ob ich zum richtigen Ergebnis komme oder nicht.	1.33	1.10	.41	1.33	1.07
6	Ich ändere meinen Lösungsweg, wenn ich bei einer Aufgabe nicht mehr weiter komme.	1.50	1.05	.35	1.58	1.06
7	Ich kann Mitschülern erklären, wie ich die Aufgabe gelöst habe.	1.70	1.05	.38	1.81	1.07
8	Ich kann Fehler in meinen Lösungen aufspüren.	1.63	0.99	.34	1.66	0.97
9	Bei Sachaufgaben lese ich den Text vollständig durch.	2.62	0.74	.23	2.68	0.73
10	Ich unterstreiche bei Sachaufgaben wichtige Informationen.	1.40	1.08	.36	1.29	1.18
11	Ich kann wichtige Informationen von unwichtigen unterscheiden.	1.69	1.05	.42	1.74	1.03
12	Ich lese mir eine Sachaufgabe mehrmals durch.	1.82	1.02	.45	1.93	1.00
13	Ich mache mir eine Zeichnung.	1.26	0.97	.34	1.79	0.92
14	Ich schreibe Frage, Rechnung und Antwort auf.	2.18	0.94	.32	1.94	1.09
15	Bei schwierigen Aufgaben lasse ich mir mehr Zeit als bei leichten Aufgaben.	1.86	1.07	.38	2.09	1.00
16	Ich lasse mir Zeit, zu entscheiden, welche Informationen der Sachaufgabe wichtig sind.	1.48	1.05	.51	1.48	1.07
17	Ich lasse mir Zeit, zu entscheiden, ob ich die gegebenen Zahlen plus-, minus-, mal- oder durchrechnen muss.	1.59	1.07	.46	1.71	1.03
18	Ich überprüfe, ob ich richtig gerechnet habe.	1.93	1.02	.44	1.92	1.04
19	Ich arbeite mich Schritt für Schritt an die Lösung heran.	2.14	0.93	.37	2.14	0.94
20	Ich schreibe jeden Schritt sorgfältig auf.	1.79	1.02	.45	1.81	0.93
21	Ich arbeite genau.	2.32	0.81	.34	2.29	0.80
22	Wenn es mit dem Rechnen nicht gleich klappt, dann fange ich noch einmal an.	1.72	1.08	.33	1.80	1.06
23	Bevor ich genau rechne, schätze ich, was ungefähr herauskommen müsste.	1.51	1.10	.40	1.57	1.06

24	Wenn ich fertig bin, schaue ich mir meinen Lösungsweg noch einmal an und überlege, ob er richtig oder falsch ist.	1.85	1.01	.53	1.88	1.02	.57
25	Beim Lösen neuer Sachaufgaben überlege ich mir, wie ich es bei anderen Aufgaben gemacht habe.	1.71	1.02	.55	1.69	1.02	.52

Anmerkung. r_{it} = Trennschärfe. Antwortformat reicht von nie bis immer (0–3); Zuordnung der Items: Vorhersage: 1, 5, 9, 10, 11, 12, 13, 14, 15; Planung: 3, 7, 16, 17; Evaluation: 2, 8, 24, 25 und Kontrolle: 4, 6, 18, 19, 20, 21, 22, 23.

^a $N_{PAC} = 261$ (Prätest). ^b $N_{RAC} = 273$ (Posttest).

Faktorenanalyse

Im nächsten Schritt wurde die faktorielle Struktur des Fragebogens mit einer konfirmatorischen Faktorenanalyse (CFA) überprüft und die Güte des Modells wurde beurteilt (Moosbrugger & Schermelleh-Engel, 2012, S. 340). An dieser Stelle sollen lediglich die Ergebnisse des retrospektiven Schülerfragebogens (Posttest) Beachtung finden. Für die Ergebnisdarstellung des prospektiven Schülerfragebogens (Prätest) wird aus ökonomischen Gründen auf Anhang E.2 verwiesen.

Die 25 Items wurden theoriegeleitet den vier Faktoren *Vorhersagen*, *Planen*, *Kontrollieren*, *Evaluieren* zugewiesen, sodass die inhaltliche Bestimmung der vier Faktoren dem Originalfragebogen folgte (vgl. Desoete, 2007, S. 710–711). Als Schätzmethode wurde die Robust-Maximum-Likelihood-Methode verwendet. Der MLR-Schätzer nimmt sowohl eine Yuan-Bentler-Korrektur der Chi-Quadrat-Statistik als auch eine Schätzung robuster Standardfehler nach Huber-White vor, was in Anbetracht der nicht normalverteilten Rohdaten nötig war (Steinmetz, 2015, S. 66). Es zeigte sich, dass die latenten Faktoren *Planen* und *Vorhersagen* sowie *Kontrollieren* und *Evaluieren* sehr hoch miteinander korrelierten (> 1.0 , siehe Tabelle 4.14) und eine nicht positiv definierte Kovarianz-Matrix zur Folge hatten.

„A completely standardized factor loading with a value greater than 1.0 is problematic if the CFA consists of congeneric indicator sets (i. e., each indicator loads on one factor only)“ (Brown, 2015, S. 162). Diese Problematik sowie ein insgesamt starker Zusammenhang zwischen den latenten Variablen ($r > 0.50$) trafen auf die Analyse zu. Gründe hierfür können vielfältig sein: Sowohl hohe Multikollinearität als auch lineare Abhängigkeiten in den Daten können mögliche Ursachen einer nicht positiv definierten Matrix sein (Brown, 2015, S. 163). Vorrangig wird aber ein fehlspezifiziertes Modell als Auslöser gesehen, wenn das spezifizierte Modell von dem Modell, das die Daten unterstützt, enorm abweicht. Daher empfiehlt Brown (2015, S. 163): „Move back into a purely exploratory analytic framework (EFA) to revamp the measurement model“.

Tabelle 4.14. Korrelations- und Kovarianzmatrix der vier Faktoren zum Posttest

	Vorhersagen	Planen	Evaluieren	Kontrollieren
Vorhersagen	1.00	0.25	0.23	0.17
Planen	1.10	1.00	0.32	0.24
Evaluieren	0.87	0.90	1.00	0.25
Kontrollieren	0.99	1.02	0.93	1.00

Anmerkung. Korrelationen (Hauptdiagonale und untere Datenmatrix); Kovarianzen (obere Datenmatrix, grau gefärbt).

Infolgedessen wurde eine explorative Faktorenanalyse durchgeführt, um die Strukturen des Fragebogens zu erfassen. Das Kaiser-Meyer-Olkin-Kriterium bestätigte, dass sich die Stichprobe für die Analysen eignete, $KMO = .87$ („meritorious“ Hutcheson & Sofroniou, 1999, S. 225). Der kleinste KMO-Wert des Items 18 lag bei .47 und somit unter dem akzeptablen Limit von .50, weshalb das Item ausgeschlossen wurde (Field, 2013, S. 684–685). Durch eine erneut ausgeführte Hauptachsenanalyse mit 24 Items reduzierte sich das Kaiser-Meyer-Olkin-Kriterium minimal auf .86. Auf Itemebene war kein KMO-Wert unter der kritischen Grenze von .50.

Gemäß dem Kaiser-Kriterium wurden sieben Faktoren extrahiert, die insgesamt 56 % der Varianz erklärten. Der Scree-Test ergab einen Faktor. Aufgrund der Diskrepanz wurde die Parallelanalyse nach Horn hinzugezogen, welche ebenfalls einen Faktor extrahierte. Da die Parallelanalyse zu den zuverlässigen Extraktionskriterien zählt und insbesondere bei nicht eindeutigen Eigenwertsverläufen zu bevorzugen ist (Eid et al., 2011, S. 913; Hayton et al., 2004, S. 201; Moosbrugger & Schermelleh-Engel, 2012, S. 332), wurde deren Faktorenzahl gefolgt und auf eins festgesetzt. Dies lässt die Schlussfolgerung zu, dass homogene Items vorliegen, die zu einer eindimensionalen Skala metakognitiver Fähigkeiten führen.

Diesen Ergebnissen zufolge lag den Items ein Konstrukt, die *metakognitiven Fähigkeiten*, zugrunde (siehe Tabelle 4.15). 15 Items wiesen mittlere Ladungen $> .40$ auf diesem Faktor auf. Keine der Itemladungen war jedoch größer als .59. Darunter waren Items aller theoriegeleiteten Konstrukte vertreten (*Vorhersage*: 3 Items; *Planung*: 4 Items; *Evaluation*: 3 Items; *Kontrolle*: 5 Items). 6 der 9 *Vorhersage*-Items luden nur sehr schwach ($< .40$) auf den Faktor. Hierbei handelte es sich um die Items, deren geringe Trennschärfen bereits auffällig waren ($r_{ii} < .40$). Bei den ausgeschlossenen Items 4, 6 und 8 handelte es sich ebenfalls um Items geringer Trennschärfen. Der Faktor erklärt lediglich 25 % der Varianz im Modell.

Tabelle 4.15. Faktorladungen der Hauptachsenanalyse für die Schülerfragebogenitems metakognitiver Fähigkeiten zum Zeitpunkt des Posttests (RAC)

Item	Faktor
1 Bevor ich mit einer Sachaufgabe beginne, überlege ich mir, ob die Aufgabe schwer oder leicht ist.	.37
2 Wenn ich fertig bin, überprüfe ich mein Ergebnis.	.48
3 Bevor ich loslege, überlege ich mir, wie ich zum Ziel kommen kann.	.55
4 Bei schwierigen Aufgaben arbeite ich genauer als bei leichten Aufgaben.	.40
5 Schon bevor ich loslege, kann ich einschätzen, ob ich zum richtigen Ergebnis komme oder nicht.	.30
6 Ich ändere meinen Lösungsweg, wenn ich bei einer Aufgabe nicht mehr weiter komme.	.37
7 Ich kann Mitschülern erklären, wie ich die Aufgabe gelöst habe.	.41
8 Ich kann Fehler in meinen Lösungen aufspüren.	.40
9 Bei Sachaufgaben lese ich den Text vollständig durch.	.49
10 Ich unterstreiche bei Sachaufgaben wichtige Informationen.	.39
11 Ich kann wichtige Informationen von unwichtigen unterscheiden.	.45
12 Ich lese mir eine Sachaufgabe mehrmals durch.	.49
13 Ich mache mir eine Zeichnung.	.29
14 Ich schreibe Frage, Rechnung und Antwort auf.	.33
15 Bei schwierigen Aufgaben lasse ich mir mehr Zeit als bei leichten Aufgaben.	.39
16 Ich lasse mir Zeit, zu entscheiden, welche Informationen der Sachaufgabe wichtig sind.	.54
17 Ich lasse mir Zeit, zu entscheiden, ob ich die gegebenen Zahlen plus-, minus-, mal- oder durchrechnen muss.	.59
19 Ich arbeite mich Schritt für Schritt an die Lösung heran.	.61
20 Ich schreibe jeden Schritt sorgfältig auf.	.57
21 Ich arbeite genau.	.47
22 Wenn es mit dem Rechnen nicht gleich klappt, dann fange ich noch einmal an.	.44
23 Bevor ich genau rechne, schätze ich, was ungefähr heraus kommen müsste.	.48
24 Wenn ich fertig bin, schaue ich mir meinen Lösungsweg noch einmal an und überlege, ob er richtig oder falsch ist.	.61
25 Beim Lösen neuer Sachaufgaben überlege ich mir, wie ich es bei anderen Aufgaben gemacht habe.	.56

Anmerkung. $N_{RAC} = 274$ (Posttest). Faktorladungen $> .40$ sind hervorgehoben; Item 18 wurde im Prätest ausgeschlossen ($KMO < .50$). Zuordnung der Items: Vorhersage: 1, 5, 9, 10, 11, 12, 13, 14, 15; Planung: 3, 7, 16, 17; Evaluation: 2, 8, 24, 25 und Kontrolle: 4, 6, 19, 20, 21, 22, 23.

Die metakognitiven Fähigkeiten der Lernenden wurden demzufolge aus den 15 Item-Kennwerten mit Faktorladungen $> .40$ errechnet. Die Berechnung eines Testwerts war zulässig, weil die Testwertbildung auf homogenen Items basierte (Kelava & Moosbrugger, 2012, S. 85).

Diskussion

Mit dem Ziel, zu überprüfen, welchen Einfluss die nach der Intervention gemessenen metakognitiven Fähigkeiten auf den Lösungserfolg bzw. die Problemlösekompetenzen nahmen, waren die Ergebnisse des *Retrospective Assessment of Children* (RAC) von Bedeutung. Der Schülerfragebogen gilt als reliabel, sodass zuverlässige Aussagen über die metakognitiven Fähigkeiten der Lernenden zu erwarten waren. Es hatte sich bewährt, die 7-stufige Skala auf eine 4-stufige Skala zu reduzieren. Es gelang den Kindern ohne sichtbare Probleme, zwischen den Stufen *nie*, *selten*, *manchmal* und *immer* zu differenzieren, da sie ihnen aus dem Alltag vertraut waren. Dass die Stufe *oft* und somit die Tendenz zur Mitte unberücksichtigt blieb, wurde nicht als störend erachtet. Es fällt jedoch auf, dass insbesondere die Vorhersage-Items geringe Trennschärfen sowie geringe Faktorladungen aufwiesen. Sehr bedenklich ist zudem, dass dem Modell ein geringes Erklärungspotenzial zugrunde liegt.

Instrument 3 zur Erhebung der metakognitiven Fähigkeiten: prospektiver/retrospektiver Schülerfragebogen II (Vorhersage-Evaluierung)

Da insbesondere die Evaluierungsfähigkeiten ein geeigneter Prädiktor zu sein scheinen, um zwischen guten, mittleren und schwachen Problemlösern zu unterscheiden (Clercq, Desoete, & Roeyers, 2000, S. 305), wurden aus zeitökonomischen Gründen neben den Vorhersagefähigkeiten die Evaluierungsfähigkeiten herausgegriffen und unmittelbar vor sowie nach dem Problemlöse-Performance-Test gemessen (vgl. EPA2000, Clercq et al., 2000). Die Kinder standen bei allen drei Textaufgaben vor der Aufgabe, ihre Vorhersage- und Evaluierungsfähigkeiten einzustufen. Aus diesem kurzen Schülerfragebogen wurde sich ein spezifisches, auf die problemhaltigen Textaufgaben abgestimmtes Urteil versprochen, während der Schülerfragebogen I ganzheitlich die metakognitiven Fähigkeiten zu erfassen suchte. Um der Schwierigkeit entgegenzuwirken, dass an den Messzeitpunkten neben dem Lösen der Textaufgaben wenig Zeit für die Erfassung metakognitiver Fähigkeiten blieb, wurde sich für ein Multiple-Choice-Antwortformat ausgesprochen (Efklides et al., 1997, S. 233). Durch ein offenes Aufgabenformat wäre erschwerend hinzugekommen, dass Grundschulkinder die

Hürde überwinden müssten, ihre Einschätzung zu verbalisieren. Diese Anforderung würde für die untersuchte Altersstufe eine nahezu unüberwindbare Barriere darstellen (Clercq et al., 2000, S. 305), weshalb sie nicht weiter verfolgt wurde. Darüber hinaus wäre nicht auszuschließen, dass der Schülerfragebogen weniger die tatsächlichen metakognitiven Fähigkeiten, sondern eher das metakognitive Wissen und die Einstellungen der jungen Lernenden abbildet (Desoete, 2007, S. 716).

Die vierstufigen Beurteilungs-Skalen lehnten sich an den Skalen von Desoete (2001, S. 45) an (siehe Abbildung 4.3). Zuerst mussten die Lernenden die Aufgabe lesen (ohne sie zu lösen) und einschätzen, ob sie diese *sehr sicher* oder *sicher* lösen bzw. nicht lösen könnten (vgl. Clercq et al., 2000, S. 306; Desoete, 2001, S. 27). Nachdem sie die Problemstellung bearbeitet hatten, nahmen sie eine erneute rückblickende Beurteilung vor. Um sicherzugehen, dass die Probanden nach der Bearbeitung der Aufgabe nicht die vorher getroffene Entscheidung revidierten, wurden zwei Fragebögen unterschiedlicher Farbe eingesetzt. Als alle mit der Bearbeitung der Textaufgaben begonnen hatten, sammelte die testdurchführende Person die Vorhersage-Urteile ein (z. B. Clercq et al., 2000, S. 306).

A Glaubst du, du wirst die Aufgabe richtig lösen? <input type="checkbox"/> Ich bin mir sehr sicher, dass ich die Aufgabe richtig lösen werde. <input type="checkbox"/> Ich bin mir sicher, dass ich die Aufgabe richtig lösen werde. <input type="checkbox"/> Ich bin mir sicher, dass ich die Aufgabe falsch lösen werde. <input type="checkbox"/> Ich bin mir sehr sicher, dass ich die Aufgabe falsch lösen werde.	B Glaubst du, du hast die Aufgabe richtig gelöst? <input type="checkbox"/> Ich bin mir sehr sicher, dass ich die Aufgabe richtig gelöst habe. <input type="checkbox"/> Ich bin mir sicher, dass ich die Aufgabe richtig gelöst habe. <input type="checkbox"/> Ich bin mir sicher, dass ich die Aufgabe falsch gelöst habe. <input type="checkbox"/> Ich bin mir sehr sicher, dass ich die Aufgabe falsch gelöst habe.
--	--

Abbildung 4.3. Vierstufige Ratingskala zur Vorhersagefähigkeit (A) und Evaluierungsfähigkeit (B).

Die Lernenden erhielten zwei Punkte, wenn ihr Urteil mit ihrer tatsächlichen Lösung übereinkam, d. h., gingen sie davon aus, dass sie *sehr sicher* waren, die Aufgabe richtig zu lösen, und sie taten dies auch, dann erhielten sie 2 Punkte. Gleiches galt für die Einschätzung, sie würden die Aufgabe *sehr sicher* falsch lösen. Das Ankreuzen von *ich bin mir sicher*, führte bei richtiger Zuweisung (entweder zur richtigen oder zur falschen Lösung) zu 1 Punkt. Gab es keine Übereinstimmung zwischen der Einschätzung und der tatsächlichen Lösung, wurden 0 Punkte erzielt. Diese Skalierung wurde in Anlehnung an den Originalfragebogen vorgenommen (Desoete, 2001, S. 27, 2007, S. 728, 730; Desoete et al., 2003, S. 190–191).

Reliabilitäten

Die Subskalen der Fragebögen genügten nicht dem Testgütekriterium der Reliabilität (siehe Tabelle 4.16). Diese waren als niedrig einzustufen (Weise, 1975, S. 219).

Tabelle 4.16. Metakognitive Fähigkeiten: Reliabilitäten der Skalen

	Gesamt	Vorhersagen	Evaluieren
Prätest	.74	.56	.73
Posttest	.71	.47	.46
Follow-up-Test	.77	.57	.54

An dieser Stelle können die dem Originalfragebogen zugrunde liegenden Reliabilitäten nicht vergleichend herangezogen werden. Grund hierfür ist, dass die dortig ermittelten Reliabilitäten aus einer Bearbeitung von 25 Aufgaben resultierten und nicht nur wie in der vorliegenden Untersuchung aus drei.

Diskussion

Die Schülerbefragung zu Beginn und zum Ende der eigenständigen Arbeitsphase stellte an sich eine gute Möglichkeit dar, die Vorhersage- bzw. Reflexionsfähigkeiten der Lernenden zu messen. Rückblickend sei negativ angemerkt, dass der Fragebogen für Grundschulkinder zu textlastig war (siehe Anhang B.6). Da die Kinder ihren Erfolg vorhersagen bzw. evaluieren mussten, genügte es nicht, dass der Schülerfragebogen nur die Items zur Erfassung der metakognitiven Fähigkeiten enthielt. Vielmehr mussten auch die einzuschätzenden Testaufgaben schriftlich fixiert werden. Aufgrund der sich hieraus ergebenden Textfülle war die Übersichtlichkeit nicht gewahrt. Hinzu kommt, dass die zur Verfügung stehende Zeit knapp bemessen war, sodass schlechte Leser dadurch benachteiligt gewesen sein könnten. Zudem wären negative Auswirkungen auf die Motivation nicht auszuschließen. Im Post- und Follow-up-Test wurde der retrospektive Schülerfragebogen gekürzt und auf der letzten Seite des Textaufgaben-Performance-Tests abgebildet (siehe Anhang B.2).

Einen weiteren Kritikpunkt stellen die Formulierungen der Itemstufen dar. Ihre Ähnlichkeit brachte unweigerlich die Gefahr der Verwechslung mit sich: Die ersten beiden waren positiv formuliert und unterschieden sich nur im Ausmaß der Sicherheit, *ich bin mir sicher* bzw. *ich bin mir sehr sicher* (siehe Abbildung 4.3). Analoges galt für die beiden letzten Stufen, nur dass zusätzlich das Wort *richtig* durch das Wort *falsch* ersetzt wurde. Es besteht die Gefahr, dass Kinder, die nach dem Lesen der Textaufgaben

unmittelbar mit der Problemlösung starten wollen, die Items nur flüchtig lesen und dadurch verwechseln. In diesem Fall wäre naheliegend, dass sie weder bereit sind, viel Zeit zu investieren noch die Einschätzung gewissenhaft vorzunehmen. Dies könnte ein Indiz für die geringen Reliabilitäten bei der Vorhersage sein. Für Folgeanalysen ist denkbar, dass das Antwortformat nicht in ganzen Sätzen erfolgt, sondern lediglich mit Schlagworten oder Smileys gearbeitet wird. Die möglichen Stufen der Einschätzung könnten dann schnell erfasst werden und der Problematik der Verwechslung entgegenwirken.

Fazit

Die Erhebung von Fremd- und Selbstberichten verfolgte das Ziel, die einzelnen Komponenten metakognitiver Fähigkeiten multimethodal zu erfassen. Es wurde erwartet, dass Grundschulkinder, die ihre Fähigkeiten häufig überschätzen und ihren Lösungsprozess selten von sich aus reflektieren, Schwierigkeiten haben, ihre metakognitiven Fähigkeiten objektiv einzuschätzen. Beide Methoden wurden eingesetzt, um zu untersuchen, inwiefern sie zu ähnlichen Ergebnissen kommen und das gleiche Konstrukt zu messen beabsichtigen. Der Fremdbericht durch die Lehrkräfte musste jedoch aus den zuvor geschilderten Gründen ausgeschlossen werden, weshalb die Intention, multimethodal zu untersuchen, scheiterte. Eine Berücksichtigung aller der Operationalisierungsmöglichkeiten hätte eine Stichprobenreduktion auf $N = 219$ Probanden ergeben, die nicht in Kauf genommen werden konnte. Faktorenanalytisch wurden überprüft, ob das Konstrukt *metakognitive Fähigkeiten* mit möglichst einem Faktor adäquat abzubilden war. Alle drei aus den Selbstberichten erhobenen Merkmale wurden einer Hauptachsenanalyse unterzogen, welche einen Faktor extrahierte. Die Ladungen waren nur für die Vorhersage- und Evaluierungsfähigkeiten der Lernenden $> .40$. Der aggregierte Faktor ging als Repräsentant des Konstruktts *metakognitive Fähigkeiten* neben den restlichen Kontrollvariablen in die multiple Regression ein.

4.5.3.5 Attitudes und Beliefs

„What one knows about a domain clearly sets the boundaries for what one can establish within it. Issues of affect are equally important, for they similarly establish boundaries and determine behavior“ (Schoenfeld, 1985, S. 154). Ob ein Lösender bei der Bearbeitung einer problemhaltigen Textaufgabe die ProblembARRIERE überwinden kann, hängt nicht nur von seinen Fähigkeiten und Fertigkeiten, die er in den Lösungsprozess einbringt, und von seinem Vorwissen ab, sondern auch von seinen *Attitudes* und *Beliefs* (Reiss et al., 2002, S. 52; Schoenfeld, 1985, S. 35, 154; Verschaffel et al., 2000, S. X). Diese affektiven Komponenten stellen nicht zu unterschätzende Einflussgrößen auf den Lernerfolg dar (Awofala, 2014, S. 280–282; Charles et al., 1987, S. 10; Gol-

din, Rösken, & Törner, 2009, S. 4; Lester, Garofalo, & Kroll, 1989a, S. 84–85; Lester & Kroll, 1990, S. 57; Reiss et al., 2002, S. 52).

Die Lernenden verknüpfen ein bestimmtes Gefühl mit dem Objekt, im vorliegenden Falle mit den Textaufgaben, sodass ein bestimmter emotionaler Zustand ausgelöst wird (vgl. Grigutsch, Raatz, & Törner, 1998, S. 7). Die erzeugten Gefühle bzw. Zustände sind subjektiv und können daher interindividuell ganz unterschiedlich ausfallen (Goldin et al., 2009, S. 4). Misserfolge, fehlendes Lob, schlechte Noten oder fehlendes Selbstvertrauen hemmen beispielsweise den Problemlösenden in seiner Motivation und Kreativität und letztlich auch in seiner Leistung (Charles et al., 1987, S. 19; Lester et al., 1989a, S. 84–85; Lester & Kroll, 1990, S. 57). Individuelle Einstellungen gegenüber Textaufgaben beeinflussen, wenn auch eher unbewusst, die Herangehensweisen an Problemstellungen und die Bereitschaft, Mathematik zu lernen und die Aufgaben zu lösen (Awofala, 2014, S. 284; Grigutsch et al., 1998, S. 4). Die Beeinflussung kann sowohl positiv als auch negativ sein, weshalb sie als mögliche Störvariable zu erfassen ist.

Instrument zur Erhebung der Attitudes und Beliefs

Im Rahmen des Schülerfragebogens wurden die *Attitude Inventory Items* von Charles et al. (1987, S. 27) herangezogen, um die Attitudes und Beliefs der Lernenden beim Lösen problemhaltiger Textaufgaben zu erfassen (vgl. Anhang B.7). Der Fragebogen kam sowohl vor dem Training als auch 13 Wochen nach dem Training zum Einsatz. Aus Gründen der Vereinheitlichung wird in Anlehnung an die Erfassung der metakognitiven Fähigkeiten nachfolgend von *Prospective Attitude Inventory Items* (PAII), Erhebung vor dem Training, bzw. *Retrospective Attitude Inventory Items* (RAII), Erhebung nach dem Training, gesprochen.

Im Fragebogen werden drei Konstrukte unterschieden: *Bereitschaft* („willingness“: Items 1, 3, 9, 10), *Durchhaltevermögen* („perseverance“: Items 2, 4, 6, 11) und *Selbstbewusstsein* („self-confidence“: Items 5, 7, 8, 12, 13) (Charles et al., 1987, S. 26). In Anlehnung an die Autoren wurde das dichotome Antwortverhalten (ja/nein) übernommen und die invers formulierten Items (2, 7, 9, 11, 12) wurden rekodiert. Im Anschluss wurde der Testscore als Summenwert aller Items gebildet (Charles et al., 1987, S. 27).

Reliabilitäten

Beiden Fragebögen lagen nach Weise (1975, S. 219) niedrige bis mittlere interne Konsistenzen zugrunde (siehe Tabelle 4.17). Die Autoren der Skala publizierten keine Reliabilitätskennwerte zum Fragebogen, weshalb an dieser Stelle lediglich die Kennwerte der vorliegenden Untersuchung aufgeführt werden können.

Tabelle 4.17. Attitudes und Beliefs: Reliabilitäten der Skalen

	Gesamt	Bereitschaft	Durchhaltevermögen	Selbstbewusstsein
PAII	.83	.65	.44	.80
RAII	.84	.70	.55	.77

Anmerkung. PAII = Prospective Attitude Inventory Items (Prätest); RAII = Retrospective Attitude Inventory Items (Posttest). Zuordnung der Items: Bereitschaft: 1, 3, 9, 10; Durchhaltevermögen: 2, 4, 6, 11; Selbstbewusstsein: 5, 7, 8, 12, 13.

Aus Tabelle 4.18 geht hervor, dass die Items 2, 4 und 11 des Konstrukts *Durchhaltevermögen* im Prätest sehr niedrige Trennschärfekoeffizienten aufwiesen ($r_{it} = < .30$, Weise, 1975, S. 219). Die Items scheinen demzufolge weniger geeignet zu sein, zwischen Probanden mit hohem und Probanden mit niedrigem Testwert zu differenzieren. Gründe hierfür können sein, dass die Grundschulkinder sozial erwünscht antworteten oder aber ihre Anstrengungsbereitschaft und ihre Frustrationstoleranz überschätzten und somit ihr Durchhaltevermögen nicht objektiv beurteilten. Dies fiel beim Prätest (PAII) stärker ins Gewicht als beim Posttest (RAII), sodass zusätzlich davon ausgegangen werden kann, dass den Problemlösenovizen eine objektive Einschätzung ihres *Durchhaltevermögens* leichter fiel, wenn sie bereits Problemlöseerfahrungen gesammelt hatten und auf diese zurückgreifen konnten. Das vierte Item war zu beiden Zeitpunkten durch niedrige Trennschärfen gekennzeichnet ($r_{it} = < .30$, Weise, 1975, S. 219). Die Ursache kann darin gesehen werden, dass es sich um ein Item mit doppelter Verneinung handelte, dessen Formulierung *ich höre nicht auf* Verständnisschwierigkeiten und letztlich Verzerrungen in den Antworten hervorbringen kann. Die Mittelwerte beider Messzeitpunkte $M_{PAII} = .72$ bzw. $M_{RAII} = .78$ (siehe Tabelle 4.18) deuteten allerdings darauf hin, dass die Kinder ihr Durchhaltevermögen sehr hoch einschätzten, eventuell überschätzten. Den Ergebnissen zufolge waren sie der Auffassung, dass sie nicht sofort aufgeben, wenn sie keine Lösung finden.

Tabelle 4.18. Analyse der Schülerfragenbogenitems Attitudes und Beliefs zum Prätest (PAII) und zum Posttest (RAII)

Item	PAII ^a			RAII ^b		
	<i>M</i>	<i>SD</i>	r_{it}	<i>M</i>	<i>SD</i>	r_{it}
1 Ich wage mich an jede knifflige Sachaufgabe heran.	.59	.49	.55	.64	.48	.53
2 Wenn ich nicht sofort die richtige Lösung finde, gebe ich auf.	.89	.31	.22	.91	.28	.44

3	Ich löse gerne knifflige Sachaufgaben.	.53	.50	.60	.49	.50
4	Ich höre nicht auf, bis ich eine Lösung gefunden habe.	.72	.45	.26	.78	.42
5	Ich bin mir sicher, dass ich die meisten kniffligen Sachaufgaben lösen kann.	.55	.50	.50	.61	.49
6	Ich kann mich lange mit kniffligen Sachaufgaben beschäftigen.	.62	.49	.47	.64	.48
7	Beim Lösen von kniffligen Sachaufgaben brauche ich meistens Hilfe von anderen.	.56	.50	.60	.66	.47
8	Ich kann die meisten kniffligen Sachaufgaben lösen.	.63	.48	.59	.73	.45
9	Ich mag keine Sachaufgaben, die ich nicht gleich verstehe.	.57	.50	.38	.70	.46
10	Ich knolle gerne.	.55	.50	.46	.58	.49
11	Ich gebe beim Lösen von Sachaufgaben schnell auf.	.84	.37	.29	.90	.31
12	Viele knifflige Sachaufgaben sind mir zu schwer.	.55	.50	.58	.63	.48
13	Ich bin gut im Lösen von kniffligen Sachaufgaben.	.50	.50	.66	.55	.50

Anmerkung. PAII = Prospective Attitude Inventory Items (Prätest); RAII = Retrospective Attitude Inventory Items (Posttest); r_{it} = Trennschärfe. Antwortformat: nein (0); ja (1); Antwortformat rekonduzierter Items (2, 7, 9, 11, 12): nein (1); ja (0). Zuordnung der Items: Bereitschaft: 1, 3, 9, 10; Durchhaltevermögen: 2, 4, 6, 11; Selbstbewusstsein: 5, 7, 8, 12, 13.

^a $N_{RAII} = 310$. ^b $N_{PAII} = 309$.

Faktorenanalyse

Auch bei der Kontrollvariablen *Attitudes* und *Beliefs* wurde mit einer konfirmatorischen Faktorenanalyse die faktorielle Struktur des Fragebogens überprüft und die Güte des Modells beurteilt (Moosbrugger & Schermelleh-Engel, 2012, S. 340). Die 13 Items wurden theoriegeleitet nach Charles et al. (1987, S. 26) den drei Faktoren *Bereitschaft*, *Durchhaltevermögen*, *Selbstbewusstsein* zugeordnet. Die Analyse wurde mit dem R package *lavaan* durchgeführt. Als Schätzmethode wurde die robuste Unweighted-Least-Squares-Schätzmethode (ULSMV-Schätzer) verwendet, da die Rohdaten dichotom vorlagen (vgl. CFA, Abschnitt 4.5.1.3). Wie die Fit-Indizes zeigen, wies das dreifaktorielle Modell vor dem Training (PAII) einen guten Modellfit auf: $\chi^2(62) = 134.856$, $p < .001$; $\chi^2/df = 2.175$, CFI = 0.948, TLI = 0.935, RMSEA = 0.062 (p

= .086)²⁰⁵ (Brown, 2015, S. 74–75; Hair et al., 2009, S. 753; Moosbrugger & Schermelleh-Engel, 2012, S. 338; Nussbeck, Eid, & Lischetzke, 2006, S. 196).

Unter Berücksichtigung des signifikanten χ^2 -Werts erfüllen die Fit-Indizes die Vorgaben bei Stichproben mit mehr als 250 Probanden und mehr als 12 gemessenen Variablen: CFI $\geq .92$, TLI $\geq .92$, RMSEA $< .07$ (Hair et al., 2009, S. 753). Die Fit-Indizes des RAII zum Zeitpunkt des Nachtests weisen einen gerade noch akzeptablen Modellfit auf: $\chi^2(62) = 143.170$, $p < .001$; $\chi^2/df = 2.309$, CFI = 0.926, TLI = 0.907, RMSEA = 0.065²⁰⁶. Die Modellfits zum Prä- und Posttest konnten durch den Ausschluss des Items mit der geringen Trennschärfe (Item 4) weder repliziert noch verbessert werden, weshalb das Item in der Analyse berücksichtigt blieb. Es kann davon ausgegangen werden, dass die Items 1, 3, 9 und 10 Indikatoren des Konstrukt *Bereitschaft*, die Items 2, 4, 6, und 11 des Konstrukt *Durchhaltevermögen* und die Items 5, 7, 8, 12 und 13 des Konstrukt *Selbstbewusstsein* sind.

Diskussion

Es wurde sich für dieses Erhebungsinstrument ausgesprochen, da es eines der wenigen darstellt, das für den Einsatz bei jungen Probanden konzipiert wurde. Einen Überblick über die Erfassung von Attitudes und Beliefs außerhalb der Grundschule gibt Pehkonen (1995, S. 25–27). Trotz des empirischen Befundes, dass im Allgemeinen die Attitudes und Beliefs bei Grundschulkindern als kontrolliert anzusehen sind (Spiel et al., 2011, S. 308, vgl. Abschnitt 4.2.4.1), wurde die Erhebung insbesondere im Kontext des Problemlösens als sinnvoll erachtet. Wenn man bedenkt, dass die Lernenden mit der Textaufgabe ein bestimmtes Gefühl assoziieren, das wiederum einen bestimmten emotionalen Zustand auslösen kann, dann ist diese Einflusskomponente beim Lösen problemhaltiger Textaufgaben als wichtig erachtet (Grigutsch et al., 1998, S. 7, vgl. Abschnitt 4.5.3.5).

Bedenklich war die nicht zufriedenstellende Reliabilität der Subskala *Durchhaltevermögen*. Ein Grund hierfür kann sein, dass sich Grundschulkinder in ihren Fähigkeiten und Leistungen überschätzen und auf diesem Merkmal eine homogene Stichprobe abbilden (Weinert, 1998, S. 119). Dies hat zur Konsequenz, dass die geschätzte Reliabilität geringer ausfällt als in heterogenen Stichproben (Eid et al., 2011, S. 820). Den Ergebnissen der konfirmatorischen Faktorenanalyse zufolge konnte das theoretische Mo-

²⁰⁵ Es fällt auf, dass die Goodness-of-Fit-Indizes des Modells unter Verwendung des WLSMV-Schätzers nahezu identisch sind mit den Goodness-of-Fit-Indizes des Modells unter Verwendung des ULSMV-Schätzers: $\chi^2(62) = 132.894$, $p < .001$, $\chi^2/df = 2.143$, CFI = 0.946, TLI = 0.932, RMSEA = 0.061.

²⁰⁶ Die Goodness-of-Fit-Indizes des Modells unter Verwendung des WLSMV-Schätzers fallen hier geringfügig schlechter aus: $\chi^2(62) = 150.559$, $p < .001$, $\chi^2/df = 2.428$, CFI = 0.907, TLI = 0.883, RMSEA = 0.068.

dell bestätigt und folglich die dem Fragebogen zugrunde liegende Struktur repliziert werden. Es ist von einem konstruktvaliden Instrument auszugehen, da es die Merkmale Attitudes und Beliefs zu erfassen scheint.

4.5.3.6 Mehrsprachigkeit

Mit der Bedeutung der Sprache bei mathematischen Lernprozessen geht eine weitere Einflussgröße auf den Lernerfolg einher: der Sprachstand der Lösenden (vgl. Abschnitt 1.5.3.1). Lernende, die migrationsbedingt mehrsprachig aufwachsen, stehen vor der Herausforderung, die an sie gerichteten schulischen Anforderungen (nicht nur im Deutschunterricht) zu bewältigen. Dies geschieht nicht problemlos, auch wenn sie über alltagskommunikative Kompetenzen verfügen (Heinze, Herwartz-Emden, Braun, & Reiss, 2011, S. 28; Heinze, Herwartz-Emden, & Reiss, 2007, S. 566). Vielmehr müssen die Lernenden die deutsche Sprache über die Verständigung hinaus erlernen, um auf dieser Sprachbasis Lernprozesse erfolgreich absolvieren zu können. Dass dies gerade bei Kindern mit Migrationshintergrund mit einer Benachteiligung einhergeht, wurde schon in der PISA-, IGLU- und LAU-Studie aufgedeckt (Deseniss, 2015, S. 50; Heinze et al., 2007, S. 566). Im Vergleich zu deutschen Kindern erzielten Kinder mit migrationsbedingter Mehrsprachigkeit bereits in der Grundschule signifikant geringere mathematische Kompetenzen (Heinze et al., 2011, S. 26–27; Schwippert, Bos, & Lanckes, 2003, S. 285). Es konnte ein bedeutsamer Zusammenhang zwischen mathematischen Kompetenzen und dem Migrationsstatus nachgewiesen werden. Die Hamburger Lernausgangsuntersuchung LAU 5 zeigte in den fünften Jahrgangsstufen ebenso deutliche Unterschiede in den Mathematikleistungen und den nonverbalen kognitiven Lernvoraussetzungen zugunsten deutscher Kinder auf (Heinze et al., 2007, S. 566). Dagegen fand Deseniss (2015, S. 340–341) heraus, dass sich der Migrationshintergrund von Fünftklässlern nachteilig auf den Lösungserfolg auswirkt. Dies wurde vorrangig in mündlichen, weniger in schriftlichen, Äußerungen deutlich.

Im Hinblick auf das Lösen problemhaltiger Textaufgaben werden insbesondere bei Lernenden mit migrationsbedingter Mehrsprachigkeit Verständnisschwierigkeiten erwartet, die sich negativ auf den Lösungserfolg auswirken könnten (vgl. Abschnitt 1.5.3). In diesem Zusammenhang sind das Deuten von Präpositionen und Komposita, der sogenannte Strukturwortschatz eines Textes, komplexe Verbformen und Satzverbindungen als nicht zu unterschätzende Hürden des Verständnisses anzusehen (Deseniss, 2015, S. 65; Duarte et al., 2011, S. 44; Kaiser & Schwarz, 2009, S. 68–69).

Instrument zur Erhebung der Mehrsprachigkeit und Diskussion

Es galt die Annahme, dass Kinder mehrsprachig aufwachsen, wenn sie mit mindestens einem Elternteil in einer anderen Sprache als Deutsch kommunizieren (vgl. Lenhard &

Schneider, 2006, S. 46). Kinder, die nur in der deutschen Sprache mit ihren Eltern kommunizieren, wachsen gemäß der Annahme einsprachig auf. Im Rahmen der Leseverständnistestung (ELFE 1–6) wurden die Kinder aufgefordert, anzukreuzen, ob sie zu Hause einsprachig oder mehrsprachig kommunizieren: „Mache bei ‚Muttersprache‘ ein Kreuz bei ‚deutsch‘, wenn du und deine Eltern deutsch sprechen. Kreuze ‚andere‘ an, wenn du mit deinen Eltern eine andere Sprache sprichst.“ (Lenhard & Schneider, 2006, S. 19). Wobei „andere“ auch anzukreuzen ist, wenn das Kind nur mit einem Elternteil deutsch spricht.

Ferner wäre möglich gewesen, den Migrationshintergrund an weiteren Parametern festzumachen. Heinze et al. (2007, S. 571) nutzten zusätzlich zur Familiensprache die Parameter Staatsangehörigkeit, Herkunft der Familie, Aussiedlungshintergrund, Geburtsort, Religion und Vorname. Sie konnten zeigen, dass eine kombinierte Erfassung von Parametern den Migrationshintergrund zuverlässiger vorhersagt, als wenn lediglich ein Parameter, z. B. die Staatsangehörigkeit, erhoben wird (Heinze et al., 2007, S. 572). Auskünfte über die von den Autoren genannten Parametern hätten die Erziehungsberechtigten, nicht jedoch die Drittklässler erteilen können. Es wurde befürchtet, dass die Bereitschaft der Erziehungsberechtigten, die Intelligenz, die mathematischen Basiskompetenzen und das Leseverständnisses ihres Kindes messen zu lassen, sinkt, wenn eine zusätzlichen Erhebung soziodemografischer Daten erfolgt. Durch die Erhebung der Fülle an Parametern soziodemografischer Daten wäre ein Akzent gesetzt worden, der nicht im Fokus der Untersuchung stand. Es wurde befürchtet, dass die Erziehungsberichtigen abgeschreckt wären und eine Teilnahme verweigern könnten, weshalb im Vorfeld die Erfassung des Leseverständnisses der Erfassung des Migrationshintergrunds vorgezogen wurde. Mit dem Ziel, möglichst viele Kinder für die Teilnahme zu gewinnen, wurde der Fokus nicht auf die Erfassung soziodemografischer Daten gelegt. Alternativ eine Befragung durch die Lehrkräfte durchzuführen, wurde aus zeit- und durchführungsökonomischen Gründen nicht in Betracht gezogen.

Hinzu kam, dass in den Stadt Schulen ein Großteil der Erziehungsberechtigungen Schwierigkeiten hatte, den in deutscher Sprache verfassten Elternbrief zu verstehen, sodass sowohl die Lehrkräfte als auch die Schulleitung unterstützend und motivierend zur Teilnahme aufriefen. Einen zusätzlichen Schwerpunkt auf die Erfassung des Migrationshintergrunds zu legen, hätte Skepsis bei den Erziehungsberechtigten hervorrufen können.

4.5.3.7 Geschlecht

Gerade weil die Forschungslage zu geschlechterspezifischen Unterschieden in der Grundschule uneindeutig ist (Stern, 1997, S. 109), sollte überprüft werden, ob das Ge-

schlecht Einfluss auf die Lösungsprozesse problemhaltiger Textaufgaben und den damit einhergehenden Erfolg nimmt.

Stern (1997, S. 108) konnte zeigen, dass Jungen sowohl im Vorschultest und beim Lösen mathematischer Textaufgaben als auch im proportionalen Denken den Mädchen überlegen waren. Darüber hinaus deckte sie auf, dass Jungen und Mädchen, die bereits im Vorschultest hohe Leistungen erzielten, sich diesen Vorteil bis in die 5. Klasse bewahren konnten (Stern, 1997, S. 110–111). Beobachtbar war auch, dass wesentlich mehr Jungen, die sich in der Vorschulzeit noch im mittleren Leistungsdrittel befanden, in der 5. Klasse das obere Leistungsdrittel erreichten als Mädchen (Stern, 1997, S. 110–111). Gründe für den Leistungsvorteil sieht Stern in der Tatsache, dass die Jungen in der Grundschulzeit von ihren bereits in der Vorschulzeit entwickelten mathematisch-numerischen Kompetenzen profitieren.

Dagegen konnten Hyde, Fennema und Lamon (1990, S. 139) diese Geschlechterdifferenzen in der Grundschule nicht bestätigen (vgl. Hyde, 2005, S. 588). Im Gegenteil, deren Metaanalyse von 100 Studien ergab, dass sich Kinder zwischen fünf und zehn Jahren nicht beim Lösen komplexer mathematischer Probleme unterschieden, der Effekt lag bei $d = 0.00$ (Hyde et al., 1990, S. 148). Moderate Geschlechterunterschiede im Problemlösen wurden erst bei 15–18-Jährigen bzw. 19–25-Jährigen festgestellt, die Effektgrößen lagen bei $d = 0.29$ bzw. $d = 0.32$ (Hyde et al., 1990, S. 148). Von Interesse war, herauszufinden, ob das Geschlecht einen Einfluss auf die Vorhersage des Lösungserfolgs und der Problemlösekompetenzen nimmt.

4.6 Versuchsdurchführung

Die Untersuchung gliederte sich in fünf Teile, die sich insgesamt über vier Messzeitpunkte von Mitte Juni 2013 bis Ende März 2014 erstreckten (siehe Tabelle 4.19). Alle Erhebungen fanden sowohl in den Experimentalgruppen als auch in der Kontrollgruppe statt. Die Gruppen unterschieden sich ausschließlich in der Interventionsmaßnahme (vgl. Abschnitt 4.3).

Um die Testbelastung für die teilnehmenden Grundschulkinder zu verteilen und den Interventionsstart direkt zu Schuljahresbeginn zu ermöglichen, fand Teil 1, die Erhebung der Störvariablen Leseverständnis, mathematische Basiskompetenzen und Intelligenz, bereits Ende der zweiten Jahrgangsstufe (MZP 0) statt. Diese Erhebung glich einer Bestandsanalyse vor den Sommerferien. Die zeitliche Verlagerung hatte den Vorteil, dass die standardisierten Tests in den Sommerferien ausgewertet und die Klassen vor Schuljahresbeginn auf Basis einer Parallelisierung den experimentellen Bedingungen zugewiesen werden konnten. Ursprünglich war jedoch vorgesehen, dass auch der vierte Prätest, der Textaufgaben-Performance-Test, am Ende des zweiten Schuljah-

res durchgeführt werden sollte. Da sich aber der vierte geplante Testtermin vor den Sommerferien nicht mehr in allen Schulen realisieren ließ, wurde der Performance-Test aus Gründen der Vergleichbarkeit in allen Klassen erst im August, der ersten Unterrichtswoche des dritten Schuljahres, erhoben. Um den geplanten Versuchsablauf nicht zu gefährden, erfolgte das Parallelisierungsverfahren auf den drei Störvariablen zum Messzeitpunkt 0.

Tabelle 4.19. Ablauf der Studie mit den einzelnen Erhebungszeitpunkten

Testinstrumente	2013				2014
	Juni	August	bis	Dezember	März
MZP 0	MZP 1	Intervention	MZP 2	MZP 3	
ELFE 1–6	x				
HRT 1–4	x				
CPM	x				
Textaufgaben–Performance–Test		x		x	x
Schülerfragebogen I		x		x	
Schülerfragebogen II		x		x	x
Lehrerfragebogen		x			

Anmerkung. MZP = Messzeitpunkt; Testinstrumente: ELFE 1–6 = Erfassung der sprachlichen Fähigkeiten und des Leseverständnisses; HRT 1–4 = Erfassung der mathematischen Basiskompetenzen; CPM = Erfassung der Intelligenz; Schülerfragebogen I = prospektive und retrospektive Erfassung der metakognitiven Fähigkeiten (Vorhersage, Planen, Kontrollieren, Evaluieren) in Bezug auf das allgemeine Lösen von Textaufgaben; Schülerfragebogen II = prospektive und retrospektive Erfassung der metakognitiven Fähigkeiten (Vorhersage, Evaluieren) in Bezug auf den Problemlöse-Performance-Test; Lehrerfragebogen = Erfassung der metakognitiven Fähigkeiten der Lernenden.

Die Lehrkräfte wurden in der ersten Teamsitzung bzw. Gesamtlehrerkonferenz des Schuljahres über die Gruppenzuweisung informiert, sodass bereits in der zweiten Unterrichtswoche in allen Klassen die Untersuchung startete. Es war zu erwarten, dass die Implementierung der Problemlösestunde alle Probanden gleichermaßen motivieren würde, da sie etwas Neuartiges darstellte, was sie aus der zweiten Jahrgangsstufe vermutlich noch nicht kannten. Das Problemlösen nahm keine Sonderrolle ein, vielmehr wurde es für alle Drittklässler zum festen Bestandteil des Mathematikunterrichts. Alle

Kinder waren im Gegensatz zu einer Wartegruppe, die keine problemhaltigen Textaufgaben löst, der gleichen Anstrengung und Herausforderung unter gleichen Rahmenbedingungen ausgesetzt (Hasselhorn & Hager, 2006, S. 344). Somit kann dem Einwand begegnet werden, dass die Kinder der nicht-trainierten Klassen einer weniger sinnvollen Lernumgebung ausgesetzt gewesen wären. Aus organisatorischen Gründen konnte die Versuchsleitung nicht alle Interventionsmaßnahmen durchführen, weshalb personengebundene Störvariablen nicht zu verhindern waren.

In der ersten Problemlösestunde (MZP 1) wurden die Kinder im Prätest aufgefordert, drei problemhaltige Textaufgaben zu bearbeiten. Das Ziel des Prätests war es, das Ausgangsniveau der Drittklässler zu erfassen, um mögliche Leistungsveränderungen im Posttest auf das Training zurückführen zu können. Auch wenn der Prätest nicht mehr in das Parallelisierungsverfahren eingehen konnte, wurde angestrebt, die Vergleichbarkeit der Treatmentbedingungen über den Prätest zu kontrollieren. In der Folgestunde des gleichen Tages wandten sich die teilnehmenden Lernenden und ihre Lehrkraft der Beantwortung eines Fragebogens zu. Der Schülerfragebogen sollte Einblicke in die individuellen Einstellungen, Selbsteinschätzung und metakognitiven Fähigkeiten der Lerner geben. Das Bestreben war es, aufzudecken, wie gut Grundschulkinder ihre Fähigkeiten und Fertigkeiten in der neuen Domäne Problemlösen einschätzen können. Demgegenüber waren die Lehrkräfte aufgefordert, die Problemlösekompetenzen ihrer Schülerinnen und Schüler einzuschätzen. Eine Woche später startete in allen Gruppen die zwölfwöchige Intervention, der Kern der Untersuchung. Das Unterrichtsarrangement variierte je nach Treatmentbedingung. Die Rahmenstruktur war für alle teilnehmenden Schüler gleich.

Nachdem zwölf unterschiedliche Trainingsaufgaben (vgl. Abschnitt 4.2.2.2) im Mathematikunterricht bearbeitet worden waren, rundete der Posttest die Intervention ab (MZP 2). Sprich: Die Lernenden waren 13 Wochen nach dem Prätest erneut aufgefordert, drei problemhaltige Textaufgaben zu lösen. Dieser Posttest sollte aufklären, ob ein Wissenstransfer erfolgte und inwiefern sich die Gruppen diesbezüglich unterschieden. In einem zweiten Posttest, dem Follow-up-Test, galt es dann, nach weiteren 13 Wochen, von Neuem drei strukturgleiche problemhaltige Textaufgaben zu bearbeiten (MZP 3). Zwischen dem ersten (MZP 2) und dem zweiten Posttest (MZP 3) erfolgte keine Intervention. Die Kinder absolvierten in diesen 13 Wochen ihren regulären Mathematikunterricht, sodass der zweite Posttest Hinweise auf nachhaltiges Lernen geben kann.

Die Testungen zu den vier Messzeitpunkten wurden von der Autorin und zwei Hilfskräften mit Unterrichtserfahrungen durchgeführt. Den Problemlöseunterricht der nicht-trainierten Klassen übernahmen die regulären Mathematiklehrkräfte. Das Training und

somit den Problemlöseunterricht der trainierten Klassen gestaltete die Autorin. Für eine ausführliche Darstellung des Trainings wird auf Abschnitt 4.4 verwiesen.

4.7 Einordnung der Versuchsteilnehmer

An der Intervention nahmen 20 dritte Klassen ($N = 366$) aus acht rheinland-pfälzischen Grundschulen teil. Die Rekrutierung der Grundschulen erfolgte im Einzugsgebiet Landau, Rhein-Pfalz-Kreis und Ludwigshafen.

Die 366 Schülerinnen und Schüler setzten sich aus 175 Mädchen (48 %) und 191 Jungen (52 %) zusammen. Das Alter der Drittklässler lag zwischen 7 und 10 Jahren ($M = 8.10$, $SD = 0.47$) mit Mittelwerten von $M = 8.07$ ($SD = 0.38$) bei den Mädchen und $M = 8.12$ ($SD = 0.54$) bei den Jungen.

Um die Effektivität der unterschiedlichen Interventionsmaßnahmen (vgl. Abschnitt 4.3) zur Förderung der Problemlösekompetenzen zu untersuchen, mussten die teilnehmenden dritten Klassen auf die vier experimentellen Bedingungen verteilt werden. Hierbei wurden die potenziellen interindividuellen Unterschiede kontrolliert und mögliche Konfundierungen ausgeschlossen, indem die Klassen den einzelnen experimentellen Bedingungen durch eine Parallelisierung auf Klassenebene zugewiesen wurden (Breaugh & Arnold, 2007; Myers & Hansen, 2012, S. 278–281). Damit sollte verhindert werden, dass sich die zu vergleichenden Gruppen signifikant in ihren Leistungsvoraussetzungen unterschieden (Stuart & Rubin, 2008, S. 155). Das Matching-Verfahren erfolgte vor dem eigentlichen Interventionsbeginn. Grundlage waren, wie in Abschnitt 4.5.3 aufgezeigt, die am Ende der zweiten Jahrgangsstufe, zum Messzeitpunkt 0, erhobenen personenbezogenen Störvariablen *Intelligenz* (siehe Abschnitt 4.5.3.1), *mathematische Basiskompetenzen* (siehe Abschnitt 4.5.3.2) und *sprachliche Fähigkeiten und Leseverständnis* (siehe Abschnitt 4.5.3.3). Diese drei Matching-Variablen wurden berücksichtigt, da sie nicht nur aus theoretischer, sondern auch aus inferenzstatistischer Sicht das Lösen problemhaltiger Textaufgaben beeinflussten (vgl. Moore, 2012, S. 462; Stuart & Rubin, 2008, S. 162).²⁰⁷ Auf Basis der Individualdaten wurden jeder Klasse ein Leseverständnisscore, ein mathematischer Basiskompetenzscore und ein Intelligenzscore zugeordnet, welche jeweils die Mittelwerte der Klasse repräsentierten. Mit dem Ziel, vergleichbare, experimentelle Bedingungen zu erzeugen, die sich in der mittleren Ausprägung der gemessenen Merkmale und im Ausmaß der Unterschiede zwischen den Klassen in den verschiedenen Bedingungen nicht unterscheiden, erfolgte die Parallelisierung auf Basis der drei Klassenmittelwerte (Eid et al., 2011, S. 59; Moore, 2012, S. 461, 2015, S. 151–152; Stuart & Rubin, 2008, S.

²⁰⁷ Die inferenzstatistisch gewonnenen Ergebnisse werden in Abschnitt 5.1.1 im Rahmen der Voranalysen zur Vergleichbarkeit der Stichprobe aufgezeigt.

156). Da der Parallelisierung intervallskalierte Kovariate zugrunde lagen, existierten keine identischen Klassen (Moore & Moore, 2013, S. 511). Die Ähnlichkeit der Klassen wurde mithilfe der Mahalanobis-Distanz bestimmt (Moore & Moore, 2013, S. 511). Sie ist der Propensity-Score-Distanz vorzuziehen, wenn weniger als fünf Kovariaten vorliegen (Stuart & Rubin, 2008, S. 162). Die Klassen wurden mit dem R package *blockTools*²⁰⁸ in homogenen Blocks (Vierlingen) gruppiert und dann randomisiert den vier experimentellen Bedingungen (Unit 1–4) zugewiesen (Greevy, Lu, Silber, & Rosenbaum, 2004, S. 264–265; Moore, 2012, S. 465–467). Bei diesem Verfahren wurden Vierlinge gebildet, welche sich bezüglich der Störvariablen möglichst ähnlich sind. Die Zuweisung der Klassen in Units und deren maximale Distanz kann Tabelle 4.20 entnommen werden. Dabei ließ sich nicht verhindern, dass sich mit zunehmender Blockbildung die maximale Distanz erhöhte und die Klassen innerhalb eines Blocks heterogener wurden.

Tabelle 4.20. Blockbildung der 20 Klassen in Units durch das Parallelisierungsverfahren auf Basis der Intelligenz, der mathematischen Basiskompetenzen und des Leseverständnisses

	Unit 1	Unit 2	Unit 3	Unit 4	Maximale Distanz
Block 1	1.1	2.1	3.1	4.1	1.586703
Block 2	1.2	2.2	3.2	4.2	1.607222
Block 3	1.3	2.3	3.3	4.3	1.720024
Block 4	1.4	2.4	3.4	4.4	2.782594
Block 5	1.5	2.5	3.5	4.5	3.711933

Anmerkung. Die Abkürzungen für die Schulen wurden anonymisiert und entsprechend den Units und der Zeilenbelegung durchnummeriert.

Die Versuchsleiterin trainierte die Unit 1 und 3. Diese Festlegung basierte auf schulischen Rahmenbedingungen. Vor Interventionsbeginn wurde bekannt, dass sich während der Maßnahme in den Klassen 1.2 und 3.1 ein Wechsel der Lehrkraft vollziehen wird. Ein Zeitvertrag und der Antritt des Mutterschutzes verwehrten den beiden Lehrerinnen die vollständige Durchführung des Experiments. Um situations- und bedingungsgebundene Störvariablen, die ein Lehrerwechsel mit sich bringen könnte, zu kontrollieren, wurde festgelegt, dass die Gruppen, in denen die Klassen 1.2 und 3.1 zu

²⁰⁸ Das R package *blockTools* (Version 0.6–2) geht auf Moore und Schnakenberg (2015) zurück.

verorten waren, zu den Trainingsklassen zählten. Die Zuweisung der Units zu den Bedingungen mit bzw. ohne kommunikative Settings erfolgte randomisiert.

4.7.1 Die Trainingsgruppen

Insgesamt nahmen 10 Klassen eine Stunde pro Woche an dem Repräsentationstraining teil. Die Geschlechterverteilung war fast ausgeglichen. Unter den 191 Kindern waren 96 Mädchen (50 %) und 95 Jungen (50 %). Weiterhin wurden die 10 trainierten Klassen halbiert, um die kommunikativen Zweiersettings zu berücksichtigen. Tabelle 4.21 können die beiden Trainingsbedingungen entnommen werden.

Tabelle 4.21. Übersicht über die Verteilung der Probanden in den Trainingsgruppen T+K+ und T+K-

	T+K+	T+K-
<i>N</i> (gesamt)	95	96
Mädchen	48	48
Jungen	47	48
Durchschnittsalter	8.09 (<i>SD</i> = 0.41)	8.04 (<i>SD</i> = 0.44)

Anmerkung. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings.

4.7.2 Die Nicht-Trainingsgruppen

10 der 20 teilnehmenden Klassen erhielten kein Repräsentationstraining. Stattdessen bearbeiteten die 175 Kinder, darunter 79 Mädchen (45 %) und 96 Jungen (55 %), die problemhaften Textaufgaben mit ihren regulären Lehrkräften im Mathematikunterricht (vgl. Abschnitt 4.3). Die Integration kommunikativer Zweiersettings führte auch in den nicht-trainierten Klassen als zusätzliche experimentelle Bedingung zur Splitting der Klassenzahl (siehe Tabelle 4.22).

Tabelle 4.22. Übersicht über die Verteilung der Probanden in den Nicht-Trainingsgruppen T-K+ und T-K-

	T-K+	T-K-
<i>N</i> (gesamt)	91	84
Mädchen	43	36

Jungen	48	48
Durchschnittsalter	8.13 ($SD = 0.43$)	8.13 ($SD = 0.58$)

Anmerkung. T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

4.8 Statistische Methoden

Die jeweilig genutzte statistische Methode wird in der Reihenfolge der zu untersuchenden Fragestellungen aufgezeigt. Ergänzend wird sichergestellt, dass die Voraussetzungen für die Durchführung der jeweiligen Methode erfüllt sind. Diese Analysen werden im Ergebnisteil ausgespart, um den Fokus auf die Ergebnisse und nicht auf die Verfahren zu richten.

4.8.1 Multivariate Varianzanalyse bei der Testung der Gleichverteilung der Leistungsvoraussetzungen

Die Testung der Gleichverteilung der Leistungsvoraussetzungen (Intelligenz, mathematische Basiskompetenzen und Leseverständnis) für die vier experimentellen Bedingungen wurde mit einer multivariaten Varianzanalyse durchgeführt (siehe Abschnitt 5.1.1). Im Gegensatz zum Parallelisierungsverfahren (siehe Abschnitt 4.7) basierten die Mittelwertvergleiche der einzelnen Gruppen nicht auf den Klassenmittelwerten der einzelnen Variablen. Vielmehr blieb die Level-2-Ebene *Klasse* unberücksichtigt, sodass alle Probanden aus fünf Klassen die entsprechende Gruppe, T+K+, T+K-, T-K+ oder T-K-, repräsentierten. Eine multivariate Varianzanalyse wurde drei separaten univariaten Varianzanalysen vorgezogen, um den Fehler der 1. Art, d. h. wahre Nullhypotesen, zu verwerfen, zu reduzieren (Field, 2013, S. 624–625; Stevens, 2009, S. 146). Sie ermöglichte, unter Berücksichtigung der Korrelationen der drei abhängigen Variablen, aufzudecken, ob sich die Gruppen hinsichtlich einer Kombination von Dimensionen unterschieden (Stevens, 2009, S. 146). Gerade vor dem Hintergrund, gleiche Leistungsvoraussetzungen auf Basis mehrerer Kontrollvariablen zu erzeugen, war die Berücksichtigung einer Kombination von Dimensionen relevant. Eine multivariate Analyse ist bei multivariaten Hypothesen separaten univariaten Analysen vorzuziehen (Keselman et al., 1998, S. 361; Warne, 2014, S. 5; vgl. Tabachnick & Fidell, 2014, S. 357). Der Empfehlung von Tabachnick und Fidell (2014, S. 310), moderat korrelierende abhängige Variablen zu berücksichtigen ($r < |.6|$), wurde nachgekommen (alle $.23 \leq r \leq .46$).

Die Anwendung von Varianzanalysen als parametrische Verfahren ist mit konkreten Voraussetzungen an die vorhandenen Daten geknüpft: 1. Die abhängige Variable muss bzw. die abhängigen Variablen müssen intervallskaliert sein, 2. das beobachtete Merkmal muss in jeder Gruppe normalverteilt sein, 3. die Varianzen der Gruppen müssen gleich sein (Varianzhomogenität) und 4. die zu vergleichenden Messwerte müssen voneinander unabhängig sein (Field, 2013, S. 168–176, 442–446; B. Rasch, Friese, Hofmann, & Naumann, 2014, S. 30–31; Stevens, 2007, S. 56–62; Tabachnick & Fidell, 2014, S. 111–125). Multivariate Varianzanalysen setzen zusätzlich eine multivariate Normalverteilung (zu 2.) und Homogenität gruppenspezifischer Varianz-Kovarianz-Matrizen der abhängigen Variablen voraus (zu 3.) (Field, 2013, S. 642; Tabachnick & Fidell, 2007, S. 86–90). Die 1., 3. und 4. Voraussetzung war erfüllt. Multivariate Normalverteilung konnte angenommen werden, weil zum einen pro Experimentalbedingung über 20 Probanden und zum anderen mit drei abhängigen Variablen eine geringe Anzahl abhängiger Variablen vorhanden waren (Tabachnick & Fidell, 2014, S. 293). Die Varianzanalyse gilt dann als robust gegenüber einer Verletzung der Normalverteilung.

Weil Mittelwertunterschiede eines Messzeitpunktes statistisch analysiert wurden, galt die Voraussetzung der Unabhängigkeit als erfüllt (B. Rasch et al., 2014, S. 31; Tabachnick & Fidell, 2007, S. 88). In diesem Fall wurde jede Klasse exakt einer konkreten Bedingung zugeordnet und deren Intelligenz, mathematische Basiskompetenzen und Leseverständnis wurden zum MZP 0 verglichen.

4.8.2 *Univariate Varianzanalyse bei der Testung der Gleichverteilung der Prätest-Leistungen*

Mit dem Fokus, herauszufinden, ob sich die vier experimentellen Bedingungen in ihren Prätest-Leistungen unterscheiden, wurden im Gegensatz zu der Testung auf Gleichverteilung der Leistungsvoraussetzungen die drei zum MZP 1 erhobenen abhängigen Variablen separaten univariaten Varianzanalysen unterzogen (siehe Abschnitt 5.1.2). Es wurde sich darauf gestützt, dass bei Forschungsfragen, bei denen univariate Effekte von Interesse sind, auch eine univariate Auswertung zu fokussieren ist (Keselman et al., 1998, S. 361–362). Da der vorliegenden Untersuchung hinsichtlich des *Erfolgs*, der *Problemlösekompetenzen* und der *Anzahl selbstgenerierter externer Repräsentationen* keine multivariaten Hypothesen zugrunde liegen, wurde dieser Empfehlung gefolgt. Der *Lösungserfolg* und die *Problemlösekompetenzen* bilden zwar zwei Facetten des gleichen Konstrukt ab, sollen aber weder zusammen noch ergänzend interpretiert werden. Vielmehr sollen sie vergleichend herangezogen werden. Das harte Kriterium (richtige oder falsche Lösung), das mit der abhängigen Variablen *Lösungserfolg* gemessen wird, ignoriert vollständig den absolvierten Lösungsprozess so-

wie die erzielten Teilleistungen der Kinder. Von der Erfassung des Prozesses wird sich versprochen, Teilerfolge oder auch kleine Effekte der Interventionsmaßnahmen identifizieren zu können, die der reine Lösungserfolg nicht zu erfassen ermöglicht (siehe Abschnitt 4.5.1.1).

Ferner unterstützen Tabachnick und Fidell (2014, S. 374) die Berichterstattung univariater Ergebnisse, sofern die Schlussfolgerungen beider Verfahren übereinstimmen. Um abzusichern, dass dies zutrifft, wurden sowohl univariate als auch multivariate Analysen durchgeführt, wenn auch im Ergebnisteil lediglich der Empfehlung, univariater Ergebnisse zu berichten, entsprochen wurde. Im Anhang D.1 werden die multivariaten Ergebnisse aufgezeigt, um die Übereinstimmung zu belegen.

Im Hinblick auf die Voraussetzungen der Varianzanalyse bleibt im vorliegenden Analyseschritt festzuhalten, dass alle abhängigen Variablen und Kontrollvariablen im Intervallskalenniveau vorlagen (Voraussetzung 1).²⁰⁹ Die abhängige Variable *Lösungserfolg* wurde mit dem Aufgabenscore zwar dichotom (richtig/falsch) erfasst, ging aber nur als aggregierter Testscore und demnach als Mittelwert über alle Aufgaben in die Varianzanalysen ein. Somit kann von Intervallskalenniveau ausgegangen werden. Aufgrund der Problemhaftigkeit der Testaufgaben war zu erwarten, dass diese vor der Intervention nur von sehr wenigen Grundschulkindern richtig gelöst werden (vgl. Abschnitt 1.4.2). Hinzu kommt, dass mangelnde Problemlösekompetenzen sichtbar und kaum externe Repräsentationen von sich aus generiert werden. Diese theoretisch fundierte Annahme bestätigte sich in den Daten und hatte linksschiefe Verteilungen zur Folge, die insbesondere bei den Prätestaufgaben erkennbar wurden. Die Varianzanalyse verhält sich dieser Verletzung gegenüber robust, insbesondere dann, wenn die zu vergleichenden Gruppengrößen nahezu identisch sind (Field, 2013, S. 444; B. Rasch et al., 2014, S. 31; Tabachnick & Fidell, 2007, S. 87).

Eine Verletzung der Normalverteilungsannahme geht häufig mit einer Verletzung der Homogenitätsannahme einher, die statistischen Tests gelten diesbezüglich als sensitiv (Stevens, 2007, S. 58, 2009, S. 230). Es bleibt auch hier festzuhalten, dass der Stichprobenumfang und die Stichprobenverteilung auf die experimentellen Bedingungen die Varianzhomogenität beeinflussen (B. Rasch et al., 2014, S. 31; Stevens, 2007, S. 57): Ist das Verhältnis der größten und der kleinsten Gruppengröße kleiner als 1.5, so können die Gruppen als gleich groß behandelt, die Varianzheterogenität vernachlässigt und die F-Statistik als robust gegenüber ungleichen Varianzen angesehen werden (Stevens, 2007, S. 57–58). In der vorliegenden Stichprobe wurde dieser kritische Wert von 1.5 nie überschritten, weshalb die Verletzung der Homogenitätsannahme unbe-

²⁰⁹ Hier wird Bezug genommen auf die in der vorherigen Diskussion aufgezeigten Voraussetzungen (siehe Diskussion der multivariaten Varianzanalyse bei der Testung der Gleichverteilung der Leistungsvoraussetzungen).

rücksichtigt blieb. Die Voraussetzung der Unabhängigkeit galten sowohl beim Vergleich der Prätest-Leistungen zum MZP 1 als auch beim Vergleich der Leistungsvo-raussetzungen (B. Rasch et al., 2014, S. 31; Tabachnick & Fidell, 2007, S. 88). Alle Grundvoraussetzungen, um Varianzanalysen anwenden zu können, waren demnach erfüllt.

4.8.3 Testung der Verletzung der UnabhängIGkeitsannahme

Der vorliegenden Untersuchung liegt eine hierarchische Datenstruktur zugrunde, weil die Stichprobenziehung auf mehreren Ebenen erfolgte (Eid et al., 2011, S. 700). Aus der Menge aller Schulen wurden im ersten Schritt einige Schulen ausgewählt, darauf folgend wurde eine Auswahl von dritten Klassen aus den zur Verfügung stehenden Klassen dieser Schulen getroffen. Die Stichprobenziehung erfolgte daher auf zwei Levels (Ebene 3: Schulen, Ebene 2: Klassen). Die Lernenden, die in Klassen und somit in Level-2-Einheiten verschachtelt sind, bilden die Ebene 1. Bei längsschnittlich erhobenen Daten handelte es sich wiederum um eine besondere Art der Verschachtelung (Snijders & Bosker, 1999, S. 166): Die Messzeitpunkte gelten als Level-1-Einheiten, die Lernenden als Level-2-Einheiten und die experimentellen Bedingungen (Gruppen) als Level-3-Einheiten.

Aus mehrebenenanalytischer Sicht wird ein ökologischer Fehlschluss begangen, wenn „man einen Zusammenhang bzw. einen Effekt, der auf der Ebene der Gruppe (Level-2-Einheiten) gefunden wurde, fälschlicherweise auf der Ebene von Individuen (Level-1-Einheiten) interpretiert“ (Eid et al., 2011, S. 701–702). Da in der vorliegenden Untersuchung die gefundenen Effekte weder auf Individual- noch auf Klassenebene interpretiert werden, sondern lediglich auf Ebene der experimentellen Bedingungen, wird das Risiko falscher Schlussfolgerungen beim Interpretieren von Zusammenhängen minimiert. Aufgrund der Tatsache, dass die Datengrundlage keine Mehrebenenanalysen zuließ, musste auf Varianzanalysen zurückgegriffen werden. Um Aussagen über die Verletzung der UnabhängIGkeitsannahme und demzufolge den Einfluss der Klasse, der Lernatmosphäre und auch des Lehrers, treffen zu können, wurde die Intra-klassenkorrelation (ICC) berechnet (siehe Abschnitt 5.1.3).

4.8.4 Univariate Varianzanalysen mit Messwiederholung bei der Testung der Merkmalsausprägungen zu verschiedenen Messzeitpunkten

Um die zeitliche Veränderung der Merkmale *Lösungserfolg*, *Problemlösekompetenzen* und *Anzahl generierter Repräsentationen* erfassen zu können, wurden diese mehrmals an denselben Probanden gemessen. Es liegen Merkmalsausprägungen vor, die direkt vor der Intervention, unmittelbar nach der Intervention und zusätzlich zu einem späte-

ren Zeitpunkt erhoben wurden (Prä-, Post- und Follow-up-Testung). Die messwiederholten Daten sind demnach voneinander abhängig, was bei der Wahl des inferenzstatistischen Instruments berücksichtigt wurde. Es wurde die Varianzanalyse mit Messwiederholung gewählt, da sie gegenüber der Varianzanalyse ohne Messwiederholung den Vorteil hat, beim Vergleich beliebig vieler Messzeitpunkte die vorhandene Abhängigkeit der Daten zu berücksichtigen und einer Alpha-Fehler-Kumulierung entgegenzuwirken (Hair et al., 2009, S. 386, 407; B. Rasch et al., 2014, S. 66).

An Varianzanalysen mit und ohne Messwiederholung sind ähnliche Voraussetzungen geknüpft. Sie stimmen in den ersten drei überein: „Intervallskaliertheit der Daten, Normalverteilung des Merkmals sowie Homogenität der Varianzen in den Stufen des Faktors bzw. der Bedingungskomponenten mehrerer Faktoren“ (B. Rasch et al., 2014, S. 71). Anstelle der Unabhängigkeit der zu vergleichenden Messwerte tritt in Analysen mit messwiederholten Daten die Voraussetzung, „dass alle Korrelationen zwischen den einzelnen Stufen des messwiederholten Faktors homogen sein müssen“ (B. Rasch et al., 2014, S. 143). Deren Prüfung ist vor dem Hintergrund essenziell, dass zwischen den einzelnen Messungen heterogene Korrelationen zu erwarten sind, wobei zwei unmittelbar benachbarte Messungen höher miteinander korrelieren als weiter auseinanderliegende (Keselman et al., 1998, S. 370). In Situationen der Verletzung dieser als Sphärizität bezeichneten Voraussetzung wurden die Freiheitsgrade mithilfe des Korrekturverfahrens nach Huynh und Feldt (wenn $\epsilon > .75$) adjustiert (Field, 2013, S. 548; B. Rasch et al., 2014, S. 74). Zum anderen wurde in den folgenden Post-hoc-Analysen die Bonferroni-Korrektur der Tukey-HSD vorgezogen, da sie bei Verletzungen der Sphärizität als robust gilt (Field, 2013, S. 547). Die konform gehenden Ergebnisse der multivariaten Analysen werden im Anhang D.2 aufgezeigt.

4.8.5 Aufgabenspezifische Analyse der Effekte auf den Lösungserfolg

Wird auf aufgabenspezifischer Ebene überprüft, ob das Training, die kommunikativen Zweiersettings oder eine Kombination beider Faktoren einen Effekt auf den Lösungserfolg ausübt, so muss berücksichtigt werden, dass mit dem Lösungserfolg eine dichotome Variable vorliegt. Dies hat zur Konsequenz, dass die Voraussetzungen für ein allgemeines lineares Modell bereits mit der Voraussetzung des metrischen Skalenniveaus nicht erfüllt war (Baltes-Götz, 2016, S. 4). Aufgrund dessen wurde auf eine generalisierte Schätzgleichung zurückgegriffen, die sich insbesondere für messwiederholte oder geclusterte Daten eignet und keine Anforderungen an eine Normalverteilung stellt (Ballinger, 2004, S. 128; Homish, Edwards, Eiden, & Leonard, 2010, S. 559–560; Liang & Zeger, 1986, S. 13). Die Voraussetzungen für eine GEE-Analyse sind vergleichsweise gering (Baltes-Götz, 2016, S. 33): Durch das Messwiederholungsdesign gilt die Annahme unabhängiger Beobachtungen als erfüllt. Zudem ist die

Stichprobengröße ausreichend groß (> 100) und die Anzahl der Beobachtungen pro Lernender mit drei als hinreichend klein einzustufen. Bezuglich der fehlenden Werte war die MCAR-Bedingung erfüllt. Folglich traf eine Verletzung der Bedingung „vollständig zufällig fehlend“ nicht zu (Wirtz, 2004, S. 112): $\chi^2 (43) = 34.38$, $p = .823$.

4.8.6 Analyse des Effekts selbstgenerierter externer Repräsentationen auf den Lösungserfolg bzw. die Problemlösekompetenzen

Nachfolgend soll exemplarisch am Kriterium Lösungserfolg gezeigt werden, dass die Voraussetzungen für die regressionsanalytischen Analyseverfahren gegeben waren. Es wird versichert, dass die Voraussetzungen auch beim Kriterium Problemlösekompetenzen erfüllt waren. Von einem analogen Bericht wird aus ökonomischen Gründen abgesehen.

Handling der potenziellen Kontrollvariablen und deren Bedeutung bei der Vorhersage des Lösungserfolgs

Die Prädiktoren, die Einfluss auf die abhängige Variable nehmen könnten, wurden zunächst nach theoretischen Gesichtspunkten ausgewählt (Cohen, Cohen, West, & Aiken, 2003, S. 144; Eid et al., 2011, S. 629; Stevens, 2009, S. 75, siehe Kapitel 4.5.2). Darüber hinaus waren deren Korrelationen mit der abhängigen Variablen sowie mit den restlichen unabhängigen Variablen ausschlaggebend (Stevens, 2009, S. 73; Tabachnick & Fidell, 2014, S. 180). Erstgenannte sollten möglichst hoch, letztgenannte möglichst klein ausfallen. Demzufolge wurde erwartet, dass ein Teil der erklärten Varianz auf die experimentelle Bedingung (Gruppe) und die Prädiktoren Geschlecht, Mehrsprachigkeit, Intelligenz, mathematische, sprachliche und metakognitive Fähigkeiten sowie Attitudes und Beliefs zurückgeführt werden kann (vgl. Seidel & Shavelson, 2007, S. 456). Ferner wurde die Prätest-Leistung (Lösungserfolg bzw. Problemlösekompetenzen) als möglicher Einflussfaktor einbezogen.

Mit dem Ziel, vorhersagen zu wollen, welcher Anteil des Lösungserfolgs auf die Repräsentationen zurückgeführt werden kann, mussten die potenziellen Einflussfaktoren, die mit dem Lösungserfolg in Zusammenhang stehen, kontrolliert werden (Bortz & Schuster, 2010, S. 342).

Es wurde sich dafür ausgesprochen, alle potenziellen Störvariablen in der Modellspezifikation zu berücksichtigen, um deren Einfluss zu kontrollieren, auch wenn dies nicht der Empfehlung, wenige Regressoren in das Modell aufzunehmen, entsprach (Ramsey & Schafer, 2013, S. 345). Aus diesem Grund wurde mit den Selektionsprinzipien der Rückwärts- und Vorwärtsmethode ergänzend untersucht, welche Prädiktoren einen eigenständigen Einfluss auf das Kriterium ausübten bzw. welche Regressoren zu selektieren waren. Beide Methoden ergaben übereinstimmend, dass unter

allen Prädiktoren die sprachlichen Fähigkeiten, die Mehrsprachigkeit und das Geschlecht keinen signifikanten Beitrag zur Varianzaufklärung leisteten (siehe Modell 1, Tabelle 4.23).

Alle neun Störvariablen sagten zusammen 61 % des Lösungserfolgs vorher (siehe Tabelle 4.23). Nach dem Ausschluss der drei Variablen (sprachliche Fähigkeiten und Leseverständnis, Mehrsprachigkeit und Geschlecht) reduzierte sich der Anteil geringfügig auf 60 % (siehe Tabelle 4.23). Dies sprach für das sparsamere Modell, auch vor dem Hintergrund, dass Prädiktoren, die Teil des Regressionsmodells sind obwohl sie keinen Einfluss auf die abhängigen Variablen ausüben, die Koeffizienten verzerrn und die Präzision der Schätzung mindern können (Cohen et al., 2003, S. 185–186; Ramsey & Schafer, 2013, S. 345). Das sparsamere Modell mit sechs Störvariablen wäre demnach dem Modell mit neun Störvariablen vorzuziehen (Ramsey & Schafer, 2013, S. 296; Stevens, 2009, S. 81).

Tabelle 4.23. Einfluss der Störvariablen auf das Kriterium Lösungserfolg

Variable	Lösungserfolg			
	Modell 1		Modell 2	
	B	95 % CI	B	95 % CI
Konstante	-.23*	[-.42, -.04]	-.22*	[-.41, -.03]
Experimentelle Bedingung	.05***	[.03, .08]	.06***	[.03, .08]
Lösungserfolg (Prätest)	.21**	[.05, .36]	.21**	[.06, .37]
Intelligenz	.01**	[.002, .02]	.01**	[.003, .02]
Mathematische Fähigkeiten	<.01**	[.001, .003]	<.01**	[.001, .003]
Metakognitive Fähigkeiten	.20***	[.16, .23]	.20***	[.16, .23]
Attitudes und Beliefs	.02***	[.01, .03]	.02***	[.01, .03]
Sprachliche Fähigkeiten	<.01	[-.001, .002]		
Mehrsprachigkeit	.03	[-.03, .08]		
Geschlecht	.01	[-.04, .06]		
R^2	.61		.60	
R^2 korrigiert	.59		.60	
F	48.19***		72.01***	

Anmerkung. N = 291. 95 % CI = Konfidenzintervall; B = Nicht standardisierter Regressionskoeffizient; R^2 = Determinationskoeffizient; R^2 korrigiert = korrigierter Determinationskoeffizient. Modell 1 inkludiert neun Störvariablen (uneingeschränktes Modell), Modell 2 sechs Störvariablen (eingeschränktes, sparsames Modell).

* $p < .05$. ** $p < .01$. *** $p < .001$.

Entgegen diesen Empfehlungen wurde sich für das vollständige Modell (Modell 1) ausgesprochen. Zum einen konnten dadurch die Störvariablen kontrolliert werden, zum anderen zeigte sich im direkten Modellvergleich, dass die Koeffizienten kaum voneinander abwichen und somit die Schätzung des angereicherten Modells trotzdem präzise scheint. Tabelle 4.23 zeigt den direkten Modellvergleich für das Kriterium *Lösungserfolg* auf, was analog auf die *Problemlösekompetenzen* übertragen werden kann. Ferner fällt auf, dass der korrigierte Determinationskoeffizient ($R^2_{\text{korrigiert}}$), der vorrangig bei einer hohen Anzahl von Prädiktoren zu beachten ist, vernachlässigbar gering vom Determinationskoeffizient abweicht.

Handlungen der großen Anzahl beobachteter Repräsentationsmerkmale

Vor dem Hintergrund, dass sich bei der Aufnahme aller 28 Repräsentationsmerkmale negative Regressionsgewichte ergaben, wurden Suppressionseffekte nicht ausgeschlossen. Der scheinbar negative Zusammenhang wurde als Anzeichen für negative Suppression gedeutet. Darüber hinaus wird jede Variable als Suppressorvariable definiert, welche die vorhergesagte Validität einer dritten Variablen oder einem Set von Variablen durch ihre Aufnahme in die Regressionsgleichung erhöht (Conger, 1974, S. 36; Velicer, 1978, S. 954). Unter Validität wird hier die Korrelation des Kriteriums mit der Prädiktorvariablen verstanden (MacKinnon, Krull, & Lockwood, 2000, S. 175). Problematisch ist dabei, dass die vorhergesagten Regressionskoeffizienten einzelner Prädiktoren in ihrer Größe schwankten, je nachdem, welche weiteren Prädiktoren in das Modell aufgenommen wurden. Teilweise ließen sich auch Vorzeichenwechsel verzeichnen, die durch die Hinzunahme weiterer unabhängiger Variablen hervorgerufen wurden (Bortz & Schuster, 2010, S. 342; Cohen et al., 2003, S. 144). Bleiben die Effekte einzelner Regressoren nahezu stabil, ungeachtet dessen, welche Modelle geschätzt werden, so gelten diese als robust und zuverlässig (Cohen et al., 2003, S. 144).

Dieses mehrfach zu beobachtende Phänomen der Suppression wird exemplarisch an den Tabellenprädiktoren aufgezeigt (siehe Tabelle 4.24). Es wäre ebenso denkbar gewesen, hierfür die Zeichnungs- oder Rechnungsmerkmale heranzuziehen. Da beide Kategorien gegenüber der Kategorie Tabelle über eine höhere Anzahl an Merkmalen verfügten, wurde sich aus ökonomischen Gründen für die Tabellenmerkmale entschieden.

Es zeigte sich, dass bei der Aufnahme der ersten beiden Merkmale (T1 und T2) in das Regressionsmodell der Regressionskoeffizient des ersten Merkmals im Vergleich zu dem Koeffizienten des zweiten Merkmals nahezu stabil blieb. Wurde im nächsten Schritt zusätzlich das fünfte Merkmal (T5) berücksichtigt, so fällt auf, dass sich β_1 drastisch reduzierte. Demgegenüber wurde ein Anstieg des ersten Koeffizienten ver-

zeichnet, wenn beispielsweise die ersten vier Tabellenmerkmale (T1, T2, T3 und T4) in das Modell aufgenommen wurden. Dies ging mit einem Vorzeichenwechsel der Koeffizienten β_2 , β_3 und β_4 einher.

Tabelle 4.24. Schrittweise multiple Regression anhand der Tabellenmerkmale zur Aufspürung von Suppressoreffekten

Prädiktoren	β_1	β_2	β_3	β_4	β_5	β_6	β_7	ΔR^2
T 1	.38***							.15
T 2		.36***						.13
T 3			.26***					.07
T 4				.12*				.02
T 5					.32***			.10
T 6						.27***		.08
T 7							.31***	.10
T1, T 2	.34*	.05						.15
T 1, T 3	.37***		.02					.15
T 1, T 4	.45***			-.13*				.16
T 1, T 5	.09				.32***			.15
T 1, T 6	.36***					.03		.15
T 1, T 7	.31***						.10	.15
T 1, T 2, T 3	.32*	.05	.02					.15
T 1, T 2, T 5	.09	.29*			.03			.15
T 1, T 2, T 6	.29†	.07				.04		.15
T 1, T 2, T 7	.19	.12					.12	.16
T 1, T 2, T 3, T 4	.57**	-.04	-.18*	-.09				.16
T 1, T 2, T 3, T 4, T 5	.58**	-.09	-.16	-.21**	.15†			.17
T 1, T 2, T 3, T 4, T 5, T 6	.53**	-.05	-.23*	-.26**	.15†	.12		.18
T 1, T 2, T 3, T 4, T 5, T 6, T 7	.47*	.00	-.24*	-.29**	.13†	.03	.18*	.19
T 2, T 3		.32***	.07					.13
T 2, T 4		.38***		-.05				.13
T 2, T 5		.13†			.27***			.14
T 2, T 6		.31***				.10†		.14
T 2, T 7		.28***					.17**	.15

Anmerkung. T1 = Nutzung: Trial and Error; T2 = Nutzung: Erarbeitung; T3 = systematisches Vorgehen; T4 = Abstraktionsgrad; T5 = Kontrolle; T6 = Konstanz eines Faktors; T7 = Dimensionen; β = standardisierten Regressionskoeffizienten; ΔR^2 = semipartialle Korrelation.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Auch wenn die Tabelle 4.24 lediglich einen Auszug einer möglichen schrittweisen Regression präsentiert, so wird hieran bereits deutlich, dass die Vorhersage des Lösungserfolgs von Suppressionseffekten beeinflusst ist. Neben der großen Anzahl an Prädiktoren wurden die hohen Korrelationen untereinander als Ursache hierfür gesehen.

Um im nächsten Schritt eine Selektion der Prädiktoren zu erwirken, wurden ebenso wie bei der Auswahl der Kontrollvariablen die Selektionsprinzipien der Rückwärts- und Vorwärtsmethode herangezogen (vgl. Field, 2013, S. 323). Beide Methoden identifizierten jedoch kontroverse Modelle, weshalb ihnen nicht gefolgt wurde.

Multiple Regressionsanalyse bei der Testung des bereinigten Effektes der Repräsentationen auf das Kriterium Lösungserfolg

Ob den Repräsentationen ein relativer Einfluss bei der Vorhersage des Lösungserfolgs zugesprochen werden kann, wurde mit einer multiplen Regressionsanalyse überprüft. Diese Analyse wurde gewählt, um den Einfluss potenzieller Störvariablen zu kontrollieren und den bereinigten Varianzanteil, der lediglich auf die Repräsentationen zurückgeht, zu bestimmen. Alle in das Modell aufgenommenen Variablen erfüllten die Voraussetzung des Variablentyps. Die abhängige Variable wurde als intervallskaliert behandelt, die unabhängigen Variablen als intervallskaliert oder kategorial (Field, 2013, S. 312).

Regressionsmodell mit 13 Prädiktoren (9 Kontrollvariablen, 4 Repräsentationsfaktoren): Unter der Berücksichtigung von 13 Prädiktoren erfüllte die Probandenzahl 291 die erforderliche Voraussetzung von 117 ($N \geq 104 + 13$), um die Güte einzelner Prädiktoren vorhersagen zu können (Tabachnick & Fidell, 2014, S. 159).

Die Analyse der Residuen mithilfe von Streudiagrammen zeigte, dass eine lineare Beziehung zwischen der abhängigen und den unabhängigen Variablen sowie Homoskedastizität angenommen werden durften. Eine Ausnahme stellte diesbezüglich die Störvariable Intelligenz dar. Es zeigte sich insgesamt ein positiver Zusammenhang zwischen der Intelligenz und dem Lösungserfolg zum Posttest, auch wenn dieser Zusammenhang weniger linear zu sein schien als bei den anderen Prädiktoren. Die trichterförmige Punktwolke der Residuen deutete auf Heteroskedastizität und eine Verletzung der Voraussetzung hin (Field, 2013, S. 348).

Nach Stevens (2009, S. 114) dürfen nicht mehr als 5 % der standardisierten Residuen $> |2|$ sein, was mit 4 % erfüllt war. In 96 % aller Fälle lagen die standardisierten Resi-

duen im Bereich von zwei Standardabweichungen, in 100 % aller Fälle innerhalb von drei Standardabweichungen. Ausreißer für das Set von Prädiktoren gab es keine. Der kritische Wert von $.19^{210}$ wurde in keinem der Fälle überschritten (Stevens, 2009, S. 104). Es konnten 2 multivariate Ausreißer bei den Prädiktoren identifiziert werden. Die Mahalanobis-Distanzen lagen in diesen beiden Fällen über dem kritischen χ^2 -Wert von 34.52 ($df = 13$, $\alpha = .001$) (Tabachnick & Fidell, 2014, S. 10, 164, 203). Die Cooks-Distanzen ($min = .00$, $max = .03$) zeigten jedoch, dass es sich bei beiden Fällen um keine einflussreichen Ausreißer > 1 handelte, weshalb sie nicht ausgeschlossen werden mussten.

Ferner wurde aufgrund der Durbin-Watson-Statistik ($d = 1.86$) davon ausgegangen, dass die Fehler voneinander unabhängig waren und keine Autokorrelation der Residualwerte bestand (Durbin & Watson, 1951). Für $N = 290^{211}$, $k = 13$ und $\alpha = .05$ ergab sich aus der Durbin-Watson-Tabelle die Untergrenze von $d_U = 1.72$ und $d_0 = 1.89$, in die die Prüfgröße fiel.²¹²

Laut dem Histogramm folgte die Verteilung der Residuen einer Normalverteilung (siehe Abbildung 4.4A). Das Histogramm ist annähernd symmetrisch und glockenförmig. Minimale Abweichungen zeigten sich um den Mittelwert. Das P-P-Verteilungsdiagramm, das die kumulierten Häufigkeiten der beobachteten Werte gegen die erwarteten kumulierten Häufigkeiten der Referenzverteilung abträgt, spiegelt ebenso diese zufriedenstellende Übereinstimmung der beobachteten und erwarteten Häufigkeiten wider (siehe Abbildung 4.4B). Die geringfügige Streuung um die Diagonale deutet auf eine minimale Abweichung von der Normalverteilung hin, die im akzeptablen Bereich liegt.

²¹⁰ Berechnungsgrundlage: $3p/n = 3(k+1)/n = 3(13+1)/291$ mit k als Anzahl der Prädiktoren (Stevens, 2009, S. 104).

²¹¹ Die Durbin-Watson-Tabelle sieht die exakte Stichprobengröße von $N = 291$ nicht vor, weshalb der Bereich für $N = 290$ herangezogen wurde.

²¹² Die Grenzwerte stammen von folgender Website: <http://web.stanford.edu/~clint/bench/dw05c.htm> (letzter Zugriff am 01.07.2016).

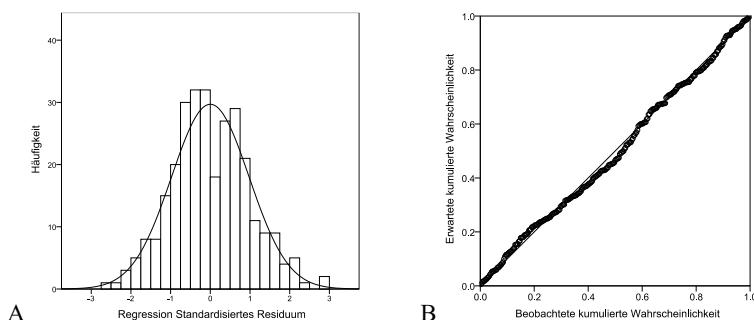


Abbildung 4.4. Residuenanalyse des Regressionsmodells mit 13 Prädiktoren (9 Kontrollvariablen und 4 Repräsentationsfaktoren). (A) Histogramm der standardisierten Residuen und (B) P-Plot (Probability-Probability-Plot) der studentisierten Residuen.

Ferner konnte keine Multikollinearität identifiziert werden. Demzufolge existierten weder hohe Korrelationen $> .80$ oder $.90$ zwischen zwei oder mehreren Prädiktoren noch Varianzinflationsfaktoren (VIF) größer als 10 oder kleiner als 0.2 (Field, 2013, S. 325). Der letztgenannte Faktor sagt voraus, ob ein starker linearer Zusammenhang eines Prädiktors mit einem oder mehreren Prädiktoren besteht.

Die Verletzung der Homoskedastizität bei dem Prädiktor Intelligenz schien sich nicht negativ auf die Modellschätzung auszuwirken (vgl. Field, 2013, S. 348–349). Er zählte zu den schwächeren Prädiktoren, hatte keine eigene Vorhersagekraft und wirkte sich kaum auf die Schätzung aus. Wurde die Intelligenz aus dem Modell ausgeschlossen, so verringerte sich die Vorhersagegüte unerheblich von $R^2 = .66$ auf $R^2 = .65$. Darüber hinaus blieben die Regressionskoeffizienten vom Ausschluss nahezu unberührt, weshalb der Prädiktor inkludiert blieb, um die Intelligenz kontrollieren zu können. Demnach wird angenommen, dass in der Summe alle Voraussetzungen, die nötig sind, um eine multiple Regression durchführen zu können, erfüllt waren. Dies hat wiederum zur Folge, dass das Modell sowohl akkurat für die Stichprobe als auch generalisierbar für die Population zu sein schien (Field, 2013, S. 350).

Regressionsmodell mit 37 Prädiktoren (9 Kontrollvariablen, 28 Repräsentationsmerkmale): Auch wenn alle 37 Prädiktoren in das Regressionsmodell aufgenommen wurden, erfüllte die Probandenzahl von 291 die erforderliche Voraussetzung von 141 ($N \geq 104 + 37$), um die Güte einzelner Prädiktoren vorhersagen zu können (Tabachnick & Fidell, 2014, S. 159).

Die Analyse der Residuen deutete auch hier darauf hin, dass eine lineare Beziehung zwischen der abhängigen Variablen und den unabhängigen Variablen angenommen werden durfte und in den meisten Fällen Homoskedastizität vorlag. Auffällig waren

neben der Intelligenz weitere Repräsentationsmerkmale, deren Verteilung trichterförmig verlief, was ein Indiz für Heteroskedastizität war.

Wenn auch hier die Bedingung berücksichtigt wird, dass nicht mehr als 5 % der standardisierten Residuen $> |2|$ sein dürfen (Stevens, 2009, S. 114), dann galt diese im vorliegenden Fall mit 3 % als erfüllt. Demzufolge lagen in 97 % aller Fälle die standardisierten Residuen im Bereich von zwei Standardabweichungen, in 100 % aller Fälle innerhalb von drei Standardabweichungen. Ausreißer für das Set von Prädiktoren gab es keine, weil der kritische Wert von .39²¹³ nie überschritten wurde (Stevens, 2009, S. 104). Multivariate Ausreißer auf Prädiktorebene wurden dagegen in 23 Fällen identifiziert. Deren Mahalanobis-Distanzen lagen in allen Fällen über dem kritischen χ^2 -Wert von 69.44²¹⁴ ($df = 37, \alpha = .001$) (Tabachnick & Fidell, 2014, S. 10, 203). Die Cooks-Distanzen ($min = .00, max = .05$) zeigten jedoch, dass es sich hierbei um keine einflussreichen Ausreißer > 1 handelte, weshalb sie nicht ausgeschlossen werden mussten.

Die Voraussetzung, dass die Residualwerte voneinander unabhängig sind, wurde auch in diesem Modell erfüllt. Wenngleich für diese hohe Anzahl an Prädiktoren keine kritische Unter- und Obergrenze der Durbin-Watson-Prüfgröße vorliegt, gilt ein Wert, der „nahe bei 2“ liegt, als Bestätigung dafür, dass die Residualwerte voneinander unabhängig sind (J. Janssen & Laatz, 2007, S. 434). Im vorliegenden Modell traf dies für die Prüfgröße $d = 1.92$ zu.

Dem Histogramm zufolge gleicht die Verteilung der Residuen einer Normalverteilung in idealer Weise (siehe Abbildung 4.5A). Sie ist sowohl symmetrisch als auch glockenförmig. Es zeigten sich kaum Abweichungen von der Normalverteilung. Aus dem P-P-Verteilungsdiagramm geht diese Übereinstimmung der beobachteten und erwarteten Häufigkeiten ebenso hervor (siehe Abbildung 4.5B). Die Streuung um die Diagonale fällt in diesem Modell deutlich geringer aus als in dem sparsamen Modell (vgl. Abbildung 4.4).

²¹³ Berechnungsgrundlage: $3p/n = 3(k+1)/n = 3(37+1)/291$ mit k als Anzahl der Prädiktoren (Stevens, 2009, S. 104).

²¹⁴ Die kritischen χ^2 -Werte konnten der Tabelle von Tabachnick und Fidell (2014, S. 10) nur für die Freiheitsgrade 30 ($\chi^2 = 59.70$) und 40 ($\chi^2 = 73.40$), nicht jedoch die Zwischenstufen abgelesen werden. Aufgrund dessen wurde zur Berechnung die Originalquelle herangezogen (Pearson & Hartley, 1958, S. 612).

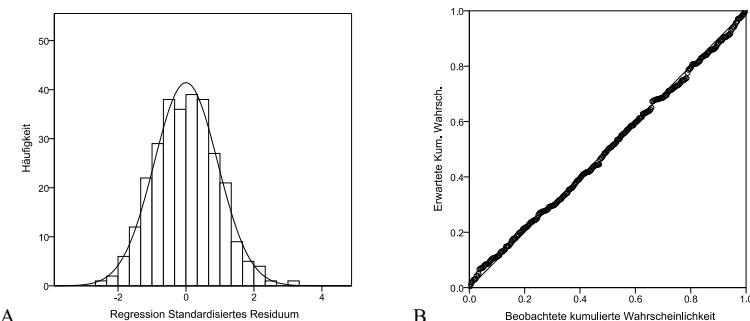


Abbildung 4.5. Residuenanalyse des Regressionsmodells mit 37 Prädiktoren (9 Kontrollvariablen und 28 Repräsentationsmerkmalen). (A) Histogramm der standardisierten Residuen und (B) P-P-Plot (Probability-Probability-Plot) der studentisierten Residuen.

Nachteilig war bei dieser Modellschätzung, dass Multikollinearitäts- und Suppressionsprobleme nicht ausgeschlossen werden konnten. Es ließen sich sowohl Prädiktoren, die sehr hoch miteinander korrelierten ($> .80$ oder $.90$), als auch Varianzinflationsfaktoren (VIF), die größer als 10 bzw. kleiner als 0.2 waren, identifizieren. Beides war ein Indiz für Multikollinearität, das sich durchaus als problematisch erwies (Field, 2013, S. 325; Hutcheson & Sofroniou, 1999, S. 82). Multikollinearität hat zur Folge, dass die Schätzung der Parameter unzuverlässig sein kann, was sich wiederum auf die Interpretation der Ergebnisse auswirken kann (Hutcheson & Sofroniou, 1999, S. 79). Geringfügige Änderungen in der Stichprobe können einen enormen Einfluss auf die Regressionsgleichung und somit auch auf die Interpretation der Ergebnisse ausüben. Da der Empfehlung, mehr Daten zu erheben, um der Multikollinearität entgegenzuwirken, nicht entsprochen werden konnte (Hutcheson & Sofroniou, 1999, S. 85), wurde sich für das Modell ausgesprochen, das die faktorenanalytisch gewonnenen Faktoren berücksichtigt. Hierdurch wurden zuverlässigere Schätzer erwartet.



5 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Datenanalyse präsentiert, die sich in Bezug auf die Fragestellungen der vorliegenden Arbeit ergeben haben. Die Ergebnisse werden für die fünf Forschungsfragen separat dargestellt und folgen den in Abschnitt 3.3.2 aufgeführten Hypothesen. Vorangestellt sind die Vor-Analysen, die erforderlich waren, um die Gruppen vergleichen zu dürfen (Abschnitt 5.1) sowie die Vor-Analysen zur Wirksamkeitshypothesenprüfung (Abschnitt 5.2). Die Datenanalyse erfolgte mithilfe von SPSS23 und R. Das allgemeine Signifikanzniveau wurde auf $\alpha = .05$ festgelegt. In Fällen der Verletzung der Sphärität wurden die Freiheitsgrade mithilfe des Korrekturverfahrens nach Huynh und Feldt (wenn $\varepsilon > .75$) adjustiert (Field, 2013, S. 548; B. Rasch et al., 2014, S. 74). Um einer Alpha-Fehler-Kumulierung entgegenzuwirken, wurde bei Mehrfachvergleichen immer eine Bonferroni-Adjustierung vorgenommen. Zur Quantifizierung der Effekte wird das partielle Eta-Quadrat η_p^2 berichtet.²¹⁵ Nach Cohen (1988, S. 285–287) wird von einem kleinen Effekt ausgegangen, wenn Werte um $\eta_p^2 = .01$ liegen, von einem mittleren Effekt, wenn die Werte $.06 < \eta_p^2 < .14$ liegen und von einem großen Effekt, wenn die Werte größer als $\eta_p^2 \geq .14$ sind.

Alle abhängigen Variablen gingen als aggregierte Mittelwerte von zwei unabhängigen Ratern in die Analyse ein. Um nachfolgend den Fokus auf die Ergebnisse zu richten, wurden die Begründungen für das jeweilige Analyseverfahren sowie die Prüfung der zu erfüllenden Voraussetzungen bereits im Methodenteil, Abschnitt 4.8, detailliert aufgezeigt, begründet und diskutiert.

Das Kapitel 5 ist in 3 Abschnitte gegliedert. Bevor im Einzelnen konkret auf die Forschungsfragen eingegangen wird, wird in Vor-Analysen die Vergleichbarkeit der Stichprobe geprüft (5.1). Diese Vor-Analysen waren vor dem Hintergrund, dass sie die erforderlichen Ausgangsvoraussetzungen absicherten, für die Interpretation der Ergebnisse essenziell. Die Ausgangsfragestellung, die in der Pilotierung untersucht wurde, steht in Abschnitt 5.2 im Vordergrund. Die Implementierung des Trainings basierte auf der Annahme, dass sich Grundschulkinder durch die Intervention animieren lassen, externe Repräsentationen zu generieren und aktiv in ihren Lösungsprozess zu integrieren. Dass diese Annahme in der Pilotierung bei sehr kleiner Stichprobe bestätigt werden konnte, wird in diesem Abschnitt aufgezeigt. Abschnitt 5.3 widmet sich der Beantwortung der vier Forschungsfragen.

²¹⁵ „ η^2 shows the proportion of variance in the DV (SS_{total}) attributable to the effect (SS_{effect})“ (Tabachnick & Fidell, 2014, S. 86).

5.1 Vor-Analysen zur Vergleichbarkeit der Stichprobe

Um die Vergleichbarkeit der Stichprobe sicherzustellen, wurde im Rahmen der Vor-Analysen untersucht, ob sich die vier experimentellen Bedingungen in ihren Leistungsvoraussetzungen unterschieden oder ob sie vergleichbar waren (Abschnitt 5.1.1). Als Maß für die individuellen Leistungsvoraussetzungen wurden die erzielten Testscores für die Intelligenz (CPM), die mathematischen Basiskompetenzen (HRT 1–4) und das Leseverständnis (ELFE 1–6) herangezogen, die zum MZP 0 vor Interventionsbeginn mit den in Abschnitt 4.5.3 vorgestellten Erhebungsinstrumenten erhoben wurden. Um deren Gleichverteilung zu testen, wurden die erzielten Klassenmittelwerte der Testscores gegenübergestellt und einer einfaktoriellen Varianzanalyse unterzogen. Das Parallelisierungsverfahren war erfolgreich, wenn sich die Gruppen zum Messzeitpunkt 1 nicht signifikant voneinander unterschieden. Die potenziellen Störvariablen sind als kontrolliert anzusehen.

Es schlossen sich Testungen der Gleichverteilung der Prätest-Leistungen (Abschnitt 5.1.2), der Verletzung der Unabhängigkeitssannahme (Abschnitt 5.1.3) wie auch der Einflüsse des Geschlechts und der Mehrsprachigkeit (Abschnitt 5.1.4), an.

5.1.1 Test der Gleichverteilung der Leistungsvoraussetzungen für die vier experimentellen Bedingungen

Das Parallelisierungsverfahren basierte auf den personengebundenen Störvariablen *Intelligenz*, *mathematische Basiskompetenzen* und *Leseverständnis*, weil inferenzstatistisch mit einer multivariaten Varianzanalyse gezeigt werden konnte, dass sich die 20 teilnehmenden Klassen signifikant in allen drei Matchingvariablen unterschieden: $V = .28$, $F(57, 963) = 1.74$, $p = .001$, $\eta_p^2 = .09$.²¹⁶ Die Ergebnisse der einfaktoriellen Varianzanalysen legten offen, dass in allen drei Leistungsvoraussetzungen signifikante Gruppenunterschiede existierten: für Intelligenz ($F(19, 331) = 1.68$, $p = .04$, $\eta_p^2 = .09$), für mathematische Basiskompetenzen ($F(19, 330) = 2.03$, $p = .01$, $\eta_p^2 = .10$) und für Leseverständnis ($F(19, 332) = 1.86$, $p = .02$, $\eta_p^2 = .10$). Die Effekte waren nach Cohen (1988, S. 287) von mittlerer Stärke ($.06 < \eta_p^2 < .14$).

Durch die Parallelisierung wiesen die personengebundenen Störvariablen in jeder experimentellen Bedingung die gleiche Verteilung auf. Wie erwartet, ist in den letzten Vierlingen eine heterogenere Verteilung der Störvariablen vorzufinden als in den restlichen Vierlingen. Es war keinesfalls verwunderlich, dass, wenn pro Blockbildung

²¹⁶ Diese sowie die folgenden multivariaten Ergebnisse basieren auf dem Pillai-Bartlett-Spur-Test, da dieser von den multivariaten Tests den robustesten Kennwert zu liefern scheint und vergleichsweise robust gegenüber Verletzungen von Voraussetzungen ist (Field, 2013, S. 643). Der Kontingenzkoeffizient Cramérs V ist ein standardisiertes χ^2 -Maß.

immer die Klassen zu einem Vierling zusammengeschlossen werden, die sich am ähnlichsten sind, gegen Ende der Verteilung heterogenere Klassen übrig bleiben und vereinigt werden.

Tabelle 5.1. Ausgangsvoraussetzungen auf Gruppenebene nach dem Parallelisierungsverfahren zum Messzeitpunkt 0

	T+				T-			
	K+		K-		K+		K-	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
CPM	29.95	1.48	29.88	1.94	29.58	1.34	28.67	0.52
ELFE	50.62	4.93	49.82	5.53	50.86	5.35	45.45	4.66
HRT	45.82	1.71	48.55	4.40	47.23	2.90	43.44	2.87

Anmerkung. $N = 5$ (Anzahl der Klassen). T+ = Training; T- = kein Training; K+ = kommunikative Zweiersettings; K- = keine kommunikativen Zweiersettings; CPM = Intelligenz; ELFE = sprachliche Fähigkeiten und Leseverständnis; HRT = mathematische Basiskompetenzen. Skala (mathematische Basiskompetenzen, Leseverständnis): 0–100; Skala (Intelligenz): 0–36.

Der deskriptiven Statistik zufolge, erzielte die Kontrollgruppe T-K- über alle drei abhängigen Variablen die niedrigsten Werte. Die Verteilungen der mathematischen Basiskompetenzen und des Leseverständnisses zeigten darüber hinaus, dass die erreichten T-Werte der vier Gruppen unter dem Durchschnitt der Normierungsstichprobe ($M = 50$) lagen (siehe Tabelle 5.1).

Inferenzstatistisch konnte gezeigt werden, dass sich die Trainings- (T+) und die Nicht-Trainingsgruppe (T-) nicht mehr signifikant in ihren Leistungsvoraussetzungen unterschieden. Gleichermaßen ließ sich auch auf die Gruppe mit kommunikativen Zweiersettings (K+) und ohne kommunikative Zweiersettings (K-) übertragen. Der Pillai-Spur-Test ergab weder einen signifikanten Haupteffekt für den Faktor *Training*, $V = .02$, $F(3, 335) = 1.74$, $p = .159$, $\eta_p^2 = .02$, noch einen signifikanten Haupteffekt für den Faktor *Kommunikation*, $V = .01$, $F(3, 335) = 1.02$, $p = .385$, $\eta_p^2 = .01$. Darüber hinaus wurde der Interaktionseffekt für *Training* und *Kommunikation* nicht signifikant, $V = .02$, $F(3, 335) = 2.35$, $p = .072$, $\eta_p^2 = .02$, sodass auch die Kinder, die an der kombinierten Interventionsmaßnahme aus Training und kommunikativen Zweiersettings teilnahmen, keinen signifikanten Leistungsvorsprung hinsichtlich der Kontrollvariablen hatten (siehe Abbildung 5.1).

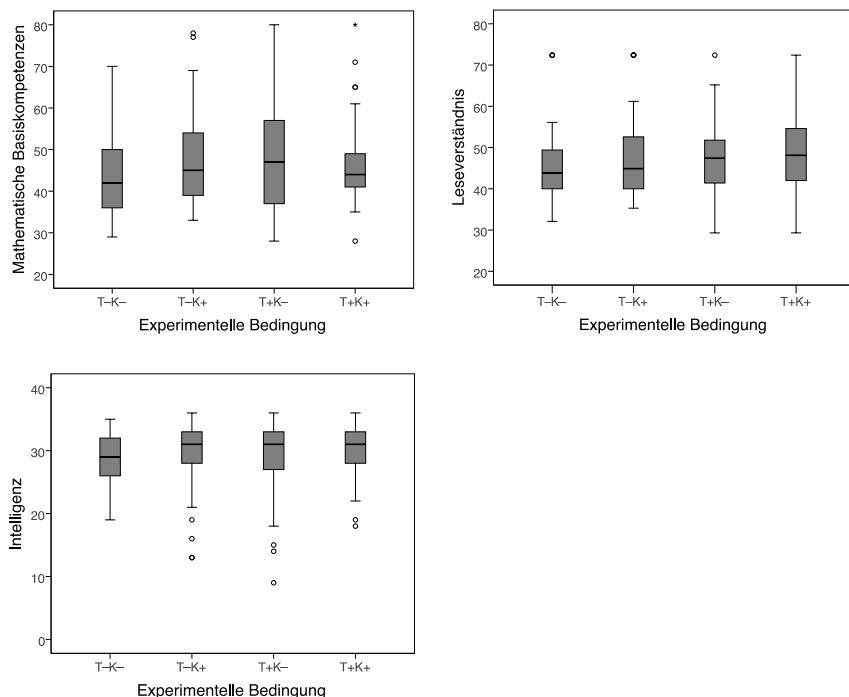


Abbildung 5.1. Vergleichbarkeit der experimentellen Bedingungen hinsichtlich der Leistungsvoraussetzungen (Intelligenz, mathematische Basiskompetenzen, Leseverständnis) nach dem Parallelisierungsverfahren zum Messzeitpunkt 0. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings. Skala (mathematische Basiskompetenzen, Leseverständnis): 0–100 (T-Werte mit $M = 50$, $SD = 10$); Skala (Intelligenz): 0–36 (Rohwerte).

Durch die gleiche Verteilung der Leistungsvoraussetzungen ist die interne Validität gegeben und wird für weitere Vergleichsanalysen angenommen. Abzusichern galt im nächsten Schritt, ob sich die Gruppen zum MZP 1 hinsichtlich ihres *Lösungserfolgs*, ihrer *Problemlösekompetenzen* und ihrer *Anzahl generierter Repräsentationen* voneinander abgrenzen oder ob diesbezüglich Vergleichbarkeit angenommen werden darf.

5.1.2 Test der Gleichverteilung der Prätest-Leistungen für die vier experimentellen Bedingungen

Zur Prüfung der Gleichverteilung wurden die Prätest-Ergebnisse des Textaufgaben-Performance-Tests herangezogen. Es wurde varianzanalytisch untersucht, ob die experimentellen Bedingungen (Gruppen) zum Zeitpunkt des Prätests in ihrem *Lösungserfolg* differierten oder ob von gleichen Populationsmittelwerten ausgegangen werden konnte. Darüber hinaus wurde der Einfluss auf die *Problemlösekompetenz* sowie die *Anzahl selbstgenerierter externer Repräsentationen* getestet.

Separate einfaktorielle Varianzanalysen deckten auf, dass sich die Trainings- (T+) und Nicht-Trainingsgruppe (T-) im *Lösungserfolg* ($F(1, 359) = 8.28, p = .004, \eta_p^2 = .02$) und in den *Problemlösekompetenzen* ($F(1, 359) = 6.83, p = .009, \eta_p^2 = .02$) signifikant voneinander unterschieden. Hinsichtlich der *Anzahl selbstgenerierter externer Repräsentationen* lag jedoch Vergleichbarkeit vor, $F(1, 359) = 0.86, p = .355, \eta_p^2 < .01$. Diese Vergleichbarkeit war für die Kommunikations- und Nicht-Kommunikationsgruppe²¹⁷ bei der *Anzahl selbstgenerierter externer Repräsentationen* nicht gegeben, $F(1, 359) = 5.17, p = .024, \eta_p^2 = .01$. Stattdessen hinsichtlich des *Lösungserfolgs*, $F(1, 359) = 1.56, p = .213, \eta_p^2 < .01$, und der *Problemlösekompetenzen*, $F(1, 359) = 2.32, p = .129, \eta_p^2 = .01$.

Der Interaktionseffekt aus den Faktoren *Training* und *Kommunikation* wies lediglich Signifikanz auf der abhängigen Variablen *Problemlösekompetenzen*, $F(1, 359) = 6.29, p = .013, \eta_p^2 = .02$, nicht aber auf der *Anzahl selbstgenerierter externer Repräsentationen*, $F(1, 359) = 0.66, p = .419, \eta_p^2 < .01$ und dem *Lösungserfolg* auf, $F(1, 359) = 2.11, p = .147, \eta_p^2 = .01$. Der deskriptiven Statistik zufolge kam eine Benachteiligung der Kontrollgruppe in Betracht, da bereits deren Leistungsvoraussetzungen (Intelligenz, mathematische Basiskompetenzen und Leseverständnis) geringfügig unter denen der experimentellen Gruppen lagen (siehe Tabelle 5.1, Abschnitt 5.1.1). Post-hoc-Analysen nach Tucky (Stevens, 2007, S. 68) bestätigten, dass sich die Kontrollgruppe hinsichtlich der Problemlösekompetenzen signifikant von allen drei Experimentalgruppen unterschied (alle $p < .05$). Weitere signifikante Gruppenunterschiede konnten nicht gefunden werden.

Infolgedessen musste davon ausgegangen werden, dass die vier experimentellen Bedingungen in Bezug auf ihre Prätest-Leistungen nicht vergleichbar waren. Um Interventionseffekte auch als solche identifizieren zu können, wurden im Nachgang die

²¹⁷ Die Bezeichnungen „Kommunikationsgruppe“ und „Gruppe mit kommunikativen Zweiersettings“ sowie die Bezeichnungen „Nicht-Kommunikationsgruppe“ und „Gruppe ohne kommunikative Zweiersettings“ werden synonym verwendet. Aus ökonomischen Gründen werden vorrangig die ersten Varianten Anwendung finden.

Gruppen einander angeglichen. Dies wurde erzielt, indem so lange das schwächste Kind aus der Kontrollgruppe und jeweils das leistungsstärkste Kind aus jeder der drei Experimentalgruppen ausgeschlossen wurde, bis deren Vergleichbarkeit abgesichert werden konnte (vgl. Felbrich, 2005, S. 149). Insgesamt war ein Ausschluss von 24 Drittklässlern, sechs Kindern pro experimentelle Bedingung, vonnöten, um die Gleichverteilung der Prätest-Leistungen sicherzustellen. Erfüllten mehrere Probanden das Ausschlusskriterium, Leistungsstärkster bzw. Leistungsschwächster zu sein, entschied der Zufallsgenerator über die Eliminierung. Folgeanalysen ergaben bei allen drei abhängigen Variablen weder signifikante Haupteffekte für die Faktoren *Training* und *Kommunikation* noch signifikante Interaktionseffekte. Dies erlaubte die Annahme, dass sich nach dem Ausschluss der 24 Probanden die vier experimentellen Bedingungen in ihren Prätest-Leistungen nicht mehr signifikant voneinander unterschieden.

5.1.3 Testung der Verletzung der Unabhängigkeitsannahme

Um die Verletzung der Unabhängigkeitsannahme zu prüfen, wurde der Fragestellung nachgegangen, welcher Varianzanteil auf Unterschiede *innerhalb*, aber auch *zwischen* den Gruppen zurückzuführen war. Je größer dieser Unterschied ist, desto wahrscheinlicher ist die Unabhängigkeit der Messwerte verletzt (Eid et al., 2011, S. 702). Die Variation zwischen Level-2-Einheiten²¹⁸ im Vergleich zur Variation zwischen Level-3-Einheiten wurde mit der Intraklassenkorrelation (ICC) erfasst (Snijders & Bosker, 1999, S. 16–21). Zu deren Berechnungen wurde das R package *lme4* (Version 1.1–7) genutzt (Bates, Maechler, Bolker, & Walker, 2015). Für die abhängige Variable *Lösungserfolg* lag die ICC bei $\rho = .016$, für die *Problemlösekompetenzen* bei $\rho = .026$ und für die *Anzahl selbstgenerierter externer Repräsentationen* bei $\rho = .012$.²¹⁹

Alle drei Kennwerte (Prätest) waren nach Hox (2010, S. 244) als klein einzustufen, weil $\rho < .05$. Hiernach war davon auszugehen, dass die Klasse nur einen kleinen Effekt auf die Kinder ausübt und die abhängige Variable (*Lösungserfolg*, *Problemlösekompetenz* bzw. *Anzahl selbstgenerierter externer Repräsentationen*) innerhalb der Klasse stark variierte, was zur Folge hatte, dass die Unterschiede zwischen den Klassen relativ klein waren (Field, 2013, S. 817–818). Eine ICC von $\rho = .016$ (Lösungserfolg im Prätest) gibt an, dass pro Lernende/Lernender 2 % der Varianz durch die Klasse, die sie/er besucht, erklärt werden kann (vgl. Bliese & Hanges, 2004, S. 402). Die Unab-

²¹⁸ In der vorliegenden Längsschnittuntersuchung bilden die Messzeitpunkte die Level-1-Einheiten, die Drittklässler die Level-2-Einheiten, die Klassen, denen die Lernenden angehören, die Level-3-Einheiten und die experimentellen Bedingungen die Level-4-Einheiten.

²¹⁹ Berechnungsgrundlage ist die Formel $\rho = \text{Var}(u_{0j}) / (\text{Var}(u_{0j}) + \text{Var}(r_{ij}))$, wobei $\text{Var}(u_{0j})$ für die Varianz zwischen den Klassen und $\text{Var}(r_{ij})$ für die Varianz innerhalb der Klassen steht. Zusammen bilden sie die Gesamtvarianz (Eid et al., 2011, S. 702).

hängigkeitsannahme galt zwar als verletzt ($\rho \neq 0$), das Risiko eine statistische Fehlentscheidung zu treffen war somit vorhanden, die Auswirkungen jedoch als gering einzustufen, da die Intraklassenkorrelationen nahe 0 lagen. Die gruppenspezifischen Fehlereinflüsse wurden daher als gering eingestuft. Eid, Gollwitzer und Schmitt (2011, S. 705) empfehlen, bei Unabhängigkeitsverletzungen Variablen aufzunehmen und zu berücksichtigen, die für systematische Unterschiede in den abhängigen Variablen zwischen den Level-2-Einheiten verantwortlich sein könnten. Aufgrund dessen wurde die hohe Anzahl an Kontrollvariablen (vgl. Abschnitt 4.5.3) erhoben. Darüber hinaus empfehlen die Autoren, bei geringen Level-2-Einheiten die Gruppenzugehörigkeit als Variable in die Regressionsanalyse aufzunehmen. Wenn die Voraussetzungen, ein hierarchisch lineares Modell zu rechnen, nicht gegeben sind, dann kann die Gruppenzugehörigkeit als fester Faktor behandelt werden, obwohl es sich streng genommen um einen zufälligen Faktor handelt (Eid et al., 2011, S. 705).

5.1.4 Test der Einflüsse des Geschlechts und der Mehrsprachigkeit

Aus darstellungsökonomischen Gründen sei für eine grafische Darstellung der Einflüsse des Geschlechts bzw. der Mehrsprachigkeit wie auch deren inferenzstatistische Absicherung auf Anhang F verwiesen.

Zum Zeitpunkt des Prätests waren sowohl der *Lösungserfolg* und die *Problemlösekompetenzen* als auch die *Anzahl selbstgenerierter externer Repräsentationen* unabhängig vom Geschlecht. Deskriptiv vorhandene Abweichungen der Geschlechter ließen sich nicht gegen den Zufall absichern. Während in der Trainings- und Nicht-Trainingsgruppe ebenso wie in der Kommunikations- und Nicht-Kommunikationsgruppe nahezu gleiche Verteilungen vorlagen, fällt auf, dass die Jungen der Gruppe T+K- einen minimal höheren Lösungserfolg als die Mädchen erzielten (Jungen: $M = 0.12$, $SD = 0.20$; Mädchen: $M = 0.07$, $SD = 0.14$), was in der Gruppe T+K+ genau umgekehrt zu beobachten war (Jungen: $M = 0.08$, $SD = 0.16$; Mädchen: $M = 0.12$, $SD = 0.19$). Die Unterschiede waren in beiden Fällen jedoch nicht signifikant.

Analoge Ergebnisse fand man bei den *Problemlösekompetenzen*. Auch hier lagen vor dem Training in den Trainings- und Nicht-Trainingsgruppen nahezu ausgewogene Problemlösekompetenzen vor (Jungen: $M_{T+} = .32$, $SD = .53$; $M_{T-} = .24$, $SD = .42$; Mädchen: $M_{T+} = .35$, $SD = .51$; $M_{T-} = .26$, $SD = .42$), während in den Kommunikationsgruppen die Mädchen und in den Nicht-Kommunikationsgruppen die Jungen minimal höhere Ausprägungen erzielten (siehe Anhang F, Tabelle F.1). Selbst die Unterschiede in den vier experimentellen Bedingungen waren unabhängig vom Geschlecht. Weder die Haupteffekte noch die Interaktionseffekte der Between-Subjects-Faktoren *Training* und *Kommunikation* waren signifikant (siehe Anhang F, Tabelle F.1).

Ferner wurde untersucht, ob es Diskrepanzen im Leistungsverhalten von Kindern, die migrationsbedingt mehrsprachig bzw. einsprachig aufwachsen, gab. Die Ergebnisse ließen einen signifikanten Einfluss der Mehrsprachigkeit auf den *Lösungserfolg* zugunsten der Kinder, die einsprachig aufwachsen, erkennen (einsprachig aufwachsend: $M = .10$, $SD = .17$; mehrsprachig aufwachsend: $M = .06$, $SD = .14$). Dieser Einfluss war losgelöst von der experimentellen Bedingung, da keine signifikanten Interaktionseffekte mit den Faktoren *Training* und *Kommunikation* vorlagen. Demzufolge unterschieden sich mehrsprachig aufwachsende Kinder von einsprachig aufwachsenden Kindern in ihrem *Lösungserfolg*, unabhängig von den Faktoren *Training* und *Kommunikation*. Die Variation der Mittelwerte war alleinig auf den signifikanten Haupteffekt des Faktors *Mehrsprachigkeit* zurückzuführen, $F(1, 326) = 4.51$, $p = .034$, $\eta_p^2 < .01$. Folglich war der Einfluss der Mehrsprachigkeit bei trainierten Kindern genauso groß wie bei untrainierten Kindern, $F(1, 326) = 0.52$, $p = .470$, $\eta_p^2 < .01$. Analoges gilt für Kinder, die der Gruppe ohne bzw. mit kommunikativen Zweiersettings angehörten, $F(1, 326) = 3.05$, $p = .082$, $\eta_p^2 < .01$. Darüber hinaus wurden für die abhängige Variable *Lösungserfolg* bis auf den Haupteffekt des Faktors Mehrsprachigkeit alle Haupt- und Interaktionseffekte der Faktoren *Training* und *Kommunikation* nicht signifikant. Für die abhängigen Variable *Problemlösekompétences* und *Anzahl selbstgenerierter externer Repräsentationen* gab es weder einen signifikanten Haupteffekt Mehrsprachigkeit noch signifikante Interaktionseffekte. Demzufolge gab es keinen Einfluss durch die Mehrsprachigkeit zum Zeitpunkt des Prättests (siehe Anhang F, Tabelle F.2).

5.2 Vor-Analyse zur Wirksamkeitshypothesenprüfung

Die Implementierung der Hauptuntersuchung basierte auf der Annahme, dass eine mehrwöchige Interventionsmaßnahme eine häufigere Nutzung von selbstgenerierten Repräsentationen bewirkt. Diese Ausgangsfragestellung, welche den Effekt der Interventionsmaßnahme untersuchte, wurde zuerst in der Pilotierung verfolgt und schließlich in der Hauptuntersuchung zu replizieren gesucht. Die Darlegung der Pilotierungsresultate und somit die Prüfung der Ausgangsfragestellung werden einleitend aufgeführt. An dieser Stelle sei erneut darauf hingewiesen, dass abweichend zur Hauptuntersuchung die Pilotierung in der vierten Jahrgangsstufe stattfand.

Ziel der Ausgangsfragestellung war es, in der Pilotierung zu überprüfen, ob eine mehrwöchige Interventionsmaßnahme Viertklässler anregt, externe Repräsentationen beim Lösen problemhaltiger Textaufgaben zu generieren und häufiger zu nutzen. Diese Wirksamkeitshypothesenprüfung soll *innerhalb* der Gruppen vorhersagekonforme Verbesserungen aufdecken (vgl. Abschnitt 3.3.2).

Für die Prüfung der Hypothese wurde für jede Testaufgabe von jedem Rater ermittelt, welche Repräsentationen (Zeichnung, Rechnung, Tabelle, Liste²²⁰, schriftsprachliche Beschreibung) generiert wurden. Im Anschluss daran wurde jeweils ein aggregierter Summenscore gebildet, der die Anzahl selbstgenerierter externer Repräsentationen pro Aufgabe abbildete. Aus diesen zwei Summenscores (Aufgabenscores) wurde dann für jeden Messzeitpunkt ein Mittelwert gebildet, sodass der Testscore als gemittelter Summenscore angibt, wie viele Repräsentationen im Durchschnitt pro Aufgabe zum jeweiligen Messzeitpunkt konstruiert wurden.

Hypothese 0: Das Repräsentationstraining befähigt Viertklässler externe Repräsentationen beim Lösen problemhaltiger Textaufgaben zu generieren und zu nutzen.

Rein deskriptiv konnte festgehalten werden, dass beide Klassen bereits nach einem kurzen Training von fünf Sitzungen mehr externe Repräsentationen generierten als davor. Insgesamt fiel auf, dass beide Klassen zu Beginn des Trainings bereits externe Repräsentationen in ihren Lösungsprozess integrierten und diesbezüglich wenig voneinander abwichen, $M_{T+K+} = 1.21$, $SD = 0.36$; $M_{T+K-} = 1.15$, $SD = 0.36$ (siehe Tabelle 5.2). Im Mittel entsprach dies ca. einer Externalisierung pro Textaufgabe. Darüber hinaus behielten sie zum dritten Messzeitpunkt, acht Wochen nach dem Training, das Generieren externer Repräsentationen bei und änderten diesbezüglich kaum ihre Vorgehensweisen. Die Ergebnisse spiegeln insgesamt wider, dass die untersuchten Viertklässler bereits von sich aus und ohne Training auf externe Repräsentationen zurückgriffen, teilweise auch schon multiple Repräsentationen generierten.

Tabelle 5.2. Deskriptive Ergebnisse der Trainingsgruppen mit und ohne Kommunikation in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+K+ ^a	1.21	0.36	1.45	0.44	1.45	0.57
T+K- ^b	1.15	0.36	1.34	0.68	1.36	0.41

Anmerkung. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings. Es existiert kein maximaler Testscore, da er die durchschnittliche Summe pro Aufgabe generierter externer Repräsentationen abbildet.

^a*N* = 19, ^b*N* = 20.

²²⁰ In der Pilotstudie wurden noch fünf Kategorien selbstgenerierter externer Repräsentationen unterschieden. Bei der Überarbeitung des Kategoriensystems wurde die Kategorie Liste aufgrund ihrer fehlenden Trennschärfe aufgelöst und je nach Merkmal entweder der Kategorie Zeichnung oder der Kategorie Tabelle zugeordnet (vgl. Abschnitt 4.2.3.1). Die Analyse der Pilotierung basiert jedoch auf dem Kategoriensystem, das die Listen beinhaltet.

Ein Vergleich der beiden Klassen im Hinblick auf die *Anzahl selbstgenerierter externer Repräsentationen* zum Prätest ergab, dass sie sich dahin gehend nicht essenziell voneinander unterschieden. Die Differenz von 0.06, 95 % CI [-.17, .30] war nicht signifikant, $t(37) = .52, p = .604$.²²¹ Die Ergebnisse spiegeln wider, dass die älteren Grundschulkinder bereits von sich aus und ohne Training auf externe Repräsentationen zurückgriffen, teilweise auch schon multiple Repräsentationen generierten ($M_{\text{Prätest}} > 1$). Eine mehrfaktorielle Varianzanalyse mit Messwiederholung auf dem ersten Faktor ergab, dass die Viertklässler, unabhängig davon, welcher Gruppe sie angehörten, über die Zeit ihre *Anzahl selbstgenerierter externer Repräsentationen* erhöhten, $F(2, 74) = 3.25, p = .044, \eta_p^2 = .08$. Geplante Kontraste legten einen signifikanten Leistungszuwachs vom Prä- zum Posttest, $F(1, 37) = 1.86, p = .034, \eta_p^2 = .12$, nicht jedoch vom Post- zum Follow-up-Test, $F(1, 37) = 0.001, p = .979, \eta_p^2 < .01$, offen. Hierbei handelte es sich über die Zeit vom Prä- zu Posttest um einen Effekt mittlerer Stärke ($.06 < \eta_p^2 < .14$, siehe Cohen, 1988, S. 287).

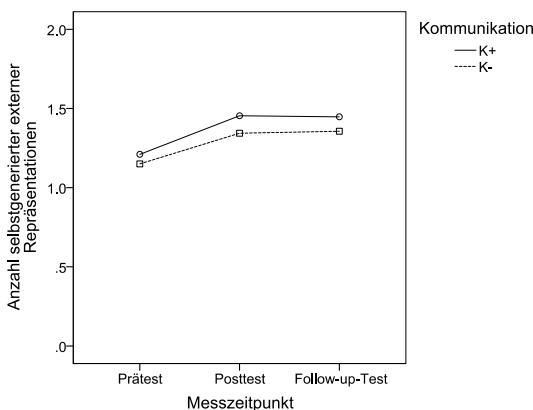


Abbildung 5.2. Mittlere Anzahl selbstgenerierter externer Repräsentationen für die Klassen mit kommunikativen Zweiersettings (K+) und ohne kommunikative Zweiersettings (K-) zu drei Messzeitpunkten.

Ob sich die Kinder in Zweiersettings über ihre Eigenproduktionen austauschten oder nicht, beeinflusste die *Anzahl selbstgenerierter externer Repräsentationen* nicht, $F(1, 37) = 0.69, p = .412, \eta_p^2 = .02$. Ferner wiesen die beiden Klassen über die Zeit ein analoges Profil auf und differierten nicht, $F(2, 74) = 0.03, p = .969, \eta_p^2 < .01$. Die Pa-

²²¹ Das durch die Bias-corrected-and-accelerated-Methode, kurz BCa-Methode, angepasste 95 %-Konfidenz-intervall wird in Klammern berichtet.

rallelität der beiden Klassen geht auch aus dem Interaktionsdiagramm hervor (siehe Abbildung 5.2).

Im Hinblick auf die Hauptuntersuchung galt es zu untersuchen, ob Kinder der Klassenstufe 4 offen und von ihrem bisherigen Mathematikunterricht möglichst unvoreingenommen an die Lösung von Textaufgaben herangehen, weshalb der Zuwachs vom Prä- zum Posttest auf Repräsentationsebene zu spezifizieren gesucht wurde. Dabei stellte man fest, dass die Kinder beider Klassen verstärkt auf Zeichnungen und Rechnungen zurückgriffen und dies auch beibehielten, wobei Ersteres in der Klasse ohne kommunikative Settings und Letzteres in der Klasse mit kommunikativen Settings dominierte (siehe Abbildung 5.3).

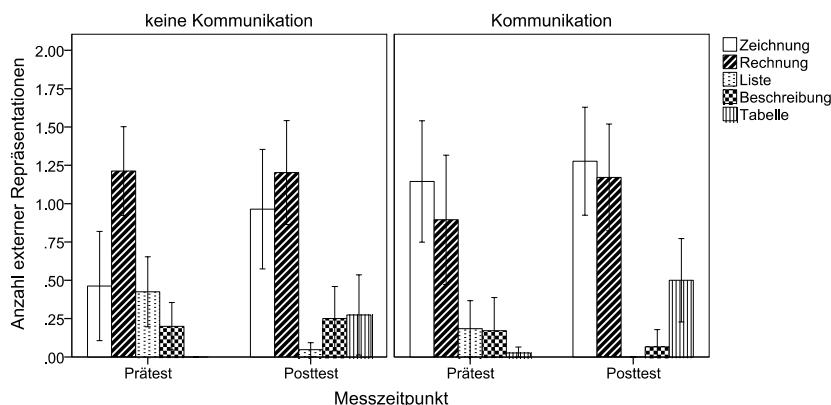


Abbildung 5.3. Mittlere Anzahl von Zeichnungen, Rechnungen, Listen, schriftsprachlichen Beschreibungen und Tabellen für die Trainingsgruppen mit und ohne kommunikative Zweiersettings zu zwei Messzeitpunkten. Die Fehlerindikatoren geben das 95 %-Konfidenzintervall an.

Den Repräsentationsformen Tabellen, Listen und schriftsprachliche Beschreibungen wurden sowohl vor als auch nach dem Kurztraining eine untergeordnete Rolle zugesprochen, wenngleich die Nutzung von Tabellen nach dem Training zunahm. Dies zeigte, dass die Klassen – wie bereits in Abschnitt 4.2.4.1 beschrieben – unterschiedliche Heran- und Vorgehensweisen präferierten, die ihnen aus dem Mathematikunterricht schon vertraut waren.

Tabelle 5.3 kann entnommen werden, dass sich die Generierung und Nutzung von Zeichnungen in der Gruppe mit Kommunikation (T+K+) im Mittel um 0,5 und in der Gruppe ohne Kommunikation (T+K-) im Mittel um 0,4 erhöhte.

Während Viertklässler der Nicht-Kommunikationsklasse nach dem Kurztraining weniger auf Beschreibungen zurückgriffen als vor dem Training, zeigte sich in der Kommunikationsklasse ein geringfügiger Anstieg.

Tabelle 5.3. Deskriptive Ergebnisse der Trainingsgruppen mit und ohne Kommunikation in der Anzahl selbstgenerierter externer Repräsentationen zu zwei Messzeitpunkten

	Zeichnungen				Rechnungen			
	Prä		Post		Prä		Post	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+K+ ^a	0.46	0.76	0.95	0.88	1.21	0.62	1.16	0.74
T+K- ^b	1.14	0.82	1.26	0.733	0.89	0.88	1.13	0.74

	Beschreibungen				Tabellen			
	Prä		Post		Prä		Post	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+K+ ^a	0.20	0.33	0.24	0.52	0.00	0.00	0.29	0.59
T+K- ^b	0.17	0.45	0.04	0.17	0.03	0.09	0.47	0.56

Anmerkung. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings. Aufgrund ihrer untergeordneten Relevanz wird die Kategorie Liste nicht explizit aufgezeigt.

^a*N* = 19. ^b*N* = 20.

Inferenzstatistische Analysen konnten aufgrund mangelnder Vergleichbarkeit zum Zeitpunkt des Prätests nicht angestellt werden.

Es bleibt festzuhalten, dass zum einen bereits ein kurzes Training von fünf Sitzungen eine häufigere Nutzung von selbstgenerierten externen Repräsentationen bewirkte und Viertklässler motivierte, externe Repräsentationen zu generieren. Demzufolge kann die Wirksamkeitshypothese als bestätigt angesehen werden und legitimiert die Durchführung der Hauptuntersuchung. Zum anderen wurde transparent, dass der Mathematikunterricht einen nicht zu unterschätzenden Einfluss auf die Heran- und Vorgehensweise der Kinder ausübt. Letztgenanntes veranlasste dazu, die Hauptuntersuchung bereits in der dritten Jahrgangsstufe durchzuführen, um gewährleisten zu können, dass die Kinder auf wenig Vor- und Erfahrungswissen im Umgang mit externen Repräsentationen zurückgreifen und neuen Heran- und Vorgehensweisen gegenüber aufgeschlossen sind (vgl. Abschnitt 4.2.4.1). Im Folgenden wird nun unter Berücksichtigung der Voruntersuchung auf die Forschungsfragen eingegangen.

5.3 Analyse der abhängigen Variablen

Im Rahmen der 1. Forschungsfrage wurde varianzanalytisch untersucht, ob sich die Interventionsmaßnahme positiv auf den Lösungserfolg, die Problemlösekompetenzen bzw. die Anzahl selbstgenerierter externer Repräsentationen auswirkte. Die Ergebnisse werden in Abschnitt 5.3.1 präsentiert. In den beiden nachfolgenden Abschnitten wird aufgezeigt, ob die Implementierung des *Repräsentationstrainings* (Abschnitt 5.3.2) bzw. der *Kommunikation* (Abschnitt 5.3.3) einen positiven Effekt auf den Lösungserfolg, die Problemlösekompetenzen bzw. die Anzahl selbstgenerierter externer Repräsentationen ausühte. Darauf folgend wurde überprüft, ob die Kombination der Faktoren *Repräsentationstraining* und *Kommunikation* einen zusätzlichen Effekt auf die vorstehend genannten abhängigen Variablen bewirkte. Die Ergebnisdarstellung erfolgt in Abschnitt 5.3.4. Aufgrund der Zusammenghörigkeit der ersten vier Forschungsfragen erfolgt im Anschluss eine kurze Zusammenfassung der gefundenen Effekte (Abschnitt 5.3.5). Abschließend wird erörtert, ob – und wenn ja, welche – externe Repräsentationen besonders geeignet waren, den Lösungserfolg bzw. die Problemlösekompetenzen vorherzusagen, und inwiefern die Kontrollvariablen einen Einfluss auf die Vorhersage nahmen (Abschnitt 5.3.6).

Bevor jeweils die inferenzstatistischen Ergebnisse präsentiert werden, erfolgt einleitend eine deskriptive Analyse der Daten. Die Ergebnisdarstellung bezieht sich zunächst immer auf den gesamten Textaufgaben-Performance-Test. Im Anschluss erfolgt eine aufgabenspezifische Analyse, welche beabsichtigt, aufzudecken, ob es Unterschiede in der Entwicklung zwischen trainierten und untrainierten Aufgabentypen gab und inwiefern der Aufgabentyp die Effekte beeinflusst. Demzufolge wurde der Aufgabentyp als zweiter Within-Subjects-Faktor mit drei Ausprägungen, *Vergleichsaufgabe*, *Aufgabe mit komplexen Informationen* und *Aufgabe zur Verhältnisverteilung* in das Design aufgenommen. Zuvor wurde in G*Power sichergestellt, dass das Hinzufügen des Faktors Aufgabentyp keinen statistischen Nachteil für die Prüfung der anderen Faktoren mit sich brachte und die Teststärke groß genug war. Die Stichprobe legitimierte die Berücksichtigung des Faktors in der Analyse.

Insgesamt konnten 15 Probanden nicht vollständig an der Interventionsstudie teilnehmen. Gründe hierfür waren zum einen krankheits- oder kurbedingt Ausfälle, zum anderen Schulwechsel oder Zurückstufungen in die zweite Klasse. 15 der 366 Kinder fehlten somit mindestens an einem Messezeitpunkt, sodass ihre Daten unvollständig vorlagen. Weil sich dadurch die Stichprobe nur um 4 % reduzierte, durften die Fälle listenweise ausgeschlossen werden (Graham, Cumsille, & Elek-Fisk, 2003, S. 90; Wirtz, 2004, S. 111). Die Missing-Data-Analyse identifizierte ein zufälliges Auftreten der fehlenden Werte. Es gab weder Zusammenhänge zwischen fehlenden Werten und

Werten anderer Variablen noch traten Muster fehlender Werte gehäuft auf. Zudem mussten 24 Drittklässler, 6 Kindern pro experimentelle Bedingung, ausgeschlossen werden, um dem gescheiterten Parallelisierungsverfahren entgegenzuwirken und trotzdem die Gleichverteilung der Prätest-Leistungen sicherzustellen (siehe Abschnitt 5.1.2).

5.3.1 Forschungsfrage 1: Effekt der Interventionsmaßnahme

Die erste Forschungsfrage widmet sich der Prüfung, ob sich die Merkmalsausprägungen des *Lösungserfolgs*, der *Problemlösekompetenzen* und der *Anzahl selbstgenerierter externer Repräsentationen* zu den einzelnen Messzeitpunkten signifikant unterscheiden oder ob alle Populationsmittelwerte der Messzeitpunkte des Within-Subjects-Faktors gleich sind.

Für die Hypothesenprüfung wurde die Richtigkeit der Testaufgaben dichotom erfasst. Ein Aufgabenscore 0 stand für ein falsches, ein Aufgabenscore 1 für ein richtiges Ergebnis. Der Mittelwert der drei aggregierten Aufgabenscores eines Messzeitpunktes bildete den entsprechenden Testscore ab. Dieser gab an, wie viele Aufgaben pro Test richtig gelöst wurden.

Hypothese 1.1: Über alle Gruppen hinweg lösen Drittklässler problemhaltige Textaufgaben nach der Interventionsmaßnahme erfolgreicher als davor.

Im Mittel steigerten alle Drittklässler über die Zeit ihren Lösungserfolg. Vor dem Training lag der durchschnittliche Testscore bei $M = .08$ ($SD = .16$). Dieser Bodeneffekt geht auch aus dem Boxplot hervor (siehe Abbildung 5.4). Die meisten Kinder erzielten einen Testscore von 0.

Der Lösungserfolg lag zum Zeitpunkt des Posttests bei $M = .38$ ($SD = .34$) und zum Zeitpunkt des Follow-up-Tests bei $M = .43$ ($SD = .35$). Dass sich unabhängig von den experimentellen Bedingungen der mittlere *Lösungserfolg* an den drei Messzeitpunkten unterschied, kann aus dem signifikanten Haupteffekt Zeit abgeleitet werden, $F(2, 646) = 235.49$, $p < .001$, $\eta_p^2 = .42$. Geplante Kontraste mit Bonferroni-Adjustierung deckten einen signifikanten Leistungszuwachs vom Prä- zum Posttest, $F(1, 323) = 310.09$, $p < .001$, $\eta_p^2 = .49$, wie auch vom Post- zum Follow-up-Test, $F(1, 323) = 6.46$, $p = .012$, $\eta_p^2 = .02$, auf. Letztgenannter Effekt war im Vergleich zu dem erstgenannten als vernachlässigbar klein einzustufen, was auch angesichts der sich nicht ändernden Mediane im Boxplot sichtbar wird (siehe Abbildung 5.4).

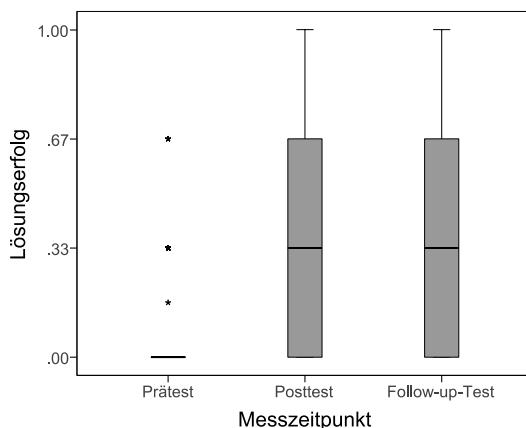


Abbildung 5.4. Mittlerer Testscore des Lösungserfolgs zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation. Skala: 0–1.

Ferner ist dem Boxplot zu entnehmen, dass zum Zeitpunkt des Post- und Follow-up-Tests symmetrische Verteilungen vorliegen, welche sich unwesentlich voneinander unterscheiden. Testscores von .33 und .66, sprich eine bzw. zwei von drei Aufgaben wurden richtig gelöst, stellten nach dem Training keine Extremwerte mehr dar, wie dies noch zum Zeitpunkt des Prätests der Fall war. Stattdessen zählten sie zu dem Bereich, der die mittleren 50 % aller Fälle umfasste. Die Aufgaben behielten für viele Kinder auch nach der Intervention noch ihre Problemhaftigkeit.

Da der Lösungserfolg über alle Gruppen hinweg signifikant von Messzeitpunkt 1 zu Messzeitpunkt 2 zunahm, konnte die Hypothese 1.1 bestätigt werden.

Aufgabenspezifische Analyse

Während bisher der Lösungserfolg gemittelt über alle drei Aufgaben im Fokus stand, wird nachfolgend eine aufgabenspezifische Analyse die Ergebnisdarstellung vervollständigen. Es soll untersucht werden, ob die Effekte für alle Aufgabentypen gleich sind. Verglichen wird, wie sich der mittlere Lösungserfolg über die Zeit für die drei Aufgaben der Aufgabentypen *Vergleichsaufgabe*, *Aufgabe mit komplexen Informationen* und *Aufgabe zur Verhältnisverteilung* entwickelte und ob sich Einflüsse der drei Aufgabentypen ableiten lassen.

Die Veranschaulichung der Zusammenhänge in Abbildung 5.5 bestätigten die bisherigen Ergebnisse dahin gehend, dass der Lösungserfolg vom Prä- zum Posttest bei allen

drei Aufgabentypen zunahm.²²² Es zeigte sich, dass den Kindern vor dem Trainingsstart das Lösen der *Vergleichsaufgabe* leichter fiel als das Lösen der beiden anderen Aufgabentypen.

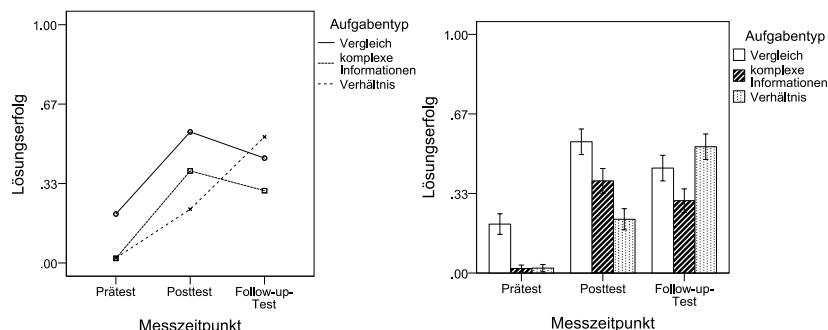


Abbildung 5.5. Mittlere Aufgabenscores des Lösungserfolgs zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation. Skala: 0–1.

Die Kinder erzielten bei der *Vergleichsaufgabe* einen mittleren Testscore $M = .21$ ($SD = .41$), während die Testscores bei den letztgenannten Aufgabentypen, der *Aufgabe mit komplexen Informationen* und der *Aufgabe zur Verhältnisverteilung*, nahezu 0 waren ($M_{komplex} = .02$, $SD = .14$; $M_{Verhältnis} = .02$, $SD = .14$). Wurde bei diesen beiden anfangs schwierigeren Aufgaben der Blick auf die Leistungsentwicklung vom Präzum Posttest gerichtet, so fiel auf, dass die Erfolge zunahmen, jedoch ein größerer Zuwachs bei der *Aufgabe mit komplexen Informationen* zu beobachten war. Dieser lag bei .36 im Vergleich zu .20 bei der *Aufgabe zur Verhältnisverteilung* (Posttest: $M_{komplex} = .38$, $SD = .48$; $M_{Verhältnis} = .22$, $SD = .41$). Wie erwartet, wurde bei beiden Aufgabentypen eine Leistungssteigerung erzielt, wobei ein Vorteil für den trainierten Aufgabentyp bestand.

Ferner wurde deutlich, dass der deskriptiv identifizierte Anstieg des Lösungserfolgs vom Post- zum Follow-up-Test lediglich auf den dritten, untrainierten Aufgabentyp, die *Aufgabe zur Verhältnisverteilung*, zurückzuführen war. Bei den beiden trainierten Aufgabentypen, der *Vergleichsaufgabe* sowie der *Aufgabe mit komplexen Informationen*, war ein geringfügiger Abfall um ca. .10, bei der *Aufgabe zur Verhältnisverteilung*

²²² Aufgrund der Tatsache, dass es sich beim Lösungserfolg um eine dichotome und keine metrische abhängige Variable handelt, erfolgt die Visualisierung des Zusammenhangs nicht durch den Boxplot, sondern durch ein Balkendiagramm (J. Janssen & Laatz, 2007, S. 663).

ein Anstieg um .31 zu verzeichnen. Die vollständigen Ergebnisse der deskriptiven Analyse können Anhang G.1 (Tabelle G.1) entnommen werden.

Inferenzstatistisch konnte gezeigt werden, dass sich die drei Aufgabentypen signifikant voneinander unterschieden (Haupteffekt Aufgabentyp: $\chi^2(2) = 52.82, p < .001$).²²³ Geplante Kontraste ergaben Unterschiede zwischen den trainierten und untrainierten Aufgabentypen, $\chi^2(1) = 20.24, p < .001$, wie auch zwischen den beiden trainierten Aufgabentypen, $\chi^2(1) = 33.51, p < .001$. Der Haupteffekt hatte über die Zeit Bestand, $\chi^2(4) = 47.00, p < .001$ (Interaktionseffekt Zeit x Aufgabentyp). Paarweise Vergleiche identifizierten, dass sich sowohl vom Prä- zum Post- als auch vom Post- zum Follow-up-Test die trainierten von den untrainierten wie auch die trainierten Aufgabentypen untereinander signifikant unterschieden. Aus ökonomischen Gründen und untergeordneter Relevanz für die vorliegende Untersuchung wird auf deren Darstellung verzichtet.

Hypothese 1.2: Über alle Gruppen hinweg verfügen Drittklässler nach der Interventionsmaßnahme über höhere Problemlösekompetenzen als davor.

Die Problemlösekompetenzen wurden kriteriumsbezogen mit einer fünfstufigen Skala erfasst, indem den individuellen Lösungsprozessen auf Grundlage der gesetzten Kriterien ein Aufgabenscore zwischen 0 und 4 zugeordnet wurde (vgl. Abschnitt 4.5.1.1). Der Mittelwert aller drei Aufgabenscores bildet sich im Testscore ab.

Wurde die Entwicklung der Problemlösekompetenzen der Drittklässler fokussiert, so ließ sich auch hier im Mittel ein Kompetenzzuwachs über die Zeit feststellen. Die Veränderung war vom Prä- zum Postest am größten einzustufen und blieb vom Post- zum Follow-up-Test nahezu unverändert, was auch aus der Abbildung 5.6 hervorgeht.

²²³ Aufgrund der dichotomen abhängigen Variablen erfolgt die Analyse mit dem Wald- χ^2 -Test (siehe Abschnitt 4.8.5).

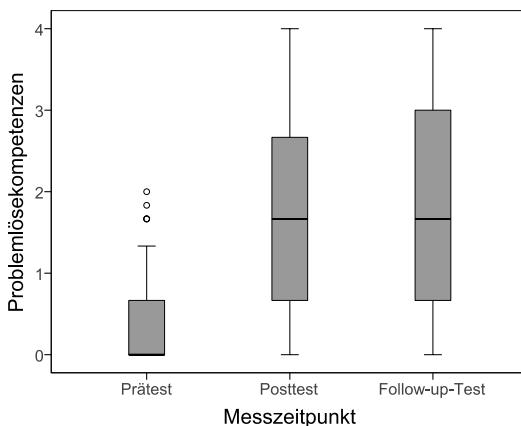


Abbildung 5.6. Mittlerer Testscore der Problemlösekompetenzen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation. Die Fehlerindikatoren geben das 95 %-Konfidenz-intervall an. Skala: 0–4.

Die univariate Varianzanalyse mit Messwiederholung identifizierte für die abhängige Variable *Problemlösekompetenzen* einen signifikanten Haupteffekt des Faktors Zeit, $F(1.99, 643.32) = 421.04, p < .001, \eta_p^2 = .57$. Der beobachtete Effekt war für die *Problemlösekompetenzen* größer als für den *Lösungserfolg*, $\eta_p^2 = .57 < \eta_p^2 = .42$ (vgl. Hypothese 1.1). Ergänzende Kontrastanalysen mit Bonferroni-Adjustierung bestätigten, dass die Probanden zum Posttest ($M = 1.72, SD = 1.22$) signifikant höhere Problemlösekompetenzen vorwiesen als zum Prätest ($M = 0.30, SD = .48$), $F(1, 323) = 591.89, p < .001, \eta_p^2 = .65$. Dagegen unterschieden sich deren Problemlösekompetenzen zwischen der zweiten und dritten Messung nur noch hinreichend signifikant voneinander, $F(1, 323) = 3.77, p = .053, \eta_p^2 = .01$. Demzufolge gilt auch die Hypothese 1.2 als bestätigt.

Aufgabenspezifische Analyse

Für alle Aufgabentypen konnte ein Anstieg der Problemlösekompetenzen vom Prä- zum Posttest bestätigt werden (siehe Abbildung 5.7). Im Boxplot wird deutlich, dass die Kinder zum Zeitpunkt des Prätests insbesondere bei der *Aufgabe mit komplexen Informationen* und der *Aufgabe zur Verhältnisverteilung* im Mittel über mangelnde Problemlösekompetenzen verfügten und nicht in der Lage waren, die Aufgabe zu lösen (siehe Abbildung 5.7).

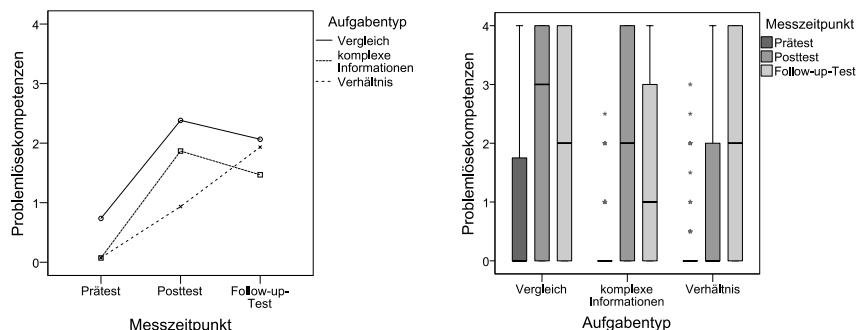


Abbildung 5.7. Mittlere Aufgabenscores der Problemlösekompetenzen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation. Skala: 0–4.

Sinkende Aufgabenscores vom Post- zum Follow-up-Test waren bei beiden trainierten Aufgabentypen (*Vergleichsaufgabe* und *Aufgabe mit komplexen Informationen*) auch für diese abhängige Variable zu beobachten. Nur bei der *Aufgabe zur Verhältnisverteilung*, die nicht trainiert wurde, erfolgte ein weiterer Kompetenzzuwachs um 1.0 (siehe Anhang G.1, Tabelle G.1).

Inferenzstatistisch ließen sich gemittelt über die Stufen des Faktors *Training* und *Kommunikation* Unterschiede zwischen den Stufenmittelwerten des Faktors *Aufgabentyp* identifizieren, $F(2, 646) = 95.96, p < .001, \eta_p^2 = .23$ (Haupteffekt Aufgabentyp). Der Effekt war nach Cohen (1988, S. 287) von großer Stärke ($\eta_p^2 > .14$). Demzufolge erzielten die Kinder unabhängig vom Messzeitpunkt bei den drei Aufgaben signifikant unterschiedliche Problemlösekompetenzen. Die Testscores der *Vergleichsaufgabe* waren größer als die der *Aufgabe mit komplexen Informationen* (Mittelwertsdifferenz von $.60, p < .001$) und die der *Aufgabe zur Verhältnisverteilung* (Mittelwertsdifferenz von $.74, p < .001$), weshalb sie als die leichteste der drei Aufgaben identifiziert werden konnte. Die *Aufgabe mit komplexen Informationen* übertraf ebenfalls die *Aufgabe zur Verhältnisverteilung* (Mittelwertsdifferenz von $.15, p = .026$).

Darüber hinaus gab es einen signifikanten Interaktionseffekt der Faktoren *Zeit* und *Aufgabentyp*, $F(3.87, 1248.79) = 46.43, p < .001, \eta_p^2 = .13$. Diese Interaktion deutete auf Unterschiede zwischen den Aufgabentypen über die Zeit hin (siehe Abbildung 5.7). Kontrastanalysen zeigten, zwischen welchen Messzeitpunkten aufgabenspezifische Unterschiede vorlagen. Die geplanten Kontraste deckten auf, dass sich die trainierten Aufgabentypen sowohl vom Prä- zum Posttest, $F(1, 323) = 93.70, p < .001, \eta_p^2 = .23$, als auch vom Post- zum Follow-up-Test, $F(1, 323) = 142.98, p < .001, \eta_p^2 = .31$, signifikant vom untrainierten Aufgabentyp unterschieden. Dass diese im

zweiten Abschnitt, vom Post- zum Follow-up-Test, stärker differierten, wird auch durch die gegensätzlichen Steigungen im Interaktionsdiagramm deutlich (siehe Abbildung 5.7). Die beiden trainierten Aufgabentypen, *Vergleichsaufgaben* und *Aufgabe mit komplexen Informationen*, unterschieden sich weder vom Prä- zum Posttest, $F(1, 323) = 1.29, p = .257, \eta_p^2 < .01$, noch vom Post- zum Follow-up-Test signifikant voneinander, $F(1, 323) = 0.46, p = .497, \eta_p^2 < .01$.

Ungeachtet dessen, welcher Gruppe die Kinder angehörten, bleibt festzuhalten, dass sie bei den trainierten Aufgabentypen über die Zeit höhere *Problemlösekompetenzen* erzielten als bei dem untrainierten Aufgabentyp. Folglich machte es einen Unterschied, ob der Aufgabentyp Bestandteil der Interventionsmaßnahme war oder nicht. Aufgaben werden erfolgreicher gelöst, wenn der zugehörige Aufgabentyp zuvor im Unterricht thematisiert wurde.

Hypothese 1.3: Über alle Gruppen hinweg generieren Drittklässler nach der Interventionsmaßnahme mehr externe Repräsentationen als davor.

Für jede Aufgabe wurde ein Summenscore berechnet, der die Anzahl der selbstgenannten externen Repräsentationen (Zeichnung, Rechnung, Tabelle, schriftsprachliche Beschreibung) abbildet. Aus diesen drei Aufgabenscores wurde dann für jeden Messzeitpunkt ein Mittelwert gebildet, sodass der Testscore als gemittelter Summenscore angibt, wie viele externe Repräsentationen im Durchschnitt pro Aufgabe zum jeweiligen Messzeitpunkt konstruiert wurden (vgl. Ausgangshypothese in Abschnitt 5.2).

Wurden die erzielten Mittelwerte rein deskriptiv miteinander verglichen, so nahm die Anzahl selbstgenerierter externer Repräsentationen vom Prä- ($M = 0.34, SD = .40$) zum Posttest ($M = 1.34, SD = .68$) zu (siehe Abbildung 5.8). Zum Follow-up-Test war nahezu keine Veränderung, weder in Form eines Abfalls noch in Form eines weiteren Anstiegs, zu verzeichnen ($M = 1.36, SD = .70$). Während die Drittklässler vor der Untersuchung selten von sich aus externe Repräsentationen generierten und in ihren Lösungsprozess integrierten, stieg mit zunehmender Zeit – unabhängig von der experimentellen Bedingung – die Anzahl an. Im Mittel wurde nach der Interventionsmaßnahme pro Aufgabe ca. eine Repräsentation generiert (siehe Abbildung 5.8).

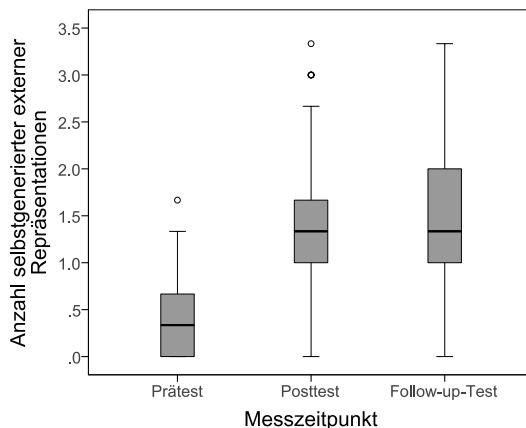


Abbildung 5.8. Mittlerer Testscore der Anzahl selbstgenerierter externer Repräsentationen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.

Inferenzstatistisch konnten die deskriptiven Ergebnisse bestätigt werden. Die Probanden generierten im Verlaufe der Zeit signifikant mehr externe Repräsentationen, als sie dies noch vor der Interventionsmaßnahme getan hatten, $F(2, 646) = 454.67, p < .001, \eta_p^2 = .59$. Dieser signifikante Unterschied wurde beim Mittelwertvergleich des ersten zum zweiten ($F(1, 323) = 651.95, p < .001, \eta_p^2 = .67$), jedoch nicht vom zweiten zum dritten Messzeitpunkt nachgewiesen ($F(1, 323) = 0.15, p = .698, \eta_p^2 < .01$). Dass lediglich die ersten beiden Messzeitpunkt die Anzahl der selbstgenerierten externen Repräsentationen beeinflusste, zeigte sich in dem enormen Zuwachs vom Prä- zum Posttest und dem ausbleibenden Zuwachs vom Post- zum Follow-up-Test (siehe Abbildung 5.8).

Aufgabenspezifische Analyse

Die für den gesamten Textaufgaben-Performance-Test gefundenen Ergebnisse ließen sich aufgabenspezifisch replizieren. Vor dem Training zogen die Kinder im Mittel kaum Repräsentationen zur Lösungsfindung heran (Prättest: $M_{\text{Vergleich}} = 0.50, SD = 0.66; M_{\text{komplex}} = 0.30, SD = 0.50; M_{\text{Verhältnis}} = 0.21, SD = 0.42$). Die Anzahl selbstgenerierter externer Repräsentationen nahm nach dem Training bei allen drei Aufgaben deutlich zu (siehe Abbildung 5.9). Im Mittel betrug der Anstieg 1.19, 1.04 bzw. 0.80, wobei der größte Wert der Vergleichsaufgabe und der kleinste Wert der Aufgabe zur Verhältnisverteilung zuzuordnen war (siehe Anhang G.1, Tabelle G.1). Demnach waren bei den trainierten Aufgabentypen die Zuwächse am größten.

Der Boxplot deckt außerdem auf, dass die mittleren 50 % aller Probanden nach dem Training zwischen 1 und 2 externen Repräsentationen bei einem trainierten Aufgabentyp generierten. Bei dem nicht-trainierten Aufgabentyp waren es zwischen 0 und 2 Repräsentationen (siehe Abbildung 5.9). In den seltensten Fällen wurden auch vier Repräsentationen konstruiert.

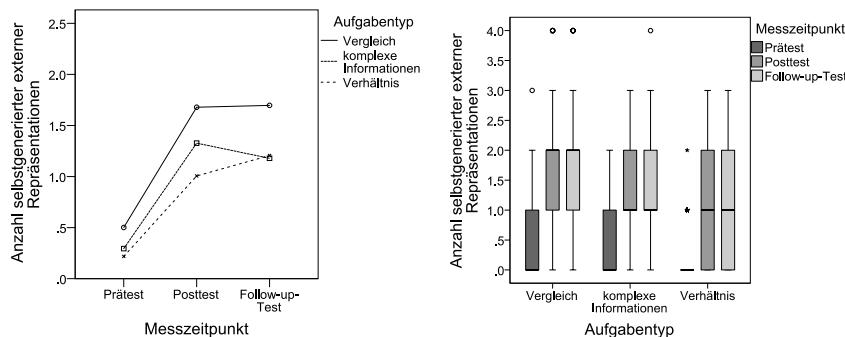


Abbildung 5.9. Mittlere Aufgabenscores der Anzahl selbstgenerierter externer Repräsentationen zu den drei Messzeitpunkten unabhängig von den Faktoren Training und Kommunikation.

Die Varianzanalyse mit Messwiederholung bestätigte, dass die Drittklässler gemittelt über die *Zeit* und die *experimentelle Bedingung* bei den drei Aufgaben signifikant unterschiedlich viele externe Repräsentationen entwickelten (Haupteffekt Aufgabentyp: $F(1.95, 630.75) = 146.60, p < .001, \eta_p^2 = .31$). Dieser Effekt hatte auch über die *Zeit* Bestand, $F(3.71, 1197.34) = 12.86, p < .001, \eta_p^2 = .04$ (Interaktionseffekt *Zeit* x *Aufgabentyp*). Geplante Kontraste legten offen, dass sich die trainierten Aufgabentypen sowohl vom Prä- zum Posttest, $F(1, 323) = 32.84, p < .001, \eta_p^2 = .09$, als auch vom Post- zum Follow-up-Test, $F(1, 323) = 14.05, p < .001, \eta_p^2 = .04$, signifikant von dem untrainierten Aufgabentyp unterschieden. Dies gilt auch für die beiden trainierten Aufgabentypen. Die *Vergleichsaufgaben* und *Aufgaben mit komplexen Informationen* unterschieden sich signifikant (Prä- zu Posttest: $F(1, 323) = 5.25, p = .023, \eta_p^2 = .02$; Post- zu Follow-up-Test: $F(1, 323) = 4.71, p = .031, \eta_p^2 = .01$).

Im Zuge der ersten Forschungsfrage gelten alle drei Hypothesen als bestätigt. Alleinig die Implementierung der Interventionsmaßnahme bewirkte eine Leistungssteigerung seitens der Drittklässler, sodass die Thematisierung problemhaltiger Textaufgaben in der Grundschule als berechtigt anzusehen ist (vgl. Abschnitt 1.1). Um jedoch Aussagen darüber treffen zu können, ob und inwiefern sich die Gruppen in ihren Merkmals-

ausprägungen und ihrer Entwicklung unterschieden, wurde im Rahmen der Forschungsfragen 2–4 untersucht.

5.3.2 Forschungsfrage 2: Effekt des Repräsentationstrainings

Im Fokus der zweiten Forschungsfrage stand, ob sich zum einen das Repräsentationstraining positiv auf den *Lösungserfolg*, die *Problemlösekompetenzen* und die *Anzahl selbstgenerierter externer Repräsentationen* von Drittklässlern auswirkt, und zum anderen, ob sich die Trainingsgruppe von der Nicht-Trainingsgruppe, auch über die Zeit, unterscheidet.

Hypothese 2.1: Drittklässler lösen problemhafte Textaufgaben erfolgreicher, wenn sie trainiert werden.

Bereits die Mittelwerte deuteten darauf hin, dass sowohl die Trainingsgruppe als auch die Nicht-Trainingsgruppe ihren Lösungserfolg über die Zeit steigerten. Während Kinder, die am Training teilnahmen (T+), ihren Lösungserfolg-Testscore im Mittel um .40 erhöhen konnten, betrug der Scoreanstieg in Nicht-Trainingsgruppen (T-) im Mittel nur .20 (siehe Tabelle 5.4).

Tabelle 5.4. Deskriptive Ergebnisse der Trainings- und Nicht-Trainingsgruppe im Lösungserfolg zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+ ^a	.10	.18	.48	.35	.54	.37
T- ^b	.07	.14	.28	.30	.30	.29

Anmerkung. T+ = Training; T- = kein Training. Skala= 0–1.

^a*N*=172. ^b*N* = 155.

Beide Gruppen erhöhten vom Prä- zum Posttest ihre Lösungserfolgsrate, wobei der größere Zuwachs in der Trainingsgruppe zu beobachten war (siehe Abbildung 5.10). Vor dem Hintergrund, dass die Skala lediglich von 0 bis 1 reicht (0 = keine Aufgabe richtig, 1 = alle Aufgaben richtig) und der Lösungserfolg das harte Beurteilungskriterium darstellt, kann insbesondere der Anstieg in den Trainingsgruppen als bedeutsam eingeschätzt werden.

Diese Überlegenheit konnte auf das Treatment zurückgeführt werden, da beiden Gruppen zum Zeitpunkt des Prätests vergleichbare Ausgangsvoraussetzungen zugrunde lagen ($M_{T+} = .10$, $SD_{T+} = .18$; $M_{T-} = .07$, $SD_{T-} = .14$). Darüber hinaus wurde erkennbar,

dass sich die Trainingsgruppe im Gegensatz zur Nicht-Trainingsgruppe zum Follow-up-Test weiterhin leicht verbesserte.

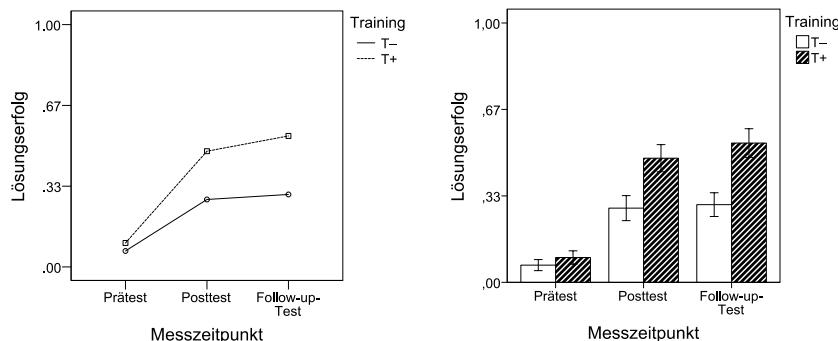


Abbildung 5.10. Mittlerer Testscore des Lösungserfolgs der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten. Skala: 0–1.

Die Varianzanalyse mit Messwiederholung bestätigte, dass die Trainingsgruppe und die Nicht-Trainingsgruppe differierten und kein paralleles Leistungsprofil aufwiesen. Es zeigte sich, dass der Haupteffekt *Training* mit der abhängigen Variablen *Lösungserfolg* in Verbindung gebracht werden konnte ($F(1, 323) = 40.60, p < .001, \eta_p^2 = .11$). Demzufolge unterschied sich die Trainingsgruppe in ihrem *Lösungserfolg* signifikant von der Nicht-Trainingsgruppe, unabhängig vom Messzeitpunkt. Dabei war die Trainingsgruppe der Nicht-Trainingsgruppe im paarweisen Vergleich überlegen (Mittelwertsdifferenz von .16, $p < .001$). Dass Drittklässler der Trainingsgruppe (T+) tendenziell höhere Leistungen erzielten als Gleichaltrige der Nicht-Trainingsgruppe (T-), geht auch aus Abbildung 5.10 hervor.

Darüber hinaus gab es einen signifikanten Interaktionseffekt der Faktoren *Zeit* und *Training*, $F(2, 646) = 21.12, p < .001, \eta_p^2 = .06$. Diese Interaktion wies auf Unterschiede zwischen beiden Gruppen über die Zeit hin. Die geplanten Kontraste ergaben, dass die Gruppen lediglich vom Prät- zum Posttest ($F(1, 323) = 24.51, p < .001, \eta_p^2 = .07$), nicht jedoch vom Post- zum Follow-up-Test, ($F(1, 323) = 1.64, p = .202, \eta_p^2 = .01$) signifikant in ihrem *Lösungserfolg* differierten. Demzufolge nahm der *Lösungserfolg* der Kinder signifikant durch die Interventionsmaßnahme zu und blieb nachhaltig bestehen. Ersteres wird im Interaktionsdiagramm durch die unterschiedlichen Steigungen sichtbar: Der Anstieg vom Prät- zum Posttest war in der Trainingsgruppe (T+) steiler als in der Nicht-Trainingsgruppe (T-). Letzteres geht aus den nicht abfallenden Leistungen der Kinder hervor (siehe Abbildung 5.10).

Aufgabenspezifische Analyse

Für alle drei Testaufgaben konnte die Steigerung des Lösungserfolgs vom Prä- zum Posttest sowohl für die Trainings- (T+) als auch für die Nicht-Trainingsgruppe (T-) bestätigt werden. Während die Gruppe T+ ihren Lösungserfolg für die drei Aufgaben im Mittel um .38, .48 bzw. .28 erhöhte, gelang das der Gruppe T- im Mittel nur um .29, .23 bzw. 12.²²⁴ Dabei fiel auf, dass die beiden erstgenannten Zuwächse für die trainierten Aufgabentypen, *Vergleichsaufgaben* und *Aufgaben mit komplexen Informationen* größer ausfielen als für den letztgenannten nicht-trainierten Aufgabentyp, *Aufgaben zur Verhältnisverteilung*. Weitere Ergebnisse der deskriptiven Statistik können Anhang G.1 (Tabelle G.2) entnommen werden.

Die für den gesamten Textaufgaben-Performance-Test identifizierte Leistungssteigerung des Lösungserfolgs vom Post- zum Follow-up-Test war lediglich auf die *Aufgabe zur Verhältnisverteilung* zurückzuführen. Aus dem Interaktionsdiagramm geht hervor, dass Kinder, losgelöst davon, ob sie der Gruppe T+ oder T- angehörten, nur bei dieser Aufgabe ihre Leistung vom Post- zum Follow-up-Test steigerten (siehe Abbildung 5.11).

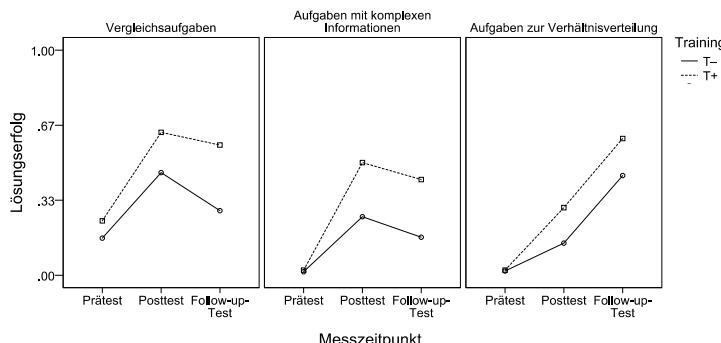


Abbildung 5.11. Mittlere Aufgabenscores des Lösungserfolgs der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten. Skala: 0–1. Trainierte Aufgabentypen = Vergleichsaufgaben, Aufgaben mit komplexen Informationen; nicht-trainierter Aufgabentyp = Aufgaben zur Verhältnisverteilung.

Bei den trainierten Aufgabentypen, der *Vergleichsaufgabe* und der *Aufgabe mit komplexen Informationen*, war dagegen ein Leistungsabfall zu beobachten. Insgesamt zeig-

²²⁴ Die Reihenfolge der Zuwächse berücksichtigt die Aufgabenreihenfolge im Textaufgaben-Performance-Test: 1. Vergleichsaufgabe, 2. Aufgabe mit komplexen Informationen, 3. Aufgabe zur Verhältnisverteilung.

ten die Gruppen über die Zeit annähernd bei jedem Aufgabentyp die gleichen Leistungsentwicklungen, immer zugunsten der Gruppe T+. Den deskriptiven Ergebnissen zufolge schien die Wirkung des Trainings auf allen Stufen des Aufgabentyps gleich zu sein, sodass der Trainingseffekt über die Zeit von der Aufgabe unabhängig und für alle drei Aufgaben gültig war (siehe Abbildung 5.11). Zugleich wurde deutlich, dass sich bei dem nicht-trainierten Aufgabentyp ein anderes Leistungsverhalten als bei den trainierten Aufgabentypen offenbarte.

Inferenzstatistisch konnten die deskriptiven Ergebnisse bestätigt werden. Weder die Interaktion der Faktoren *Training* und *Aufgabentyp*, $\chi^2(2) = 2.60, p = .273$, noch die Interaktion über die *Zeit* (Zeit x Training x Aufgabentyp), $\chi^2(4) = 3.08, p = .544$, wurden signifikant. Folglich war der Trainingseffekt unabhängig von dem zu bearbeitenden Aufgabentyp.

Hypothese 2.2: Drittklässler erzielen einen höheren Problemlösekompetenzzuwachs, wenn sie trainiert werden.

Rein deskriptiv wurde auch mit dem Fokus auf die Problemlösekompetenzen ein Zuwachs in der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) deutlich. Vor der Interventionsmaßnahme lag der erreichte Testscore in beiden Gruppen nahe dem Extremwert 0, sodass entweder die Seite des Knobelhefts leer war, die Informationen der Aufgabe nur herausgeschrieben, nicht aber mit ihnen gearbeitet wurde, oder die Daten bearbeitet wurden, ohne dass ein Problemverständnis sichtbar wurde (vgl. Abschnitt 4.5.1.1). Hierunter fiel außerdem, wenn lediglich eine Lösungszahl, jedoch eine falsche, notiert war. Während die Trainingsgruppe (T+) ihren Testscore vom Prä- zum Posttest mehr als Versechsfachen konnte, gelang es der Nicht-Trainingsgruppe (T-) um knapp das Fünffache. Der Problemlöse-Testscore stieg im erstgenannten Falle durchschnittlich um 1.83, im letztgenannten um 0.95 (siehe Tabelle 5.5).

Tabelle 5.5. Deskriptive Ergebnisse der Trainings- und Nicht-Trainingsgruppe in den Problemlösekompetenzen zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+ ^a	0.35	0.52	2.18	1.14	2.35	1.24
T- ^b	0.25	0.41	1.20	1.11	1.25	1.13

Anmerkung. T+ = Training; T- = kein Training. Skala: 0–4.

^a*N*=172. ^b*N*=155.

Vom Post- zum Follow-up-Test ergab sich in beiden Gruppen keine Verschlechterung (siehe Abbildung 5.12). Wird die Beurteilungsskala der Problemlösekompetenzen her-

angezogen, so stand der von der Trainingsgruppe (T+) zum Zeitpunkt des Posttests erzielte Testscore von 2 im Mittel a) für einen Hefteintrag, der mindestens ein gewisses Verständnis der Problematik aufzeigte, b) für die Nutzung einer geeigneten Repräsentation, die entweder nicht vollständig oder falsch ausgeführt und somit kein oder ein falsches Ergebnis erzielt wurde, oder c) für die richtig notierte Lösungszahl, aber der Lösungsweg fehlte oder nicht nachvollziehbar war (vgl. Abschnitt 4.5.1.1).

Es wurde deutlich, dass hypothesenkonform durch die Interventionsmaßnahmen im Mittel eine Kompetenzsteigerung erfolgte. Die unterschiedlichen Steigungen zwischen den ersten beiden Messzeitpunkten waren ein Indiz für eine signifikante Interaktion, während die nahezu parallele Entwicklung zwischen dem zweiten und dritten Messzeitpunkt auf eine nicht signifikante Interaktion hindeutete (siehe Abbildung 5.12). Zieht man ergänzend den Boxplot heran, so werden die unterschiedlichen Verteilungen der erzielten Testscores sichtbar (siehe Abbildung 5.12). Zum Zeitpunkt des Post- und Follow-up-Tests zählte der kleinste zu erreichende Testscore min = 0 bei der Nicht-Trainingsgruppe zu den mittleren 50 % aller Fälle, wohingegen er bei der Trainingsgruppe nur noch zu den mittleren 25 % aller Fälle gehörte. In der Summe geht aus den Veranschaulichungen hervor, dass die Drittklässler vor der Interventionsmaßnahme über ähnlich ausgeprägte Problemlösekompetenzen verfügten, die Gruppe trainierter Kinder aber eine deutlich größere Kompetenzentwicklung gegenüber der Gruppe untrainierter Kinder erwirken konnten.

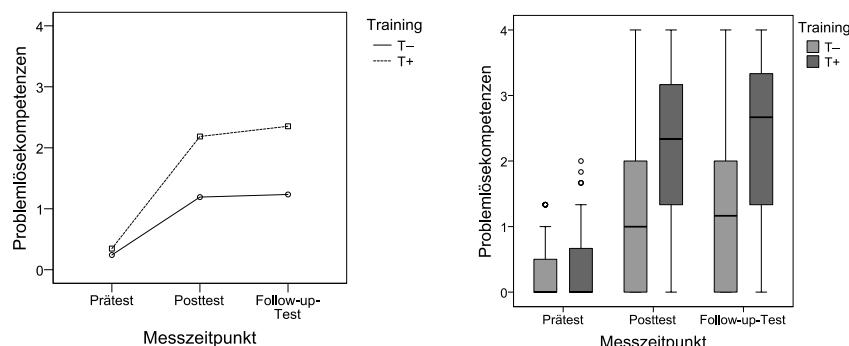


Abbildung 5.12. Mittlerer Testscore der Problemlösekompetenzen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten. Skala: 0–4.

Die gleichsinnige, vorhersagekonforme Veränderung für die Problemlösekompetenzen konnte durch die univariate Varianzanalyse mit Messwiederholung bekräftigt werden. Trainings- (T+) und Nicht-Trainingsgruppe (T-) unterschieden sich signifikant in ihren

Problemlösekompetenzen, $F(1, 323) = 74.87, p < .001, \eta_p^2 = .19$ (Haupteffekt Training). Im paarweisen Vergleich war die Trainingsgruppe der Nicht-Trainingsgruppe überlegen (Mittelwertsdifferenz von $.74, p < .001$).

Dies spiegelte sich auch in der Veränderung über die Zeit wider, $F(1.99, 643.32) = 45.91, p < .001, \eta_p^2 = .12$ (Interaktionseffekt der Faktoren Zeit x Training). Mithilfe der geplanten Kontraste konnte gezeigt werden, dass nur die Kompetenzzuwächse vom ersten zum zweiten Messzeitpunkt, zugunsten der Trainingsgruppe, signifikant voneinander abwichen, $(F(1, 323) = 60.20, p < .001, \eta_p^2 = .16)$. Demzufolge hatten die Gruppe der Kinder, die durch ein Training angeregt wurden, externe Repräsentationen zu generieren und in ihren Lösungsprozess zu integrieren, einen stärkeren Zugewinn an Problemlösekompetenzen zu verzeichnen als die Gruppe der Kinder, die keine Unterstützung durch ein Repräsentationstraining erfuhr (siehe Abbildung 5.12).

Aufgabenspezifische Analyse

Für alle drei Testaufgaben zeichnete sich für die Problemlösekompetenzen die gleiche Entwicklung wie für den Lösungserfolg ab. Bei allen drei Aufgaben war vom Prä- zum Posttest ein Kompetenzzuwachs zu beobachten. Dieser fiel beim trainierten Aufgabentyp, den *Vergleichsaufgaben*, am stärksten aus (siehe Abbildung 5.13).

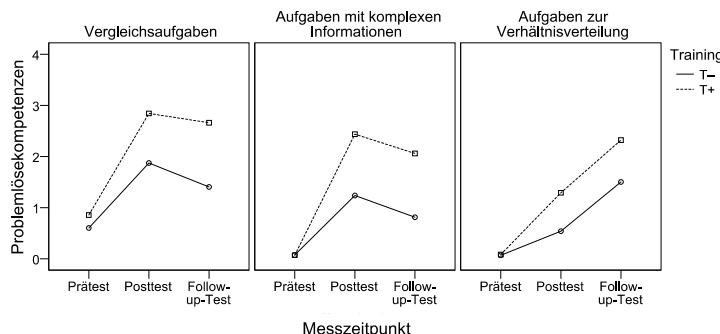


Abbildung 5.13. Mittlere Aufgabenscores der Problemlösekompetenzen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten. Skala: 0–4.

Zusätzlich bestätigte sich bei allen drei Aufgaben, dass die trainierten Kinder den nicht-trainierten überlegen waren. Wird in den Interaktionsdiagrammen der Zeitraum vom Post- zum Follow-up-Test betrachtet, so fällt auf, dass der für den gesamten Test identifizierte Zuwachs vom Post- zum Follow-up-Test letztlich nur auf den nicht-

trainierten Aufgabentyp, die *Aufgabe zur Verhältnisverteilung*, zurückzuführen war. Die Diagramme decken auf, dass die Steigung für diesen Zeitraum bei den trainierten Aufgabentypen leicht abfiel und nur bei dem nicht-trainierten Aufgabentyp stark zunahm. Detaillierte Ergebnisse der deskriptiven Statistik können im Anhang G.1 (Tabelle G.5) eingesehen werden.

Neben der Tatsache, dass sowohl das Training als auch der Aufgabentyp einen eigenständigen Einfluss auf die Problemlösekompetenzen ausübten, ergab die Varianzanalyse mit Messwiederholung einen signifikanten Interaktionseffekt 1. Ordnung (Training x Aufgabentyp): $F(2, 646) = 4.76, p = .009, \eta_p^2 = .02$. Demnach übten die Faktoren *Training* und *Aufgabentyp* zusätzlich zu den signifikanten Haupteffekten einen gemeinsamen Einfluss auf die Problemlösekompetenzen aus. Konkret bedeutet dies, dass unabhängig von der Zeit die Wirkung des Trainings auf die Problemlösekompetenzen nicht auf allen Stufen des Aufgabentyps, sprich allen drei Aufgaben, gleich war. Vor dem Hintergrund, dass die Kompetenzentwicklung über die Zeit und somit die Interaktionen mit dem Faktor *Zeit* im Fokus der Untersuchung stand, werden die an dieser Stelle erforderlichen Kontrastanalysen aus ökonomischen Gründen ausgespart. Aufgrund ihrer Bedeutung für anschließende Untersuchungen werden die Ergebnisse im Anhang G.2.1 ausgeführt.

Über die Zeit gesehen hatte dieser Effekt keinen Bestand (Interaktionseffekt 2. Ordnung (Zeit x Training x Aufgabentyp): $F(3.87, 1248.79) = 2.00, p = .093, \eta_p^2 < .01$). Die nicht signifikante Interaktion zweiter Ordnung zeigte auf, dass die Wechselwirkung zwischen *Zeit* und *Training* von den Stufen des *Aufgabentyps*, sprich den drei Aufgaben, unabhängig war. Die Visualisierung der Dreifachinteraktion in Abbildung 5.13 deutet an, dass für jede der drei Stufen des Faktors *Aufgabentyp* eine Interaktion der Faktoren *Zeit* und *Training* vorlag.

Hypothese 2.3: Drittklässler generieren nach der Interventionsmaßnahme mehr externe Repräsentationen, wenn sie trainiert werden.

Die Betrachtung der Mittelwerte ergab, dass sowohl die trainierte Gruppe (T+) als auch die nicht trainierte Gruppe (T-) im Mittel die Anzahl selbstgenerierter externer Repräsentationen erhöhte (siehe Tabelle 5.6).

Die Problemlösenovizen entwickelten vor der Interventionsmaßnahme selten von sich aus externe Repräsentationen ($M_{T+} = 0.36, SD_{T+} = 0.40$ bzw. $M_{T-} = 0.32, SD_{T-} = 0.39$), was offenlegte, dass sie externalisierte Denk- und Erkenntniswerkzeuge weniger als Unterstützungsmöglichkeit beim Lösen problemhaltiger Textaufgaben in Betracht zogen bzw. sie ihnen in diesem Kontext weniger vertraut waren.

Tabelle 5.6. Deskriptive Ergebnisse der Trainings- und Nicht-Trainingsgruppe in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+ ^a	0.36	0.40	1.56	0.62	1.64	0.63
T- ^b	0.32	0.39	1.11	0.66	1.05	0.65

Anmerkung. T+ = Training; T- = kein Training. Es existiert kein maximaler Testscore, da er die durchschnittliche Summe pro Aufgabe generierter externer Repräsentationen abbildet.

^a*N* = 172, ^b*N* = 155.

Deskriptiv zeigte sich, dass die Kinder der Trainingsgruppe (T+) nach dem Training im Durchschnitt pro Aufgabe mehr externe Repräsentationen generierten als Kinder der Nicht-Trainingsgruppe (T-) (siehe Abbildung 5.14).

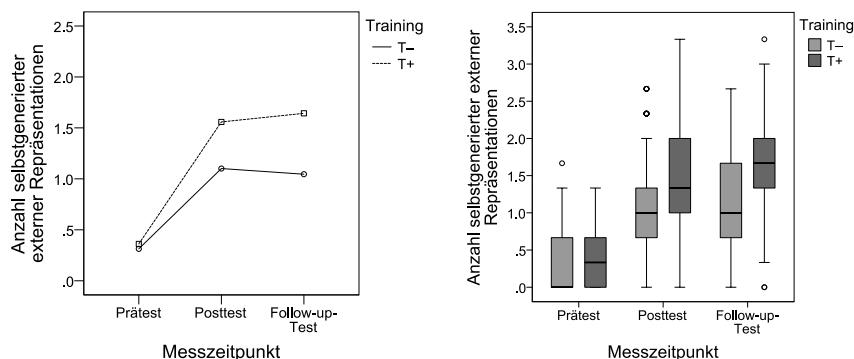


Abbildung 5.14. Mittlerer Testscore der Anzahl selbstgenerierter externer Repräsentationen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.

Den Beobachtungen zufolge erstellten Lernende der Trainingsgruppe nach der Interventionsmaßnahme im Mittel $M = 1.56$ ($SD = 0.62$) externe Repräsentationen pro Aufgabe, Lernende der Nicht-Trainingsgruppe im Mittel $M = 1.11$ ($SD = 0.66$). Dies ist ein Hinweis dafür, dass Kinder der Trainingsgruppe nach der Intervention durchaus auch multiple Repräsentationen generierten und ihre Lösung nicht ausschließlich auf eine Repräsentation stützten, wie es verstärkt bei der Gruppe der nicht-trainierten Kinder auftrat. Dies geht auch aus dem Boxplot hervor (siehe Abbildung 5.14). Im Mittel konstruierten 50 % aller trainierten Kinder nach der Interventionsmaßnahme zwischen

1 und 2 Repräsentationen pro Aufgabe, während der Interquartilbereich der Nicht-Trainingsgruppe im Mittel um 1 Repräsentation streut.

Die Varianzanalyse mit Messwiederholung bestätigte den signifikanten Unterschiede zwischen den beiden Gruppen, $F(1, 323) = 67.15, p < .001, \eta_p^2 = .17$ (Haupteffekt *Training*). Der Faktor *Training* nahm unabhängig vom Faktor *Zeit* und *Kommunikation* Einfluss auf die Anzahl selbstgenerierter externer Repräsentationen. Gemäß dem paarweisen Vergleich war die Trainingsgruppe der Nicht-Trainingsgruppe überlegen (Mittelwertsdifferenz von $.37, p < .001$).²²⁵

Der signifikante Interaktionseffekt der beiden Faktoren *Training* und *Zeit* gab an, dass sich die beiden Gruppen auch über die Zeit in ihrer Anzahl selbstgenerierter externer Repräsentationen unterschieden, $F(2, 646) = 27.82, p < .001, \eta_p^2 = .08$ (siehe Abbildung 5.14). Die statistische Bedeutsamkeit der Unterschiede zwischen den Gruppen konnte durch Kontrastanalysen für den Zeitraum vom ersten zum zweiten Messzeitpunkt, $F(1, 323) = 27.72, p < .001, \eta_p^2 = .08$, nicht jedoch für den vom zweiten zum dritten bestätigt werden, $F(1, 323) = 3.48, p = .063, \eta_p^2 = .01$. Es lässt sich ableiten, dass am Training teilnehmende Kinder nach der Interventionsmaßnahme signifikant mehr externe Repräsentationen generierten als Kinder, die die Aufgaben im regulären Mathematikunterricht bearbeiteten (siehe Abbildung 5.14).

Aufgabenspezifische Analyse

Der Effekt, dass nach der Interventionsmaßnahme sowohl in der Trainingsgruppe als auch in den Nicht-Trainingsgruppe die Anzahl selbstgenerierter externer Repräsentationen zunahm, spiegelte sich bei allen drei Aufgabentypen wider (siehe Abbildung 5.15). Die größte Zunahme war bei dem trainierten Aufgabentyp *Vergleichsaufgaben* zu verzeichnen. Die deskriptiven Ergebnisse sind im Detail im Anhang G.1 (Tabelle G.8) einzusehen. Es zeigte sich, dass die Kinder auch beim untrainierten Aufgabentyp im Mittel mehr externe Repräsentationen generierten, als sie dies noch vor der Interventionsmaßnahme getan hatten. Im Unterschied zu den bisherigen aufgabenspezifischen Analysen fiel auf, dass die Steigerung der Anzahl selbstgenerierter externer Repräsentationen vom Post- zum Follow-up-Test nicht nur auf die *Aufgabe zur Verhältnisverteilung*, sondern ebenso auf die *Vergleichsaufgabe* zurückzuführen war. Trainierte Kinder konstruierten bei dieser Aufgabe auch noch 13 Wochen nach der Interventionsmaßnahme (MZP 3) tendenziell mehr Repräsentationen, als sie dies noch zuvor gemacht hatten (siehe Abbildung 5.15).

²²⁵ Um einer Alpha-Fehler-Kumulierung entgegenzuwirken, wurde die Bonferroni-Korrektur vorgenommen.

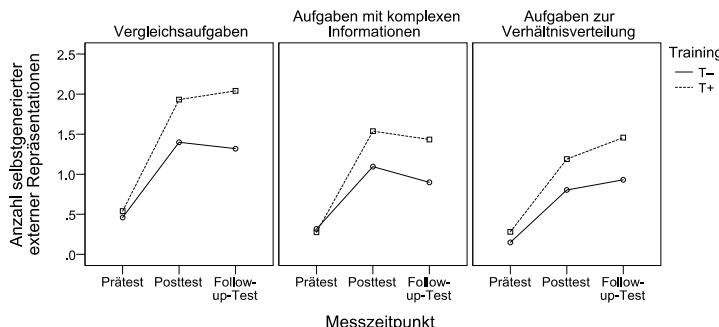


Abbildung 5.15. Mittlere Aufgabenscores der Anzahl selbstgenerierter externer Repräsentationen der Trainings- (T+) und der Nicht-Trainingsgruppe (T-) zu den drei Messzeitpunkten.

Zuvor konnte anhand der signifikanten Haupteffekte der Faktoren *Training* und *Aufgabentyp* gezeigt werden, dass jeder der Faktoren einen eigenständigen Einfluss auf die Anzahl selbstgenerierter externer Repräsentationen ausübt. Die Einflüsse von *Training* und *Aufgabentyp* waren jedoch unabhängig voneinander, d. h. der Einfluss des Trainings auf die Anzahl selbstgenerierter externer Repräsentationen war bei allen Aufgaben vergleichbar groß. Dies kann aus der nur hinreichend signifikanten Interaktion der Faktoren *Training* und *Aufgabentyp* geschlossen werden, $F(1.95, 630.75) = 2.96$ $p = .054$, $\eta_p^2 = .01$.

Auch über die Zeit gesehen konnte der Effekt nicht bestätigt werden (Interaktionseffekt 2. Ordnung der Faktoren Zeit x Training x Aufgabentyp: $F(3.71, 1197.34) = 1.08$, $p = .361$, $\eta_p^2 < .01$). Dem nicht signifikanten Interaktionseffekt zufolge war die Wechselwirkung zwischen *Zeit* und *Training* von den Stufen des *Aufgabentyps*, sprich von den drei Aufgaben, unabhängig. Dies erbrachte den Nachweis, dass der Trainingseffekt über die Zeit für alle Aufgaben gültig war.

Zusammenfassend bleibt festzuhalten, dass trainierte Drittklässler nach der Interventionsmaßnahme in ihren Leistungen untrainierten Drittklässlern überlegen waren. Sie lösten problemhafte Textaufgaben signifikant erfolgreicher, erzielten einen statistisch bedeutsamen höheren Problemlösekompetenzzuwachs und generierten signifikant mehr externe Repräsentationen, sodass die Hypothesen 2.1, 2.2 und 2.3 als bestätigt gelten. Ferner zeigte sich, dass die Effekte des Trainings über die Zeit von den Aufgabentypen unabhängig waren.

5.3.3 Forschungsfrage 3: Effekt der Kommunikation

Die dritte Forschungsfrage fokussiert den Effekt der kommunikativen Zweiersettings. Es gilt zu untersuchen, ob sich die Gruppe, in der sich die Gleichaltrigen untereinander austauschen durften, in ihrem *Lösungserfolg*, ihren *Problemlösekompetenzen* bzw. ihrer *Anzahl selbstgenerierter externer Repräsentationen* von der Gruppe, in der sich Gleichaltrige nicht untereinander austauschen durften, abgrenzt, und wie deren beider Entwicklung verläuft.

Hypothese 3.1: Drittklässler lösen problemhaltige Textaufgaben nach der Interventionsmaßnahme erfolgreicher, wenn sie sich in kommunikativen Settings austauschen.

Mit Blick auf die Ausgangsvoraussetzungen beider Gruppen zum Zeitpunkt des Prätests lässt sich festhalten, dass die Lernenden größtenteils nicht in der Lage waren, die drei problemhaltigen Textaufgaben richtig zu lösen ($M_{K+} = .09$, $SD_{K+} = .17$; $M_{K-} = .08$, $SD_{K-} = .15$) (siehe Tabelle 5.7). Nach der Interventionsmaßnahme konnte vergleichbar zu den trainierten und nicht-trainierten Klassen ein Leistungszuwachs in beiden Gruppen (K+ und K-) verzeichnet werden. Während anfangs weder die Kommunikationsgruppe (K+) noch die Nicht-Kommunikationsgruppe (K-) im Mittel eine Aufgabe korrekt löste, gelang ihnen nach der Interventionsmaßnahme im Mittel eine von drei erfolgreichen Problemlösungen, was einem Testscore von .33 entspricht. Im direkten Vergleich war der Zuwachs in der Gruppe mit kommunikativen Settings (K+) geringfügig höher. Ihr Scoreanstieg lag bei .33, hingegen bei .27 in der Gruppe K-.

Tabelle 5.7. Deskriptive Ergebnisse der Kommunikationsgruppe und der Nicht-Kommunikationsgruppe im Lösungserfolg zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
K+ ^a	.09	.17	.42	.33	.46	.35
K- ^b	.08	.15	.35	.35	.39	.36

Anmerkung. K+ = kommunikative Zweiersettings; K- = keine kommunikativen Zweiersettings. Skala: 0–1.

^a $N=167$. ^b $N=160$.

Inferenzstatistisch ließen sich gemittelt über die Stufen des Faktors *Training* Unterschiede zwischen den Stufenmittelwerten des Faktors *Kommunikation* identifizieren, $F(1, 323) = 4.93$, $p = .027$, $\eta_p^2 = .02$ (Haupteffekt Kommunikation). Der Effekt ist nach Cohen (1988, S. 285) jedoch als klein einzustufen. Die paarweisen Vergleiche

indizierten einen leichten Vorteil zugunsten der Kommunikationsgruppe (K+) (Mittelwertsdifferenz von .55, $p = .027$).²²⁶

Entgegen der Erwartungen verlief die Leistungsentwicklung beider Gruppen hinsichtlich des Lösungserfolges nahezu vergleichbar (siehe Abbildung 5.16).

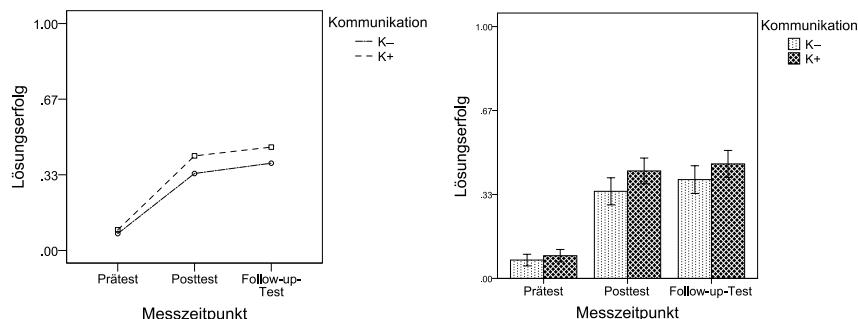


Abbildung 5.16. Mittlerer Testscore des Lösungserfolgs der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte. Skala: 0–1.

Des Weiteren ergab die Varianzanalyse mit Messwiederholung für die Interaktion der Faktoren *Zeit* und *Kommunikation* kein signifikantes Ergebnis, $F(2, 646) = 1.96$, $p = .142$, $\eta_p^2 = .01$. Demzufolge unterschieden sich die beiden Gruppen unabhängig von den Messzeitpunkten zwar in ihrem Lösungserfolg, jedoch existierte über die Zeit hinweg kein Vorteil für die Gruppe mit kommunikativen Zweiersettings (K+). Im Interaktionsdiagramm wird deutlich, dass in beiden Gruppen ein Zuwachs erfolgte, dieser jedoch nahezu parallel verlief, was auf keinen Effekt hindeutete und durch fehlende Signifikanz bestätigt wurde. Die Hypothese musste folglich verworfen werden.

Aufgabenspezifische Analyse

Insgesamt lösten die Drittklässler unmittelbar nach der Interventionsmaßnahme die *Vergleichsaufgabe* und die *Aufgabe mit komplexen Informationen* erfolgreicher als noch vor dem Training, wobei zum Follow-up-Test hin der Erfolg wieder leicht abnahm. Der *Aufgabe zur Verhältnisverteilung* lag ein anderer Verlauf zugrunde als bei den beiden anderen Aufgabentypen: Wenngleich die Kinder zum Zeitpunkt des Posttests erfolgreicher waren als vor der Interventionsmaßnahme und der erzielte Testscore um .17 (K+) bzw. .22 (K-) anstieg, so sank im Anschluss deren Erfolgsscore nicht wie

²²⁶ Um einer Alpha-Fehler-Kumulierung entgegenzuwirken, wurde die Bonferroni-Korrektur vorgenommen.

bei den restlichen Aufgaben, sondern stieg um weitere .25 (K-) bzw. .37 (K+) (siehe Anhang G.1, Tabelle G.3).

Zum Zeitpunkt des Prätests erzielten die Gruppen bei der *Aufgabe mit komplexen Informationen* und der *Aufgabe zur Verhältnisverteilung* im Vergleich zur *Vergleichsaufgabe* vergleichbar niedrige Erfolgsraten (siehe Abbildung 5.17). Kinder, die sich unter Gleichaltrigen austauschen durften, lösen nach der Interventionsmaßnahme die trainierten Aufgaben, die *Vergleichsaufgabe* und die *Aufgabe mit komplexen Informationen*, erfolgreicher als Kinder, die sich nicht unter Gleichaltrigen austauschen durften.

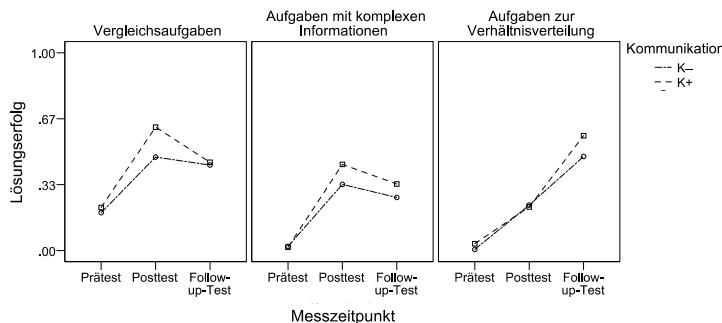


Abbildung 5.17. Mittlere Aufgabenscores des Lösungserfolgs der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte. Skala: 0–1.

Vorangestellte Analysen machten deutlich, dass die Faktoren *Kommunikation* und *Aufgabentyp* eigenständige Einflüsse auf den Lösungserfolg nahmen. Aufgrund des nicht signifikanten Interaktionseffektes ist davon auszugehen, dass diese unabhängig voneinander sind ($\chi^2(2) = 2.97, p = .226$). Folglich ist der Einfluss der Kommunikation auf den Lösungserfolg bei allen Aufgaben vergleichbar groß. Jedoch ergab sich ein signifikanter Interaktionseffekt über die *Zeit* (Interaktionseffekt 2. Ordnung der Faktoren *Zeit* x *Kommunikation* x *Aufgabentyp*: $\chi^2(4) = 12.46, p = .014$). Hieraus lässt sich schließen, dass die Interaktion der Faktoren *Zeit* und *Kommunikation* von den Stufen des *Aufgabentyps* abhängig war.

Wenn man die Tatsache berücksichtigt, dass die Interaktion 1. Ordnung der Faktoren *Zeit* und *Kommunikation* nicht signifikant war, dann weiß man durch die signifikante Interaktion 2. Ordnung, dass die Interaktion der Faktoren *Zeit* und *Kommunikation* nicht auf allen Stufen des *Aufgabentyps* ohne Wirkung war. Konkret übten die Faktorstufen *Vergleichsaufgabe*, *Posttest* und *Kommunikationsgruppe* eine gemeinsame und gleichzeitige Wirkung auf den Lösungserfolg aus, $\chi^2(1) = 9.02, p = .003$, $OR = 3.10$.

Demzufolge differierten die Gruppe K+ und K- bei der Vergleichsaufgabe zum Zeitpunkt des Posttests signifikant in ihrem Lösungserfolg. Weitere signifikante Effekte blieben aus.

Hypothese 3.2: Drittklässler erzielen einen höheren Problemlösekompetenzzuwachs, wenn sie sich in kommunikativen Settings austauschen.

Wenn die Problemlösekompetenzen zum Prätest in Tabelle 5.8 betrachtet werden, dann stehen die niedrigen Testscores für mangelnde Ausgangsvoraussetzungen. Die Kommunikationsgruppe (K+) erzielte zum Prätest $M_{K+} = 0.32$ ($SD_{K+} = 0.47$) von 4 Punkten, wohingegen die Nicht-Kommunikationsgruppe (K-) $M_{K-} = 0.28$ ($SD_{K-} = 0.48$) von 4 Punkten erreichte.

Tabelle 5.8. Deskriptive Ergebnisse der Kommunikationsgruppe und der Nicht-Kommunikationsgruppe in den Problemlösekompetenzen zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
K+ ^a	0.32	0.47	1.83	1.20	2.00	1.26
K- ^b	0.28	0.48	1.60	1.25	1.65	1.34

Anmerkung. K+ = kommunikative Zweiersettings; K- = keine kommunikativen Zweiersettings. Skala: 0–4.

^a $N = 167$. ^b $N = 160$.

Nach der Interventionsmaßnahme war ein Kompetenzzuwachs in beiden Gruppen vorzufinden. Dieser fiel rein deskriptiv bei den Kindern, die sich unter Gleichaltrigen austauschen durften, größer aus (Posttest: $M_{K+} = 1.83$, $SD_{K+} = 1.20$ im Vergleich zu $M_{K-} = 1.60$, $SD_{K-} = 1.25$). Zudem bleibt festzuhalten, dass der Kompetenzzuwachs über die Zeit bis zum Follow-up-Test nicht absank, sogar minimal anstieg (siehe Abbildung 5.18). Insgesamt war in beiden Gruppen eine positive Leistungsentwicklung zu beobachten, wobei der Kommunikationsgruppe (K+) ein Vorteil zugesprochen werden konnte.

Die Varianzanalyse mit Messwiederholung brachte den Nachweis, dass signifikante Unterschiede zwischen den Gruppe K+ und K- hinsichtlich der Problemlösekompetenzen existierten, $F(1, 323) = 7.68$, $p = .006$, $\eta^2_p = .02$. Der Kompetenzzuwachs war in der Kommunikationsgruppe größer als in der Nicht-Kommunikationsgruppe (Mittelwertsdifferenz .24, $p = .006$) (siehe Abbildung 5.18). Wie auch die deskriptiven Werte in Tabelle 5.8 verdeutlichen, spiegelte sich die Entwicklung beider Gruppen über die Zeit in einem signifikanten Interaktionseffekt der Faktoren *Kommunikation* und *Zeit*

wider, $F(1.99, 643.32) = 4.58, p = .011, \eta_p^2 = .01$. Dieser Effekt ist jedoch nach Cohen (1988, S. 285) als niedrig einzustufen ($\eta_p^2 \leq .01$).

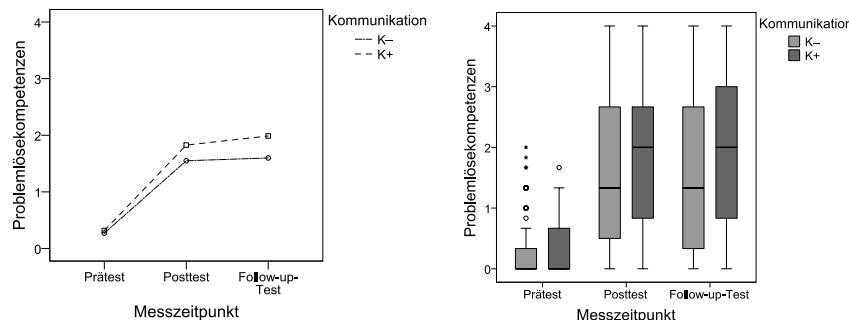


Abbildung 5.18. Mittlerer Testscore der Problemlösekompetenzen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppen (K-) über drei Messzeitpunkte. Skala: 0–4.

Festzuhalten bleibt, dass sich ein Vorteil für die Entwicklung derjenigen ergab, die der Kommunikationsgruppe (K+) angehörten (siehe Abbildung 5.18). Signifikant wurde die Kompetenzsteigerung für den Zeitraum vom Prä- zum Posttest, $F(1, 323) = 4.06, p = .045, \eta_p^2 = .01$, nicht jedoch für den Zeitraum vom Post- zum Follow-up-Test, $F(1, 323) = 1.05, p = .306, \eta_p^2 < .01$. Die Hypothese 3.2 gilt demzufolge als bestätigt.

Aufgabenspezifische Analyse

Vor dem Hintergrund, dass der Lösungserfolg und die Problemlösekompetenzen sehr eng zusammenhängen, war es nicht verwunderlich, dass sich die Befunde, die zuvor für die abhängige Variable *Lösungserfolg* gefunden wurden, auf die abhängige Variable *Problemlösekompetenzen* replizieren ließen. Den Interaktionsdiagrammen zufolge gab es für alle drei Aufgabentypen zwischen beiden Faktoren vom Prä- zum Posttest einen positiv linearen Zusammenhang, der bei der *Aufgabe zur Verhältnisverteilung* vom Post- zum Follow-up-Test Bestand hatte. Für den Zeitraum vom Post- zum Follow-up-Test lag bei den trainierten Aufgabentypen ein negativ linearer Zusammenhang vor (siehe Abbildung 5.19).

Wenn die Kompetenzzuwächse, die unmittelbar nach dem Training erzielt wurden, untereinander verglichen werden, dann fielen diese für die trainierten Aufgaben größer aus als für die untrainierten. Die nahezu parallelen Verläufe lassen vermuten, dass innerhalb eines Aufgabentyps keine Unterschiede zwischen den K+ und K- Gruppen feststellbar waren. Für eine ausführliche Darstellung der deskriptiven Ergebnisse wird auf Anhang G.1 (Tabelle G.6) verwiesen.

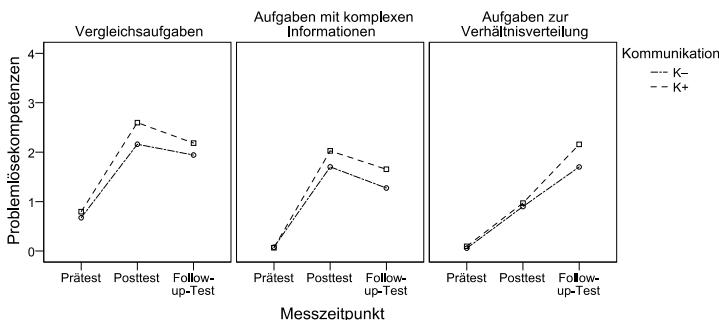


Abbildung 5.19. Mittlere Aufgabenscores der Problemlösekompetenzen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte. Skala: 0–4.

Dass die Drittklässler bei den drei Testaufgaben unterschiedliche Problemlösekompetenzen erzielten, jedoch der Faktor *Kommunikation* diesen Effekt kaum beeinflusste, ließ sich inferenzstatistisch bestätigen. Weder die Interaktion 1. Ordnung (Kommunikation x Aufgabentyp), $F(2, 646) = 0.22, p = .804, \eta_p^2 < .01$, noch die Interaktion 2. Ordnung (Zeit x Kommunikation x Aufgabentyp), $F(3.87, 1248.79) = 2.08, p = .084, \eta_p^2 = .01$, wurden signifikant. Über den gesamten Zeitraum betrachtet zeigte die Gruppe K+ und die Gruppe K- bei jedem der drei Aufgabentypen ein nahezu paralleles Kompetenzprofil. Aufgrund fehlender Signifikanz war davon auszugehen, dass die voneinander abweichende Entwicklung der beiden Gruppen über die Zeit, welche aus dem signifikanten Interaktionseffekt *Zeit* und *Kommunikation* hervorging, unabhängig von der zu bearbeitenden Aufgabe war.

Hypothese 3.3: Drittklässler, die sich in kommunikativen Settings austauschen, unterscheiden sich in der Anzahl selbstgenerierter externer Repräsentationen nicht von Drittklässlern, die sich nicht in kommunikativen Settings austauschen.

Die Drittklässler beider Gruppen griffen vor der Interventionsmaßnahme selten von sich aus auf die Konstruktion externer Repräsentationen zurück (siehe Tabelle 5.9). Die Anzahl lag im Mittel bei $M_{K+} = 0.39 (SD_{K+} = 0.40)$ bzw. $M_{K-} = 0.29 (SD_{K-} = 0.38)$. Nach der Interventionsmaßnahme nahm der Rückgriff auf externe Repräsentationen als Denk- und Erkenntniswerkzeuge zu, der Unterschied beider Gruppen war jedoch klein, $M_{K+} = 1.39 (SD_{K+} = 0.64)$ bzw. $M_{K-} = 1.30 (SD_{K-} = 0.71)$.

Tabelle 5.9. Deskriptive Ergebnisse der Kommunikationsgruppe und der Nicht-Kommunikationsgruppe in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
K+ ^a	0.35	0.40	1.39	0.64	1.40	0.66
K- ^b	0.29	0.38	1.30	0.71	1.33	0.75

Anmerkung. K+ = kommunikative Zweiersettings; K- = keine kommunikativen Zweiersettings. Es existiert kein maximaler Testscore, da er die durchschnittliche Summe pro Aufgabe generierter externer Repräsentationen abbildet.

^a*N* = 167. ^b*N* = 160.

Die deskriptivstatistischen Ergebnisse deuten auf ein parallel verlaufendes Kompetenzprofil hin, sodass keine signifikante Interaktion erwartet wurde. Die Veranschaulichungen der Zusammenhänge im Boxplot sowie im Interaktionsdiagramm stützen diese Vermutung. Außer, dass die Werte der Gruppe K- stärker streuten, waren die Mediane beider Gruppen nach der Interventionsmaßnahme fast identisch (siehe Abbildung 5.20).

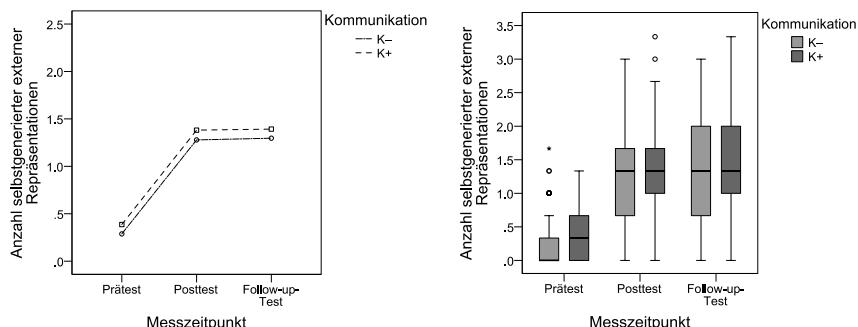


Abbildung 5.20. Mittlerer Testscore der Anzahl selbstgenerierter externer Repräsentationen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte.

Varianzanalytisch ergab der signifikante Haupteffekt *Kommunikation*, $F(1, 323) = 4.93$, $p = .027$, $\eta_p^2 = .02$, dass Unterschiede zwischen den beiden Gruppen vorhanden waren. Über die Zeit hinweg bestätigte sich dieser Unterschied nicht, $F(2, 646) < 0.01$, $p = .996$, $\eta_p^2 < .001$ (Interaktionseffekt der Faktoren

Zeit x Kommunikation). Vielmehr deutete der nicht signifikante Interaktionseffekt der Faktoren *Kommunikation* und *Zeit* eindeutig auf eine parallele Entwicklung in der Generierung und Nutzung externer Repräsentationen hin.

Folglich nahm die Kommunikation einen eigenständigen Einfluss auf die Anzahl selbstgenerierter externer Repräsentationen und war unabhängig von der Zeit. Der auf die Kommunikation zurückzuführende Effekt war bei allen drei Messzeitpunkten gleich groß.

Die Hypothese 3.3, dass die Wechselwirkung der Faktoren *Kommunikation* und *Zeit* keinen Einfluss auf die Anzahl selbstgenerierter externer Repräsentationen nimmt, gilt somit als bestätigt.

Aufgabenspezifische Analyse

Den Interaktionsdiagrammen zufolge gab es innerhalb eines Aufgabentyps keine Unterschiede zwischen der Kommunikationsgruppe (K+) und der Nicht-Kommunikationsgruppe (K-). Sowohl bei der *Vergleichsaufgabe* und der *Aufgabe mit komplexen Informationen* als auch bei der *Aufgabe zur Verhältnisverteilung* wird in Abbildung 5.21 der parallele Verlauf sichtbar. Die Geraden beider Gruppen weisen vergleichbare Steigungen auf.

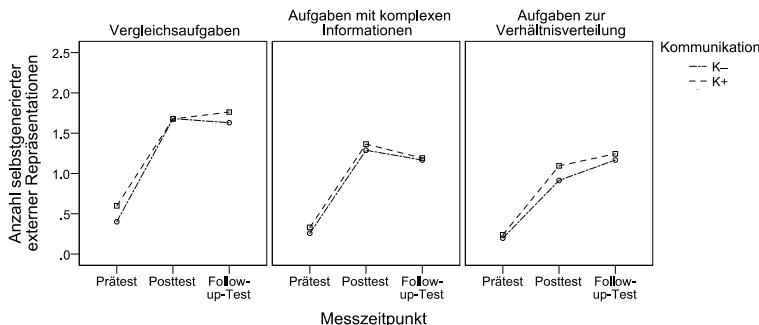


Abbildung 5.21. Mittlere Aufgabenscores der Anzahl selbstgenerierter externer Repräsentationen der Kommunikations- (K+) und der Nicht-Kommunikationsgruppe (K-) über drei Messzeitpunkte.

Die Vergleiche zwischen den Aufgabentypen zeigten, dass die Kinder bei allen drei Aufgaben nach der Interventionsmaßnahme mehr Repräsentationen generierten als davor. Die Anzahl nahm vom Prä- zum Posttest für die *Vergleichsaufgabe* um 1.09 (K+) bzw. 1.28 (K-), für die *Aufgabe mit komplexen Informationen* um 1.03 (K+) bzw. 1.04 (K-) und für die *Aufgabe zur Verhältnisverteilung* um 0.87 (K+) bzw. 0.72 (K-)

zu. Vom Post- zum Follow-up-Test stieg bei der *Aufgabe zur Verhältnisverteilung* die Anzahl selbstgenerierter externer Repräsentationen erneut um 0.14 (K+) bzw. 0.26 (K-), während sie bei der *Aufgabe mit komplexen Informationen* zurückging. Bei der *Ver gleichsaufgabe* konnte die Gruppe K+ die Anzahl minimal steigern, dagegen sank sie in der Gruppe K- geringfügig (siehe Anhang G.1, Tabelle G.9).

Die Parallelität der beiden Gruppen bezüglich der Anzahl selbstgenerierter externer Repräsentationen spiegelte sich auch auf aufgabenspezifischer Ebene wider. Der nicht signifikante Interaktionseffekt der Faktoren *Kommunikation* und *Aufgabentyp* gibt an, dass der Aufgabeneinfluss auf die Problemlösekompetenzen bei der Kommunikations- und der Nicht-Kommunikationsgruppe gleich war. Weder die Interaktion 1. Ordnung (*Kommunikation* x *Aufgabentyp*: $F(1.95, 630.75) = 0.45, p = .634, \eta_p^2 < .01$) noch die Interaktion 2. Ordnung (*Kommunikation* x *Aufgabentyp* x *Zeit*: $F(3.71, 1248.79) = 1.54, p = .193, \eta_p^2 = .01$) waren signifikant.

Es bleibt festzuhalten, dass sich die Kommunikationsgruppe (K+) und die Nicht-Kommunikationsgruppe (K-) hypothesenkonform über die Zeit hinweg in ihren *Problemlösekompetenzen*, nicht jedoch in der *Anzahl selbstgenerierter externer Repräsentationen* signifikant voneinander unterschieden. Dagegen konnte der erwartete Gruppenunterschied in Bezug auf den *Lösungserfolg* nicht nachgewiesen werden. Das Hinzu ziehen der aufgabenspezifischen Ergebnisse legte dar, dass der Kommunikationseffekt über die Zeit von den drei Aufgabentypen unabhängig war.

5.3.4 Forschungsfrage 4: Effekt unterschiedlicher Gruppen

Ob sich die Lernenden der vier experimentellen Bedingungen in ihren Leistungen (Lösungserfolg, Problemlösekompetenzen, Anzahl selbstgenerierter externer Repräsentationen) voneinander unterscheiden, stand im Zentrum des Interesses der vierten Forschungsfrage. Es galt zu untersuchen, ob die Gruppe, die eine Kombination aus Training und Kommunikation erhält, über die Zeit dominiert.

Hypothese 4.1: Trainierte Drittklässler, die sich zusätzlich unter Gleichaltrigen austauschen, lösen problemhafte Textaufgaben nach der Interventionsmaßnahme am erfolgreichsten.

Hypothese 4.2: Drittklässler, die weder trainiert wurden noch sich unter Gleichaltrigen austauschen, steigern nach der Interventionsmaßnahme ihren Lösungserfolg am geringsten.

Die Drittklässler der vier experimentellen Bedingungen zeigten hypothesenkonform kein paralleles Lösungsprofil. Von allen Gruppen erreichten die Kinder der Bedingung T+K+ vom Prä- zum Posttest mit .41 den größten und die Kinder der T-K- Bedingung mit .19 den geringsten Leistungsanstieg (siehe Tabelle 5.10).

Tabelle 5.10. Deskriptive Ergebnisse der vier experimentellen Bedingungen im Lösungserfolg zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+K+ ^a	.10	.18	.51	.33	.57	.35
T+K- ^b	.10	.18	.44	.37	.51	.38
T-K+ ^c	.08	.15	.32	.31	.34	.31
T-K- ^d	.05	.12	.24	.29	.26	.27

Anmerkung. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings. Skala: 0–1.

^a*N* = 85. ^b*N* = 87. ^c*N* = 82. ^d*N* = 73.

Dies spiegelte sich auch in den prozentualen Häufigkeiten wider (siehe Abbildung 5.22). 18 % der Drittklässler der Gruppe T+K+ lösten nach der Intervention keine der drei Aufgaben richtig, hingegen blieben 51 % der Drittklässler der Gruppe T-K- erfolglos.

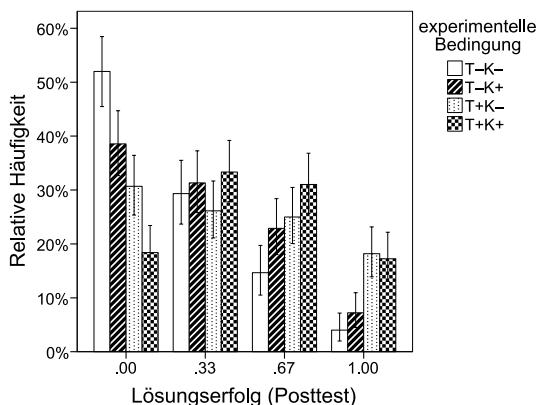


Abbildung 5.22. Relative Häufigkeiten des Lösungserfolgs-Testscores zum Zeitpunkt des Posttests.
 T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

Auffällig war auch, dass die Wahrscheinlichkeit, einen hohen Testscore (.67 bzw. 1.00) zu erreichen, in der Kontrollgruppe T-K- wesentlich geringer war als in den an-

deren drei experimentellen Bedingungen. Nach der Interventionsmaßnahme kam die Gruppe der trainierten Kinder, die sich austauschen durften (T+K+), im Vergleich zu den anderen Bedingungen deutlich häufiger zu zwei korrekten Problemlösungen (siehe Abbildung 5.22).

Beim Lösen aller drei problemhaltigen Textaufgaben waren lediglich 4 % der Kinder der Kontrollgruppe erfolgreich. In den trainierten Gruppen T+K- und T+K+ wurde dieser Testscore von rund 20 % der Probanden erzielt.

Dem Boxplot zufolge lässt sich unmittelbar nach der Interventionsmaßnahme die Gruppe T-K- als die schwächste Gruppe identifizieren (siehe Abbildung 5.23). Die Gruppen T+K+ und T+K- wiesen zum Zeitpunkt des Posttests eine gleiche Verteilung der Daten auf. Ihre erzielten Scores streuten stärker als in der Gruppe T+K+, welche zum Posttest klar dominierte. 13 Wochen nach der Interventionsmaßnahme (Follow-up-Test) glichen sich die Verteilungen der trainierten Klassen, wobei der Median der Gruppe T+K+ am größten war.

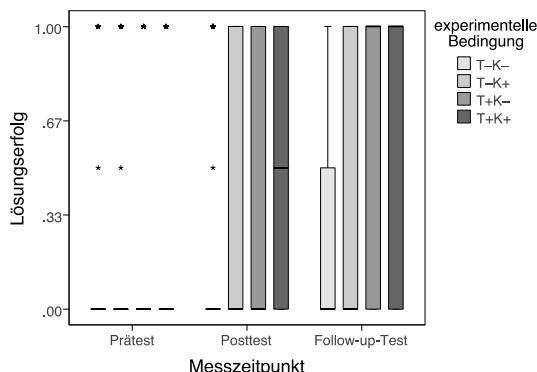


Abbildung 5.23. Verteilung des mittleren Lösungserfolgs für die vier experimentellen Bedingungen. Skala: 0–1. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

Der deskriptiv identifizierte Vorteil zugunsten der Kinder, die trainiert wurden und sich austauschen durften (T+K+), konnte varianzanalytisch nicht bestätigt werden (Interaktionseffekt 1. Ordnung (Training x Kommunikation): $F(1, 323) = 0.08, p = .781, \eta_p^2 < .001$; Interaktionseffekt 2. Ordnung (Training x Kommunikation x Zeit): $F(2, 646) = 0.03, p = .968, \eta_p^2 < .001$). Die nicht signifikante Interaktion erster Ordnung gibt an, dass das Repräsentationstraining nicht zu höherem Lösungserfolg führte,

wenn zusätzlich intersubjektiver Austausch gefördert wurde. Dass die Interaktion zwischen *Zeit* und *Training* von den Stufen der *Kommunikation* unabhängig war, geht aus der nicht signifikanten Interaktion zweiter Ordnung hervor und wird in Abbildung 5.24 am ähnlichen Verlaufsmuster sichtbar. Hieraus folgt, dass sich die vier Gruppen im Mittel nicht signifikant voneinander unterschieden, weshalb die Hypothesen 4.1 und 4.2 abzulehnen sind.

Insofern waren Kinder, die sich zusätzlich zum Training unter Gleichaltrigen austauschen durften ($T+K+$), trainierten Kindern, die sich nicht unter Gleichaltrigen austauschen durften ($T+K-$), über die Zeit im Lösungserfolg nicht überlegen. Folglich lösten trainierte Kinder mit und ohne Austausch den Textaufgaben-Performance-Test nahezu übereinstimmend. Analoges konnte bei den Nicht-Trainingsgruppen beobachtet werden: Die Veränderung des Lösungserfolgs über die Zeit war bei untrainierten Kindern, die sich austauschen durften ($T-K+$), annähernd gleich groß wie bei Kindern, die sich nicht austauschen durften ($T-K-$) (siehe Abbildung 5.24). In beiden Fällen war der beobachtete minimale Vorteil für die Kommunikationsgruppen nicht statistisch bedeutsam.



Abbildung 5.24. Grafische Darstellung der Interaktion 2. Ordnung (Zeit x Training x Kommunikation) für den Lösungserfolg. Für jede Stufe des Faktors Training (kein Training/Training) ist die bedingte Interaktion Zeit x Kommunikation dargestellt. Skala: 0–1.

Aufgabenspezifische Analyse

Die aufgabenspezifischen Analysen machten transparent, dass die vier experimentellen Bedingungen innerhalb eines Aufgabentyps ein ähnliches Lösungsverhalten aufwiesen. Es fiel auf, dass die *Vergleichsaufgabe* die leichteste Testaufgabe darstellte. Im Vergleich zu den anderen beiden Aufgaben wurde sie vereinzelt bereits vor der Interventionsmaßnahme erfolgreich gelöst. Die mittleren Testscores zum Zeitpunkt des

Prätests streuten zwischen .15 und .26, wohingegen sie bei den beiden anderen Testaufgaben lediglich zwischen .00 und .04 streuten. Darüber hinaus fällt auf, dass die Gruppe T+K- ihre Erfolgsrate vom Post- zum Follow-up-Test erneut steigerte (siehe Abbildung 5.25). Für eine detaillierte, tabellarische Darstellung der deskriptiven Ergebnisse wird auf Anhang G.1 (Tabelle G.4) verwiesen.

Bei allen Aufgabentypen nahm der Lösungserfolg vom Prä- zum Posttest zu, fiel zum Follow-up-Test bei der *Vergleichsaufgabe* (mit Einschränkung auf die Gruppe T+K-) und der *Aufgabe mit komplexen Informationen* wieder ab, während die Kinder, gleich welcher der vier experimentellen Bedingungen sie angehörten, bei der *Aufgabe zur Verhältnisverteilung* ihren Erfolg zum Follow-up-Test hin erneut steigern konnten.

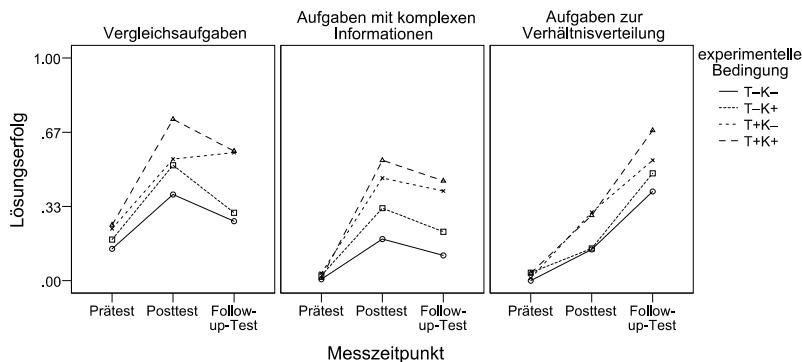


Abbildung 5.25. Grafische Darstellung der Interaktion 3. Ordnung (Zeit x Training x Kommunikation x Aufgabentyp) für den Lösungserfolg. Skala: 0–1. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

Aus dem Interaktionsdiagramm geht hervor, dass die Kinder bei jeder Aufgabe vom Prä- zum Posttest einen Zuwachs erzielten, der bis auf die *Aufgabe zur Verhältnisverteilung* immer zugunsten der Gruppe T+K+ ausfiel. Der zuvor berichtete Befund, dass sich die vier Gruppen, auch über die Zeit gesehen, im Mittel nicht signifikant voneinander unterschieden, hatte für jeden Aufgabentyp Bestand. Die Wirkung der Interventionsmaßnahme auf den Lösungserfolg war bei allen Aufgaben nahezu gleich, was auf keinen Effekt hindeutete und durch fehlende Signifikanz bestätigt wurde (Interaktion 2. Ordnung (Training x Kommunikation x Aufgabentyp): $\chi^2(2) = 2.97, p = .226$, Interaktion 3. Ordnung (Zeit x Training x Kommunikation x Aufgabentyp): $\chi^2(4) =$

.90, $p = .925$). Demzufolge schien der Aufgabentyp den Effekt der Interventionsmaßnahmen über die Zeit nicht signifikant zu beeinflussen.

Hypothese 4.3: Trainierte Drittklässler, die sich zusätzlich unter Gleichaltrigen austauschen, erzielen nach der Interventionsmaßnahme den größten Problemlösekompetenzzuwachs.

Hypothese 4.4: Drittklässler, die weder trainiert werden noch sich unter Gleichaltrigen austauschen, erzielen nach der Interventionsmaßnahme den geringsten Problemlösekompetenzzuwachs.

Die beim Lösungserfolg beobachtete Leistungsentwicklung wurde auch bei den Problemlösekompetenzen erkennbar. Alle vier Gruppen erhöhten durch die Interventionsmaßnahme ihre Problemlösekompetenzen. Die deutlichste Zunahme war hypothesenkonform in der Trainingsgruppe mit kommunikativen Zweiersettings (T+K+) und die niedrigste in der Kontrollgruppe T-K- zu beobachten (Posttest: $M_{T+K+} = 2.27$, $SD_{T+K+} = 1.05$; $M_{T-K-} = 1.00$, $SD_{T-K-} = 1.00$). Darüber hinaus fiel auf, dass die Trainingsgruppen mit und ohne Kommunikation in ihrem Kompetenzzuwachs annähernd vergleichbar waren, ein leichter Vorteil jedoch zugunsten der Gruppe mit Kommunikation bestand (Mittelwertsdifferenzen von 1.94 bzw. 1.74). Zudem hoben sich die beiden Gruppen T+K+ und T+K- von den Nicht-Trainingsgruppen T-K+ und T-K- ab (Mittelwertsdifferenzen von 1.08 bzw. 0.82) (siehe Tabelle 5.11).

Tabelle 5.11. Deskriptive Ergebnisse der vier experimentellen Bedingungen in den Problemlösekompetenzen zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+K+ ^a	0.33	0.49	2.27	1.05	2.52	1.14
T+K- ^b	0.36	0.56	2.10	1.22	2.19	1.32
T-K+ ^c	0.30	0.46	1.38	1.17	1.46	1.16
T-K- ^d	0.18	0.35	1.00	1.01	1.01	1.05

Anmerkung. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings. Skala: 0–4.

^a $N = 85$. ^b $N = 87$. ^c $N = 82$. ^d $N = 73$.

Die Analyse der Problemlösekompetenzen zum Zeitpunkt des Posttests ergab, dass nur noch 2 % der Kinder der Gruppe T+K+ und 3 % der Gruppe T+K- einen Problemlösekompetenz-Testscore von *min* = 0.00 aufwiesen und demzufolge ihre Kompetenzen durch das Training steigern konnten. Demgegenüber erzielten noch 28 % (T-K+) bzw.

33 % (T-K-) der Kinder der Nicht-Trainingsgruppen den Minimalwert $min = 0.00$. Der Maximalwert $max = 4.00$, der besagt, dass bei allen drei Aufgaben mithilfe einer geeigneten externen Repräsentation das richtige Ergebnis ermittelt werden konnte, wurde von drei experimentellen Bedingungen (T+K+: 5 %, T+K-: 9 %, T-K+: 5 %) erreicht, nicht jedoch von der Kontrollgruppe (T-K-: 0 %).

Die Förderung der Problemlösekompetenzen schien wie erwartet in den Trainingsgruppen effektiver zu sein als in der Kontrollgruppe T-K-, wobei auch die Nicht-Trainingsgruppe, in der kommunikative Zweiersettings stattfanden (T-K+), teilweise zufriedenstellende Testscores erzielte. Der Boxplot zeigt auf, dass sich nach der Interventionsmaßnahme sowohl die Verteilungen der Trainingsgruppen (T+K+ und T+K-) als auch die Verteilungen der Nicht-Trainingsgruppen (T-K+ und T-K-) ähneln. Tendenziell ging aus beiden Vergleichen ein Vorteil für die Gruppen mit kommunikativen Zweiersettings hervor, sodass die vier Stufen der experimentellen Bedingung in eine Rangfolge gebracht werden konnten. Aufsteigend ergab sich: T-K-, T-K+, T+K-, T+K+ (siehe Abbildung 5.26).

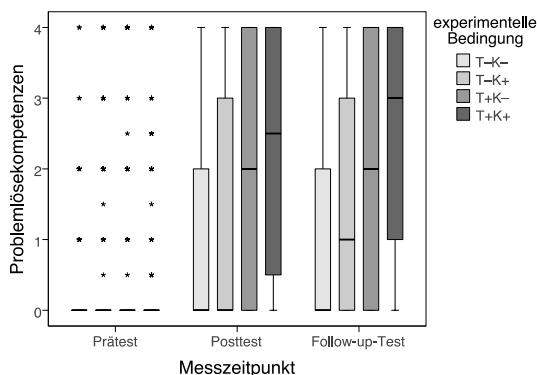


Abbildung 5.26. Verteilung der mittleren Problemlösekompetenzen für die vier experimentellen Bedingungen. Skala: 0–4. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

Demnach gelangen untrainierten Kindern, die sich in kommunikativen Zweiersettings austauschen durften (T-K+), über die Zeit in etwa gleich große Kompetenzzuwächse wie untrainierten Kindern, die sich nicht austauschen durften (T-K-) (siehe Abbildung 5.27). Der minimal zu beobachtende Vorsprung für die Gruppe mit kommunikativen Zweiersettings bestand bereits zum Prätest und konnte zum Posttest nur geringfügig ausgebaut werden. Hervorzuheben ist, dass trainierte Kinder, gleich ob mit oder ohne

kommunikative Zweiersettings, fähig waren, ihre Problemlösekompetenzen über die Zeit annähernd in gleicher Weise auszubauen. Zum Follow-up-Test konnten beide Trainingsgruppen T+K+ und T+K- ihren Zuwachs leicht ausbauen, während es den Nicht-Trainingsgruppen T-K- und T-K+ nicht gelang.

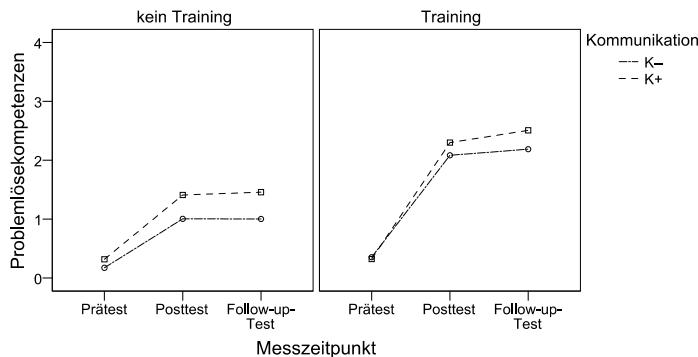


Abbildung 5.27. Grafische Darstellung der Interaktion 2. Ordnung (Zeit x Training x Kommunikation) für die Problemlösekompetenzen. Für jede Stufe des Faktors Training (kein Training/Training) ist die bedingte Interaktion Zeit x Kommunikation dargestellt. Skala: 0–4.

Den Ergebnissen der Varianzanalyse mit Messwiederholung zufolge unterschieden sich losgelöst von den Messzeitpunkten die vier Gruppen nicht signifikant voneinander, $F(1, 323) = 0.85, p = .358, \eta_p^2 < .01$ (Interaktionseffekt der Faktoren Training x Kommunikation). Folglich war davon auszugehen, dass der Effekt des *Trainings* auf die Problemlösekompetenz nicht durch die *Kommunikation* bestärkt wurde. Zwar zeigte sich deskriptiv in der zeitlichen Entwicklung eine Überlegenheit der Gruppe T+K+ und eine Unterlegenheit der Gruppe T-K-, die Diskrepanz wurde jedoch nicht signifikant (Interaktion 2. Ordnung (Training x Kommunikation x Zeit): $F(1.99, 643.32) = 0.09, p = .914, \eta_p^2 < .01$). Die Interaktionsdiagramme lassen ähnliche Verlaufsmuster für die Gruppen mit Kommunikation und die Gruppen ohne Kommunikation deutlich werden (siehe Abbildung 5.27). Demzufolge war die Interaktion zwischen *Zeit* und *Training* von den Stufen der *Kommunikation* in Bezug auf die Problemlösekompetenzen unabhängig. Eine Kombination der Faktoren *Training* und *Kommunikation* über die *Zeit* brachte keinen Zugewinn in den Problemlösekompetenzen.

Obwohl deskriptiv gezeigt werden konnte, dass die Kontrollgruppe T-K- nach der Interventionsmaßnahme erwartungsgemäß den geringsten und die Gruppe T+K+ den höchsten Kompetenzzuwachs erreichte, müssen aufgrund fehlender Signifikanz die Hypothesen 4.3 und 4.4 verworfen werden.

Aufgabenspezifische Analyse

Ziel der aufgabenspezifischen Analyse war es, zu überprüfen, ob der *Aufgabentyp* die Interaktion der Faktoren *Training* und *Kommunikation* bzw. die Interaktion der Faktoren *Zeit*, *Training* und *Kommunikation* beeinflusste. Der grafischen Darstellung der Interaktion 3. Ordnung (siehe Abbildung 5.28) zufolge war die signifikante Interaktion der Faktoren *Zeit* und *Training* von den Faktoren *Kommunikation* und *Aufgabentyp* unabhängig.

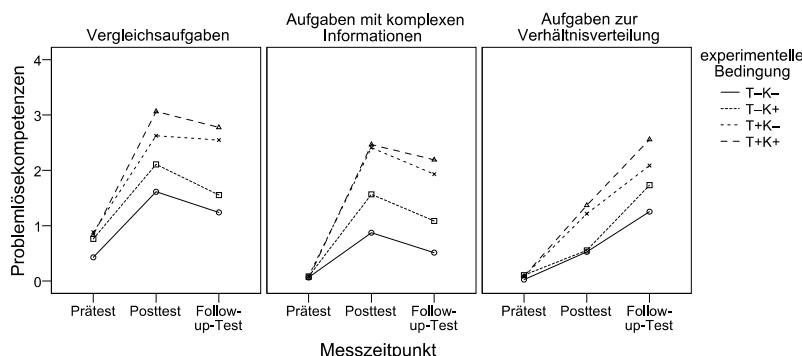


Abbildung 5.28. Grafische Darstellung der Interaktion 3. Ordnung (Zeit x Training x Kommunikation x Aufgabentyp) für die Problemlösekompetenzen. Skala: 0–4. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

Die Irrelevanz des Faktors *Kommunikation* wird in Abbildung 5.28 durch die Parallelität der Linien der Trainingsgruppen (T+K+ und T+K-) als auch die Parallelität der Linien der Nicht-Trainingsgruppen (T-K+ und T-K-) sichtbar. Die Kompetenzentwicklung innerhalb der vier experimentellen Bedingungen war über die Zeit innerhalb jedes Aufgabentyps homogen.

Der Effekt der Interventionsmaßnahme ging am deutlichsten bei den trainierten Aufgabentypen, der *Vergleichsaufgabe* und der *Aufgabe mit komplexen Informationen*, hervor. Die Zuwächse vom Prä- zum Posttest betragen bei der *Vergleichsaufgabe* 1.15, 1.38, 1.72 und 2.18, bei der *Aufgabe mit komplexen Informationen* .79, 1.45, 2.35 und 2.38, wobei die Reihenfolge der Zuwächse der Reihenfolge der Gruppen T-K-, T-K+, T+K- und T+K+ entspricht (siehe Anhang G.1, Tabelle G.7). Bei der untrainierten Aufgabe zur *Verhältnisverteilung* steigerten alle vier Gruppen ihre Problemlösekompetenzen nahezu stetig, wobei die Gruppe T-K+ ihren Zuwachs vom Post- zum

Follow-up-Test sogar erneut steigern konnte. Auf Basis der deskriptiven Ergebnisse wird geschlossen, dass sich keine Gruppe bei einer bestimmten Aufgabe in ihren Problemlösekompetenzen von einer der anderen Gruppen unterschied und somit die Wirkung der Interventionsmaßnahme auf allen Stufen der *Aufgabentypen* gleich war.

Varianzanalytisch konnte gezeigt werden, dass weder die Interaktion zweiter Ordnung (Training x Kommunikation x Aufgabentyp), $F(12, 646) = 1.27, p = .281, \eta_p^2 < .01$, noch die Interaktion dritter Ordnung signifikant wurde, $F(3.87, 1248.79) = 1.20, p = .309, \eta_p^2 < .01$. Insofern war der Effekt der experimentellen Bedingung auf die Problemlösekompetenzen auch über die Zeit hinweg vom Aufgabentyp unbeeinflusst.

Hypothese 4.5: Trainierte Drittklässler, die sich zusätzlich unter Gleichaltrigen austauschen, unterscheiden sich beim Lösen problemhaltiger Textaufgaben nach der Interventionsmaßnahme nicht in der Anzahl selbstgenerierter externer Repräsentationen von trainierten Drittklässlern, die sich austauschen.

Hypothese 4.6: Drittklässler, die weder trainiert werden noch sich unter Gleichaltrigen austauschen, generieren gleich viele externe Repräsentationen wie untrainierte Drittklässler, die sich austauschen.

Die Drittklässler externalisierten vor der Interventionsmaßnahme selten ihre Lösungs-ideen bzw. -wege, was zur Folge hatte, dass deren Lösungsfundung hauptsächlich mental ablief: Alle vier Gruppenmittelwerte lagen um $M = 0.33$ (siehe Tabelle 5.12), sodass davon auszugehen war, dass im Mittel bei einer von drei Aufgaben eine externe Repräsentation generiert wurde.

Tabelle 5.12. Deskriptive Ergebnisse der vier experimentellen Bedingungen in der Anzahl selbstgenerierter externer Repräsentationen zu allen Messzeitpunkten

	Prä		Post		Follow-up	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
T+K ^a	0.37	0.40	1.59	0.57	1.64	0.62
T+K ^b	0.31	0.39	1.52	0.67	1.65	0.64
T-K ^c	0.35	0.41	1.17	0.64	1.15	0.60
T-K ^d	0.26	0.38	1.03	0.67	0.95	0.69

Anmerkung. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings. Der Testscore bildet die durchschnittliche Summe der pro Aufgabe generierten externen Repräsentationen ab, weshalb kein maximaler Testscore existiert.

^a $N = 85$, ^b $N = 87$, ^c $N = 82$, ^d $N = 73$.

Der Rückgriff auf externalisierte Denk- und Erkenntniswerkzeuge stieg vom Prä- zum Posttest an, dominierte jedoch in den Trainingsgruppen T+K- und T+K+, die untereinander wiederum unwesentlich in ihrer Anzahl selbstgenerierter externer Repräsentationen differierten (T+K+: $M = 1.59$, $SD = 0.57$; T+K-: $M = 1.52$, $SD = 0.67$). Auch in diesem Falle verlief die Entwicklung hypothesenkonform. Kinder der Trainingsgruppen generierten im Durchschnitt mehr externe Repräsentationen als Kinder der Nicht-Trainingsgruppen, wobei es einen vernachlässigbar kleinen Einfluss der *Kommunikation* gab.

Darüber hinaus konnten beide Trainingsgruppen T+K+ und T+K- ihre Anzahl zum Follow-up-Test hin geringfügig erhöhen, wohingegen die Anzahl in den Nicht-Trainingsgruppen T-K+ und T-K- nahezu gleich blieb bzw. minimal sank. Aus prozentualer Sicht fällt auf, dass Kinder der Nicht-Trainingsgruppe (T-K- und T-K+) hauptsächlich niedrige Testscores zwischen .00 und 1.33 erlangten (siehe Abbildung 5.29).

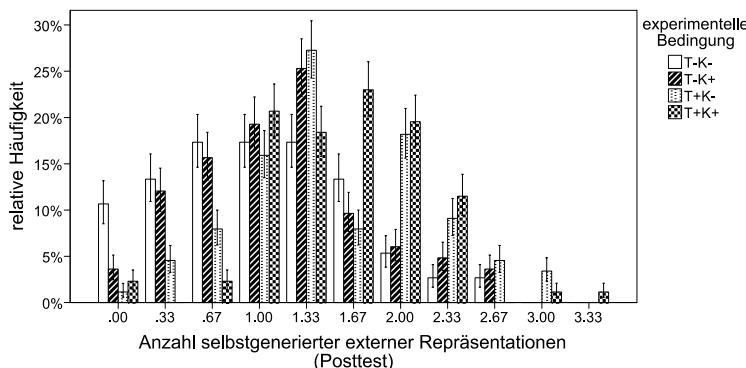


Abbildung 5.29. Relative Häufigkeit der Anzahl selbstgenerierter externer Repräsentationen zum Posttest
 T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

Dagegen erzielten Kinder der Trainingsgruppen (T+K+ und T+K-) schwerpunktmäßig Testscores im Bereich von 1.33 und 2.33 (siehe Abbildung 5.29). Letztere waren eher in der Lage, mehrere externe Repräsentationen kombiniert einzusetzen und für die Problemlösung zu nutzen.

Die deskriptiven Ergebnisse, dass die Trainings- und die Nicht-Trainingsgruppen differierten, jedoch der Faktor *Kommunikation* diesen Effekt kaum beeinflusste, ließen sich inferenzstatistisch bestätigen. Weder die Interaktion erster Ordnung (Training x Kommunikation: $F(1, 323) = 1.06$, $p = .305$, $\eta_p^2 < .01$) noch die Interaktion

zweiter Ordnung (Zeit x Training x Kommunikation: $F(2, 646) = 1.00$, $p = .368$, $\eta_p^2 < .01$) wurden signifikant. Über den gesamten Zeitraum betrachtet zeichnete sich für die Trainingsgruppe, die sich unter Gleichaltrigen austauschten (T+K+), ein paralleles Lösungsprofil zu der Trainingsgruppe, die sich nicht unter Gleichaltrigen austauschten (T+K-), ab. Die Gruppen erstellten nahezu gleich viele externe Repräsentationen über die Zeit (siehe Abbildung 5.30). Diese Parallelität konnte auch auf Nicht-Trainingsgruppen übertragen werden. Die Gruppe T-K+ generierte über die Zeit nicht mehr externe Repräsentationen als die Gruppe T-K-. Den Interaktionsdiagrammen sind die vergleichbaren, parallelen Verlaufsmuster für Gruppen mit Kommunikation und ohne Kommunikation zu entnehmen (siehe Abbildung 5.30). Für die Anzahl selbstgenerierter externer Repräsentationen war die Interaktion zwischen Zeit und Training von den Stufen der Kommunikation unabhängig, sodass eine Kombination der Faktoren Training und Kommunikation über die Zeit keinen Zugewinn in der Anzahl selbstgenerierter externer Repräsentationen erwirkte. Die Hypothesen 4.5 und 4.6 gelten daher als bestätigt.

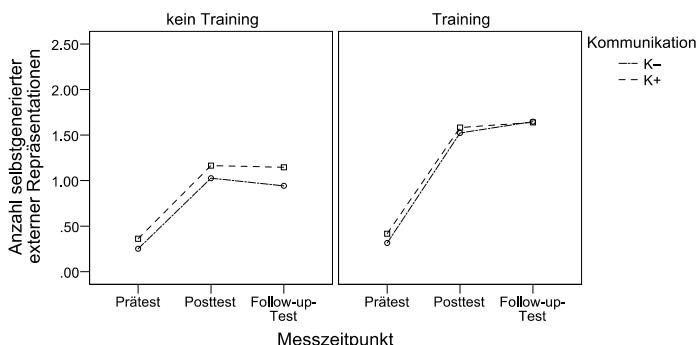


Abbildung 5.30. Grafische Darstellung der Interaktion 2. Ordnung (Zeit x Training x Kommunikation) für die Anzahl selbstgenerierter externer Repräsentationen. Für jede Stufe des Faktors Kommunikation (ja oder nein) wurde ein Diagramm für die bedingte Interaktion Zeit x Training angefertigt.

Aufgabenspezifische Analyse

Auch bei der Anzahl selbstgenerierter externer Repräsentationen waren für alle drei Aufgabentypen über die Zeit gleiche Entwicklungen zu beobachten. Unterschiede, die auf den Faktor *Aufgabentyp* zurückzuführen waren, wurden daher nicht erwartet. Kinder aller vier experimentellen Bedingungen konstruierten nach der Interventionsmaßnahme mehr externe Repräsentationen, als sie dies noch davor hatten. Bei der

Aufgabe zur Verhältnisverteilung nahm die Anzahl 13 Wochen nach der Interventionsmaßnahme in allen Gruppen und bei der *Vergleichsaufgabe* in allen Gruppen bis auf die Kontrollgruppe T-K- erneut zu, während sie bei der *Aufgabe mit komplexen Informationen* leicht zurückging (siehe Abbildung 5.31). Die deskriptiven Ergebnisse werden in tabellarischer Form im Anhang G.1 (Tabelle G.10) präsentiert.

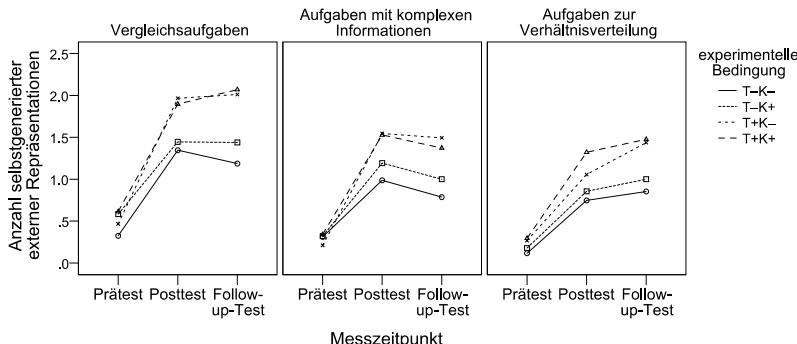


Abbildung 5.31. Grafische Darstellung der Interaktion 3. Ordnung (Zeit x Training x Kommunikation x Aufgabentyp) für die Anzahl selbstgenerierter externer Repräsentationen.

Die Varianzanalyse mit Messwiederholung ergab auch bei der abhängigen Variablen *Anzahl selbstgenerierter externer Repräsentationen* weder eine signifikante Interaktion zweiter Ordnung (Training x Kommunikation x Aufgabentyp: $F(1.95, 630.75) = 0.93$, $p = .393$, $\eta_p^2 < .01$) noch eine signifikante Interaktion dritter Ordnung (Zeit x Training x Kommunikation x Aufgabentyp: $F(3.70, 1197.34) = 1.08$, $p = .362$, $\eta_p^2 < .01$). Demzufolge nahm der Aufgabentyp keinen Einfluss auf die Verwendung selbstgenerierter externer Repräsentationen über die Zeit.

5.3.5 Zusammenfassung bisheriger Ergebnisse

Bevor die 5. Forschungsfrage im Fokus der Ergebnispräsentation steht, werden die zentralen Befunde aufgrund ihrer inhaltlichen Verknüpfung in Tabelle 5.13 gebündelt. Vor dem Hintergrund, dass der Aufgabentyp keinen Einfluss auf die für diese Untersuchung bedeutsamen Interaktionen über die Zeit nahm und der Faktor nur als Prüfung eines möglichen Einflusses einbezogen wurde, wird er bei der weiteren Zusammenfassung der Ergebnisse aus ökonomischen Gründen ausgespart.

Die Untersuchung zeigte, dass Drittklässler durch die Teilnahme am Repräsentationstraining (T+) ihren *Lösungserfolg* sowie ihre *Problemlösekompetenzen* im Laufe der

Zeit steigern konnten und Gleichaltrigen, die keine Unterstützung durch ein Repräsentationstraining erfuhren (T-), überlegen waren. Das Training bewirkte zudem eine häufigere *Generierung und Nutzung externer Repräsentationen*. Kinder, die externe Repräsentationen als Denk- und Erkenntniswerkzeuge in ihren Lösungsprozess integrierten, waren 1) erfolgreicher und erzielten 2) einen höheren Kompetenzzuwachs als Kinder, die nicht auf externe Repräsentationen zurückgriffen. Erstes wurde aufgrund der signifikanten Korrelation zwischen der *Anzahl selbstgenerierter externer Repräsentationen* und dem *Lösungserfolg* zum Zeitpunkt des Posttests ($r = .38$, 95 % CI²²⁷ [.29, .47], $p < .001$) angenommen. Für Letzteres sprach die signifikante Korrelation zwischen der *Anzahl selbstgenerierter externer Repräsentationen* und den *Problemlösekompetenzen* zum Zeitpunkt des Posttests ($r = .54$, 95 % CI [.45, .61], $p < .001$). Zum Zeitpunkt des Prätests waren die zugrunde liegenden Zusammenhänge kleiner (Erfolg: $r = .21$, 95 % CI [.11, .31], $p < .001$; Problemlösekompetenzen: $r = .38$, 95 % CI [.29, .48], $p < .001$). Demzufolge bestand ein positiver Zusammenhang zwischen der *Anzahl selbstgenerierter externer Repräsentationen* und dem *Lösungserfolg* bzw. den *Problemlösekompetenzen*.

Darüber hinaus wirkte der intersubjektive Austausch unter Gleichaltrigen über die Zeit lediglich einen Effekt auf die *Problemlösekompetenzen*. Kinder, die sich in Zweiersettings über ihren Lösungsprozess austauschen durften, erwirkten über die Zeit einen höheren *Problemlösekompetenzzuwachs* als Kinder, die sich nicht in Zweiersettings austauschen durften. Dies galt entgegen den Erwartungen nicht für den *Lösungserfolg*. Der fehlende Einfluss der Kommunikation auf die *Anzahl selbstgenerierter externer Repräsentationen* war wiederum hypothesenkonform.

Tabelle 5.13. Übersicht über Haupt- und Interaktionseffekte der 3x2x2-Varianzanalyse mit Messwiederholung auf dem ersten Faktor für alle drei abhängigen Variablen

Faktoren	MS	F	p	η_p^2
Lösungserfolg				
Zeit (Z) ^a	32.98	235.49	< .001	.42
Training (T) ^b	2.02	40.60	< .001	.11
Kommunikation (K) ^b	.25	4.93	.027	.02
Z x T ^a	2.96	21.12	< .001	.06
Z x K ^a	.27	1.96	.142	.01

²²⁷ Das durch die Bias-corrected-and-accelerated-Methode, kurz BCa-Methode, angepasste 95 %-Konfidenzintervall wird in Klammern berichtet.

T x K ^b	.004	0.08	.781	< .01
Z x T x K ^a	.01	0.03	.968	< .01
Problemlösekompetenzen				
Zeit (Z) ^c	686.73	421.04	< .001	.57
Training (T) ^b	44.46	74.87	< .001	.19
Kommunikation (K) ^b	4.56	7.68	.006	.02
Z x T ^c	74.88	45.91	< .001	.12
Z x K ^c	7.47	4.58	.011	.01
T x K ^b	.81	0.85	.358	< .01
Z x T x K ^c	.29	0.09	.914	< .01
Anzahl selbstgenerierter externer Repräsentationen				
Zeit (Z) ^a	325.61	454.67	< .001	.59
Training (T) ^b	10.99	67.15	< .001	.17
Kommunikation (K) ^b	4.93	4.93	.027	.02
Z x T ^a	19.93	27.82	< .001	.08
Z x K ^a	.00	0.00	.996	< .01
T x K ^b	.17	1.06	.305	< .01
Z x T x K ^a	.72	1.00	.368	< .01

Anmerkung. *MS* = quadratischer Mittelwert. Aufgrund der Verletzungen der Sphärizität wurde für die Problemlösekompetenzen die Huynh-Feldt-Adjustierung verwendet.

^a*df* = 2, 646. ^b*df* = 1, 323. ^c*df* = 1.99, 643.32.

Lernende, die sich zusätzlich zum Repräsentationstraining austauschen durften, hatten gegenüber trainierten Kindern ohne Austausch entgegen den Erwartungen keinen Vorteil im *Lösungserfolg* und in den *Problemlösekompetenzen*. Gleiches gilt für die Nicht-Trainingsgruppe. Der Trainingseffekt über die Zeit war unabhängig von der Kommunikation und somit für alle Stufen der Kommunikation gültig. Die Interaktion 2. Ordnung war für keine der abhängigen Variablen signifikant (siehe Tabelle 5.13). Hypothesenkonform trifft jedoch zu, dass der Faktor *Kommunikation* die Wirkung der Interventionsmaßnahme über die *Zeit* auf die *Anzahl selbstgenerierter externer Repräsentationen* nicht beeinflusste.

5.3.6 Forschungsfrage 5: Effekt selbstgenerierter externer Repräsentationen

Welcher Anteil des Lösungserfolgs bzw. der Problemlösekompetenzen auf selbstgenerierte externe Repräsentationen zurückgeführt werden kann, steht im Interesse der Forschungsfrage 5. In diesem Zusammenhang wurde überprüft, ob sich gute Prädiktoren für die Vorhersage des *Lösungserfolgs* und der *Problemlösekompetenzen* identifizieren lassen. Das hierarchische Vorgehen ermöglicht die Beantwortung der Frage danach, ob die Repräsentationsmerkmale einen zusätzlichen Beitrag in der Vorhersage leisten (Cohen et al., 2003, S. 160).

Hypothese 5.1: Das Generieren externer Repräsentationen beeinflusst positiv den Lösungserfolg zum Zeitpunkt des Posttests.

Die Kontrolle möglicher Einflussfaktoren auf den Lösungserfolg erfolgte durch eine hierarchische Regression, wobei der Anteil erklärter Varianz, der auf die *experimentelle Bedingung* zurückzuführen war, zuerst berücksichtigt wurde. Die Korrelation zwischen diesem Prädiktor und dem Lösungserfolg betrug $r = .30$ ($p < .001$).

Im zweiten Schritt folgten die Kontrollvariablen, die aus theoretischer Sicht den Lösungserfolg beeinflussten (Cohen et al., 2003, S. 158–160; Tabachnick & Fidell, 2014, S. 179): *Lösungserfolg* (Prätest), *Geschlecht*, *Mehrsprachigkeit*, *Intelligenz*, *mathematische, sprachliche und metakognitive Fähigkeiten* sowie *Attitudes und Beliefs*. Prioritäten in der Hierarchie wurden hier keine gesetzt, weshalb die Regressoren simultan aufgenommen wurden.

Die Korrelationen aller Kontrollvariablen mit der abhängigen Variablen sowie die Interkorrelationen sind Tabelle 5.14 zu entnehmen.

Tabelle 5.14. Korrelationsmatrix der Kontrollvariablen und der abhängigen Variablen Lösungserfolg

	<i>M</i>	<i>SD</i>	AV	1	2	3	4	5	6	7	8	9
AV	.38	.35	1.00									
1. Experimentelle Bedingung	1.43	1.07	.31	1.00								
2. Lösungserfolg (Prätest)	.10	.19	.45	.17	1.00							
3. Intelligenz	29.46	4.47	.38	.09	.15	1.00						
4. Sprachliche Fähigkeiten	37.90	24.33	.32	.17	.20	.24	1.00					
5. Mathematische Basiskompetenzen	37.07	28.33	.49	.10	.36	.28	.46	1.00				

6. Mehrsprachigkeit	.57	.50	.18	.15	.13	.10	.11	.09	1.00
7. Geschlecht	.46	.50	-.05	.09	-.04	.02	-.02	-.18	-.07 1.00
8. Metakognitive Fähigkeiten	-.02	.87	.67	.21	.39	.33	.22	.38	.17 -.04 1.00
9. Attitudes und Beliefs	8.78	3.47	.28	-.03	.22	.14	.13	.34	-.03 -.14 .02 1.00

Anmerkung. $N = 291$. AV = abhängige Variable Lösungserfolg; Sprachliche Fähigkeiten: Gesamttestscore ELFE 1–6; math. Basiskompetenzen = mathematische Basiskompetenzen: Gesamttestscore HRT 1–4; Experimentelle Bedingung = T+K+, T+K-, T-K+, T-K-; Geschlecht: Jungen (0); Mädchen (1); Mehrsprachigkeit: mehrsprachig (0); einsprachig (1).

Im dritten und letzten Schritt wurden die *Repräsentationsmerkmale* in das Modell integriert. Hierbei handelte es sich um die vier Faktoren, die mithilfe der Hauptachsenanalyse extrahiert wurden (vgl. Abschnitt 4.5.2).

Tabelle 5.15 sind die Ergebnisse der multiplen Regression zu entnehmen. Demnach erklärte die *experimentelle Bedingung*, der die Drittklässler angehörten, 10 % der Varianz des Lösungserfolgs, $R^2 = .10$, $F(1, 289) = 31.55$, $p < .001$. Darüber hinaus gingen auf die im zweiten Schritt aufgenommenen *Kontrollvariablen* weitere 51 % der Varianz zurück, $R^2 = .51$, $F(8, 281) = 10.43$, $p < .001$. Der Zuwachs in R^2 war signifikant. Die Repräsentationsmerkmale, welche im dritten Schritt in die Regressionsgleichung aufgenommen wurden, erklärten zusätzliche 5 % der Varianz, $R^2 = .05$, $F(4, 277) = 41.05$, $p < .001$. Letztgenannte Veränderung in R^2 gab den Varianzanteil des Lösungserfolgs wieder, den die Repräsentationsmerkmale zusätzlich zur Gruppe und den Kontrollvariablen erklärten, d. h., der Einfluss der Gruppe wie auch der Kontrollvariablen war aus dem Einfluss der Repräsentationsmerkmale herauspartialisiert. Dieser Zugewinn war statistisch signifikant ($p < .001$).

Tabelle 5.15. Hierarchische Regressionsanalyse von Kontrollvariablen und Repräsentationsmerkmalen auf den Lösungserfolg

Prädiktoren	B	SE B	β	ΔR ²
Experimentelle Bedingung	.03	.01	.08 [†]	.10***
Lösungserfolg (Prätest)	.17	.08	.09 [*]	
Intelligenz	.01	.00	.08 [†]	
Leseverständnis	.00	.00	.02	
Mathematische Basiskompetenzen	.00	.00	.13**	
Mehrsprachigkeit	.03	.03	.04	

Geschlecht	-.01	.03	.02	
Metakognitive Fähigkeiten	.18	.02	.43***	
Attitudes und Beliefs	.01	.00	.14***	.51***
Zeichnungen	.07	.01	.21***	
Schriftsprachliche Beschreibungen	.02	.01	.05	
Rechnungen	.04	.01	.10*	
Tabellen	.05	.01	.15***	.05***
Konstante	-.04	.10		
			$R^2 = .66$	
			korrigiertes	$R^2 = .64$
				$R = .81***$

Anmerkung. Die Tabelle zeigt in Anlehnung an Tabachnick und Fidell (2014, S. 214) den nicht standariserten Regressionskoeffizienten (B) und die Konstante, den standardisierten Regressionskoeffizienten (β), die semipartialle Korrelation ΔR^2 und R , R^2 sowie das korrigierte R^2 auf, nachdem alle unabhängigen Variablen inkludiert waren (Schritt 3).

† $< .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Insgesamt erklärte das Modell mit allen unabhängigen Variablen 66 % der Varianz des Lösungserfolgs, welches sich gegen den Zufall absichern ließ, $R^2 = .66$, $F(13, 277) = 41.05$, $p < .001$. Wenn aufgrund der hohen Anzahl der Prädiktoren der korrigierte Varianzanteil berücksichtigt wurde, dann wich diese Vorhersage mit $R^2 = .64$ nur minimal vom Determinationskoeffizienten $R^2 = .66$ ab. Dies deutete darauf hin, dass knapp zwei Drittel des Lösungserfolgs von den unabhängigen Variablen vorhergesagt werden konnte.

Mit Blick auf die Repräsentationsprädiktoren *Zeichnungen*, *Rechnungen*, *Tabellen* und *schrift-sprachliche Beschreibungen* bleibt festzuhalten, dass sie zwar einen Einfluss auf den Lösungserfolg ausübten, dieser jedoch sehr moderat ausfiel. Die Zeichnungen gingen als stärkster ($\beta = .21$) und schriftsprachliche Beschreibungen als schwächster Prädiktor ($\beta = .05$) des Erfolgs hervor (siehe Tabelle 5.15). Eine eindeutige Tendenz in der Wahl der Heran- und Vorgehensweise ließ sich nicht identifizieren. Auch wenn die Depiktion als stärkster Prädiktor galt, fand die Deskription, vorrangig durch die Erstellung von *Rechnungen* und *Tabellen*, ihre Anwendung. Insgesamt gilt die Hypothese 5.1 als bestätigt.

Darüber hinaus fiel auf, dass das Geschlecht als einzige nicht signifikant mit dem Lösungserfolg korrelierende Variable ($r = -.05$, $p = .221$) die Vorhersage des Lösungserfolgs nicht beeinflusste. Demzufolge unterschieden sich Mädchen und Jungen nach dem Training nicht signifikant in ihrem Lösungserfolg, auch wenn der negativen Kor-

relation zufolge die Jungen rein deskriptiv einen minimal höheren Lösungserfolg als die Mädchen erreichten. Um dennoch sagen zu können, welcher Anteil der Varianz unabhängig vom Geschlecht auf die selbstgenerierten externen Repräsentationen zurückgeht, wurde der Prädiktor trotz untergeordnetem Stellenwert berücksichtigt.

Aufgabenspezifische Analyse

Trotz der Tatsache, dass sich gemittelt über den gesamten Textaufgaben-Performance-Test keine eindeutig bevorzugten Denk- und Erkenntniswerkzeuge identifizieren ließen und ihre Vorhersagefähigkeit von 5 % als niedrig einzustufen ist, war es von Interesse, herauszufinden, ob der Aufgabentyp die Wahl der externen Repräsentationen beeinflusst, und wenn ja, inwiefern. Gemäß den bisherigen Ergebnissen war für alle Repräsentationsprädiktoren außer den schriftsprachlichen Beschreibungen, ein signifikanter Einfluss auf den Lösungserfolg zu beobachten.

Im Zuge der aufgabenspezifischen Analyse wurde das zuvor verwendete Kriterium, der Lösungserfolg-Testscore, durch den jeweiligen Aufgabenscore der *Vergleichsaufgabe*, *Aufgabe mit komplexen Informationen* bzw. *Aufgabe zur Verhältnisverteilung* ersetzt. Dadurch ergaben sich drei separate Analysen. Aufgrund der Tatsache, dass der Lösungserfolg jeder Aufgabe dichotom gemessen wurde (richtig/falsch), wurde die logistische Regression als Analyseverfahren herangezogen. Um die logistischen Regressionsmodelle mit dem multiplen linearen Regressionsmodell vergleichen zu können, wurden die Modellspezifikationen aus der vorangegangenen Analyse übernommen. Sprich die hierarchische Vorgehensweise wurde gewählt, bei der dieselben Prädiktoren in derselben Reihenfolge in das Modell aufgenommen wurden (1. Schritt: experimentelle Bedingung, 2. Schritt: Kontrollvariablen, 3. Schritt: Repräsentationsmerkmale). Anpassungen mussten lediglich bei dem Prädiktor Lösungserfolgs-Score des Prätests vorgenommen werden. Während zuvor die Analyse auf dem über drei Aufgaben aggregierten Testscore des Prätests basierte, wurde er in der aufgabenspezifischen Analyse durch den jeweiligen Aufgabenscore des Prätests ersetzt. Für die Modelle wurden akzeptable Gültigkeitstests ermittelt, die erforderlichen Voraussetzung waren erfüllt (siehe Abschnitt 4.8.5).

Der Prozentsatz der Drittklässler, die auch wirklich die Aufgabe erfolgreich gelöst hatten, der sie anhand der geschätzten Wahrscheinlichkeit zugeordnet waren, lag im Basismodell ohne Prädiktoren für die *Vergleichsaufgabe* bei 53 %, für die *Aufgabe mit komplexen Informationen* bei 63 % und für die *Aufgabe zur Verhältnisverteilung* bei 78 %. Dies bedeutet, dass im Basismodell ohne Prädiktoren der Lösungserfolg bei der Vergleichsaufgabe von 144 Probanden richtig und von 129 Probanden falsch vorhergesagt wurde. Durch die Aufnahme der *experimentellen Bedingung* im 1. und der *Kontrollvariablen* im 2. Schritt konnten insgesamt 81 % (73 % bzw. 84 %) aller Fälle rich-

tig klassifiziert werden.²²⁸ Eine erneute Zunahme der Klassifikationsgüte fiel durch die Aufnahme der *Repräsentationsmerkmale* im 3. Schritt, konform zu den Ergebnissen für den gesamten Test, gering aus. Der Anteil stieg durch die Repräsentationsmerkmale minimal auf 84 % (80 % bzw. 85 %). Hieraus kann abgeleitet werden, dass die vier externen Repräsentationen beim untrainierten Aufgabentyp keinen signifikanten Zu-gewinn darstellten, $\chi^2(4) = 6.34, p = .175$, hingegen bei den trainierten Aufgabentypen schon (Vergleichsaufgabe: $\chi^2(4) = 37.67, p < .001$; Aufgabe mit komplexen Informati-onen: $\chi^2(4) = 29.38, p < .001$).

Wenn die einzelnen Aufgabentypen fokussiert wurden (siehe Anhang G.2.2, Tabelle G.1), dann fiel auf, dass beim Aufgabentyp *Vergleichsaufgaben* eine erfolgreiche Problemlösung zum Zeitpunkt des Posttests in enger Verbindung mit dem Erstellen von *Tabellen*, $OR = 2.67$ ²²⁹ stand, gefolgt von *Rechnungen*, $OR = 1.94$. Der Parameter $OR = 2.67$ gibt an, dass sich das Wahrscheinlichkeitsverhältnis um den Faktor 2.67 ändert, wenn der Prädiktor *Tabelle* um eine Einheit erhöht wird und die restlichen Prädiktoren konstant bleiben (vgl. Baltes-Götz, 2012, S. 11). *Zeichnungen* und *schrift-sprachliche Beschreibungen* galten ebenfalls als signifikante Prädiktoren, wenngleich ihr positiver Zusammenhang mit dem Lösungserfolg niedriger ausfiel ($OR_Z = 1.69$ bzw. $OR_B = 1.65$).

Aufgaben mit komplexen Informationen wurden besonders erfolgreich mit *Zeichnun- gen* gelöst, $OR = 2.75$ ($p < .001$). *Rechnungen* und *Tabellen* nahmen mit $OR = .95$ bzw. $OR = .97$ nahezu keinen Einfluss auf das Kriterium, was mit fehlender Signifi-kanz bestätigt wurde (siehe Anhang G.2.2, Tabelle G.1).

Bei dem Aufgabentyp *Aufgabe zur Verhältnisverteilung* fiel auf, dass lediglich ein po-sitiver Zusammenhang zwischen *Tabellen* und dem Lösungserfolg zum Posttest ($OR = 1.52$) bzw. *Rechnungen* und dem Lösungserfolg zum Posttest ($OR = 1.30$) be-stand. Beide Zusammenhänge mit dem Kriterium wurden durch Signifikanz bestätigt ($p < .05$).

²²⁸ Die Reihenfolge der Prozentsätze repräsentiert die Aufgabenreihenfolge: 1. Vergleichsaufgaben, 2. Aufgabe mit komplexen Informationen und 3. Aufgabe zur Verhältnisverteilung.

²²⁹ Während in der linearen Regressionsanalyse der Determinationskoeffizient oder die erklärte Vari-anz als gängiges Effektstärkenmaß Verwendung findet, wird in der logistischen Regressionsanalyse bevorzugt auf die Interpretation der einzelnen Koeffizienten *OR*, den Odds-Ratio, zurückgegriffen (Eid et al., 2011, S. 786–787). Sie gelten als Zusammenhangsmaß bei kategorialen Variablen (Eid et al., 2011, S. 773). Ein Wert von $OR = 1$ suggeriert keinen Zusammenhang, ein Wert von $OR > 1$ einen positiven Zusammenhang und ein Wert von $OR < 1$ einen negativen Zusammenhang zwi-schen dem Prädiktor und dem Kriterium. Je größer das Odds-Ratio, desto größer ist die Chance, dass die Kategorie mit dem Wert *max* = 1, hier Erfolg, im Gegensatz zur Kategorie mit dem Wert *min* = 0, Misserfolg, erzielt wird.

Insgesamt bleibt festzuhalten, dass die Stärke des Zusammenhangs zwischen den Repräsentationsprädiktoren und dem Kriterium je nach zu bearbeitendem Aufgabentyp unterschiedlich ausfiel. Während bei der *Vergleichsaufgabe* der Lösungserfolg in engem Zusammenhang mit *Tabellen* und *Rechnungen* stand, stand er bei der *Aufgabe mit komplexen Informationen* in engem Zusammenhang mit *Zeichnungen*. Bei dem untrainierten Aufgabentyp fielen die positiven Zusammenhänge geringer aus, waren aber für *Tabellen* und *Rechnungen* signifikant.

Nachfolgend wurde überprüft, welcher Anteil des Problemlösezuwachses auf die Generierung und Nutzung externer Repräsentationen zurückgeführt werden konnte und in welchem Maße sich dieser Anteil von dem des Lösungserfolges unterschied. Die Hypothesenprüfung basierte zunächst auf dem Testscore des Posttests. Abschließend erfolgten aufgabenspezifische Analysen.

Hypothese 5.2: Das Generieren externer Repräsentationen beeinflusst positiv die Problemlösekompetenzen zum Zeitpunkt des Posttests. Hypothese

5.3: Der Einfluss selbstgenerierter externer Repräsentationen ist auf die Problemlösekompetenzen zum Zeitpunkt des Posttests größer als auf den Lösungserfolg.

Analog zur Hypothesenprüfung 5.1 wurde mithilfe einer hierarchischen Regression überprüft, inwiefern die selbstgenerierten externen Repräsentationen das Kriterium, die Problemlösekompetenzen, bedeutsam beeinflussten, wenn die Störvariablen als kontrolliert anzusehen waren (Cohen et al., 2003, S. 160). Da bei der Erfassung der Problemlösekompetenzen der Prozess, und nicht nur das Produkt, das Endergebnis, Beachtung fand, wurde ein größerer Einfluss der selbstgenerierten externen Repräsentationen auf die Problemlösekompetenzen als auf den Lösungserfolg erwartet.

Zudem wurde angenommen, dass auch bei der Vorhersage der Problemlösekompetenzen ein Großteil der erklärten Varianz auf die Kontrollvariablen zurückgeführt werden kann. In das Regressionsmodell aufgenommen wurde im ersten Schritt die *experimentelle Bedingung*, gefolgt von den Prädiktoren *Problemlösekompetenzen (Prätest)*, *Intelligenz*, *sprachliche*, *mathematische* und *metakognitive Fähigkeiten*, *Geschlecht*, *Mehrsprachigkeit* sowie *Attitudes und Beliefs*. Wenngleich der Prädiktor *Geschlecht* nicht signifikant mit der abhängigen Variablen korrelierte ($r = .01$, $p = .422$, siehe Tabelle 5.16), wurde er trotz alledem berücksichtigt, um eine geschlechtsunabhängige Vorhersage treffen zu können. Die *metakognitiven Fähigkeiten* gingen auch im Rahmen der Problemlösekompetenzen mit dem aus der Hauptachsenanalyse gewonnenen Faktor in das Regressionsmodell ein. Die Korrelationen mit dem Kriterium und die Interkorrelationen mit dem Prädiktor Problemlösekompetenzen (Prätest) sind Tabelle 5.16 zu entnehmen. Für alle weiteren Interkorrelationen sei auf Tabelle 5.14, Abschnitt 5.3.6, verwiesen. Im dritten Schritt wurden die aus der Hauptachsenanalyse gewonne-

nen *Repräsentationsprädiktoren* in das Regressionsmodell aufgenommen (vgl. Hypothesenprüfung 5.1).

Tabelle 5.16. Korrelationsmatrix der Kontrollvariablen und der abhängigen Variablen Problemlösekompetenzen

	<i>M</i>	<i>SD</i>	AV	1	2	3	4	5	6	7	8	9
AV	1.70	1.25	1.00	.42	.46	.39	.36	.48	.17	.01	.61	.30
2. Kompetenzen (Prätest)	1.43	1.07		.18	1.00	.18	.22	.41	.12	-.01	.40	.23

Anmerkung. $N = 291$. 1 = experimentelle Bedingung (T+K+, T+K-, T-K+, T-K-); 2 = Problemlösekompetenzen (Prätest); 3 = Intelligenz; 4 = sprachliche Fähigkeiten (Gesamttestscore ELFE 1–6); 5 = mathematische Basiskompetenzen (Gesamttestscore HRT 1–4); 6 = Mehrsprachigkeit: mehrsprachig (0); einsprachig (1); 7 = Geschlecht: Jungen (0); Mädchen (1); 8 = metakognitive Fähigkeiten; 9 = Attitudes und Beliefs.

Den Ergebnissen zufolge erklärte die experimentelle Bedingung, der die Drittklässler angehörten, 17 % der Varianz des Lösungserfolgs, $R^2 = .17$, $F(1, 289) = 60.30$, $p < .001$ (siehe Tabelle 5.17). Der Erklärungszuwachs durch die im zweiten Schritt aufgenommenen Kontrollvariablen betrug weitere 44 % der Varianz, $R^2 = .44$, $F(8, 281) = 40.50$, $p < .001$. Dieser Zuwachs in R^2 war signifikant.

Die Repräsentationsmerkmale, welche im dritten Schritt in die Regressionsgleichung aufgenommen wurden, erklärten zusätzliche 13 % der Varianz, $R^2 = .13$, $F(4, 277) = 34.67$, $p < .001$. Letztgenannte Veränderung in R^2 gibt den Varianzanteil der Problemlösekompetenzen wieder, den die Repräsentationsmerkmale zusätzlich zur Gruppe und zu den Kontrollvariablen erklärten, d. h., der Einfluss der Gruppe sowie der Kontrollvariablen war aus dem Einfluss der Repräsentationsmerkmale herauspartielliert. Dieser Zugewinn war ebenfalls statistisch signifikant ($p < .001$). Insgesamt erklärte das Modell mit allen unabhängigen Variablen 74 % der Varianz der Problemlösekompetenzen, welches gegen den Zufall abgesichert werden konnte, $R^2 = .74$, $F(13, 277) = 61.90$, $p < .001$. Wenn aufgrund der hohen Anzahl der Prädiktoren der korrigierte Varianzanteil berücksichtigt wurde, dann wichen diese Vorhersage mit $R^2 = .73$ nur minimal vom Determinationskoeffizienten $R^2 = .74$ ab, sodass knapp drei Viertel der Problemlösekompetenzen von den unabhängigen Variablen vorhergesagt werden.

Tabelle 5.17. Hierarchische Regressionsanalyse von Kontrollvariablen und Repräsentationsmerkmalen auf die Problemlösekompetenzen

Prädiktoren	B	SE B	β	ΔR ²
Experimentelle Bedingung	.16	.04	.14***	.17***
Problemlösekompetenzen (Prätest)	.12	.07	.06	
Intelligenz	.02	.01	.08*	
Leseverständnis	.00	.00	.04	
Mathematische Basiskompetenzen	.01	.00	.11**	
Mehrsprachigkeit	.06	.08	.03	
Geschlecht	.04	.08	.01	
Metakognitive Fähigkeiten	.47	.05	.32***	
Attitudes und Beliefs	.05	.01	.14***	.44***
Zeichnungen	.38	.04	.32***	
Schriftsprachliche Beschreibungen	.14	.04	.11**	
Rechnungen	.26	.04	.20***	
Tabellen	.26	.04	.20***	.13***
Konstante	.03	.30		
				R ² = .74
				korrigiertes R ² = .73
				R = .86***

Anmerkung. Die Tabelle zeigt in Anlehnung an Tabachnick und Fidell (2014, S. 214) lediglich den nicht standariserten Regressionskoeffizienten (B) und die Konstante, den standardisierten Regressionskoeffizienten (β), die semipartiale Korrelation ΔR² und R, R² sowie das korrigierte R² auf, nachdem alle unabhängigen Variablen inkludiert waren (Schritt 3).

* < .10. * p < .05. ** p < .01. *** p < .001.

Auf Repräsentationsebene bleibt festzuhalten, dass alle Repräsentationsprädiktoren *Zeichnungen*, *Rechnungen*, *Tabellen* und *schriftsprachliche Beschreibungen* die Problemlösekompetenzen signifikant beeinflussten. Darunter schienen die Zeichnungen als stärkster ($\beta = .32$) und die schriftsprachlichen Beschreibungen als schwächster Prädiktor ($\beta = .11$) die Problemlösekompetenzen zu beeinflussen (siehe Tabelle 5.17). Rechnungen und Tabellen konnte mit $\beta = .20$ ein vergleichbarer Einfluss zugesprochen werden. Wenngleich die depiktionale Repräsentation in ihrer Vorhersagekraft dominierte, war die Relevanz der deskriptionalen Repräsentationen, insbesondere der Rechnungen und Tabellen, nicht zu vernachlässigen. Insgesamt konnte keine eindeutige Tendenz der Lernenden in der Heran- und Vorgehensweise beobachtet werden.

Vor dem Hintergrund, dass die selbstgenerierten externen Repräsentationen 13 % der Problemlösekompetenzen und 5 % des Lösungserfolgs vorhersagen, gelten die Hypothesen 5.2 und 5.3 als bestätigt.

Aufgabenspezifische Analyse

Bisherigen Analysen zufolge konnten den Zeichnungen beim Lösen problemhaltiger Textaufgaben der größte und schriftsprachlichen Beschreibungen der kleinste Einfluss auf die Problemlösekompetenzen zugesprochen werden, wenn auch beide signifikant waren. Zeichnungen trugen fast dreimal mehr als die Beschreibungen zur Vorhersage der Problemlösekompetenzen bei. Rechnungen und Tabellen wiesen das gleiche Potenzial in der Vorhersage auf (siehe Hypothesenprüfung 5.2).

Um zu überprüfen, ob die Dominanz der Zeichnungen und die Relevanz der anderen Repräsentationsformen bezüglich der Vorhersage der Problemlösekompetenzen vom Aufgabentyp abhängig waren, wurden aufgabenspezifische Analysen durchgeführt. Das Kriterium war nicht mehr der erzielte Testscore der Problemlösekompetenzen, sondern jeweils einer der drei Aufgabenscores der *Vergleichsaufgabe*, der *Aufgabe mit komplexen Informationen* oder der *Aufgabe zur Verhältnisverteilung*. Mit dem Ziel, die Regressionsmodelle zu vergleichen, blieben auch hier die Wahl der Kontrollvariablen und Repräsentationsmerkmale sowie der Modellschätzung durch ein hierarchisches Vorgehen unverändert (siehe Hypothesenprüfung 5.1–5.3). Die Kontrollvariable Problemlösekompetenzen zum Zeitpunkt des Prätests (Testscore) wurde ersetzt durch den entsprechenden Prätest-Aufgabenscore.

Die aufgabenspezifischen Analysen legten dar, dass die Gruppenzugehörigkeit der Kinder eine unterschiedliche Wirkung auf die Problemlösekompetenzen nach der Interventionsmaßnahme ausübt (siehe Anhang G.2.2, Tabelle G.2). Ihr Einfluss auf den untrainierten Aufgabentyp war mit $R^2 = .06$ moderat, wohingegen er bei den trainierten Aufgabentypen stärker ins Gewicht fiel, insbesondere bei der Aufgabe mit komplexen Informationen $R^2 = .16$ (Vergleichsaufgabe: $R^2 = .10$). Der Anteil erklärter Varianz, der auf die Kontrollvariablen zurückzuführen war, schwankte zwischen $\Delta R^2 = .21$ und $\Delta R^2 = .34$. Diesbezüglich zeigte sich, dass die metakognitiven Fähigkeiten bei allen drei Aufgaben der einflussreichste Prädiktor war. Darüber hinaus beeinflussten die Intelligenz bei der Vergleichsaufgabe, das Geschlecht bei der Aufgabe mit komplexen Informationen und die mathematischen Basiskompetenzen bei der Aufgabe zur Verhältnisverteilung sowie der Aufgabe mit komplexen Informationen signifikant die Problemlösekompetenzen. Die Attitudes und Beliefs erwiesen sich bei der Vergleichsaufgabe und der Aufgabe zur Verhältnisverteilung als signifikanter Einflussfaktor. Die standardisierten Regressionsgewichte können Anhang G.2.2 (Tabelle G.2) entnommen werden.

Wenn der Blick auf die Repräsentationsmerkmale gerichtet wird, dann wird deutlich, dass der Anteil erklärter Varianz, der auf die selbstgenerierten externen Repräsentationen zurückzuführen war, beim Aufgabentyp *Aufgaben zur Verhältnisverteilung* mit $\Delta R^2 = .06$ ($F(13, 259) = 15.22, p < .001$) am geringsten ausfiel. An zweiter Stelle lagen die *Vergleichsaufgaben* mit $\Delta R^2 = .09$ ($F(13, 259) = 22.64, p < .001$) und die Spitzenposition übernahmen die *Aufgaben mit komplexen Informationen* mit $\Delta R^2 = .22$ ($F(13, 259) = 28.47, p < .001$). Demzufolge war beim letztgenannten Aufgabentyp im direkten Vergleich zu den anderen beiden Aufgabentypen die Relevanz der selbstgenerierten externen Repräsentationen am größten, beim untrainierten Aufgabentyp am geringsten. Dies ging konform mit den aufgabenspezifischen Ergebnissen für den Lösungserfolg.

Die separaten multivariaten Regressionsanalysen ergaben, dass der Aufgabentyp die Vorhersage der Problemlösekompetenzen zu beeinflussen scheint (siehe Anhang G.2.2, Tabelle G.2). Während beim Aufgabentyp 1, der *Vergleichsaufgaben*, Tabellen ($\beta = .22$) knapp vor Rechnungen ($\beta = .20$) und Zeichnungen ($\beta = .18$) als stärkster Prädiktor gelten, dominierten Zeichnungen ($\beta = .51$) beim Aufgabentyp 2, den *Aufgaben mit komplexen Informationen*. Beim Aufgabentyp 3, den *Aufgaben zur Verhältnisverteilung*, nahmen Zeichnungen dagegen keinen signifikanten Einfluss auf die Vorhersage ($\beta = .04$). Hier dominierten die Rechnungen ($\beta = .22$), gefolgt von den Tabellen ($\beta = .13$) und den schriftsprachlichen Beschreibungen ($\beta = .10$).

Durch die aufgabenspezifische Analyse wurde transparent, dass die Vorhersagekraft der Repräsentationen vom zu bearbeitenden Aufgabentyp abhängig war.



6 Diskussion

In Kapitel 6 werden die Ergebnisse der empirischen Untersuchung und ihre didaktischen Konsequenzen vor dem Hintergrund theoretischer Ansätze diskutiert. Um den Ausgangspunkt der Diskussion herauszustellen, werden zunächst die zentralen Ziele und Fragestellungen der Interventionsstudie zusammenfassend reflektiert und durch die Ergebnisse der Pilotierung gestützt (Abschnitt 6.1). Der weitere Aufbau des Kapitels folgt der Darstellung des Ergebnisteils: Zunächst wird im Rahmen der Forschungsfrage 1 der Effekt der Interventionsmaßnahme diskutiert (Abschnitt 6.2). Danach werden sowohl die Effekte des Trainings (Abschnitt 6.3) als auch die Effekte der kommunikativen Zweiersettings (Abschnitt 6.4) separat aufgezeigt und erörtert. Im Anschluss daran steht der Zusammenhang der beiden Faktoren *Training* und *komunikative Zweiersettings* im Vordergrund der Diskussion (Abschnitt 6.5). Abschließend wird der Fokus auf die Effekte, die auf die externen Repräsentationen zurückgeführt werden können, gerichtet (Abschnitt 6.6). In Abschnitt 6.7 werden die methodischen Überlegungen reflektiert und diskutiert, sofern dies noch nicht im Kapitel 4 erfolgte. Im Anschluss daran werden die Implikationsmöglichkeiten für die Praxis (Abschnitt 6.8) aufgezeigt. Die Schlusszusammenfassung und der Ausblick runden den Diskussionsanteil ab (Abschnitt 6.9).

Es bleibt darauf hinzuweisen, dass die Grenzen der Studie sowie Ansätze für zukünftige Forschungsvorhaben nicht in einem eigens dafür vorgesehenen Kapitel diskutiert werden. Es wurde für sinnig und nachvollziehbarer erachtet, dies an entsprechender Stelle aufzugreifen, mitzudiskutieren und nicht vom jeweiligen Kontext zu lösen.

6.1 Ziele und Fragestellung der Interventionsstudie

Aus psychologischer Sicht können externe Repräsentationen insbesondere dann, wenn sie vom Lösenden selbst generiert werden, zum lösungsunterstützenden Denk- und Erkenntniswerkzeug werden. Sie fungieren als Gedächtnissstütze, auf die sich der Lösende immer wieder beziehen kann, anhand der Veränderungen und Umstrukturierungen aktiv vollzogen, aber auch wieder rückgängig gemacht werden können. Prozesse des Denkens und Lösens werden gestärkt, sobald der Lösende die Rolle des Beobachters einnimmt und den Lösungsprozess von außen betrachtet, neu durchdenkt und reflektiert (siehe Abschnitt 1.5.1). Unbestritten bleibt, dass der Lösende die Textaufgaben verstehen muss, um die wesentlichen von den unwesentlichen Informationen separieren und adäquat abbilden zu können. Es stellt sich die Frage danach, inwieweit ein Repräsentationstraining die Lernenden unterstützt, adäquate Repräsentationen zu konstruieren, mit deren Hilfe die Bewältigung problemhaltiger Textaufgaben gelingen

kann. In der vorliegenden Arbeit wurde anhand eines experimentellen Designs zunächst untersucht, ob sich die Kinder auf das Generieren externer Repräsentationen einlassen, und wenn ja, welche Denk- und Erkenntniswerkzeuge sie in ihrem Lösungsprozess konstruieren. Zudem wurde überprüft, ob sich das Repräsentationstraining positiv auf den Lösungserfolg und die Problemlösekompetenzen auswirkt und ob das Repräsentationstraining gegenüber dem regulären Mathematikunterricht Vorzüge bietet. Aufgrund der Annahme, dass externe Repräsentationen im intersubjektiven Austausch zur Kommunikations- und Argumentationsstütze werden können (siehe Abschnitt 1.5.5.2–1.5.5.3), von denen beide Kommunikationspartner profitieren können, wurde überprüft, ob sich die Implementierung kommunikativer Zweiersettings positiv in der Leistungsentwicklung widerspiegelt. Zudem wurde untersucht inwiefern der jeweilige Aufgabentyp die Vorhersage beeinflusst.

Ausgangsfragestellung

In der Pilotierung wurde untersucht, ob eine fünfwochige Interventionsmaßnahme eine häufigere Generierung und Nutzung externer Repräsentationen bewirkt, konkret, ob sich die Häufigkeit der Nutzung vom Prä- zum Posttest signifikant verändert. Ein Prä-Post-Vergleich der erzielten Leistungen (Lösungserfolg, Problemlösekompetenzen) wurde in diesem Rahmen nicht verfolgt (siehe Abschnitt 4.2). Da angenommen wird, dass die Nutzung externer Repräsentationen unabhängig von der Schwierigkeit der Aufgabe ist, sofern sie nicht sehr leicht oder sehr schwer sind, wurde in der Pilotierung lediglich diese inferenzstatistische Auswertung verfolgt.

Die Pilotstudie konnte die Hypothese bestätigen, dass das Repräsentationstraining die Lernenden anregte, externe Repräsentationen zu generieren und in ihren Lösungsprozess zu integrieren. Im Prä-Post-Vergleich spiegelte sich eine signifikante Zunahme der Anzahl selbstgenerierter externer Repräsentationen wider.

Auffällig war, dass die beiden vierten Klassen der Pilotierung bereits vor dem Training externe Repräsentationen als Denk- und Erkenntniswerkzeuge herangezogen und un aufgefordert in ihren Lösungsprozess integriert hatten. Dies schien der empirischen Beobachtung, dass Problemlösenovizen selten von sich aus externe Repräsentationen generieren (vgl. Abschnitt 1.4.2), zu widersprechen. Dies könnte bedeuten, dass die Probanden der Pilotierung nicht als Problemlösenovizen anzusehen sind, sondern bereits auf ein Repräsentationsrepertoire zurückgreifen konnten, dass sie auszubauen fähig waren. Der direkte Vergleich beider Stichproben macht deutlich, dass die Dritt klässler der Hauptuntersuchung zum Zeitpunkt des Prätests im Mittel lediglich $M = 0.34$ ($SD = .02$) externe Repräsentationen pro Aufgabe generierten, was bei insgesamt drei Testaufgaben bedeutet, dass im Mittel lediglich eine von drei Aufgaben mit hilfe einer externen Repräsentation gelöst wurde. Demgegenüber lag die Anzahl bei

den Viertklässlern der Pilotstudie im Mittel bei $M = 1.18$ ($SD = .36$). Infolgedessen generierten die Jahrgangsjahre vor der Interventionsmaßnahme im Mittel eine externe Repräsentation pro Aufgabe. Dies bestätigt die Ergebnisse von Hohn (2012, S. 113–114, 117), dass Drittklässler seltener externe Repräsentationen generieren als Viertklässler (vgl. Abschnitt 1.4.2). Im Vergleich zu den Ergebnissen von Hohn dominierten die Probanden der vorliegenden Untersuchung in der Konstruktion von Denk- und Erkenntniswerkzeugen.

Es zeigte sich, dass in diesem kurzen Interventionszeitraum die Lernenden bedeutsam mehr auf Tabellen und Zeichnungen zurückgriffen, als dies vorher der Fall war. Wenn man aber berücksichtigt, dass bereits vor der Interventionsmaßnahme einzelne externe Repräsentationen favorisiert wurden (vgl. Abschnitt 5.2), dann kann dieser Effekt nicht generalisiert werden. Zudem spricht gegen eine Generalisierung, dass zum einen die Stichprobe der Pilotstudie lediglich aus zwei Klassen bestand und Klassen- und Lehrereffekte als nicht kontrolliert anzusehen waren. Zum anderen, dass weder die Ausgangsvoraussetzung der Lernenden noch weitere für den Lösungserfolg bedeutsame Einflussfaktoren erhoben wurden. Aus diesen Gründen sind die signifikanten Ergebnisse und daraus resultierende Interpretationen nur als Tendenzen zu verstehen, die es in der Hauptstudie zu replizieren galt.

Einen weiteren kritischen Punkt stellt die Variation in der Anzahl an Eigenproduktionen in den Reflexionsphasen dar. Es gab Unterrichtsstunden, in denen drei, vier oder auch fünf Eigenproduktionen in der Reflexionsphase thematisiert wurden. Begründend hierfür war die Identifikation der passenden Anzahl zu thematisierender Eigenproduktionen im Hinblick auf die Hauptuntersuchung. Demzufolge lief das Training selten identisch ab. Es variierte in seinem zeitlichen Umfang, was eine zusätzliche Störvariable darstellt, um den signifikanten Effekt interpretieren zu können. Zusammenfassend bleibt jedoch festzuhalten, dass die Pilotierung neben all ihren Grenzen wesentlich zur Klärung unzähliger methodischer Fragen sowie zur Optimierung und Abstimmung der einzelnen Trainingsbausteine beitrug. Sie zeigte, dass sich die Kinder kreativ auf die Generierung externer Repräsentationen einließen, was ein Repräsentationstraining zu rechtfertigen scheint.

Die Beobachtung, dass die Kommunikations- und die Nicht-Kommunikationsklasse kaum voneinander differierten, lediglich unterschiedliche externe Repräsentationen fokussierten, ist ein Indiz dafür, dass sich die Implementierung von kommunikativen Settings hypotheseskonform nicht auf die Anzahl selbstgenerierter externer Repräsentationen auswirkt (vgl. Hypothese 1.3 und 3.3).

6.2 Forschungsfrage 1: Effekt der Interventionsmaßnahme

Im Rahmen der ersten Forschungsfrage wurde über alle Gruppen hinweg untersucht, ob die Implementierung problemhaltiger Textaufgaben über die Zeit einen Effekt auf den Lösungserfolg, die Problemlösekompetenzen und die Anzahl selbstgenerierter externer Repräsentationen ausübt.

Bevor die Ergebnisse jedoch diskutiert werden, soll an dieser Stelle aufgezeigt werden, warum die Haupteffekte interpretiert werden dürfen, obwohl signifikante Interaktionseffekte vorliegen. Dies ist problematisch, wenn die Bedeutsamkeit des Haupteffekts beispielsweise auf den hohen Zellmittelwert in einer Bedingungskombination zurückzuführen ist (Eid et al., 2011, S. 430). Aus diesem Grund wurde sichergestellt, dass der signifikante Effekt nicht nur auf einem großen bedingten Haupteffekt in einer Bedingungskombination beruht. Wie aus den grafischen Darstellungen der Mittelwertsunterschiede hervorgeht, gab es einen Effekt über die Zeit, der sich für alle drei abhängigen Variablen in allen Stufen der experimentellen Bedingungen zeigte (siehe Abbildung 6.1). Demnach darf geschlussfolgert werden, dass der Faktor Zeit in allen Stufen der experimentellen Bedingung einen Effekt hat. Alle Effekte mit $\eta_p^2 > .42$ sind nach Cohen (1988, S. 287) als groß einzustufen.

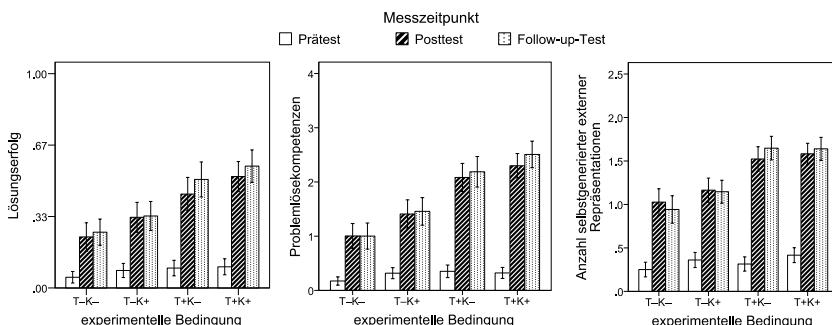


Abbildung 6.1. Mittelwertsunterschiede für den Lösungserfolg, die Problemlösekompetenzen und die Anzahl selbstgenerierter externer Repräsentationen für alle experimentellen Bedingungen. T+K+ = Training, kommunikative Zweiersettings; T+K- = Training, keine kommunikativen Zweiersettings; T-K+ = kein Training, kommunikative Zweiersettings; T-K- = kein Training, keine kommunikativen Zweiersettings.

Den Ergebnissen zufolge haben im Mittel alle Gruppen vom Einsatz problemhaltiger Textaufgaben im Mathematikunterricht profitiert. Nach der Intervention waren die Probanden im Mittel häufiger in der Lage, die Problemaufgaben richtig zu lösen, ver-

fügten über höhere Problemlösekompetenzen und integrierten signifikant mehr externe Repräsentationen in ihren Lösungsprozess als davor.

Dies lässt den Schluss zu, dass die Aufgabendomäne problemhafte Textaufgaben durchaus Berechtigung in der Grundschule findet und deren Lösen ebenso wie die Generierung von Denk- und Erkenntniswerkzeugen trainiert werden kann. Die Interventionsmaßnahme gilt demzufolge als wirksam und effektiv (vgl. Abschnitt 4.3). Dass die Effekte bei allen Aufgabentypen nachgewiesen werden konnten, ist ein Indiz dafür, dass das Training nicht nur einzelne Textaufgabentypen fördert, sondern beim Lösen problemhafter Textaufgaben aufgabenunspezifisch Wirkung zeigt. Folglich haben sich die fünf zentralen Trainingsbausteine beim Lösen problemhafter Textaufgaben bewährt (siehe Abschnitt 4.4.4): Das Hineinversetzen, Nachvollziehen und Verstehen fremder Repräsentationen, deren Benennung und Beschreibung sowie das kollektive Aushandeln lösungsunterstützender Repräsentationsmerkmale hilft den Lernenden ein Repräsentations-Repertoire zu entwickeln und weiterzuentwickeln.

Weitere Studien müssen jedoch zeigen, inwiefern die gefundenen Effekte auf andere Inhaltsbereiche, außerhalb des Textaufgabenlösens, übertragbar sind. Möglicherweise sind problemhafte Textaufgaben in besonderer Weise für das Lösen auf Basis externer Repräsentationen geeignet. Ferner ist auch denkbar, dass die Wirkung der Interventionsmaßnahme vom Alter der Lernenden beeinflusst ist. Möglicherweise sind insbesondere Grundschulkinder offen für neue Heran- und Vorgehensweisen, wohingegen ältere Lernende verschlossener sind und auf ihnen vertraute, arithmetische Vorgehensweisen zurückgreifen.

Die aufgabenspezifischen Analysen bestätigten bisherige Befunde dahin gehend, dass sich gemittelt über die experimentellen Bedingungen sowohl der Lösungserfolg und die Problemlösekompetenzen als auch die Anzahl selbstgenerierter externer Repräsentationen zwischen den Aufgabentypen unterscheiden. Hieraus lässt sich einerseits ableiten, dass den Aufgaben unterschiedliche Schwierigkeit zugrunde liegt, was die Ergebnisse der auf den Pilotierungsaufgaben basierenden Itemanalyse stützt. Andererseits liegen die Vermutungen nahe, dass entweder die Aufgaben nicht in gleicher Weise für das Lösen mit externen Repräsentationen geeignet sind oder der Einsatz externer Repräsentationen bei einigen Aufgaben weniger naheliegend und lösungsunterstützend scheint als bei anderen Aufgaben.

Ist der Fokus auf die einzelnen Aufgaben gerichtet, so fällt auf, dass unter allen drei Aufgabentypen die Vergleichsaufgabe als die leichteste und die Aufgabe zur Verhältnisverteilung als die schwerste identifiziert werden konnte, wenngleich alle ihre Problemhaftigkeit beibehielten, was an den geringen Lösungsraten deutlich wird. Die bisherigen empirischen Befunde sind bezüglich der Vergleichsaufgaben heterogen: So wohl Lösungswahrscheinlichkeiten von 38 % (R. Rasch, 2009, S. 82) als auch Lö-

sungswahrscheinlichkeiten von 15 % konnten bei Gleichaltrigen nachgewiesen werden (Stern, 1998, S. 16 s. a. Abschnitt 4.5.1.2). In der vorliegenden Untersuchung schafften vor der Interventionsmaßnahme 20 % der Drittklässler eine erfolgreiche Problemlösung, sodass sie mit der Stichprobe von Stern vergleichbar sind.

Über die Zeit gesehen konnten signifikante Unterschiede im Lösungserfolg, den Problemlösekompetenzen und der Anzahl selbstgenerierter externer Repräsentationen zwischen den Aufgabentypen, die in der Intervention bearbeitet wurden, und den Aufgabentypen, die nicht in der Intervention bearbeitet wurden, beobachtet werden. Zwischen den trainierten Aufgabentypen gab es in Zusammenhang mit dem Lösungserfolg und den Problemlösekompetenzen keine signifikanten Unterschiede.²³⁰ Beides ist ein Indiz dafür, dass die Wirkung der Interventionsmaßnahme bei trainierten, in der Intervention behandelten Aufgabentypen größer ausfällt als bei untrainierten, nicht behandelten Aufgabentypen. Die Effekte zwischen dem Prä- und Posttest sind nach Cohen (1988, S. 287) von mittlerer bzw. großer Stärke (Lösungserfolg: $\eta_p^2 = .09$; Problemlösekompetenzen: $\eta_p^2 = .23$).

Darüber hinaus fiel auf, dass die *Aufgabe zur Verhältnisverteilung* signifikant seltener mit externen Repräsentationen gelöst wurde, als die beiden anderen trainierten Aufgabentypen. Hierfür wären mehrere Gründe denkbar:

1. Nicht auszuschließen ist, dass die Drittklässler ohne Unterstützung durch die Interventionsmaßnahme nicht in der Lage waren, a) die Problemstellung experimentell zu bewältigen bzw. b) ihre Lösungsgedanken und -ansätze zu externalisieren. Dies führt zum Schluss, dass eine Transferwirkung auf untrainierte Aufgabengruppen ausgeblieben wäre.
2. Es kommt ebenso in Betracht, dass bei diesem Aufgabentyp arithmetische Züge dominieren, die wiederum voraussetzen, dass die gegebenen Zahlenwerte entsprechend den Anforderungen aufgeteilt werden und Wissen über Zahlbeziehungen angewandt wird (vgl. Abschnitt 4.5.1.2). Angesichts dessen wäre der Aufgabentyp mit Rechnungen bevorzugt, eventuell mit Tabellen, weniger aber mit Zeichnungen und Beschreibungen zu lösen, wohingegen das Kriterium der Repräsentationsvielfalt verletzt wäre (siehe Abschnitt 4.1.2). Favorisieren Kinder bei diesem Aufgabentyp eine Repräsentationsform, kann bzw. darf nicht erwartet werden, dass sie vielfältige Repräsentationen generieren und auf Formen zurückgreifen, die sich als wenig geeignet erweisen. Der ausbleibende Effekt würde dann in unmittelbarem Zusammenhang mit der Aufgabe zur Ver-

²³⁰ Für die abhängige Variable Anzahl selbstgenerierter externer Repräsentationen wurde ein signifikanter Unterschied zwischen den trainierten Aufgabentypen gefunden. Der Effekt ist jedoch mit $\eta_p^2 = .02$ bzw. $\eta_p^2 = .03$ als gering einzustufen (Cohen, 1988, S. 285).

hältnisverteilung stehen und wäre nicht auf weitere untrainierte Aufgabentypen übertragbar.

3. Wenn die Nutzung externer Repräsentationen bei der Aufgabe zur Verhältnisverteilung für die Kinder nicht naheliegend schien, dann wäre denkbar, dass sie dazu tendierten, die Aufgabe im Kopf zu lösen. Kam hinzu, dass sie ihre im Kopf produzierte Lösungszahl nicht hinterfragten und als korrekt annahmen, weil dieser Aufgabentyp im Unterricht unreflektiert blieb (Chinn & Brewer, 1993, S. 20–21, s. a. Abschnitt 4.5.1.2), so war keine Änderung im Leistungsverhalten und demnach auch keine Steigerung der Anzahl selbstgenerierter externer Repräsentationen zu erwarten. Selten weichen Kinder von ihrem Lösungsvorgehen ab, wenn sie davon ausgehen die Aufgabe erfolgreich gelöst zu haben (R. Rasch, 2001b, S. 257, s. Kapitel 4.5.1.2).

Für Drittens spricht, dass bei der *Aufgabe zur Verhältnisverteilung* nach der Interventionsmaßnahme im Vergleich zu den beiden trainierten Aufgabentypen deutlich häufiger die Lösung im Kopf berechnet wurde. Die Generierung externer Repräsentationen blieb zum Zeitpunkt des Prätests bei der Aufgabe mit komplexen Informationen (72 %) und der Aufgabe zur Verhältnisverteilung (79 %) häufiger aus als bei der Vergleichsaufgabe (58 %). Nach der Interventionsmaßnahme lösten die *Vergleichsaufgabe* nur noch 11 %, die *Aufgabe mit komplexen Informationen* nur noch 17 % der Kinder ohne externe Repräsentationen. Bei der *Aufgabe zur Verhältnisverteilung* waren es jedoch mit 32 % fast doppelt bzw. dreimal so viel Kinder, die das Lösen im Kopf ohne Externalisierungen beibehielten. Wenn Lernende ihren Lösungsprozess externalisierten, griffen sie fast ausschließlich auf deskriptionale Repräsentationen zurück: 3 % wählten einen zeichnerischen, 42 % einen rechnerischen, 34 % einen tabellarischen und 26 % der Kinder einen beschreibenden Zugang. Die Summe überschreitet knapp die 100 %, weil teilweise mehr als eine Repräsentation generiert wurde. Dass die Lernenden hauptsächlich deskriptionale Repräsentationen favorisierten, spricht für eine untergeordnete Relevanz depiktionaler Repräsentationen bei diesem Aufgabentyp und folglich für den unter 2. aufgeführten Grund.

Die Ursache dafür, dass die *Aufgabe zur Verhältnisverteilung* signifikant seltener mit externen Repräsentationen gelöst wurde als die trainierten Aufgabentypen, darf aus vorstehenden Befunden nicht in einer ausbleibenden Transferwirkung gesehen werden. Die Wahl des untrainierten Aufgabentyps, Aufgaben zur Verhältnisverteilung, scheint eine mögliche Transferwirkung der Interventionsmaßnahme in Bezug auf das Generieren externer Repräsentationen zu beeinflussen, sogar zu verhindern. Es wäre in der weiteren Forschung vonnöten, genauer auf aufgabenspezifische Effekte zu achten.

6.3 Forschungsfrage 2: Effekt des Repräsentationstrainings

Im Rahmen der 2. Forschungsfrage wird die Wirkung des Repräsentationstrainings auf den Lösungserfolg, die Problemlösekompetenzen und die Anzahl selbstgenerierter externer Repräsentationen überprüft. Die Bestätigung der Hypothese, dass Kinder, die durch ein Training angeleitet und gefördert wurden, externe Repräsentationen zu generieren und für ihre Lösungsfindung zu nutzen, erfolgreichere Problemlöser sind als Kinder, die diese spezielle Förderung nicht erhalten, betont das vielseitige Potenzial externer Repräsentationen und deren Einsatz im Unterricht. Ferner ist die signifikante Zunahme an selbstgenerierten externen Repräsentationen ein Hinweis dafür, dass die angestrebten Ziele durch das Training erreicht wurden. Die Kinder ließen sich auf die problemhaften Textaufgaben sowie die damit verbundenen Handlungsauforderungen ein und griffen mehr und mehr auf ihre individuellen Ressourcen zurück, welche sie zugleich externalisierten. In diesem Zusammenhang konnte gezeigt werden, dass die Anzahl selbstgenerierter externer Repräsentationen hoch mit den Problemlösekompetenzen korreliert und somit ein enger Zusammenhang zwischen beiden Variablen besteht. Wenn vergleichend herangezogen wird, dass die Korrelation mit dem Lösungserfolg geringer ausfällt als mit den Problemlösekompetenzen, dann lässt sich aus Sicht der *Representational Change Theory* ableiten, dass häufig mit einer inadäquaten externen Repräsentation gestartet wird, jedoch nicht jedem Lösenden eine erfolgreiche Umstrukturierung gelingt, die ihn zur Einsicht führt (vgl. Abschnitt 1.3.3.2). Das Training und somit die einzelnen Trainingsbausteine (siehe Abschnitt 4.4.4) verhelfen einem Großteil der Kinder, die inadäquate Repräsentation so zu verändern und umzustrukturieren, dass sie Teileinsichten (*partial insights*) und demzufolge Teilkompetenzen erzielen, die sie der Lösung näherbringen, was sich im Prozess, aber noch nicht im Endprodukt widerspiegelt.

Den Ergebnissen zufolge unterschieden sich die Trainings- und Nicht-Trainingsgruppen in ihrem Lösungserfolg, ihren Problemlösekompetenzen und der Anzahl selbstgenerierter externer Repräsentationen vom Prä- zum Posttest signifikant voneinander. Die Gruppen konnten vom Post- zum Follow-up-Test ihren Leistungszuwachs aufrechterhalten oder noch geringfügig ausbauen, was für ein nachhaltiges Training spricht. Die aufgabenspezifische Analyse deckte allerdings auf, dass dieser Effekt vom Post- zum Follow-up-Test fast ausschließlich auf die Aufgabe zur Verhältnisverteilung des untrainierten Aufgabentyps zurückzuführen war. Während bei den beiden trainierten Aufgabentypen ein geringer, für Interventionsmaßnahmen üblicher Leistungsabfall zu beobachtet war, der aufgrund fehlender Signifikanz für eine nachhaltige Wirkung steht, zeigten die Kinder beim untrainierten Aufgabentyp ausschlaggebende Steigerungen. Gemittelt über alle drei Testaufgaben ergab sich daher eine nahezu konstante bzw. steigende Entwicklung vom Post- zum Follow-up-Test.

Es stellt sich nun die Frage, warum bei dem Aufgabentyp, der in der Interventionsmaßnahme unberücksichtigt blieb und demzufolge nie mit den Kindern reflektiert wurde, eine Steigerung im Lösungserfolg, den Problemlösekompetenzen wie auch der Anzahl selbstgenerierter externer Repräsentationen erfolgte. Gründe hierfür können zum einen auf curriculare Inhalte zurückgeführt werden. Es ist naheliegend, dass der reguläre Mathematikunterricht in diesem Zeitraum verstärkt das Verständnis von Zahlbeziehungen schulte (KMK 2005b, S. 9) und die Kinder ihr darüber angeeignetes Wissen nutzten und anwandten (vgl. Abschnitt 4.5.1). Dass die Kinder schwerpunktmäßig Rechnungen bei diesem Aufgabentyp generierten, untermauert diese Vermutung. Es ist denkbar, dass sie im Zuge der Zeit ihre Rechenkompetenzen erweitern konnten und zum Zeitpunkt des Follow-up-Tests, in der Mitte des dritten Schuljahres, in der Lage waren, die Aufgabe rechnerisch zu bewältigen, wohingegen sie zum Prätest, zu Beginn des dritten Schuljahres, hierzu noch nicht in der Lage waren. Die Tatsache, dass der Effekt sowohl in der Trainings- als auch der Nicht-Trainingsgruppe beobachtbar war, stützt die Vermutung, dass aufgrund des regulären Mathematikunterrichts mehr Kinder in der Lage waren, die Aufgabe erfolgreich zu bewältigen. Der Vorteil zugunsten der Trainingsgruppe könnte ein Indiz dafür sein, dass das im Training anvisierte Schulen des Strukturblicks dazu verhalf, dass bereits zum Posttest mehr Kinder die zugrunde liegenden Aufgabenstrukturen erfassen und die Probleme erfolgreich lösen konnten. Wenn Kinder einmal durchschaut und verstanden haben, wie ProblembARRIEREN mithilfe erinnerter Vorerfahrungen überwunden werden können, dann können sie dies auf beliebige strukturgleiche Aufgaben anwenden. In diesem Fall würde die Aufgabe ihren Problemcharakter verlieren.

Ausgeschlossen werden kann, dass die Kinder die Zahlenwerte erinnerten und lediglich reproduzierten. Im Vergleich zum Prä- und Posttest variierten die Zahlenwerte in der Aufgabenstellung des Follow-up-Tests, sodass eine reine Reproduktion der Lösungszahl zu einem falschen Ergebnis geführt und folglich keine Steigerung bewirkt hätte. Diese Ursache kommt somit nicht in Betracht. Insgesamt bleibt festzuhalten, dass aufgrund der Datenlage der Effekt nicht erklärt werden kann. Interviews im Anschluss an die Problembewältigung hätten hierüber Aufschluss geben können. Ohne diese bleibt die Ursache für diesen Effekt unerklärt und spekulativ.

Die Studie zeigt außerdem, dass der Einfluss des Repräsentationstrainings über die Zeit auf die Problemlösekompetenzen ($\eta^2 = .12$) größer war als auf den Lösungserfolg ($\eta^2 = .06$), auch wenn in beiden Fällen Effekte mittlerer Stärke vorliegen ($.06 < \eta^2 < .14$, Cohen, 1988, S. 287). Wird lediglich der Leistungsanstieg vom Prä- zum Posttest fokussiert, bleibt dieser Unterschied zugunsten der Problemlösekompetenzen ($\eta^2 = .15$ im Vergleich zu $\eta^2 = .10$ für den Lösungserfolg) bestehen. Dieses Ergebnis ist in Anbetracht der Charakteristika problemhaltiger Textaufgaben nicht verwunderlich (vgl.

Abschnitt 1.4.1). Sie bilden einen Aufgabentypen ab, der insbesondere Problemlösenovizen herausfordert und an ihre Grenzen stoßen lässt. Empirische Studien deckten – wie in Abschnitt 1.4.2 dargelegt – auf, dass Lernende Unterstützung benötigen, um sowohl adäquate Repräsentationen entwickeln als auch inadäquate Repräsentationen zielführend umstrukturieren zu können. Dies stellt einen langfristigen Prozess dar, der mit vielfältigen Erfahrungen einhergehen muss (vgl. Abschnitt 1.5.5.3–1.5.5.4 und 3.2). Aufgrund dessen kann in dieser Aufgabendomäne eine zwölfwöchige Intervention nicht alle Lernenden direkt zum Erfolg führen, erst recht nicht in gleichem Maße. Weil die abhängige Variable Problemlösekompetenzen den Prozess beurteilt und demzufolge ermöglicht, bereits Teilkompetenzen zu erfassen, werden hypothesenkonform größere Effekte als unmittelbar auf den Lösungserfolg festgestellt (vgl. Abschnitt 4.5.1.1). In Anbetracht dessen, dass mit dem Lösungserfolg lediglich die Richtigkeit des Ergebnisses gemessen wird, ist der wirkte signifikante Effekt jedoch als bedeutsam einzustufen. Bereits durch eine Interventionsmaßnahme, die 12 Problembewältigungen aus sechs Aufgabentypen umfasst, können nachhaltige Effekte erzielt werden, die sich sowohl im Lösungsprozess als auch im Produkt, dem Ergebnis, widerspiegeln. Es wird prognostiziert, dass durch ein längeres Training, beispielsweise über ein Schuljahr und über ein breiteres Spektrum an Problemaufgaben hinweg, größere Effekte hervorgerufen werden können.

Wenn die Gruppen fokussiert werden, die nicht am Repräsentationstraining partizipierten, sondern die problemhaltigen Textaufgaben im Mathematikunterricht mit der regulären Lehrkraft bearbeiteten, dann bleibt festzuhalten, dass auch sie ihrer Leistungen und Anzahl selbstgenerierter externer Repräsentationen über die Zeit signifikant steigern konnten. Dass der Zuwachs in der Trainingsgruppe signifikant größer ausfiel, wird auf die unterschiedlichen Interventionsmaßnahmen zurückgeführt. Es stellte sich die Frage danach, worin sich die beiden Maßnahmen unterscheiden und wie sie sich voneinander abgrenzen lassen. Hierzu wurden die von den Lehrkräften angefertigten Stundenprotokolle und Tafelbilder sowie die Knobelhefte der Kinder aus den Nicht-Trainingsgruppen herangezogen.

In der Reflexionsphase des Trainings lernten die Kinder nicht nur facettenreiche Heran- und Vorgehensweisen kennen, sondern sie lernten vor allem auch, diese nachzuvollziehen, zu verstehen, zu interpretieren sowie miteinander zu vernetzen und ineinander zu überführen (siehe Trainingsbausteine, Abschnitt 4.4.4). Es liegt die Vermutung nahe, dass insbesondere durch die Thematisierung verschiedener externer Repräsentationen Möglichkeiten aufgezeigt werden, wie die Überwindung von Problembarrieren auf vielfältigen Wegen gelingen kann. Repräsentationsformen wie Zeichnungen, Tabellen, Rechnungen und schriftsprachliche Beschreibungen sowie das kollektive Aushandeln lösungsunterstützender Repräsentationsmerkmale zum Gegenstand der

Reflexionsphase zu machen, beugt zusätzlich der Fokussierung einzelner Denk- und Erkenntniswerkzeuge vor und schafft eine breite Basis möglicher Vor- und Herangehensweisen, anhand der sich ein Repräsentationsrepertoire entwickeln kann. Hierin wird ein bedeutsamer Unterschied zu der Vergleichsgruppe, der Nicht-Trainingsgruppe, gesehen. Weder die Stundenprotokolle und Tafelbilder der Lehrkräfte noch die Eigenproduktionen der Kinder wiesen auf ein vergleichbares Vorgehen im Unterricht hin. Vielmehr geht aus den Materialien hervor, dass eine externe Repräsentation, meist eine Zeichnung oder eine Rechnung, generiert wurde. Innerhalb einer Klasse unterschieden sich die Knobelhefte der Kinder selten in den Ausführungen, was als Hinweis für eine frontale Bearbeitung der Textaufgaben gedeutet wird. Vereinzelt wurde in den Knobelheften Phasen des eigenständigen oder auch kooperativen Lernens sichtbar. Auffällig war, dass die individuellen Notizen und Lösungsideen häufig durch Streichen oder Radieren entfernt und durch eine „Musterlösung“, die auch aus dem Tafelbild der Lehrkraft hervorging, ersetzt wurde. Demzufolge wird der Hauptunterschied beider Interventionsmaßnahmen vorrangig in der Unterrichtsorganisation, sprich der Sozialform und der Eigenverantwortlichkeit, aber auch in der Thematisierung verschiedener Repräsentationen gesehen. Es geht daher konform mit den Befunden von Heinrich, Pawlitzki und Schuck (2013, S. 450), dass in den Reflexionsphasen des Problemlöseunterrichts die Lehrersteuerung dominierte und selten Lösungsansätze der Kinder aufgegriffen und bewusst gemacht wurden oder über strategische Fehler diskutiert wurde. Den Beobachtungen zufolge wurde im Unterricht der Nicht-Trainingsgruppe nicht explizit trainiert, verschiedene Repräsentationen zu konstruieren, zu vergleichen, Bezüge herzustellen, Vor- und Nachteile abzuwägen sowie diese Erfahrungen des Voneinander-Lernens in einer der nächsten Problemlösungen aufzugeben und auszuprobieren.

Die in der Fachdidaktik kontrovers diskutierte Frage, ob das Einbeziehen unterschiedlicher Lösungswege lernförderliche oder -hemmende Funktion einnimmt (vgl. Abschnitt 1.5.5.4), kann auf Basis der Untersuchung positiv evaluiert werden. Sie werden als Lernunterstützung wahrgenommen, weil sie im Training nicht nebeneinanderstehend abgearbeitet wurden. Vielmehr stand das Herstellen von Bezügen, das Schaffen eines Bewusstseins für facettenreiche Vorgehensweisen sowie das Abwägen und Diskutieren der Vor- und Nachteile im Vordergrund (vgl. Abschnitt 1.5.5.3). Da externe Repräsentationen sowohl zur Aufgabe als auch zu dem Lösenden passen müssen (vgl. Abschnitt 1.5.6), bietet gerade die Vielfalt unterschiedlicher Heran- und Vorgehensweise die Möglichkeit, verschiedene Lerntypen anzusprechen und ein adäquates Werkzeug zur Verfügung zu stellen, welches wiederum in Folgeaufgaben exploriert werden kann.

Unklar bleibt jedoch, ob in den Nicht-Trainingsgruppen bei der „Musterlösung“ auf eine von der Lehrkraft vorgefertigte Lösung oder auf die Eigenproduktion eines Kindes zurückgegriffen wurde. Demzufolge kann die Wirkung des Trainings nicht explizit auf das Reflektieren selbstgenerierter externer Repräsentationen zurückgeführt werden. Weitere Untersuchungen müssen zeigen, ob ein Training, das auf selbstgenerierten oder vorgefertigten Repräsentationen basiert, unterschiedliche Effekte zur Folge hat.

Ein Fokus der Studie war es, in den Nicht-Trainingsklassen den regulären Problemlöseunterricht abzubilden und zu erfassen. Um einen möglichst unbeeinflussten und realen Unterrichtskontext zu schaffen, wurden Mathematiklehrkräfte als Experten für diese Manipulation angesehen, sodass sich durchaus pro Nicht-Trainingsklasse andere durchführende Personen ergaben (Hager, 2008, S. 724). Nach Hager (2008, S. 724) sollen „die miteinander zu vergleichenden Trainingsmaßnahmen einander in keiner Weise angepasst werden“ und vielmehr die pro Training durchführende Personen als „Paket“ angesehen werden. Die Festsetzung weniger Rahmenbedingungen sollte sicherstellen, dass die regulären Lehrkräfte in der Gestaltung ihres Unterrichts so wenig wie möglich eingeschränkt sind. Um eine zusätzliche Beeinflussung zu verhindern, blieb ihnen die Zielsetzung der Untersuchung verborgen. Es wird angenommen, dass sich der Einfluss der unterrichtenden Lehrkräfte bei zehn Lehrkräften nahezu herauspartialisiert.

Vor dem Hintergrund, dass die Lehrkraft als systematische Störvariable gilt (Eid et al., 2011, S. 57), die unbewusst Einfluss auf die Motivation und letztlich auf die Leistung der Kinder nehmen kann, darf die dadurch hervorgerufene Wirkung nicht ignoriert werden. Es ist nicht auszuschließen und daher als Schwäche der Untersuchung anzusehen, dass sich die Versuchsleiterpersönlichkeit als situationsgebundener Störfaktor auf die abhängigen Variablen auswirkt (Eid et al., 2011, S. 57), auch wenn davon auszugehen ist, dass die Mathematiklehrkräfte aufgrund ihrer freiwilligen Teilnahme an der Untersuchung motiviert waren, die problemhaften Textaufgaben in ihren Mathematikunterricht bestmöglich zu integrieren.

Um dieser Schwäche der Untersuchung entgegenzuwirken, sollte bei einer Wiederholung des Experimentes ermöglicht werden, dass beispielsweise nur eine Lehrkraft alle Klassen unterrichtet. Bei einer großen Anzahl an Klassen wäre es denkbar, die Hälfte aller Trainingsklassen und die Hälfte aller Nicht-Trainingsklassen parallel zu unterrichten, und im Anschluss daran die anderen beiden Hälften. Durch die Splitting der Untersuchung kann nicht verhindert werden, dass die Gruppen, die in der zweiten Erhebungsphase an der Untersuchung teilnehmen, in ihrer allgemeinen Denk- und Lernentwicklung weiter fortgeschritten sind als die Kinder, die zuerst an der Untersuchung teilnehmen. Dies wird kontrollierbar, indem die Trainings- und Nicht-

Trainingsgruppen halbiert und immer zwei gegensätzliche Hälften parallel unterrichtet werden. Um die Gefahr zu kontrollieren, dass die Versuchsleitung (unbewusst oder bewusst) durch ihr Verhalten während der Durchführung des Trainings die Ergebnisse hypothesenkonform verzerrt (Versuchsleitereffekt), ist auch denkbar, dass alle Nicht-Trainingsklassen von derselben Lehrkraft unterrichtet werden. Hierbei sollte es sich idealerweise weder um die Versuchsleitung noch die reguläre Lehrkraft handeln.

In der vorliegenden Untersuchung ist demzufolge nicht auszuschließen, dass sich der Problemlöseunterricht in der Nicht-Trainingsgruppe durchaus auch hätte höher auf die Leistungen auswirken können, wenn eine fremde Lehrkraft den Unterricht absolviert hätte. Der Einfluss, den eine externe Lehrkraft ausübt, wäre dadurch als kontrolliert anzusehen. Ferner hätte das Professionswissens der Lehrkräfte, explizit deren mathematikdidaktisches Wissen sowohl über lehrbezogene als auch lernprozessbezogene Anforderungen, erhoben werden können (Blömeke et al., 2008, S. 51; Hoth, 2016, S. 29). Ob dieses Wissen in Zusammenhang mit den Leistungsentwicklungen der Kinder steht, hätte dann überprüft und kontrolliert werden können.

Aus den genannten Gründen soll an dieser Stelle weniger der Vergleich zu den restlichen experimentellen Bedingungen hervorgehoben werden, sondern, dass auch dieser Gruppe eine signifikante Leistungssteigerung im Lösungserfolg, in ihren Problemlösekompetenzen und der Anzahl selbstgenerierter externer Repräsentationen gelungen ist.

6.4 Forschungsfrage 3: Effekt der Kommunikation

Im Fokus der dritten Forschungsfrage stehen die Wirkung der kommunikativen Zweiersettings auf den Lösungserfolg, die Problemlösekompetenzen und die Anzahl selbstgenerierter externer Repräsentationen. Es galt, zu überprüfen, ob Kinder, die sich in kommunikativen Zweiersettings über ihren Lösungsprozess austauschten, erfolgreichere Problemlöser sind als Kinder, denen diese Austauschmöglichkeit nicht geboten wurde. Ein Effekt der kommunikativen Zweiersettings auf die Anzahl selbstgenerierter externer Repräsentationen wurde dagegen nicht erwartet.

Insgesamt bleibt festzuhalten, dass der Einfluss der Kommunikation auf die Problemlösekompetenzen und auf die Anzahl selbstgenerierter externer Repräsentationen hypothesenkonform verlief. Die Möglichkeit, sich unter Gleichaltrigen auszutauschen, stand somit nicht in direktem Zusammenhang mit einer höheren Anzahl externer Repräsentationen. Kinder, die sich in Zweiersettings austauschten, erzielten über die Zeit höhere Problemlösekompetenzen als Kinder, die sich nicht austauschten. Eine Übertragbarkeit dieses Effekts auf den Lösungserfolg konnte inferenzstatistisch nicht gegen den Zufall abgesichert werden, auch wenn die deskriptiven Daten eine lernförderliche Wirkung prognostizierten.

Um die von den kommunikativen Zweiersettings ausgehenden Effekte auf die Leistungsentwicklung interpretieren zu können, ist es von essenzieller Bedeutung, sowohl die Anforderungen, die das Kommunizieren an die Interaktionspartner stellt, als auch die Voraussetzungen, die beide Grundschulkinder in die Settings mitbringen, zu berücksichtigen. Empirischen Ergebnissen zufolge können Lernende von einem Austausch unter Gleichaltrigen nur profitieren, wenn sich beide auf das Gespräch einlassen. Dies impliziert, dass gegenseitig auf die Bedürfnisse und Fragen des Gegenübers eingegangen wird, aber auch das Gesagte verstanden und verinnerlicht wird (vgl. Abschnitt 2.4). Es muss ihnen dabei gelingen mit und über die externen Repräsentationen zu kommunizieren, was wiederum voraussetzt, dass die Interaktionspartner über argumentative, soziale und metakognitive Fähigkeiten verfügen (vgl. Abschnitt 2.4.3). Vor dem Hintergrund, dass argumentative und begründende Tätigkeiten zu den höchsten Kompetenzstufen zählen, den Drittklässlern aber jegliche Erfahrung und Übung darin fehlte und sie unmittelbar nach der eigenständigen Lösungsphase mit einer weiteren Herausforderung, dem kommunikativen Austausch, konfrontiert waren, waren die Befunde erwartungsgemäß. Die Tatsache, dass die Kommunikation einen Zuwachs der Problemlösekompetenzen bewirkt, nicht aber im Lösungserfolg, lässt den Schluss zu, dass zwar Teilkompetenzen erworben wurden, diese jedoch noch nicht zwingend in Erfolg münden. Wird beachtet, dass der Effekt der Kommunikation auf die Problemlösekompetenzen mit $\eta_p^2 = .02$ als klein einzustufen ist (Cohen, 1988, S. 285), ist es wenig verwunderlich, dass der Lösungserfolg, der als hartes Kriterium gilt, von einem *Kommunikationstraining* unbeeinflusst zu bleiben scheint. Das Voneinander- und Miteinander-Lernen benötigt Zeit und wird mitunter als schwierig und beschwerlich charakterisiert (Oswald, 1994, S. 11–12, vgl. Abschnitt 2.4). Es liegt daher die Vermutung nahe, dass ein Interventionszeitraum von 12 Wochen nicht ausreichte, um bedeutsame Wirkungen auf den Lösungserfolg zu erzielen. Vielmehr können die Ergebnisse dahin gehend interpretiert werden, dass die kommunikativen Settings lernförderliche Prozesse angestoßen und angebahnt haben. Dass sich dies bereits positiv auf die Problemlösekompetenzen auswirkte, konnte die vorliegende Untersuchung zeigen.

Wenn man in den Mittelpunkt stellt, dass die Grundschulkinder als Novizen des Argumentierens und Begründens ein Medium benötigen, auf das sie ihre Verständigung stützen können (vgl. Abschnitt 1.5.5), dann beeinflussen die von den beiden Interaktionspartnern generierten externen Repräsentationen maßgeblich den stattgefundenen Austausch. Die Kommunikation mit und über die externen Repräsentationen hängt folglich entscheidend davon ab, welche externen Repräsentationen konstruiert wurden, konkret, in welcher Qualität diese vorlagen und ob die Struktur adäquat abgebildet wurde. Es war zu beobachten, dass sich die Drittklässler die geschilderten Sachsituationen ausschließlich experimentell erschlossen hatten (vgl. Abschnitt 2.2). Ihr individueller

Erkundungsprozess war geprägt vom Generieren von Beispielen (Philipp, 2013, S. 89), welche im Idealfall Zusammenhänge und Beziehungen identifizierten (vgl. Abschnitt 2.2.3). Da sich die Fähigkeiten und Fertigkeiten im Umgang mit Darstellungsmitteln nur langsam aufbauen (vgl. Abschnitt 2.2.4) und hierfür nach Freudenthal (1979, S. 197) ausreichend Selbsttätigkeit erforderlich ist, legt nahe, dass den kommunikativen Zweiersettings in den ersten Trainingswochen Repräsentationen zugrunde lagen, die selten die adäquate Struktur abbildeten und häufig noch nicht zur richtigen Lösung führten. Daher ist nicht auszuschließen, dass die Denk- und Erkenntniswerkzeuge der Lernenden erst im Zuge der Interventionsmaßnahme zu „Träger[n] mathematischer Phänomene“ wurden (Philipp, 2013, S. 2, vgl. Abschnitt 2.2.3). Diesem Schluss zufolge, hätte den Lernenden in den ersten Wochen eine adäquate Basis der Kommunikation gefehlt, auf die sie sich hätten stützen können.

Als weitere Ursache für den geringen bzw. ausbleibenden Effekt der kommunikativen Settings kommt in Betracht, dass die Kinder zwar lösungsunterstützende Repräsentationen generierten, sie diese aber noch nicht richtig deuten und/oder verstehen konnten. Ohne den Strukturblick und das richtige Deuten zu trainieren, war zu erwarten, dass die Kinder die Zusammenhänge und Beziehungen nicht zu identifizieren wussten und somit der Austausch als weniger gewinnbringend einzuschätzen war. Demzufolge ist denkbar, dass die Kinder erst mit zunehmender Lösungserfahrung und nach mehreren gemeinsamen Reflexionsphasen in der Lage waren, die externe Repräsentation ihres Partners richtig zu interpretieren (vgl. Abschnitt 1.5.5). Wird hiervon ausgegangen, dann konnte in den ersten Interventionswochen, in denen das Repräsentationstraining wohl eher erst geringe Wirkungen oder sogar noch keine Wirkung zeigte, kaum eine gewinnbringende Kommunikation initiiert werden. Folglich lagen die selbstgenerierten externen Repräsentationen zu Beginn der Intervention noch nicht in der Qualität vor, dass sie von den Lernenden eigenständig und ohne Hilfestellung als Kommunikationsgrundlage genutzt werden konnten. Berücksichtigt man dies, hat streng genommen in den ersten Wochen noch kein Kommunikationstraining stattgefunden.

Losgelöst von den bisherigen Standpunkten ist denkbar, dass es den ungeübten Grundschulkindern sowohl schwerfiel, in Eigenregie ihre Denk- und Erkenntniswerkzeuge und damit verbundenen Gedankengänge nachvollziehbar offenzulegen als auch sich auf fremde, neuartige Repräsentationen ihres Partners einzulassen. Es kommt die Schwierigkeit hinzu, dass junge Lernende beim Erklären häufig egozentrisch agieren, was sich u. a. darin zeigt, dass sie davon ausgehen, dass es ihrem Gesprächspartner auf Anhieb gelingt, das Gesagte zu verstehen und nachzuvollziehen (Piaget, 1972, § 4). Problematisch ist, dass die kommunikative Phase, mit all ihren möglichen Schwierigkeiten, nicht explizit in einem unmittelbar daran anknüpfenden Klassengespräch oder Interview aufgegriffen und evaluiert wurde. Es wurde weder explizit geübt, wie man

seine Sicht der Dinge nachvollziehbar schildern kann noch wie man sicherstellen kann, dass der Partner verstanden hat, was man eigentlich sagen wollte. Erst eine Woche später, in der Folgestunde, wurde auf vier exemplarische externe Repräsentationen der Kinder näher eingegangen, sodass eine lange Distanz zwischen den individuellen kommunikativen Settings und dem Klassengespräch lag. Zudem fand der Austausch am Ende jeder Unterrichtsstunde statt, sodass nicht sichergestellt werden kann, inwieviel diese Austauschphase von den Kindern effektiv und zielgerichtet genutzt wurde. An dieser Stelle von einem *Kommunikationstraining* zu sprechen, ist gewagt und wird deshalb vermieden.

Um Aussagen über die Qualität der Kommunikationsphasen treffen zu können und folglich beurteilen zu können, ob die Lernenden sich auf die Kommunikation einließen, ob sie auf der Ebene des Beschreibens innehielten, ob sie lediglich ihre beiden Lösungen akzeptierten und unreflektiert nebeneinander stehen ließen, auch wenn sie unterschiedlich waren, oder ob sie den Austausch ausgiebig nutzten, wäre eine Videografie oder ein anschließendes Interview mit den Interakteuren erforderlich gewesen. Auf dieser Basis hätte nachvollzogen werden können, ob es den Drittklässlern in der Reflexion gelungen ist, sowohl das Hauptaugenmerk auf die Struktureigenschaften anstelle der Oberflächenmerkmale zu richten als auch die externen Repräsentationen richtige zu deuten (vgl. Abschnitt 1.5.5.3). Da die in der Untersuchung erfassten Daten keine detaillierten Rückschlüsse auf Ablauf und Inhalte der Dialoge sowie die Effektivität der kommunikativen Zweiersettings erlaubten, kann der ausbleibende Effekt auf den Lösungserfolg bzw. der kleine Effekt auf die Problemlösekompetenzen nicht eindeutig interpretiert und einer Ursache zugeordnet werden.

Vor dem Hintergrund, dass das Videografieren der Zweiersettings geplant war, aber pro Klasse im Mittel nur fünf Einverständniserklärungen vorlagen, musste die Videografie verworfen werden und es musste letztlich in Kauf genommen werden, dass der genaue Ablauf und die Organisation der Kommunikationsphasen nicht charakterisiert werden können. Interviews wären aus zeitökonomischen Gründen und in Anbetracht der zusätzlichen Belastung für die Schulen nicht realisierbar gewesen. In Folgeuntersuchungen sollte die Videografie und Interviews berücksichtigt und stärker fokussiert werden, um mehr über die Gedanken der Kinder beim Lösen mithilfe externer Repräsentationen zu erfahren und mögliche Handicaps zu identifizieren. Es ist daher nicht auszuschließen, dass neben den mangelnden Vorerfahrungen und dem hohen Anspruch, den die Kommunikation mit sich bringt, sowohl der äußere, organisatorische Rahmen als auch die Motivation und Bereitschaft der Kinder den Effekt negativ beeinflussen.

Als weiterer Punkt ist der Einsatz des Gesprächsleitfadens anzuführen. Es bleibt ungeklärt, ob der zur Verfügung gestellte Gesprächsleitfaden von den Lernenden als Unter-

stützung wahrgenommen wurde, ob er sich als zu schwer oder ungeeignet erwies oder aber ob dieser von den Lernenden in der Phase unberücksichtigt blieb. Alle drei Interpretationen sind denkbar, deuten aber auf konträre Ursachen hin, die auf Basis der Daten unbeantwortet bleiben. In Folgeuntersuchungen muss unabdingbar dessen Einsatz evaluiert werden.

Eine weitere Schwäche der Untersuchung stellt die Ungewissheit, wie die Lehrkräfte der Gruppe T-K+ den Austausch in ihren Unterricht integrierten, dar. Die Stundenprotokolle ergaben entgegen den Erwartungen keine eindeutigen, aussagekräftigen Hinweise auf die Implementierung kommunikativer Zweiersettings, weshalb sie in diesem Zusammenhang als Blackbox zu behandeln sind. Diese Problematik lässt sich in ähnlicher Weise auf die Gruppe T-K- übertragen, weil nicht mit Sicherheit ausgeschlossen werden kann, dass die Lehrkräfte Kommunikationsphasen in ihren Unterricht integrierten. Im Vorfeld wurde überlegt, ein Item mit diesem Fokus im Lehrerfragebogen zu erheben. Dagegen sprach, dass ein solches Item die Lehrkräfte für den Einsatz kommunikativer Settings sensibilisieren könnte und dadurch die Manipulation gefährdet wäre. Da sich aus den Stundenprotokollen konkrete Hinweise in diese Richtung versprochen wurden, wurde davon abgesehen.

Vereinzelt wiesen die Stundenprotokolle der Lehrkräfte der Kontrollgruppe T-K- auch Kommunikationsphasen unter Gleichaltrigen auf, weshalb der geringe Effekt der Kommunikation auch darauf zurückgeführt werden kann, dass in den Nicht-Kommunikationsgruppen Austauschphasen unter Gleichaltrigen im Unterricht integriert waren. Erschwerend kommt hinzu, dass hierzu fehlende oder unkonkrete Angaben in den Protokollen kein Garant dafür sind, dass Partnerarbeit in deren Unterricht auch tatsächlich unberücksichtigt blieb. Die systematischen Störeinflüsse aufseiten der Bedingungsmanipulation *Kommunikation* können daher auch als Ursache für die nahezu parallele Entwicklung der beiden Gruppen K+ und K- angesehen werden. Es ist demzufolge nicht auszuschließen, dass gegensinnige Konfundierungen die Hypothese nur scheinbar widerlegen bzw. geringe Effekte erzeugen. Ein zusätzlicher Fragebogen nach der Erhebung hätte zwar Aufklärung geboten, die Belastung der teilnehmenden Lehrkräfte jedoch erneut erhöht.

Was allerdings aufgrund der Eigenproduktionen der Kinder sowie der Stundenprotokolle der Lehrkräfte ausgeschlossen werden konnte, ist, dass die Kommunikation in den Klassen der Gruppe T-K+ auf unterschiedlichen Repräsentationen basierte. Diese Schlussfolgerung resultiert auf der Beobachtung, dass alle Knobelhefte einer Klasse nahezu ausnahmslos die gleiche externe Repräsentation enthielten. Das Anfertigen zusätzlicher Notizen außerhalb des Knobelheftes war nicht gestattet, sodass diese Einflussgröße mit hoher Wahrscheinlichkeit ausgeschlossen werden kann.

Dass der Faktor *Kommunikation* die Anzahl selbstgenerierter externer Repräsentationen unbeeinflusst ließ, ist hypothesenkonform. Lediglich Austauschphasen unter Gleichaltrigen zu implementieren, bewirkt kein automatisch häufigeres Zurückgreifen auf externe Repräsentationen. Dieses Ergebnis stützt die Annahme, dass eine Steigerung der Anzahl selbstgenerierter externer Repräsentationen durch das Training, und nicht durch die Kommunikation, bewirkt wurde.

Aufgrund vorliegender Ergebnisse kann zusammenfassend, wenn auch nur eingeschränkt, bestätigt werden, dass kommunikative Settings unter Gleichaltrigen als Lernunterstützung gelten und einen Kompetenzzuwachs bewirken können. Die Schlussfolgerung, dass der kommunikative Austausch unter Gleichaltrigen per se mit einem Zuwachs der Problemlösekompetenzen einhergeht, muss aufgrund der geringen Effektstärke vorsichtig interpretiert werden.

6.5 Forschungsfrage 4: Effekt der Gruppe

Ob die beiden Faktoren *Training* und *Kommunikation* einen gemeinsamen Einfluss auf den Lösungserfolg, die Problemlösekompetenzen bzw. die Anzahl selbstgenerierter externer Repräsentationen ausüben, wurde im Rahmen der 4. Forschungsfrage untersucht. Dabei stand im Vordergrund, zu prüfen, ob sich eine bestimmte Kombination von Faktoren auf die abhängigen Variablen auswirkt.

Bisherige Ergebnisse zeigen, dass das *Training* sowie die *Kommunikation*, einen eigenständigen Einfluss auf die abhängigen Variablen hatten, wenngleich der positive Einfluss der *Kommunikation* nur bei den Problemlösekompetenzen beobachtbar war (siehe Abschnitt 6.3–6.4). Entgegen den Hypothesen war den Ergebnissen zufolge die Gruppe, die sich zusätzlich zum Training austauschte (T+K+), der reinen Trainingsgruppe (T+K-) in ihren Leistungen nicht überlegen. Der Einfluss des Trainings ist letztlich in Klassen mit und ohne kommunikative Zweiersettings aufgrund fehlender Signifikanz als gleich groß anzusehen. Selbiges wurde für die Nicht-Trainingsgruppe festgestellt: Klassen mit und ohne kommunikative Zweiersettings unterschieden sich in diesem Zusammenhang nicht signifikant voneinander. Hiernach üben die Faktoren *Kommunikation* und *Training* keinen gemeinsamen Einfluss auf die abhängigen Variablen aus, was sich aufgabenspezifisch replizieren ließ. Rein deskriptiv zeichnet sich ansatzweise eine Überlegenheit der Klassen mit Kommunikation ab, weshalb es sich in Folgeanalysen zu prüfen lohnt, ob ein längerer Interventionszeitraum zu hypothesenkonformen Ergebnissen führt (vgl. Abschnitt 6.4). Ferner wäre zu untersuchen, ob die Kommunikation einen bedeutsameren Einfluss nimmt, wenn sie explizit trainiert wird.

Ursachen, warum die Gruppen mit kommunikativen Settings den Gruppen ohne kommunikativen Settings nicht überlegen waren, lassen sich mehrere vermuten. Beim Ver-

gleich der Trainingsklassen (T+K+, T+K-) ist zu berücksichtigen, dass Kinder der Gruppe T+K+ mit *zwei* für sie eher neuartigen und ungewohnten Aufgaben konfrontiert waren, erstens dem Konstruieren adäquater externer Repräsentationen und zweitens dem Austausch darüber, wohingegen Kinder der Gruppe T+K- lediglich der ersten Herausforderung ausgesetzt waren. Es ist denkbar, dass Letztere davon profitierten, ihre Konzentration auf die Konstruktion externer Repräsentationen zu bündeln, und sich daher keine Unterschiede zwischen beiden Gruppen ergaben. Ferner ist in der Gruppe T+K+ nicht auszuschließen, dass die Kopplung für einen Teil der Lernenden einem Gewinn, für den anderen Teil einer Überforderung glich, sodass sich der negative Effekt und der positive Effekt gegenseitig aufhoben. Würde man dieser Interpretation folgen, so wäre einerseits aufgrund des deskriptiv identifizierten Vorteils für die Gruppe T+K+ davon auszugehen, dass mehr Kinder von der Interaktion *Training* und *Kommunikation* profitierten sowie das Voneinander- und Miteinander-Lernen wie auch Helfen und Kollaborieren bei ihnen gut gelungen war (vgl. Abschnitt 2.4). Andererseits ist nicht zu vernachlässigen, dass die Initiierung der Kommunikation von trainierten Kindern, denen die Konstruktion einer adäquaten Repräsentation verwehrt blieb bzw. die keine Lösung gefunden hatten, als Bloßstellung empfunden werden konnte und demnach aus motivationaler Sicht hemmend war. Das Gefühl, versagt zu haben und dies vor dem Gesprächspartner offenlegen zu müssen, ist keinesfalls förderlich für das Lernverhalten und als kontraproduktiv zu deuten. Losgelöst davon besteht auch hier die Möglichkeit, die fehlende Signifikanz auf den zur Verfügung gestellten Gesprächsleitfaden zurückzuführen, vor allem dann, wenn er von den Lernenden nicht als Unterstützung, sondern als hemmend empfunden wurde (siehe Abschnitt 6.4). Folgeanalysen müssten den Fokus Gesprächsleitfaden aufgreifen und analysieren.

Ergänzend zur Forschungsfrage 3 sei an dieser Stelle angemerkt, dass insbesondere der intersubjektive Austausch im Unterricht der Gruppe T+K+ hätte stärker forciert werden müssen, beispielsweise in Form eines eigenständigen Kommunikationstrainings oder einer gezielten Reflexion der kommunikativen Settings. Dies geht konform mit der Tatsache, dass die Kommunikation im Unterrichtskonzept mehr an Bedeutung hätte erfahren müssen, während sie in der Untersuchung eher eine Randerscheinung am Ende der Unterrichtsstunde darstellte. Die Unterrichtskonzeption hatte an dieser Stelle ein Stück weit dazu beigetragen, dass die Kommunikation nicht den erforderlichen Stellenwert erfahren hat. Im Gegensatz zu den externen Repräsentationen wurde die Kommunikation nicht explizit in einem dafür spezifizierten Training fokussiert oder reflektiert. Es wurde nicht ausdrücklich thematisiert, wie man über seine individuellen Lösungsprozesse berichtet, seine Gedanken strukturiert darlegt, sodass das Gegenüber sie nachvollziehen kann. Zudem wurde versäumt an vielfältigen Situationen das Argumentieren und Begründen zu üben und bewusst als solches hervorzuheben sowie

über dabei auftretende Schwierigkeiten zu diskutieren. Auch wenn dies implizit im Klassengespräch innerhalb der Reflexionsphase im Rahmen des Repräsentationstrainings verfolgt wurde, waren die Lernenden in der Austauschphase am Ende des Unterrichts stets auf sich allein gestellt. Es bestand die Gefahr, dass Lernende, die keinen Nutzen im Austausch sahen und sich nicht auf den Austausch einließen, weder voneinander profitierten noch an Expertise dazu gewannen.

Hätte man einzelne Kommunikationsbausteine in das Klassengespräch integriert, indem beispielsweise freiwillige Teams gebeten worden wären, ihr Gespräch zu rekonstruieren und für die anderen offenzulegen (z. B. „Zu welcher Übereinstimmung seid ihr gekommen?“), so hätten die Settings gemeinsam erörtert und mitunter zentrale Bausteine des Gesprächsleitfadens forciert werden können. Die Kinder hätten exemplarisch Anreize für die eigene Kommunikation erfahren, die sie in der nächsten Austauschphase hätten nachahmen können. Konkret hätten die Begründungen von den Gleichaltrigen auf Nachvollziehbarkeit und auf Stichhaltigkeit geprüft und hinterfragt werden können. Von besonderem Interesse hätten dabei Zweierteams sein können, deren Gespräch auf unterschiedlichen Lösungen basierte. Nach der *Phase des Hineinversetzens, Nachvollziehens und Verstehens* und der *Phase der Benennung der jeweils genutzten externen Repräsentationen* (Trainingsbausteine 1 und 2, siehe Abschnitt 4.4.4) hätten die Argumentationsprozesse im Sinne von „Wer hat Recht?“ im Rahmen des 3. Trainingsbausteins von den Kindern analysiert und diskutiert werden können. Phasen des Helfens oder auch Kollaborierens wären transparent geworden (vgl. Abschnitt 2.4.3). Zudem hätten mögliche Umstrukturierungen gemeinsam ausgehandelt und diskutiert werden können (Trainingsbaustein 4).

Das gemeinsame Erörtern von Kommunikationsprozessen wäre ebenso wie die Analyse der externen Repräsentationen lohnenswert gewesen. Dadurch wäre zum einen den Kommunikationsphasen mehr Wertschätzung zugesprochen worden und zum anderen wäre den Kindern signalisiert worden, dass der Austausch eine Bereicherung darstellen kann. Ferner hätte bereits ein stärkerer Fokus auf das Argumentieren und Begründen in der Trainingsphase sichergestellt werden müssen, damit dieser Kompetenzbereich stärker ins Bewusstsein der Grundschulkinder hätte rücken können. Es wird als enormes Potenzial angesehen, die Kommunikation zum festen Bestandteil des Unterrichts zu machen. Auch die deskriptiven Ergebnisse der Nicht-Trainingsklassen deuten auf eine lernförderliche Wirkung der kommunikativen Settings hin, wenngleich es zu betonen gilt, dass keine signifikanten Unterschiede gefunden wurden. Warum sich die Gruppe T-K+ nicht signifikant von der Gruppe T-K- abhebt, kann vor dem Hintergrund der in Abschnitt 6.4 geschilderten Situation nicht eindeutig benannt werden.

6.6 Forschungsfrage 5: Effekt selbstgenerierter externer Repräsentationen

Im Fokus der fünften Forschungsfrage wird der Einfluss der selbstgenerierten externen Repräsentationen auf den Lösungserfolg und die Problemlösekompetenzen analysiert. Hierzu war es erforderlich, den Anteil erklärter Varianz, der auf potenzielle Störvariablen zurückzuführen ist, herauszupartialisieren.

Mit Blick auf die selbstgenerierten externen Repräsentationen bleibt festzuhalten, dass sie 5 % des Lösungserfolgs und 13 % der Problemlösekompetenzen vorhersagen. In beiden Fällen sind die potenziellen Störvariablen als kontrolliert anzusehen. Blickt man lediglich auf die beiden Zahlenwerte, so ist auf den ersten Blick der Effekt, der auf die externen Repräsentationen zurückgeht als niedrig einzustufen, insbesondere bei der abhängigen Variablen Lösungserfolg. Es kommt die Frage danach auf, ob das Training die Konstruktion und Nutzung externer Repräsentationen fördert, wenn nur ein minimaler Teil des Leistungszuwachses tatsächlich auf die selbstgenerierten externen Repräsentationen zurückgeführt werden kann.

Diesem Argument kann entgegengesetzt werden, dass unter Einbezug der bisherigen empirischen Befunde das Ergebnis zu relativieren ist. Zum einen ist nicht zu vernachlässigen, dass Drittklässler aufgrund ihrer mangelnden Vorerfahrungen sowohl beim Lösen problemhaltiger Textaufgaben als auch beim Konstruieren adäquater Denk- und Erkenntniswerkzeuge als Novizen gelten, die noch nicht auf ein facettenreiches Repräsentationsrepertoire zurückgreifen können. Damit Externalisiertes individuelle Denk- und Lösungsprozesse stützen und Raum für individuelle Entdeckungen schaffen kann (siehe Abschnitt 1.5.5.1), benötigen die Grundschulkinder vielfältige Lösungserfahrungen, auf die sie zurückgreifen können. Zum anderen muss berücksichtigt werden, dass nicht jede ihrer Konstruktionsideen und -bemühungen unmittelbar in adäquaten Repräsentationen mündet, die das Ablesen der Lösung gewährleisten (vgl. Abschnitt 1.3.3). Die Lernenden sind vielfältigen, für sie neuen, komplexen Anforderungen und Lernprozessen ausgesetzt, was mit einer hohen kognitiven Belastung einhergeht. Sie müssen die Problemstellung und die dargebotenen Informationen zuerst verstehen, um sich mental ein Bild von der Situation machen zu können, welches es im nächsten Schritt zu externalisieren gilt. Sie stehen vor der Aufgabe, zwischen lösungsrelevanten und irrelevanten Merkmalen zu differenzieren und entsprechend zu selektieren, sowie auf dieser Basis das erforderliche Vorwissen zu aktivieren (siehe Abschnitt 1.5.5.1). Geht ihr Problemlöseprozess mit der Tatsache einher, dass sie eine inadäquate Repräsentation generieren, so müssen sie dies als erstes erkennen, um überhaupt Änderungen in Betracht ziehen zu können. Gelingt dies nicht, so bleiben notwendige Prozesse des neu Durchdenkens, systematischen Veränderns, Umstrukturierens und Wiederrufens bereits vollzogener Schritte sowie letztlich der Erfolg aus. Wenn aber der Produc-

zent seinen Lösungsprozess und sein Ergebnis reflektiert, dann kann die externe Repräsentation zur Gedächtnis- und Lösungsunterstützung werden. Vorstehendes spiegelt sich unter anderem im hohen Einfluss der metakognitiven Fähigkeiten wider. Kinder, die metakognitiv aktiv sind, zählen zu den erfolgreicheren Problemlösern. Die Untersuchung bestätigt diesen Einfluss.

Hinzu kommt, dass sechs verschiedene Aufgabengruppen trainiert wurden, die alle unterschiedliche Heran- und Vorgehensweise ermöglichen und favorisieren. Werden dann keine Parallelen zu vorangegangenen Problembewältigungen erkannt und Werkzeuge eher intuitiv konstruiert, so fällt es den Lernenden schwer, die Lösung zu erarbeiten. Die Kinder müssen sich immer wieder aufs Neue in die vorliegende Struktur eindenken und versuchen, diese zu verstehen und lösungsunterstützend abzubilden. Fehlt ihnen der Strukturblick, so fällt es ihnen schwer, die zugrunde liegenden Operationen aufzuspüren und zu entschlüsseln (siehe Abschnitt 1.5.5.3). Es bedarf einer langfristigen Förderung dieser vielfältigen, komplexen Prozesse. Die Ergebnisse spiegeln erwartungskonform wider, dass in einem Trainingszeitraum von 12 Wochen lediglich Weichen gestellt werden können, indem der Strukturblick geschult wird, die Generierung externer Repräsentationen angeregt und angebahnt, Handlungsmaßnahmen an die Hand gegeben, die Umstrukturierung inadäquater Repräsentationen in vielfältigen Situationen thematisiert, die Vorzüge einzelner Repräsentationen aufgezeigt und diskutiert sowie alternative Heran- und Vorgehensweisen zum Gegenstand der Reflexion gemacht werden.

Das Repräsentationsrepertoire, das sich die Kinder im Interventionszeitraum aufbauen, muss stetig weiterentwickelt werden. Blickt man aus der Perspektive, dass hierfür eine langfristige Förderung vonnöten ist, auf die Zahlenwerte, dann ist der Erklärungsanteil, der auf die externen Repräsentationen zurückzuführen ist, als groß einzustufen. Es ist bereits nach nur 12 Trainingsstunden gelungen, 5 % des Lösungserfolgs bzw. 13 % der Problemlösekompetenzen mithilfe externer Repräsentationen vorherzusagen. Vor diesem Hintergrund und der Tatsache, dass sich Teilerfolge und Teileinsichten stärker im Prozess als im Ergebnis widerspiegeln, ist nachvollziehbar, dass der Anteil erklärter Varianz bei der abhängigen Variable Problemlösekompetenz hypothesenkonform höher ist als auf den Lösungserfolg.

Greift man eine weitere Perspektive auf, so kommt beachtenswerterweise hinzu, dass fitte und leistungsstarke Drittklässler nach dem Training durchaus in der Lage waren, die Lösung im Kopf zu produzieren, nichts mehr externalisieren mussten und letztlich auch nur die Lösungszahl notierten. Es ist in diesen Fällen davon auszugehen, dass die Kinder eine mentale Repräsentation konstruierten und gegebenenfalls notwendige Umstrukturierungen mental vollzogen. Diese Leistungen wurden zwar mit einem maximalen Lösungserfolg-Testscore nicht jedoch mit einem maximalen Problemlöse-

kompetenz-Testscore erfasst (vgl. Abschnitt 4.5.1.1). Der deskriptiven Statistik zufolge traf dies für 29 Drittklässler (knapp 9 %) zu: Sie lösten alle mindestens eine Aufgabe richtig, ohne etwas zu externalisieren, das Aufschluss über den Lösungsprozess geben könnte. Würde man den Einfluss der mentalen Repräsentationen mitberücksichtigen, wäre bei den Problemlösekompetenzen ein höherer Anteil erklärter Varianz, der auf die selbstgenerierten Repräsentationen zurückgeführt werden kann, zu erwarten.

Wenn man sich die Frage danach stellt, ob sich gute Prädiktoren zur Vorhersage des Lösungserfolgs bzw. der Problemlösekompetenzen identifizieren lassen, dann ergibt sich für den gesamten Textaufgaben-Performance-Test eine leichte Dominanz zugunsten der Zeichnungen, gefolgt von den Tabellen und den Rechnungen. Schriftsprachliche Beschreibungen nehmen eine eher untergeordnete Rolle ein, was mit ihrer Schwierigkeit und ihrem hohen Anspruch in Zusammenhang steht. Hypothesenkonform konnte gezeigt werden, dass es nicht *die* Repräsentationsform gibt, die Einsicht und Erfolg garantiert, sondern dass der Aufgabentyp die Wahl beeinflusst (vgl. Abschnitt 3.3.2). Während bei der *Aufgabe mit komplexen Informationen* eindeutig die Präferenz für die depiktionalen Repräsentationen vorherrschte, dominierte bei der *Vergleichsaufgabe* der tabellarische und rechnerische Zugang, bei der *Aufgabe zur Verhältnisverteilung* der rechnerische. Wobei der Zusammenhang bei dem untrainierten Aufgabentyp am geringsten ausfiel. Die Ergebnisse sind ein Indiz dafür, dass es mit dem Training gelungen ist, die Kinder bei der Wahl einer zur Aufgabe passenden Repräsentation zu unterstützen. Während bei allen Aufgaben vielfältige Repräsentationen konstruiert werden können, entspricht die Präferenz der Kinder bei jeder der drei Aufgaben einem für die Aufgabentypen geeigneten Werkzeug. Dass der Effekt bei beiden trainierten Aufgabentypen größer ausfiel als bei den untrainierten, deutet auf eine höhere Wirksamkeit des Trainings in trainierten Aufgabentypen hin. Dieser Effekt kann jedoch, wie zuvor in Abschnitt 6.2 geschildert, von der ausgewählten *Aufgabe zur Verhältnisverteilung* beeinflusst sein. Es ist nicht auszuschließen, dass bei einem anderen untrainierten Aufgabentyp, wie beispielsweise den *Bewegungsaufgaben*, höhere Transfereffekte zu erwarten wären, weil die Aufgabe eine größere *Repräsentationsvielfalt* zulässt (vgl. Abschnitt 4.1.2). In Folgeanalysen müsste in Form von Aufgabenanalysen geklärt werden, ob sich alle Aufgabentypen in gleicher Weise eignen, externe Repräsentationen zu konstruieren. Dieses Kriterium müsste in Situationen zwingend erfüllt sein, wenn der Einfluss externer Repräsentationen untersucht wird.

Ferner ergaben die Analysen, dass den erhobenen Störvariablen hohe Relevanz zuzusprechen ist. Die Variablen *experimentelle Bedingung, Intelligenz, mathematische, sprachliche und metakognitive Fähigkeiten, Prätest-Leistungen* wie auch *Mehrsprachigkeit* und *Geschlecht* sagten 61 % des Lösungserfolgs bzw. 61 % der Problemlösekompetenzen vorher. Die metakognitiven Fähigkeiten gelten sowohl beim Lösungser-

folg als auch bei den Problemlösekompetenzen als stärkster Prädiktor, was ihre hohe Bedeutsamkeit für den Lernerfolg untermauert (vgl. Abschnitt 4.5.3.4). Demzufolge können die bisherigen Befunde zahlreicher Studien, dass Kinder, die ihr Ergebnis als wahr annehmen, ohne es zu hinterfragen, zu überprüfen oder zu reflektieren, tendenziell zu den leistungsschwächeren Problemlösern zählen, repliziert werden. Ferner konnte die Bedeutung der *Attitudes* und *Beliefs* und der mathematischen Fähigkeiten für beide abhängigen Variablen bestätigt werden. Ob Kinder Negatives und Misserfolge oder aber Herausforderung und Erfolge mit Textaufgaben assoziieren, scheint den Lösungserfolg sowie die Problemlösekompetenzen signifikant zu beeinflussen (vgl. Abschnitt 4.5.3.5). Gleiches kann auf die mathematischen Basiskompetenzen übertragen werden (vgl. Abschnitt 4.5.3.2): Leistungsstarke Rechner greifen auf sicherere Rechenkompetenzen zurück und erzielen beim Lösen problemhaltiger Textaufgaben höhere Lernerfolge als leistungsschwache Rechner. Einen signifikanten Vorteil für intelligente Kinder konnte für die Problemlösekompetenzen, ein hinreichend signifikanter für den Lösungserfolg identifiziert werden (vgl. Abschnitt 4.5.3.1). Der empirische Befund, dass dem Vorwissen eine höhere Relevanz als der Intelligenz zugesprochen wird, hat sich auch in der vorliegenden Untersuchung als zutreffend herausgestellt. Der untergeordnete Einfluss der Intelligenz geht somit konform mit den Ergebnissen der SCHOLASTIK-Studie und der LOGIK-Studie. Zusätzlich trug der Problemlöseerfolg zum Zeitpunkt des Prättests signifikant zum Anteil erklärter Varianz bei. Übereinstimmend mit den Ergebnissen von Hyde et al. (1990, S. 139) blieb das Geschlecht ohne signifikanten Einfluss, womit der von Stern identifizierte Vorteil für die Jungen beim Lösen problemhaltiger Textaufgaben nicht repliziert werden konnte. Entgegen den Erwartungen war der Einfluss der migrationsbedingten Mehrsprachigkeit auf beide abhängige Variablen nicht signifikant. Auch der mehrfach bestätigte Zusammenhang von sprachlichen Fähigkeiten und des Leseverständnisses und Lösungserfolg konnte in der Untersuchung nicht nachgewiesen werden (vgl. Abschnitt 4.5.3.3). Die Ergebnisse zeigen, dass sich die Erhebung der hohen Anzahl an Kontrollvariablen als essenziell erwiesen hat, auch wenn sie aus Sicht der Durchführungsökonomie eine hohe Belastung für die Schulen, Lehrkräfte und Kinder darstellte. Bis auf wenige Ausnahmen konnte ihnen ein signifikanter Einfluss zugesprochen werden.

Letztlich können keine konkreten Aussagen darüber getroffen werden, welche Rolle dem Arbeitsgedächtnis bei der Entwicklung externer Repräsentationen zugesprochen werden kann (vgl. Imbo & Vandierendonck, 2007). Mit der WISC-IV, der *Wechsler Intelligence Scale for Children*, hätte das Arbeitsgedächtnis mit 3 Untertests im Rahmen der Intelligenztestung miterhoben werden können (Petermann & Petermann, 2011, S. 13, vgl. Abschnitt 4.5.3.1). Interessant wäre es, in Folgeanalysen herauszufin-

den, ob es einen Zusammenhang zwischen dem Arbeitsgedächtnis und dem erfolgreichen Lösen problemhaltiger Textaufgaben gibt.

6.7 Methodische Überlegungen

Hinsichtlich der *Durchführung der Interventionsmaßnahme* bleibt festzuhalten, dass sich sowohl die dritte Jahrgangsstufe als auch die Trainingsdauer von 12 Stunden bewährt haben, um problemhafte Textaufgaben im Mathematikunterricht zu verankern. Die Drittklässler hatten an unterschiedlichen Aufgabentypen die Möglichkeit, Erfahrungen beim Generieren externer Repräsentationen zu sammeln und ihre Fähigkeiten dahin gehend auszubauen. Um den Unterrichtsanlässen des Kommunizierens und Argumentierens mehr Gewicht zuzusprechen, wäre es hilfreich gewesen, deren Förderung bewusster in den Unterricht zu integrieren. Beispielsweise hätten Gespräche unter Gleichaltrigen oder aufgetretene Schwierigkeiten im Plenum aufgegriffen, besprochen, diskutiert und gemeinsam reflektiert werden können, vergleichbar mit den externen Repräsentationen im Training. Ferner hätte die Implementation der Videografie die Güte der stattgefundenen Kommunikation, die Stärken und möglichen Schwächen des Gesprächsleitfadens aber auch die geringen bzw. ausbleibenden Effekte der Kommunikation aufzudecken und zu erklären befähigt. Für Folgeanalysen ist interessant, welchen Effekt kommunikative Settings auf das Argumentieren ausüben, wenn die Kinder zuvor ein Repräsentationstraining absolviert haben und somit über dieses Wissen und diese Kompetenzen bereits verfügen. Dieses Design würde ausschließen, dass der gefundene Effekt von den anfänglichen Repräsentationskompetenzen der Kinder beeinflusst ist. Darüber hinaus könnte untersucht werden, ob die Jahrgangsstufe oder der trainierte Inhaltsbereich, die Textaufgaben, die Effekte beeinflusst.

Im Hinblick auf die Messzeitpunkte zeigte sich rückblickend, dass der Messzeitpunkt zum Ende des zweiten Schuljahres enorme Konsequenzen für die Untersuchung nach sich zog. Die Intention war, die Leistungsvoraussetzungen der Kinder zu erheben, um mithilfe des Parallelisierungsverfahrens die Gruppenbildung vorzunehmen. Dies hätte wie ursprünglich geplant den Vorteil gehabt, dass 4 Störvariablen, die Intelligenz, die mathematischen und sprachlichen Fähigkeiten und die Problemlösekompetenzen kontrolliert gewesen wären und ein Trainingsstart zu Beginn des dritten Schuljahres hätte erfolgen können. Zudem hätten Erinnerungseffekte zum Posttest nahezu komplett ausgeschlossen werden können, da zwischen dem Prä- und Posttest 18 Wochen (6 Ferienwochen und 12 Interventionswochen) gelegen hätten. Die vorgezogene Erhebung der Störvariablen blieb jedoch für die Stichprobenzusammensetzung nicht ohne Folgen.

Insgesamt gingen in die Ermittlung der Klassenmittelwerte die Individualdaten von 347, 348 bzw. 351 Kindern ein. Die leicht variierende Anzahl der Probanden ist auf die Erhebungen an drei unterschiedlichen Tagen zurückzuführen. Nacherhebungen waren aufgrund von Krankheitsfällen und aufgrund des Ferienbeginns nur bedingt möglich. Schließt man von der Gesamtstichprobe ($N = 366$) die Probanden aus, die am Trainingsexperiment nicht vollständig teilnahmen, d. h., wenn ihnen mindestens ein Messzeitpunkt fehlte, dann lagen der Analyse nur noch 351 Kinder zugrunde. Somit reduzierte sich die Stichprobe um 4 %. Auf den ersten Blick scheint dieser Ausschluss der Probanden vernachlässigbar klein zu sein. Die deskriptive Analyse zeigte jedoch, dass 12 Probanden zwar an der Vorerhebung im 2. Schuljahr (MZH 0), aber nicht vollständig am Trainingsexperiment teilnahmen. Demzufolge gingen deren Individualdaten in die Klassenmittelwerte ein und beeinflussten die parallelisierte Zuweisung der Klassen zu den experimentellen Bedingungen, obwohl sie aufgrund unvollständiger Daten aus der späteren Vergleichsanalyse ausgeschlossen wurden. Ursachen hierfür waren zum einen Zurückstufungen in die zweite Klasse oder Schulwechsel einzelner Kinder, zum anderen aber auch krankheitsbedingte Ausfälle oder Kuraufenthalte an mindestens einem Erhebungszeitpunkt. Acht Probanden kamen im dritten Schuljahr neu hinzu, weshalb deren Intelligenz, Leseverständnis und mathematischen Basiskompetenzen am Ende des 2. Schuljahres nicht erhoben werden konnte. Folglich wurden ihre Leistungsvoraussetzungen bei der Parallelisierung nicht berücksichtigt, obwohl sie am Experiment vollständig teilnahmen und alle drei Testungen absolvierten.

Im Nachhinein hätte die Problematik der fehlenden bzw. überschüssigen Werte bei der Klassenmittelwertbestimmung verhindert werden können, wenn die erste Messung (MZH 0) nicht am Ende des zweiten, sondern zu Beginn des dritten Schuljahres stattgefunden hätte. Letztlich verfehlte die parallelisierte Verteilung der Klassen auf die experimentellen Bedingungen das angestrebte Ziel, gleiche Ausgangsvoraussetzungen zu schaffen. Mit dem Wissen, dass zwölf Kinder an der Vorerhebung (MZH 0), jedoch nicht an der Intervention (MZH 1–3), teilnahmen, hätte sich durch deren Ausschluss eine abweichende Parallelisierung ergeben (siehe Anhang H). Ein Ausschluss der Kinder, die nur am Messzeitpunkt 0 und nicht an der Intervention teilnahmen, zog in allen Units und folglich in allen experimentellen Gruppierungen Veränderungen nach sich. Keine der vier Gruppierungen setzte sich aus den fünf Klassen zusammen, wie es die der Untersuchung zugrunde liegende Parallelisierung ergab (vgl. Abbildung 4.20, Abschnitt 4.7). Es war daher anzunehmen, dass der Versuch, die Kovariaten für die Gruppenbildung auszubalancieren, gescheitert war.

Einen weiteren Diskussionspunkt stellt die tatsächliche Berücksichtigung der drei Kovariaten dar. Die Planung sah vor, die Prätest-Leistung der 20 Klassen als zusätzliche Matchingvariable in das Parallelisierungsverfahren aufzunehmen. Aufgrund der

Problematik, dass einzelne Klassen die geplante Erhebung vor den Sommerferien streichen mussten, wurde im Sinne der Vergleichbarkeit eine Erhebung aller Klassen im neuen Schuljahr realisiert. Die zeitlich verzögerte Erhebung zog nach sich, dass die Prätest-Leistung unter Einhaltung des Zeitplans nicht mehr als Matchingvariable berücksichtigt werden konnte. Es liegt die Vermutung nahe, dass deren Aufnahme in die Analyse die Vergleichbarkeit hätte sicherstellen können.

Als ungünstig erwies sich auch, dass die randomisierte Zuweisung der Klassen zu den experimentellen Bedingungen übernommen wurde, ohne diese kritisch zu hinterfragen. Dass gerade die leistungsschwächste Gruppe die Kontrollgruppe repräsentiert, hätte im Vorfeld verhindert werden können, gar verhindert werden müssen.

Einfluss der Testinstrumente der Kontrollvariablen auf die Genauigkeit der Schätzung

Im Allgemeinen ist nicht auszuschließen, dass die Genauigkeit der Schätzung von den Messinstrumenten der Störvariablen beeinflusst war. „Bei sehr reliablen Messinstrumenten ist die Verschätzung der Regressionskoeffizienten relativ gering“ (Eid et al., 2011, S. 675). In Anbetracht der vorhandenen internen Konsistenzen können verzerrte Schätzungen jedoch nicht ausgeschlossen werden. Ein weiterer Grund für Verzerrungen wird darin gesehen, dass relevante, den Lösungserfolg beeinflussende Variablen unberücksichtigt blieben. Es wurden zwar im Vorfeld sehr viele Störvariablen bedacht und berücksichtigt, es ist aber durchaus nicht auszuschließen, dass weitere relevante Prädiktoren wie beispielsweise das Arbeitsgedächtnis unbeachtet blieben. Beispielsweise ist naheliegend, dass die durchführende Lehrkraft den Lösungserfolg beeinflusst. Da es sich hierbei um eine kategoriale Variable handelt, hätten die 11 verschiedenen Lehrkräfte (10 reguläre Mathematiklehrkräfte und die Versuchsleiterin) zehn Dummykodierungen zur Folge gehabt. Die Anzahl der Prädiktoren wäre dadurch enorm angewachsen. Ferner könnte die Motivation der Lernenden oder auch der Lehrkraft den Lösungserfolg beeinflussen. Letztgenanntes tritt ein, wenn die Lehrkraft den Textaufgaben bzw. dem Training kritisch gegenübersteht und dies den Kindern gegenüber äußert. Auch wenn dies weder in den Nicht-Trainingsklassen noch in den Trainingsklassen direkt beobachtbar war, kann es nicht gänzlich ausgeschlossen werden.

Folgeanalysen sind zwingend erforderlich, um herauszufinden, ob in zusätzlich gezogenen Stichproben die Güte der Prognose repliziert werden kann (Eid et al., 2011, S. 630).

6.8 Implikationen für die Praxis

Wenn man den Bildungsplan für die Primarstufe heranzieht, dann zählen Textaufgaben zum festen Bestandteil des Mathematikunterrichts, wenn es um das Rechnen „in Kontexten“ geht (KMK 2005b, S. 9).

Die Untersuchung machte sich zum Grundsatz, das Lösen von Textaufgaben weniger als Tätigkeit, sondern mehr als Fähigkeit zu fördern (Heinze, 2007, S. 4). Es wurde sich durch das Repräsentationstraining versprochen, die Lernenden zu intentionalem Handeln zu befähigen, konkret, dass sie durch die Interventionsmaßnahme erlernen, ihren Lösungsprozess selbstbestimmt zu wählen und zu steuern, aber auch ihr Handeln zielorientiert zu hinterfragen (A. Meyer & Fischer, 2013, S. 184). Beides hat sich in der Untersuchung als essenziell erwiesen. Dass sie hierzu nicht schon vor der Interventionsmaßnahme in der Lage waren, zeigten die niedrigen Lösungsraten in der Prätestmessung, die konform gehen mit den bisherigen empirischen Erkenntnissen (vgl. Abschnitt 1.4.2, 5.3.1). Die Ergebnisse der vorliegenden Interventionsstudie sind ein Indiz dafür, dass Kinder, die an der Interventionsmaßnahme partizipierten, ihre Handlungskompetenzen ausbauen konnten.

Es wird beim Problemlösen als wichtig erachtet, dass sich die Kinder aktiv und eigenständig mit der Problemstellung auseinandersetzen und dabei eigene Lösungsideen entwickeln. Es hat sich gezeigt, dass es den Kindern gelungen ist, mathematische Phänomene aufzuspüren. Sich ein Problem explorativ zu erschließen ermöglicht ihnen, eigene Wege zu gehen, sich intensiv mit der eigenen Lösungsidee auseinanderzusetzen, aber auch, die Zone der nächsten Entwicklung zu erreichen (vgl. Abschnitt 2.2.3). Die Kinder lassen sich durch das Training anregen, ihre mentale Repräsentation zu externalisieren, was sich darin zeigte, dass sie nach dem Training häufiger auf externe Repräsentationen zurückgriffen als davor. Der vorhandene Zusammenhang lässt sich dadurch stützen, dass Lernende sich nur auf die Konstruktion externer Repräsentationen einlassen, wenn sie sie explizit als Lösungsunterstützung und nicht als Mehraufwand wahrnehmen (vgl. Abschnitt 1.5.6.1). Wenn Letzteres zutreffend wäre, d. h. dass sie externe Repräsentationen als Aufbürdung empfunden hätten, dann hätte die Anzahl selbstgenerierter externer Repräsentationen abgenommen, was allerdings nicht der Fall war.

In Übereinstimmung mit der *Representational Change Theory* konnte die vorliegende Studie zeigen, dass die Konstruktion einer Repräsentation den Erfolg nicht garantiert. Wenngleich die Anzahl selbstgenerierter externer Repräsentationen über die Zeit zunahm, führte dies nicht unweigerlich zur richtigen Lösung. Schwierigkeiten im Verstehensprozess können die Ursache dafür sein (vgl. Abschnitt 1.5.3). Auf Basis des *integrativen Modells des Text- und Bildverständnisses* nach Schnottz (2014, S. 83) ist

davon auszugehen, dass das Training die Lernenden befähigt, sowohl eine adäquate mentale Repräsentation der Textoberflächenstruktur als auch eine adäquate propositionale Repräsentation, welche dem Text zugrunde liegende Relationen und Strukturen herausstellt, zu konstruieren (vgl. Abschnitt 1.5.2). Das hieraus resultierende mentale Modell konnte externalisiert und durch repräsentationale Veränderungen der Lösung angenähert werden. Diese Folgerung basiert auf der Annahme, dass eine erfolgreiche Problembewältigung nur gelingt, wenn der Text verstanden und adäquat repräsentiert wird, was wiederum die Ergebnisse der vorliegenden Untersuchung widerspiegeln: Drittklässler konnten im Mittel ihren Lösungserfolg bzw. ihre Problemlösekompetenzen mithilfe externer Repräsentationen steigern.

Um zu vollständigem Verständnis zu gelangen und letztlich erfolgreich zu sein, muss die Bedingung der adäquaten Strukturabbildung erfüllt sein, sprich, die Problemerkmale müssen zutreffend abgebildet werden. Dass der Effekt auf die Problemlösekompetenzen größer war als auf den Lösungserfolg, lässt den Schluss zu, dass nicht allen Konstrukteuren eine adäquate Strukturabbildung gelungen war. Demzufolge benötigen Kinder Unterstützung, wenn es fehlerbehaftete externe Repräsentationen umzustrukturieren gilt. Hierzu müssen sie zuerst ihren Fehler bzw. ihre Fehldeutung erkennen und annehmen, um zielführende repräsentationale Veränderungen vornehmen zu können. Das Umstrukturieren inadäquater Repräsentationen gilt es im Unterricht an vielfältigen Situationen zu trainieren und zu fördern, sodass externe Repräsentationen zum Denk- und Erkenntniswerkzeug werden und die Lösung ablesbar wird. Im Unterricht kann dies durch das Schulen und Trainieren des Strukturblicks erreicht werden, was losgelöst vom vorliegenden Inhaltsbereich beispielsweise auch beim Umgang mit Mustern und Strukturen oder der Geometrie von zentraler Bedeutung ist (Lüken, 2012, S. 218). Darüber hinaus kann das Lockern von Bedingungen oder das Zerlegen von Gruppierungen als Möglichkeit repräsentationaler Veränderungen deutlich gemacht und als Lösungsstrategie angeboten werden (vgl. Abschnitt 1.3.3). Die Thematisierung beider Möglichkeiten hat sich als lösungsunterstützend erwiesen. Inadäquate Repräsentationen zum Gegenstand des Unterrichts zu machen, war für die Kinder zunächst unüblich, aber zugleich reizvoll. Mit Begeisterung suchten sie nach Wegen, die vorhandene Struktur zu verändern, um sich der Problemlösung anzunähern und den Problemraum einzugrenzen. Es hat sich bewährt, unvollendete oder fehlerhafte Repräsentationen gemeinsam zu diskutieren, dabei die Barriere aufzuspüren, mögliche Handlungen und repräsentationale Veränderungen abzuleiten und sich im Argumentieren zu erproben. Solche inadäquaten Repräsentationen sollten bei einer Trainingsimplementierung durchaus Berücksichtigung finden, weil sie sich auch eignen, um fehlerhafte repräsentationale Veränderungen aufzuspüren und rückgängig zu machen. Ein gemeinsamer Austausch über mögliche Veränderungsprozesse schuf ein stärkeres Be-

wusstsein und führt zu einem fundierten Erfahrungswissen, auf das die Lernenden immer wieder zurückgreifen konnten. Es wird erwartet, dass sich dies auch positiv auf das Selbstbewusstsein der Kinder, das Vertrauen in die eigenen Fähigkeiten und zugleich die Motivation auswirkt.

Dass Kinder, die eine vorgefertigte Repräsentation zum Lösen eines Problems erhalten, nicht zwingend erfolgreicher sind, steht dem gegenüber (Reuter, 2016, S. 118, 122). Der Unterschied, dass eine vorgefertigte Repräsentation in einer Situation zu helfen und in der anderen zu stören scheint, wird vor allem darin gesehen, dass sie im Fall der vorliegenden Untersuchung in der Reflexions- und in der Untersuchung Reutlers in der Erarbeitungsphase im Fokus stand. Für den Vorteil der Reflexionsphase spricht, dass die Kinder die Aufgabe bereits gelöst hatten und als Rezipienten vor der Aufgabe standen, die Repräsentation zu reflektieren und zu überdenken (vgl. Abschnitt 1.5.5.2–1.5.5.3). Im Gegensatz dazu standen die Lernenden der Untersuchung Reuters (2016) während des Lösungsprozesses vor der Herausforderung, die vorgefertigten Repräsentationen mit ihrer mentalen Repräsentationen in Einklang zu bringen und zum Lösen zu nutzen. Obwohl es sich bei ihnen um eine adäquate Repräsentation handelte, half sie nicht zwingend, die Barriere zu überwinden.

Für den Unterricht lässt sich aus der Interventionsstudie ableiten, dass es sich als günstig erweist, auf unterschiedliche Weise vollzogene repräsentationale Veränderungen, zum Unterrichtsgegenstand zu machen. Der Gefahr, die mit der Thematisierung unterschiedlicher Lösungswege einhergeht, weil sie als neuer Lernstoff zu verstehen sind, weil sie nicht für sich sprechen und anspruchsvolle Übersetzungsprozesse erfordern (vgl. Abschnitt 1.5.5), kann entgegengewirkt werden, indem für die Reflexion jeder Eigenproduktion ausreichend Zeit anberaumt wird, sodass sich die Kinder ein denken, Vermutungen äußern, aber auch Fragen stellen können. Hinzu kommt, dass das Herstellen von Bezügen und Vergleichen zwischen den einzelnen Lösungsbeispielen wesentlich ist, um Parallelitäten der Strukturabbildung wahrnehmen zu können. Wird dies herausgestellt, so bieten die unterschiedlichen Zugänge eine Lernumgebung ab, die auch Leistungsschwächere stützen und unterschiedliche Lerntypen ansprechen kann. Heterogene Lerngruppen können miteinander-, aber zugleich auch voneinander lernen.

Der Vorteil zugunsten der trainierten Gruppe deutet darauf hin, dass das Training Wirkung zeigt und die Kinder anregt, ihre mentalen Repräsentationen zu externalisieren, und sie dadurch für sich selbst als auch für Dritte sichtbar zu machen. Das Training sensibilisiert sie für unterschiedliche Zugänge und wirkt sich positiv auf den Lösungserfolg und die Problemlösekompetenzen aus. Auf Basis des positiv evaluierten Unterrichtskonzepts können Unterrichtsmaterialien im Sinne des Problems der Woche entwickelt werden, sowie Lehrkräften eine Sammlung externer Repräsentationen zur Verfügung gestellt werden.

Ferner bleibt festzuhalten, dass die Interventionsmaßnahme auch einen Beitrag zum Argumentieren leistete. Die Kinder waren am Ende der Interventionsmaßnahme durchaus in der Lage, experimentelle Beweise zu vollziehen. Folglich ist es ihnen gelungen mit ihren selbstgenerierten externen Repräsentationen den Problemraum einzuschränken, sowie beispielbasiert zu begründen und zu schließen (vgl. Abschnitt 2.2.3). Sie nutzten entsprechend ihrer Voraussetzungen ihre mathematischen Entdeckungen, um argumentativ tätig zu werden. Es hat sich bewährt, in der Reflexionsphase die Argumentationen der Kinder aufzugreifen und gegebenenfalls zu vervollständigen. Die Untersuchung bestätigt, dass nur selten Verallgemeinerungen ablesbar waren und die Kinder demzufolge kaum auf inhaltlich-anschaulicher Ebene beweisend agierten (vgl. Abschnitt 2.2.4). Die Reflexionsphase erwies sich jedoch als besonders geeignet, Verallgemeinerungen anzustoßen, wenn die selbstgenerierte externe Repräsentation die Gelegenheit dazu bot. Aus diesem Grund unterstützt das Training nicht nur das Generieren externer Repräsentationen, sondern zugleich auch beispielbezogenes Begründen. In einem gewissen Maße gelingt es mit Unterstützung durch die Lehrkraft, die Wirkung von Operationen verallgemeinernd herauszustellen.

Dass Vorstehendes in der Reflexionsphase des Unterrichts gelungen ist, war im intersubjektiven Austausch unter Gleichaltrigen nicht von Anfang an zu beobachten. Den Kindern fiel zunächst schwer, nur auf Basis ihrer selbstgenerierten externen Repräsentationen und ohne zusätzliche Unterstützung die eigenen mathematischen Entdeckungen und Zusammenhänge zu formulieren. Im Zuge der Zeit ist es ihnen mehr und mehr gelungen mit und über ihre Denk- und Erkenntniswerkzeuge zu kommunizieren. Dies bestätigt die Annahme, dass Schüler-Schüler-Interaktionen Zeit benötigen, um ihr Potenzial auszuschöpfen (vgl. Abschnitt 2.4). Folglich scheinen sich Kompetenzzuwächse auf Basis der Implementation von kommunikativen Settings nur langsam anzubahnen. Dies erklärt, dass eine geringfügige Wirkung auf die Problemlösekompetenzen, aber keine Wirkung auf den Lösungserfolg zu verzeichnen war. Wenn man die bisherigen Befunde aufgreift, würde man vermuten, dass eine langfristige Implementation intersubjektiven Austauschs lernförderlich ist, und dass auf Basis externer Repräsentationen das Problemlösen und Argumentieren vernetzt werden kann.

6.9 Zusammenfassung und Ausblick

Fokus der vorliegenden Arbeit war, zu evaluieren, ob das Lösen problemhaltiger Textaufgaben mithilfe selbstgenerierter externer Repräsentationen erfolgreicher gelingen und zu einer Kompetenzsteigerung bei Grundschulkindern führen kann. Selbstgenerierte externe Repräsentationen dienen dazu den Denk- und Lösungswegen eine Struktur zu geben, von denen die Kinder selbst aber auch ihre Kommunikationspartner profitieren. Das Potenzial externer Repräsentationen wird insbesondere darin gesehen,

dass sie zum einen Gedächtnisstütze für den Produzenten und zum anderen Gesprächs- und Argumentationsstütze für den Austausch unter Gleichaltrigen sein können. Um den niedrigen Lösungsraten der Grundschulkinder beim Lösen problemhaltiger Textaufgaben entgegenzuwirken, wurde sowohl das Generieren und Nutzen externer Repräsentationen als auch das Kommunizieren mit und über externe Repräsentationen in der Interventionsstudie gefördert. Beide Fördermaßnahmen wurden erstmals gemeinsam in einem experimentellen Design untersucht, was sich als vielversprechend und gewinnbringend herausstellte.

Das Repräsentationstraining zeigte positive und nachhaltige Wirkungen auf die Leistungen der Kinder. Die Effektstärken sind unmittelbar nach der Interventionsmaßnahme als groß einzustufen. Es ist den Kindern durch das Training gelungen, sich das Potenzial externer Repräsentationen zunutze zu machen. Die Trainingsgruppe war sowohl in den Leistungen und dem Kompetenzerwerb als auch in der Nutzung externer Repräsentationen der Nicht-Trainingsgruppe überlegen. In Anbetracht der Komplexität der Textaufgaben konnten die Kinder beachtliche Steigerungen erzielen. Die Drittklässler hatten sich bei allen sechs bearbeiteten Aufgabentypen sowohl auf die Konstruktion als auch auf die Reflexion externer Repräsentationen eingelassen. Sie generierten zu den Problemstellungen passende Repräsentationen, die zur Lösungsfundung beitrugen, und zeigten dabei variierende Vorgehensweisen.

Dass das Generieren und Nutzen externer Repräsentationen kommunikative Tätigkeiten initiiert, die sich positiv auf die Problemlösekompetenzen auswirken können, konnte ebenso beobachtet werden. Allerdings waren die Effekte der kommunikativen Settings auf die Problemlösekompetenzen als klein einzustufen. Effekte auf den Lösungserfolg waren nicht vorhanden. Dies indiziert, dass durch die Implementierung kommunikativer Settings ein Kompetenzerwerb angebahnt werden kann, welcher aber noch verstärkt werden kann.

Kinder, die beide Fördermaßnahmen erhielten, d. h. am Repräsentationstraining teilnahmen wurden und sich unter Gleichaltrigen austauschen durften, hatten keinen Vorteil gegenüber Kindern, die nur eine Fördermaßnahme erhielten. Von dem Repräsentationstraining und den kommunikativen Settings ging keine gemeinsame Wirkung auf die Leistungen und Kompetenzentwicklung der Kinder aus.

Für die Unterrichtspraxis lässt sich ableiten, dass weder die Generierung externer Repräsentation noch die Möglichkeit, sich unter Gleichaltrigen auszutauschen, per se Erfolg garantiert. Die Lernenden benötigen vielmehr Unterstützung und vielfältige Möglichkeiten, Sachverhalte experimentell zu erschließen. Dass es lohnenswert ist, an den experimentell gewonnenen Entdeckungen und somit an den individuellen Ressourcen der Lernenden anzuknüpfen, sie zum Lerngegenstand des Unterrichts zu machen, hat die vorliegende Untersuchung gezeigt. Die Ergebnisse bestätigen die Befunde, dass

das Generieren externer Repräsentationen bereits in der Grundschule trainier- und erlernbar ist und die Kinder aufgabenbezogen zu einer gezielten Auswahl an Repräsentationen befähigt. Dies schließt für die Grundschule bestätigend ein, dass es übergeordnete Strategien und Handlungskompetenzen zu trainieren gilt, und nicht einzelne Aufgabentypen. Die Trainingsbausteine haben sich beim Lösen vielfältiger problemhaltiger Textaufgaben bewährt, da sie neben dem Verstehen der Sachsituation auch den Auf- und Ausbau eines Repräsentations-Repertoires anvisieren. Angefangen vom Nachvollziehen, Hineinversetzen und Verstehen mathematischer Entdeckungen eröffnet das Training den Lernenden neue Zugänge zum Problemlösen. Das Schulen des Strukturblicks, des richtigen Deutens oder des Aufeinander-Beziehens scheint dabei zu helfen, übergeordnete Kompetenzen zu erwerben, die immer wieder aufs Neue anzuwenden und zu erweitern sind. Die trainierten Kinder haben gezeigt, dass sie vielfältige Entdeckungen modifizieren und auf neue Situationen übertragen können. Die Erfahrung, dass nicht das Ergebnis, sondern der Weg dorthin zählt, wird von den Lernenden positiv und entlastend wahrgenommen. Sie zu sensibilisieren, welche Möglichkeiten es gibt, den eigenen Überlegungen und Gedanken eine Struktur zu geben, scheinen essenzielle Schritte auf dem Weg zum eigenständigen Problemlösen, gar zum eigenständigen Mathematiktreiben zu sein. Damit leistet die Studie auf der Basis der psychologischen *Representational Change Theory* einen Beitrag für die Förderung und Weiterentwicklung der Problemlösekompetenzen der Kinder. Bisher wurde der Generierung und Nutzung vielfältiger externer Repräsentationen ein eher untergeordneter Stellenwert zugesprochen. Wenn überhaupt, waren sie Teil von Strategietrainings, die sich vor allem auf Zeichnungen konzentrierten. Dagegen ist gerade der Ansatz, vielfältige externe Repräsentationen und damit verbunden, unterschiedliche Denk- und Erkenntniswerkzeuge in den Fokus des Trainings zu stellen, nachweislich sinnvoll und lohnenswert.

Betrachtet man die Förderung des intersubjektiven Austauschs, so wird deutlich, dass es sich um einen Kompetenzbereich handelt, in dem die Kinder Zeit und Unterstützung benötigen, um argumentierend und begründend tätig zu werden. Dies geht mit der Beobachtung einher, dass unterschiedliche Akzentuierungen im Unterricht unterschiedliche Wirkungen hervorrufen bzw. unterschiedlich viel Zeit beanspruchen, um Wirkungen zu zeigen.

In Anbetracht der geringen bzw. ausbleibenden Effekte, die auf die kommunikativen Settings oder die Kopplung beider Fördermaßnahmen zurückgehen, ergeben sich vielfältige Anknüpfungspunkt für weitere Studien. Beispielsweise wäre zu erwarten, dass ein längerer Interventionszeitraum größere Wirkungen erzielen würde, vor allem hinsichtlich der Kommunikation. Die Kinder könnten nach und nach auf einen Grundstock an Repräsentationskompetenzen und folglich auf ein sich fortschreitend entwi-

ckelndes Repräsentations-Repertoire zurückgreifen, auf das sie ihre Kommunikation stützen können. Kinder, die anfangs noch Schwierigkeiten haben, ihren Gedanken eine Struktur zu geben und diese zu externalisieren, wird es erwartungsgemäß schwer fallen, auf dieser Basis zu kommunizieren. Die Ergebnisse der Untersuchung lassen vermuten, dass vielfältige Lösungserfahrungen erforderlich sind, damit das Externalisierte tatsächlich zum Kommunikationsmittel werden kann. Um dies zu überprüfen und die Rekonstruktion der tatsächlich abgelaufenen Kommunikationsprozesse zu ermöglichen, scheint die Videographie eine geeignete Methode zu sein. Sie ermöglicht im Rahmen einer längsschnittlichen Analyse aufzudecken, ab wann die selbstgenerierten externen Repräsentationen zum Kommunikationsmittel werden, welche Entwicklungsprozesse die Kinder hierbei durchlaufen, wodurch diese Prozesse unterstützt werden können, aber auch welche Unterschiede zwischen den Teams beobachtbar werden bzw. welche Faktoren die Entwicklungen beeinflussen. Im Falle, dass es nicht gelingt, mit und über die externen Repräsentationen zu kommunizieren, können Gründe hierfür identifiziert und abgeleitet werden, was als weiterer Erkenntnisgewinn gesehen sind. Die videografierten Kommunikationsprozesse ermöglichen aufzuklären, ob sich die Gleichaltrigen egozentrisch verhalten oder sich gegenseitig helfen und Prozesse des Kollaborierens sichtbar werden.

Darüber hinaus wird ein zusätzlicher Nutzen einer solchen Untersuchung darin gesehen, dass der Gesprächsleitfaden und dessen Rolle im Gesprächsverlauf evaluiert werden kann. Welche Rolle dem Gesprächsleitfaden in den kommunikativen Settings zugesprochen werden kann und inwiefern dieser möglicherweise zu optimieren ist, wäre zu evaluieren. Erst mithilfe von Interviews oder videografierten Kommunikationssequenzen können die erzielten Effekte mit dem Gesprächsleitfaden in Verbindung gebracht werden.

Ferner kann auf Basis der vorliegenden Arbeit und des *integrativen Modelles des Text- und Bildverständens* die innere, mentale Sprache, die den Externalisierungsprozess begleitet, in Folgeuntersuchungen erforscht werden. Eine Methode, die geeignet ist, um mehr über die individuellen Lösungsgedanken und konstruierten Denk- und Erkenntniswerkzeuge zu erfahren, stellen Schülerinterviews dar. Die dadurch gewonnenen Daten und Informationen können beispielsweise Aufschluss darüber geben, welche Gedanken und Überlegungen das Kind beim Konstruktionsprozess begleitet haben, welche Strukturen des zu repräsentierenden Objekts gewählt werden, welche Umstrukturierungen nötig sind, um zu Teileinsichten oder zur Einsicht zu gelangen oder auch welche Barrieren konkret die Einsicht verhindern. Bei der Generierung multipler Repräsentationen kann zusätzlich transparent werden, welche externe(-n) Repräsentation(-en) zielführend war(-en) und inwiefern Verknüpfungen derer die Lösungsfindung ermöglichten. Einzelinterviews für die Evaluierung der kommunikativen Settings her-

anzuziehen, kämen ergänzend in Betracht, weil dadurch die Einschätzung der Lernenden, ob die kommunikativen Settings lernförderliche oder lernhemmende Wirkung haben, als zusätzlicher Einflussfaktor kontrollierbar wird.

Darüber hinaus könnten die erhobenen Daten in einer größer angelegten Studie Berücksichtigung finden. Es wäre denkbar, das experimentelle Design derart auszubauen, dass durch eine größere Stichprobe die Mehrebenenstruktur auch analytisch berücksichtigt werden kann. Ein Fokus könnte hierbei auf die Kontrollgruppe gerichtet sein. Im Vorfeld könnte mithilfe eines Expertenratings erhoben werden, welche Merkmale den regulären Problemlöseunterricht kennzeichnen und ausmachen. Auf dieser Basis könnte ein Kontrolltraining implementiert werden, das nur von einer Lehrkraft durchgeführt wird, wodurch dem Versuchsleitereffekt entgegengewirkt werden kann.

In der vorliegenden Untersuchung erfolgte eine Beschränkung auf sechs unterschiedliche Aufgabentypen problemhaltiger Textaufgaben. Wenn man das gesamte Spektrum möglicher Problemaufgaben berücksichtigt, dann stellt sich die Frage danach, ob die Effekte auf andere Aufgabentypen bzw. weitere mathematische Inhaltsbereichen übertragbar sind. Folgeanalysen könnten an den bisherigen Befunden anknüpfen und darauf aufbauend Transferwirkungen auf andere Inhaltsbereiche untersuchen. Ferner kann dies dazu beitragen, die in der vorliegenden Arbeit identifizierten Repräsentationsmerkmale an anderen Aufgabentypen zu prüfen und anzureichern.

Die Arbeit hat gezeigt, dass Kinder, die im Unterricht gefordert und gefördert werden, eigenständig Probleme zu bewältigen und die dabei auftretenden Barrieren zu überwinden, nach der Interventionsmaßnahme Erfolge und Kompetenzsteigerungen erzielen, zu denen sie vor der Interventionsmaßnahme noch nicht in der Lage waren. Es wurde von ihnen etwas verlangt, was sie nicht ohne Weiteres konnten, was sie aber dazu befähigt, alles zu leisten. Auch wenn die Kinder anfangs noch nicht wahrhaben können, dass sie fähig sind, die Herausforderung zu meistern, gelingt es ihnen durch das Generieren und Nutzen externer Repräsentationen sowie das Kommunizieren darüber, Brücken zu schlagen, die die Überwindung der Barriere ermöglichen und sie zum Erfolg führen kann. Auf ihre Lösungserfahrungen und ihr Repertoire an Denk- und Erkenntniswerkzeuge können sie in neuen Situationen immer wieder aufs Neue zurückgreifen und davon profitieren. Damit hat die Arbeit Möglichkeiten aufgezeigt, wie Kinder dazu befähigt werden können, über sich hinauszuwachsen, indem sie ihre eigenen Ressourcen und Fähigkeiten aufgreifen und weiterentwickeln.

Literatur

Adams, B. C., Bell, L. C., & Perfetti, C. A. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes*, 20(3), 307–323. <https://doi.org/10.1080/01638539509544943>

Adibnia, A., & Putt, I. J. (1998). Teaching problem solving to year 6 students: A new approach. *Mathematics Education Research Journal*, 10(3), 42–58. <https://doi.org/10.1007/BF03217057>

Aebli, H. (1980). *Denken: Das Ordnen des Tuns. Band I: Kognitive Aspekte der Handlungstheorie*. Stuttgart: Klett.

Amarel, S. (1966). On the mechanization of creative processes. *IEEE Spectrum*, 3(4), 112–114. <https://doi.org/10.1109/MSPEC.1966.5216589>

Anderson, J. R. (1996). *Kognitive Psychologie* (2. Aufl.). Heidelberg: Spektrum.

Awofala, A. O. A. (2014). Examining personalisation of instruction, attitudes toward and achievement in mathematics word problems among nigerian senior secondary school students. *International Journal of Education in Mathematics, Science and Technology*, 2(4), 273–288.

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Aufl., S. 206–226). New York: Cambridge University Press.

Baddeley, A. (1997). *Human memory: Theory and practice*. Hove, UK: Psychology Press.

Baker, M. J. (1994). A model for negotiation in teaching-learning dialogues. *Journal for Artificial Intelligence in Education*, 5(2), 199–254.

Balacheff, N. (1991). The benefits and limits of social interaction: The case of mathematical proof. In A. J. Bishop, S. Mellin-Olsen, & J. van Dormolen (Hrsg.), *Mathematical knowledge: Its growth through teaching* (S. 175–192). Dordrecht: Kluwer Academic Publishers.

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, 7(2), 127–150. <https://doi.org/10.1177/1094428104263672>

Baltes-Götz, B. (2012). *Logistische Regressionsanalyse mit SPSS*. Abgerufen von <https://www.uni-trier.de/fileadmin/urt/doku/logist/logist.pdf> (letzter Zugriff am 06.11.2016)

Baltes-Götz, B. (2016). *Generalisierte lineare Modelle und GEE-Modelle in SPSS Statistics*. Abgerufen von https://www.uni-trier.de/fileadmin/urt/doku/gzlm_gee/gzlm_gee.pdf (letzter Zugriff am 06.11.2016)

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1), 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>

Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedien. Erfassung, Beschreibung und Vermittlung wirksamer metakognitiver Strategien und Regulationsaktivitäten*. Münster: Waxmann.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/doi:10.18637/jss.v067.i01>

Bauer, L. (1993). Das operative Prinzip als umfassendes, allgemeingültiges Prinzip für das Mathematiklernen? Didaktisch-methodische Überlegungen zum Mathematikunterricht in der Grundschule. *Zentralblatt für Didaktik der Mathematik*, 93(2), 76–83.

Bauersfeld, H. (2002). Interaktion und Kommunikation. *Grundschule*, 34(3), 10–15.

Baumert, J., & Lehmann, R. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: Deskriptive Befunde*. Opladen: Leske + Budrich.

Bayrhuber, H., Leuders, T., Bruder, R., & Wirtz, M. (2010). Repräsentationswechsel beim Umgang mit Funktionen – Identifikation von Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells (Projekt HEUREKO). In E. Klieme, D. Leutner, & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 28–39). Weinheim: Beltz.

Becker, P., Schaller, S., & Schmidtke, P. (1980). *Coulored progressive matrices. Manual* (2. Aufl.). Weinheim: Beltz.

Berk, L. E. (2011). *Entwicklungspsychologie* (5. Aufl.). München: Pearson Education.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, 17(5), 564–577. <https://doi.org/10.1016/j.learninstruc.2007.09.007>

Bezold, A. (2009). *Förderung von Argumentationskompetenzen durch selbstdifferenzierende Lernangebote: eine Studie im Mathematikunterricht der Grundschule* (Bd. 47). Hamburg: Dr. Kovač.

Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods*, 7(4), 400–417. <https://doi.org/10.1177/1094428104268542>

Blömeke, S., Seeber, S., Lehmann, R., Kaiser, G., Schwarz, B., Felbrich, A., & Müller, C. (2008). Messung des fachbezogenen Wissens angehender Mathematiklehrkräfte. In S. Blömeke, G. Kaiser, & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare* (S. 49–88). Münster: Waxmann.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (2006). *Bildungsstandards Mathematik: konkret: Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen*. Berlin: Corneelsen.

Blum, W., & Kirsch, A. (1989). Warum haben nichttriviale Lösungen von $f'' = f$ keine Nullstellen? Beobachtungen und Bemerkungen zum „inhaltlich-anschaulichen Beweisen“. In H. Kautschitsch & W. Metzler (Hrsg.), *Anschauliches Beweisen: 7. und 8. Workshop zur „Visualisierung in der Mathematik“ in Klagenfurt im Juli 1987 und 1988* (S. 199–210). Wien: Hölder-Pichler-Tempsky.

Blum, W., & Kirsch, A. (1991). Preformal proving: Examples and reflections. *Educational Studies in Mathematics*, 22(2), 183–203. <https://doi.org/10.1007/BF00555722>

Boaler, J. (2003). Studying and capturing the complexity of practice: The case of the dance of agency. In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Hrsg.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education Held Jointly With The 25th Conference of PME-NA* (Bd. 1, S. 3–16). Honolulu: IGPME.

Boaler, J. (2008). Promoting „relational equity“ and high mathematics achievement through an innovative mixed ability approach. *British Educational Research Journal*, 34(2), 167–194.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of railside school. *Teachers' College Record*, 110(3), 608–645.

Bölte, S. (2013). Testpsychologische Verfahren. In G. Lehmkuhl, F. Poustka, M. Holtmann, & H. Steiner (Hrsg.), *Lehrbuch der Kinder- und Jugendpsychiatrie* (S. 346–357). Göttingen: Hogrefe.

Booker, G., Bond, D., Sparrow, L., & Swan, P. (2014). *Teaching primary mathematics* (5. Aufl.). Frenchs Forest: Pearson.

Boonen, A. J. H., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research*, 68, 15–26. <https://doi.org/10.1016/j.ijer.2014.08.001>

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler: Limitierte Sonderausgabe* (4. Aufl.). Berlin: Springer.

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.

Brander, S., Kompa, A., & Peltzer, U. (1989). *Denken und Problemlösen: Einführung in die kognitive Psychologie* (2. Aufl.). Opladen: Westdeutscher Verlag.

Brandstädter, J. (1990). Evaluationsforschung: Probleme der wissenschaftlichen Bewertung von Interventions- und Reformprojekten. *Zeitschrift für Pädagogische Psychologie*, 4(4), 215–227.

Branford, B. (1913). *Betrachtungen über mathematische Erziehung vom Kindergarten bis zur Universität*. (R. Schimmmack & H. Weinreich, Übers.). Leipzig: B. G. Teubner.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717–726. [https://doi.org/10.1016/S0022-5371\(72\)80006-9](https://doi.org/10.1016/S0022-5371(72)80006-9)

Bransford, J. D., & Stein, B. S. (1993). *The ideal problem solver: A guide for improving thinking, learning, and creativity* (2. Aufl.). New York: W.H. Freeman.

Brebaugh, J. A., & Arnold, J. (2007). Controlling nuisance variables by using a matched-groups design. *Organizational Research Methods*, 10(3), 523–541. <https://doi.org/10.1177/1094428106292895>

Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Durán, R., Reed, B. S., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal*, 34(4), 663–689. <https://doi.org/10.2307/1163353>

Breuer, F. (2010). *Reflexive Grounded Theory: Eine Einführung für die Forschungspraxis* (2. Aufl.). Wiesbaden: Springer.

Bromm, W. (2009). Beispiele für den Mathematikunterricht der Klasse 4. Lösen problemhaltiger Textaufgaben. *Grundschulmagazin*, 7(5), 39–42.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2. Aufl.). New York: Guilford Press.

Bruder, R. (2000). Problemlösen im Mathematikunterricht – ein Lernangebot für alle? *Mathematische Unterrichtspraxis*, 20(1), 2–11.

Bruder, R. (2003). *Methoden und Techniken des Problemlösenlernens. Material im Rahmen des BLK-Programms „Sinus“ zur „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“*. Universität Kiel. Abgerufen von <http://www.math-learning.com/files/Skript.pdf> (letzter Zugriff am 06.11.2016)

Bruder, R., & Collet, C. (2011a). Entwicklung und Erprobung eines Unterrichtskonzeptes zum Problemlösenlernen im Mathematikunterricht – Wirkungsanalysen bei den Lehrenden und Lernenden. In H. Bayrhuber, U. Harms, B. Muszynski, B. Ralle, M. Rothgangel, L.-H. Schön, ... H.-G. Weigand (Hrsg.), *Empirische Fundierung in den Fachdidaktiken* (Bd. 1, S. 27–46). Münsster: Waxmann.

Bruder, R., & Collet, C. (2011b). *Problemlösen lernen im Mathematikunterricht*. Berlin: Cornelsen.

Bruder, R., & Weigand, H.-G. (2005). Problemlösen, Verstehen, Anwenden... aber bitte diskret! *Mathematik lehren*, 129, 4–9.

Bruner, J. S. (1971). Über kognitive Entwicklung. In J. S. Bruner, R. R. Olver, & P. M. Greenfield (Hrsg.), *Studien zur kognitiven Entwicklung: eine kooperative Untersuchung am „Center for Cognitive Studies“ der Harvard-Universität* (S. 21–53). Stuttgart: Klett.

Bruner, J. S. (1973). *Der Prozess der Erziehung* (3. Aufl.). Berlin: Berlin-Verlag.

Brunner, E. (2014a). *Mathematisches Argumentieren, Begründen und Beweisen*. Berlin: Springer.

Brunner, E. (2014b). Verschiedene Beweistypen und ihre Umsetzung im Unterrichtsgespräch. *Journal für Mathematik-Didaktik*, 35(2), 229–249. <https://doi.org/10.1007/s13138-014-0065-6>

Bühler, K. (1908). Tatsachen und Probleme zu einer Psychologie der Denkvorgänge: II. Über Gedanken Zusammenhänge. *Archiv für die gesamte Psychologie*, 12, 1–23.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.

Bullheller, S., & Häcker, H. (2002). *Coulored Progressive Matrices (CPM). Evaluation und Neunormierung* (2. Aufl.). Frankfurt am Main: Pearson.

Büttner, G. (2008). *Handbuch der pädagogischen Psychologie*. (W. Schneider & M. Hasselhorn, Hrsg.). Göttingen: Hogrefe.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2), 233–246. <https://doi.org/10.1111/j.2044-8279.1992.tb01017.x>

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, 10(2), 151–170. [https://doi.org/10.1002/\(SICI\)1099-0720\(199604\)10:2<151::AID-ACP380>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U)

Charles, R. I., & Lester, F. (1982). *Teaching problem solving. What, why, & how*. Palo Alto, CA: Dale Seymour Publications.

Charles, R. I., Lester, F. K., & O'Daffer, P. G. (1987). *How to evaluate progress in problem solving*. Virginia: National Council of Teachers of Mathematics.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1–49. <https://doi.org/10.2307/1170558>

Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 14–27. <https://doi.org/10.1037/0278-7393.30.1.14>

Clercq, A. D., Desoete, A., & Roeyers, H. (2000). Epa2000: A multilingual, programmable computer assessment of off-line metacognition in children with mathematical-learning disabilities. *Behavior Research Methods, Instruments, & Computers*, 32(2), 304–311. <https://doi.org/10.3758/BF03207799>

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education*, 23(1), 2–33. <https://doi.org/10.2307/749161>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). New York: Psychology Press.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304–1312. <https://doi.org/10.1037/0003-066X.45.12.1304>

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/Correlation analysis for the behavioral sciences* (3. Aufl.). Mahwah, NJ: Lawrence Erlbaum.

Collet, C. (2009). *Förderung von Problemlösekompetenzen in Verbindung mit Selbstregulation*. Münsster: Waxmann.

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, 34(1), 35–46. <https://doi.org/10.1177/001316447403400105>

Cooper, T. J., & Warren, E. (2008). The effect of different representations on years 3 to 5 students' ability to generalise. *ZDM*, 40(1), 23–37. <https://doi.org/10.1007/s11858-007-0066-8>

Cox, R. (1999). Representation construction, externalised cognition, and individual differences. *Learning and Instruction*, 9(4), 343–363. [https://doi.org/10.1016/S0959-4752\(98\)00051-6](https://doi.org/10.1016/S0959-4752(98)00051-6)

Cox, R., & Brna, P. (1995). Supporting the use of external representation in problem solving: The need for flexible learning environments. *Journal for Artificial Intelligence in Education*, 6(2-3), 239–302.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)

Csíkos, C., Szitányi, J., & Kelemen, R. (2012). The effects of using drawings in developing young children's mathematical word problem solving: A design experiment with third-grade Hungarian students. *Educational Studies in Mathematics*, 81(1), 47–65. <https://doi.org/10.1007/s10649-011-9360-z>

Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology*, 5(4), 331–343. [https://doi.org/10.1016/0193-3973\(84\)90006-6](https://doi.org/10.1016/0193-3973(84)90006-6)

Danner, D. (2015). Reliabilität – die Genauigkeit einer Messung. *Mannheim, GESIS-Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines)*. https://doi.org/10.15465/sdm-sg_011

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: a review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, 6, 1–13. <https://doi.org/10.3389/fpsyg.2015.00348>

de Bock, D., Verschaffel, L., Janssens, D., van Dooren, W., & Claeys, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modelling non-linear geometry problems. *Learning and Instruction*, 13(4), 441–463.

de Corte, E., Verschaffel, L., & de Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, 77(4), 460–470. <https://doi.org/10.1037/0022-0663.77.4.460>

Dedekind, B. (2012). „*Darstellen in der Mathematik“ als Kompetenz aufbauen*. Handreichungen des Programms SINUS an Grundschulen, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Universität Kiel. Abgerufen von http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Handreichung_Dedekind.pdf (letzter Zugriff am 06.11.2016)

Deseniss, A. (2015). *Schulmathematik im Kontext von Migration: Mathematikbezogene Vorstellungen und Umgangsweisen mit Aufgaben unter sprachlich-kultureller Perspektive*. Wiesbaden: Springer.

Desoete, A. (2001). *Off-line metacognition in children with mathematics learning disabilities* (Dissertation). University of Gent, Gent. Abgerufen von <https://biblio.ugent.be/publication/522137/file/1874176> (letzter Zugriff am 25.02.2016)

Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology*, 5(3), 705–730.

Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. *Metacognition and Learning*, 3(3), 189–206. <https://doi.org/10.1007/s11409-008-9026-0>

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities*, 34(5), 435–447. <https://doi.org/10.1177/002221940103400505>

Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology*, 95(1), 188–200. <https://doi.org/10.1037/0022-0663.95.1.188>

Devlin, K. J. (2003). *Das Mathe-Gen oder wie sich das mathematische Denken entwickelt und warum Sie Zahlen ruhig vergessen können*. (D. Zimmer, Übers.). München: Dt. Taschenbuch-Verlag.

Diekmann, A. (2007). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. Reinbek: Rowohlt-Taschenbuch-Verlag.

Dienes, Z. P., & Golding, E. W. (1970). *Methodik der modernen Mathematik: Grundlagen für Lernen in Zyklen*. Freiburg: Herder.

Dörfler, W. (1979). Vorwort. In W. Dörfler & R. Fischer (Hrsg.), *Beweisen im Mathematikunterricht: Vorträge des 2. Internationalen Symposiums für „Didaktik der Mathematik“ vom 26.09. bis 29.09.1978 in Klagenfurt* (S. 9–11). Wien: Hölder-Pilcher-Tempsky.

Dörfler, W. (1989). Didaktischer Einsatz mathematischer Software: Tabellenkalkulationsprogramme. In W. Walsch (Hrsg.), *Kleincomputer und Mathematikunterricht* (S. 78–85). Halle: Martin-Luther-Universität Halle-Wittenberg.

Dörfler, W. (2006). Diagramme und Mathematikunterricht. *Journal für Mathematik-Didaktik*, 27(3-4), 200–219. <https://doi.org/10.1007/BF03339039>

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (5. Aufl.). Berlin: Springer.

Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.

Dreher, A. (2013). Den Wechsel von Darstellungsformen fördern und fordern oder vermeiden? Über ein Dilemma im Mathematikunterricht. In J. Sprenger, A. Wagner, & M. Zimmermann (Hrsg.), *Mathematik lernen, darstellen, deuten, verstehen. Didaktische Sichtweisen vom Kindergarten bis zur Hochschule*. (S. 215–225). Wiesbaden: Springer.

Duarte, J., Gogolin, I., & Kaiser, G. (2011). Sprachlich bedingte Schwierigkeiten von mehrsprachigen Schülerinnen und Schülern bei Textaufgaben. In S. Prediger & E. Özil (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit. Stand und Perspektiven der Forschung und Entwicklung in Deutschland* (S. 35–53). Münster: Waxmann.

Düll, K. (2009a). Sachaufgaben mit Hilfe einer Tabelle darstellen und lösen. Kinder stellen Aufgaben dar. *Grundschulmagazin*, 7(5), 27–31.

Düll, K. (2009b). *Sachrechnen in der Grundschule: Kinder stellen sich Aufgaben dar, 1.-4. Schuljahr*. München: Oldenbourg.

Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Berlin: Springer.

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, 38(1/2), 159–177. <https://doi.org/10.2307/2332325>

Duval, R. (1991). Structure du raisonnement deductif et apprentissage de la démonstration. *Educational Studies in Mathematics*, 22(3), 233–261. <https://doi.org/10.1007/BF00368340>

Duval, R. (1999). Representation, vision, and visualization: Cognitive functions in mathematical thinking. Basic issues for learning. In F. Hitt & M. Santos (Hrsg.), *Proceedings of the Twenty-first Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education XXI* (Bd. 1, S. 3–26). Columbus, OH: Eric Publications. Abgerufen von ERIC Database. (ED466379)

Efkides, A., Papadaki, M., Papantoniou, G., & Kiosseoglou, G. (1997). Effects of cognitive ability and affect on school mathematics performance and feelings of difficulty. *The American Journal of Psychology*, 110(2), 225–258.

Eichler, K.-P. (2008). Skizzen als Hilfen beim Lösen von Sachaufgaben. *Praxis Grundschule*, 31(5), 29–30.

Eichler, K.-P. (2014). Zeichnen und Skizzieren. Handwerk und mehr. *Grundschulunterricht Mathematik*, 3, 4–7.

Eichler, K.-P. (2015). Skizzen als Werkzeug zum Lösen von Sachaufgaben – planmäßig entwickeln und erfolgreich nutzen. In R. Rink (Hrsg.), *Von Guten Aufgaben bis Skizzen Zeichnen: Zum Sachrechnen im Mathematikunterricht der Grundschule* (Bd. 31, S. 49–70). Baltmannsweiler: Schneider Verlag Hohengehren.

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden: Lehrbuch*. (2. Aufl.). Weinheim: Beltz.

Fehse, E. (2001). *Unterstützung von Kohärenzbildung beim kooperativen und individuellen Lernen mit externen Repräsentationen* (Dissertation). Universität Freiburg, Freiburg. Abgerufen von <https://www.freidok.uni-freiburg.de/data/220> (letzter Zugriff am 06.11.2016)

Felbrich, A. (2005). *Kontrastierung als effektive Lerngelegenheiten zur Vermittlung von Wissen über Repräsentationsformen am Beispiel des Graphen einer linearen Funktion* (Dissertation). Technische Universität, Berlin. Abgerufen von <http://dx.doi.org/10.14279/depositonce-1257> (letzter Zugriff am 06.11.2016)

Fetzer, M. (2007). *Interaktion am Werk: Eine Interaktionstheorie fachlichen Lernens, entwickelt am Beispiel von Schreibanlässen im Mathematikunterricht der Grundschule*. Bad Heilbrunn: Julius Klinkhardt.

Fetzer, M. (2009). Schreibe Mathe und sprich darüber - Schreibanlässe als Möglichkeit, Argumentationskompetenzen zu fördern. *Praxis der Mathematik in der Schule*, 51(30), 21–25.

Fetzer, M. (2011). Wie argumentieren Grundschulkinder im Mathematikunterricht? Eine argumentationstheoretische Perspektive. *Journal für Mathematik-Didaktik*, 32(1), 27–51. <https://doi.org/10.1007/s13138-010-0021-z>

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4. Aufl.). London: SAGE.

Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention* (2. Aufl.). Göttingen: Hogrefe.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16(4), 625–641. <https://doi.org/10.1080/10705510903203573>

Forman, E. (1989). The role of peer interaction in the social construction of mathematical knowledge. *International Journal of Educational Research*, 13(1), 55–70. [https://doi.org/10.1016/0883-0355\(89\)90016-5](https://doi.org/10.1016/0883-0355(89)90016-5)

Franke, M., & Ruwisch, S. (2010). *Didaktik des Sachrechnens in der Grundschule* (2. Aufl.). Heidelberg: Spektrum.

Freseemann, O. (2014). *Schwache Rechnerinnen und Rechner fördern. Eine Interventionsstudie an Haupt-, Gesamt- und Förderschulen*. Wiesbaden: Springer.

Freseemann, O., & Breucker, T. (2014). Förderung flexibler Übersetzungsprozesse. *Grundschulunterricht Mathematik*, 1, 8–12.

Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Hrsg.), *Complex problem solving: The european perspective* (S. 3–25). Hillsdale: Psychology Press.

Freudenthal, H. (1973a). *Mathematik als pädagogische Aufgabe* (Bd. 1). Stuttgart: Klett.

Freudenthal, H. (1973b). *Mathematik als pädagogische Aufgabe* (Bd. 2). Stuttgart: Klett.

Freudenthal, H. (1974). Die Stufen im Lernprozeß und die heterogene Lerngruppe im Hinblick auf die Middenschool. *Neue Sammlung*, 14, 161–169.

Freudenthal, H. (1979). Konstruieren, Reflektieren, Beweisen in phänomenologischer Sicht. In W. Dörfler & R. Fischer (Hrsg.), *Beweisen im Mathematikunterricht: Vorträge des 2. Internationalen Symposiums für „Didaktik der Mathematik“ vom 26.09. bis 29.09.1978 in Klagenfurt* (S. 183–200). Wien: Hölder-Pilcher-Tempsky.

Fricke, A. (1970). Operative Lernprinzipien im Mathematikunterricht der Grundschule. In A. Fricke & H. Besuden (Hrsg.), *Mathematik. Elemente einer Didaktik und Methodik. Der Unterricht in der Grundschule* (S. 79–116). Stuttgart: Klett.

Fricke, A. (1987). *Sachrechnen: das Lösen angewandter Aufgaben*. Stuttgart: Klett.

Fritz, A., Ricken, G., Gerlach, M., & Schmidt, S. (2007). *Kalkulie: Diagnose- und Trainingsprogramm für rechenschwache Kinder. Teil 1: Fertigkeitsspezifische Voraussetzungen*. Berlin: Cornelsen.

Fuchs, M. (2006). *Vorgehensweisen mathematisch potentiell begabter Dritt- und Viertklässler beim Problemlösen: empirische Untersuchungen zur Typisierung spezifischer Problembearbeitungsstile*. Münster: LIT.

Gagatsis, A., & Elia, I. (2004). The effects of different modes of representation on mathematical problem solving. In M. Johnson-Höines & A. B. Fuglestad (Hrsg.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 2, S. 447–454). Bergen: Bergen University College Press. Abgerufen von ERIC Database. (ED489751)

Gallin, P. (2010). Dialogisches Lernen. Von einem pädagogischen Konzept zum täglichen Unterricht. *Grundschulunterricht Mathematik*, 2, 4–10.

Gallin, P., & Ruf, U. (1993). Sprache und Mathematik in der Schule. Ein Bericht aus der Praxis. *Journal für Mathematik-Didaktik*, 14(1), 3–33. <https://doi.org/10.1007/BF03339303>

Garnham, A. (1981). Mental models as representations of text. *Memory & Cognition*, 9(6), 560–565. <https://doi.org/10.3758/BF03202350>

Geier, M. (2012). Der regelmäßige Einsatz von Problemaufgaben im Mathematikunterricht in Grundschulen. In M. Ludwig & M. Kleine (Hrsg.), *Beiträge zum Mathematikunterricht 2012: Vorträge auf der 46. Tagung für Didaktik der Mathematik vom 05.03. bis 09.03.2012 in Weingarten* (Bd. 1, S. 285–288). Münster: WTM.

Gerhard, S., & Glaser, B. (2014). Kinder lernen Sachaufgaben zeichnen. *Grundschulunterricht Mathematik*, 3, 16–19.

Gerstenberger-Ratzeburg, B. (2016). *Intelligenz und Teddybären: Was Ihnen mit begabten Kindern alles blühen kann*. Norderstedt: BoD.

Gerster, H.-D., & Schultz, R. (2004). *Schwierigkeiten beim Erwerb mathematischer Konzepte im Anfangsunterricht. Bericht zum Forschungsprojekt Rechenschwäche – Erkennen, Beheben, Vorbeugen*. Pädagogische Hochschule Freiburg. Abgerufen von http://www.schuleundgesundheit.hessen.de/fileadmin/content/Themen/as/forschungsbericht_dyslex.pdf (letzter Zugriff am 06.11.2016)

Gibb, E. G. (1956). Children's thinking in the process of subtraction. *The Journal of Experimental Education*, 25(1), 71–80.

Gick, M. L., & Lockhart, R. S. (1995). Cognitive and affective components of insight. In R. J. Sternberg & J. E. Davidson (Hrsg.), *The Nature of Insight* (S. 197–228). Cambridge: MIT Press.

Gläser-Zirkuda, M. (2015). Qualitative Auswertungsverfahren. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung: Strukturen und Methoden* (2. Aufl., S. 119–130). Wiesbaden: Springer.

Gogolin, I. (2009). Zweisprachigkeit und die Entwicklung bildungssprachlicher Fähigkeiten. In I. Gogolin & U. Neumann (Hrsg.), *Streitfall Zweisprachigkeit – The Bilingualism Controversy* (S. 263–280). Wiesbaden: Springer.

Gogolin, I., Kaiser, G., Roth, H.-J., Deseniss, A., Hawighorst, B., & Schwarz, I. (2004). *Mathematiklernen im Kontext sprachlich-kultureller Diversität (DFG Go 614/6)* (Abschlussbericht). Universität Hamburg. Abgerufen von <https://www.ew.uni-hamburg.de/ueber-die-fakultaet/personen/gogolin/pdf-dokumente/mathe-bericht.pdf> (letzter Zugriff am 06.11.2016)

Goldin, G. A., Rösken, B., & Törner, G. (2009). Beliefs - No longer a hidden variable in mathematical teaching and learning processes. In J. Maasz & W. Schloeglmann (Hrsg.), *Beliefs and Attitudes in Mathematics Education* (S. 1–18). Rotterdam: Sense Publishers.

Goldin, G. A., & Shteingold, N. (2001). Systems of representations and the development of mathematical concepts. In A. A. Cuoco & F. R. Curcio (Hrsg.), *The roles of representation in school mathematics* (S. 1–23). Reston, VA: NCTM.

Goos, M., Galbraith, P., & Renshaw, P. (1996). When does student talk become collaborative mathematical discussion? In P. C. Clarkson (Hrsg.), *Technology in Mathematics Education. Proceedings of the 19th Annual Conference of the Mathematics Education Research Group of Australasia in Melbourne*. (S. 237–244). Melbourne: MERGA.

Götze, D. (2008). „Mathematische Gespräche unter Kindern“ – Zum Einfluss sozialer Interaktion von Grundschulkindern beim Lösen komplexer Aufgaben. *Journal für Mathematik-Didaktik*, 29(1), 70–71. <https://doi.org/10.1007/BF03339363>

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163–189. <https://doi.org/10.1146/annurev.psych.48.1.163>

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Hrsg.), *Handbook of psychology* (Bd. 2, S. 87–114). Hoboken, NJ: John Wiley & Sons.

Graumann, C. F. (1960). *Grundlagen einer Phänomenologie und Psychologie der Perspektivität* (Bd. 2). Berlin: De Gruyter.

Greefrath, G. (2004). Offene Aufgaben mit Realitätsbezug. Eine Übersicht mit Beispielen und erste Ergebnisse aus Fallstudien. *Mathematica didactica*, 27(2), 16–38.

Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Hrsg.), *Transfer on trial: Intelligence, cognition, and instruction* (S. 99–167). Norwood, NJ: Ablex Publishing Corporation.

Greevy, R., Lu, B., Silber, J. H., & Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics*, 5(2), 263–275. <https://doi.org/10.1093/biostatistics/5.2.263>

Grevsmühl, U. (1995). *Mathematik für Grundschullehrer: ein Fernstudienlehrgang: didaktisches Begeleitheft zu E1–E4*. Tübingen: Deutsches Institut für Fernstudien.

Griesel, H. (1973). *Die neue Mathematik für Lehrer und Studenten. Bd. 2, Größen, Bruchzahlen, Sachrechnen*. Hannover: Schroedel.

Grigutsch, S., Raatz, U., & Törner, G. (1998). Einstellungen gegenüber Mathematik bei Mathematiklehrern. *Journal für Mathematik-Didaktik*, 19(1), 3–45. <https://doi.org/10.1007/BF0338859>

Groß, J. (2013). *Analyse von Lösungsprozessen beim Bearbeiten problemhaltiger Textaufgaben durch Grundschulkinder* (Dissertation). Universität Koblenz-Landau, Landau. Abgerufen von Deutschen Nationalbibliothek. (IDN 1053269048/04)

Groves, R. M., Jr, F. J. F., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. Hoboken, NJ: Wiley.

Guberman, R., & Leikin, R. (2012). Interesting and difficult mathematical problems: changing teachers' views by employing multiple-solution tasks. *Journal of Mathematics Teacher Education*, 16(1), 33–56. <https://doi.org/10.1007/s10857-012-9210-7>

Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. <https://doi.org/10.1037/h0063487>

Gürtler, T., Perels, F., Schmitz, B., & Bruder, R. (2002). Training zur Förderung selbstregulativer Fähigkeiten in Kombination mit Problemlösen in Mathematik. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schule und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 222–239). Weinheim: Beltz.

Gysin, B. (2013). Kinder erkennen Strukturen. Eine praxisorientierte Annäherung an eine herausfordernde mathematische Kompetenz. In J. Sprenger, A. Wagner, & M. Zimmermann (Hrsg.), *Mathematik lernen, darstellen, deuten, verstehen. Didaktische Sichtweisen vom Kindergarten bis zur Hochschule*. (S. 125–138). Wiesbaden: Springer.

Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest (HRT 1-4): Erfassung mathematischer Basiskompetenzen im Grundschulalter*. Göttingen: Hogrefe.

Hager, W. (2000a). *Evaluation psychologischer Interventionsmaßnahmen: Standards und Kriterien: ein Handbuch*. Bern: Huber.

Hager, W. (2000b). Zur Wirksamkeit von Interventionsprogrammen: Allgemeine Kriterien der Wirksamkeit von Programmen in einzelnen Untersuchungen. In W. Hager, J.-L. Patry, & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionen. Standards und Kriterien: Ein Handbuch* (S. 153–168). Bern: Hans Huber.

Hager, W. (2008). Evaluation von pädagogisch-psychologischen Interventionsmaßnahmen. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S. 721–732). Göttingen: Hogrefe.

Hager, W., & Hasselhorn, M. (2000). Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In W. Hager, J.-L. Patry, & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionen. Standards und Kriterien: Ein Handbuch* (S. 41–85). Bern: Hans Huber.

Hahn, H., & Janott, S. (2011). Förderung der mathematischen Kompetenz des Darstellens. *Grundschulunterricht Mathematik*, 2, 15–18.

Hahn, H., & Janott, S. (2012). Eigene Lernwege zur Bearbeitung einer Problemaufgabe – Die Vielfalt von Lösungsstrategien würdigen. *Sache – Wort – Zahl*, 127(40), 28–39.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7. Aufl.). Prentice Hall: Pearson.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development*, 35(4), 193–217. <https://doi.org/10.1159/000277167>

Hanna, G. (1989a). More than formal proof. *For the Learning of Mathematics*, 9(1), 20–23.

Hanna, G. (1989b). Proofs that prove and proofs that explain. In G. Vergnaud, J. Rogalski, & M. Artigue (Hrsg.), *Proceedings of the 13th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 2, S. 45–51). Paris: Eric Publications. Abgerufen von ERIC Database. (ED411141)

Hardy, I. (2007). Die Förderung von Problemlösekompetenzen im Unterricht. Ergebnisse der Lernforschung und Umsetzung im Schulunterricht. In *Das Lehrerhandbuch: der pädagogische Ratgeber für Lehrerinnen und Lehrer. Grundwerk* (S. 2–16). Berlin: Raabe.

Hartmann, B. (2008). Mathematik lernen und Sprachkompetenz entwickeln. In I. Esslinger-Hinz & H. Hahn (Hrsg.), *Kompetenzen entwickeln. Unterrichtsqualität in der Grundschule steigern: Entwicklungslinien und Forschungsbefunde* (2. Aufl., S. 151–156). Hohengehren: Schneider.

Hasdorf, W. (1976). Erscheinungsbild und Entwicklung der Beweglichkeit des Denkens bei älteren Vorschulkindern. In J. Lompscher (Hrsg.), *Verlaufsqualitäten der geistigen Tätigkeit* (S. 13–75). Berlin: Volk und Wissen.

Hasselhorn, M. (1995). Kognitive Trainings: Grundlagen, Begrifflichkeiten und Desiderate. In W. Hager (Hrsg.), *Programme zur Förderung des Denkens bei Kindern: Konstruktion, Evaluation und Metaevaluation* (S. 14–40). Göttingen: Hogrefe.

Hasselhorn, M., & Hager, W. (1998). Kognitive Trainings auf dem Prüfstand. Welche Komponenten charakterisieren erfolgreiche Fördermaßnahmen? In M. Beck (Hrsg.), *Evaluation als Maßnahme der Qualitätssicherung: Pädagogisch-psychologische Interventionen auf dem Prüfstand* (S. 85–98). Tübingen: Dgvt.

Hasselhorn, M., & Hager, W. (2006). Kognitives Training. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 341–349). Weinheim: Beltz.

Hasselhorn, M., & Hager, W. (2008). Transferwirkungen kognitiver Trainings. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 381–390). Göttingen: Hogrefe.

Hayes, J. R. (1981). *The complete problem solver* (2. Aufl.). Hillsdale, NJ: Lawrence Erlbaum.

Hayton, J. C., Allen, D. G., & Scarfello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205. <https://doi.org/10.1177/1094428104263675>

Hefendehl-Hebeker, L., & Hußmann, S. (2010). Beweisen – Argumentieren. In T. Leuders (Hrsg.), *Mathematik-Didaktik: Praxishandbuch für die Sekundarstufe I und II* (5. Aufl., S. 93–106). Berlin: Cornelsen.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18–32.

Heinrich, F., Bruder, R., & Bauer, C. (2015). Problemlösen lernen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 279–301). Berlin: Springer.

Heinrich, F., Pawlitzki, A., & Schuck, L.-D. (2013). Problemlöseunterricht in der Grundschule. In G. Greefrath, F. Käpnick, & M. Stein (Hrsg.), *Beiträge zum Mathematikunterricht 2013: Beiträge zur 47. Jahrestagung der Gesellschaft für Didaktik der Mathematik vom 04.–08. März in Münster* (Bd. 1, S. 448–451). Münster: WTM.

Heinze, A. (2007). Problemlösen im mathematischen und außermathematischen Kontext. *Journal für Mathematik-Didaktik*, 28(1), 3–30. <https://doi.org/10.1007/BF03339331>

Heinze, A., Herwartz-Emden, L., Braun, C., & Reiss, K. (2011). Die Rolle von Kenntnissen der Unterrichtssprache beim Mathematiklernen. Ergebnisse einer quantitativen Längsschnittstudie in der Grundschule. In S. Prediger & E. Özdil (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit. Stand und Perspektiven der Forschung und Entwicklung in Deutschland* (S. 11–33). Münster: Waxmann.

Heinze, A., Herwartz-Emden, L., & Reiss, K. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik*, 53(4), 562–581.

Heinze, A., & Reiss, K. (2004). Mathematikleistung und Mathematikinteresse in differenzieller Perspektive. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 234–249). Münster: Waxmann.

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education*, 23(3), 242–273. <https://doi.org/10.2307/749120>

Hengartner, E. (1992). Für ein Recht der Kinder auf eigenes Denken. *Die neue Schulpraxis*, (7/8), 15–27.

Hirt, U., & Wälti, B. (2010). *Lernumgebungen im Mathematikunterricht: natürliche Differenzierung für Rechenschwäche bis Hochbegabte* (2. Aufl.). Seelze-Velber: Kallmeyer.

Hirt, U., Wälti, B., & Wollring, B. (2010). Lernumgebungen für den Mathematikunterricht in der Grundschule: Begriffsklärung und Positionierung. In U. Hirt & B. Wälti (Hrsg.), *Lernumgebungen im Mathematikunterricht: Natürliche Differenzierung für Rechenschwäche bis Hochbegabte* (2. Aufl., S. 11–14). Seelze-Velber: Kallmeyer.

Hoffman, M., & Powell, A. (1989). Mathematical and commentary writing: Vehicles for student reflection and empowerment. *Mathematics Teaching*, 126, 55–57.
<https://doi.org/http://dx.doi.org/doi:10.7282/T3B56H46>

Hoffmann, A. (2011). „Ein Huhn und ein Schwein haben zusammen 6 Beine!“ Viertklässler entwickeln, nutzen und wählen Darstellungen aus für die Bearbeitung eines mathematischen Problems. *Grundschulunterricht Mathematik*, 2, 11–14.

Hohn, K. (2012). *Gegeben, Gesucht, Lösung? Selbstgenerierte Repräsentationen bei der Bearbeitung problemhaltiger Textaufgaben* (Dissertation). Universität Koblenz-Landau, Landau. Abgerufen von Deutsche Nationalbibliothek. (IDN 1028021070/34)

Holland, G. (2007). *Geometrie in der Sekundarstufe: Entdecken – Konstruieren – Deduzieren. Didaktische und methodische Fragen* (3. Aufl.). Hildesheim: Franzbecker.

Homish, G. G., Edwards, E. P., Eiden, R. D., & Leonard, K. E. (2010). Analyzing family data: A GEE approach for substance use researchers. *Addictive behaviors*, 35(6), 558–563.
<https://doi.org/10.1016/j.addbeh.2010.01.002>

Hörmann, H. (1994). *Meinen und Verstehen: Grundzüge einer psychologischen Semantik* (4. Aufl.). Frankfurt am Main: Suhrkamp.

Hoth, J. (2016). *Situationsbezogene Diagnosekompetenz von Mathematiklehrkräften: Eine Vertiefungsstudie zur TEDS-Follow-Up-Studie*. Wiesbaden: Springer.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2. Aufl.). New York, NY: Taylor & Francis.

Humphrey, G. (1963). *Thinking*. New York: Wiley Huppmann.

Hußmann, S., & Leuders, T. (2006). Hintergründe und Konzepte. In W. Brandt (Hrsg.), *Konzepte und Aufgaben zur Sicherung von Basiskompetenzen: mit Material zum Umgang mit den Kernlehrplänen Mathematik* (S. 67–89). Stuttgart: Klett.

Hussy, W. (1993). *Denken und Problemlösen* (Bd. 8). Stuttgart: Kohlhammer.

Hussy, W. (2013). Psychologie als eine empirische Wissenschaft. In W. Hussy, M. Schreier, & G. Echterhoff (Hrsg.), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (2. Aufl., S. 1–49). Berlin: Springer.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: SAGE.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581–592. <https://doi.org/10.1037/0003-066X.60.6.581>

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107(2), 139–155.

Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, 96(4), 284–309. <https://doi.org/10.1016/j.jecp.2006.09.001>

Jahnke, H. N. (1978). *Zum Verhältnis von Wissensentwicklung und Begründung in der Mathematik -- Beweisen als didaktisches Problem*. (Bd. 10). Bielefeld: Universität Bielefeld.

Jahnke, H. N., & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 331–355). Berlin: Springer.

Janssen, B. (2009). Kreative Methoden für einen lebendigen Mathematikunterricht. *Praxis der Mathematik in der Schule*, 51(30), 33–36.

Janssen, J., & Laatz, W. (2007). *Statistische Datenanalyse mit SPSS für Windows: Eine anwendungsorientierte Einführung in das Basisprogramm und das Modul Exakte Tests* (6. Aufl.). Berlin: Springer.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4(1), 71–115. https://doi.org/10.1207/s15516709cog0401_4

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.

Johnson-Laird, P. N. (2005). Mental models and thought. In K. J. Holyoak & R. G. Morrison (Hrsg.), *The Cambridge Handbook of Thinking and Reasoning* (S. 185–208). Cambridge: University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Erlbaum.

Johnson-Laird, P. N., & Garnham, A. (1980). Descriptions and discourse models. *Linguistics and Philosophy*, 3(3), 371–393. <https://doi.org/10.1007/BF00401691>

Jonkisz, E., Moosbrugger, H., & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebögen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 27–74). Berlin: Springer.

Jörissen, S., & Schmidt-Thieme, B. (2015). Darstellen und Kommunizieren. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 385–408). Berlin: Springer.

Kaiser, G., & Schwarz, I. (2009). Können Migranten wirklich nicht rechnen? Zusammenhänge zwischen mathematischer und allgemeiner Sprachkompetenz. *Friedrich Schülerheft Migration*, 68–69.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Aufl., S. 247–262). New York: Cambridge University Press.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, 22(3), 374–419. [https://doi.org/10.1016/0010-0285\(90\)90008-R](https://doi.org/10.1016/0010-0285(90)90008-R)

Käpnick, F. (1998). *Mathematisch begabte Kinder: Modelle, empirische Studien und Förderungsprojekte für das Grundschulalter*. Frankfurt am Main: Lang.

Käpnick, F. (2014). *Mathematiklernen in der Grundschule*. Berlin: Springer.

Kauffeld, S., & Grote, S. (2007). „Alles Leben ist Problemlösen“: Das Kasseler-Kompetenz-Raster. In E. Schäfer, M. Buch, I. Pahls, & J. Pfitzmann (Hrsg.), *Arbeitsleben! Arbeitsanalyse, Arbeitsgestaltung, Kompetenzentwicklung. Festschrift für Ekkehart Frieling* (S. 309–334). Kassel: Kassel University Press.

Kautschitsch, H. (1984). Der Videofilm – Geeignetes Mittel zur Visualisierung und Entwicklung mathematischer Begriffe und Modelle. In *Schriftenreihe Didaktik der Mathematik, Universität Klagenfurt* (Bd. 13, S. 59–98). Wien: Hölder Pichler Tempsky.

Kelava, A., & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 75–102). Berlin: Springer.

Kelle, U., & Kluge, S. (2010). *Vom Einzelfall zum Typus*. Wiesbaden: VS.

Kendeou, P., van den Brock, P., White, M. J., & Lynch, J. (2007). Comprehension in preschool and early elementary children: Skill development and strategy interventions. In D. S. McNamara (Hrsg.), *Reading comprehension strategies: Theories, interventions, and technologies* (S. 27–45). New York: Psychology Press.

Kershaw, T. C., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 3–13. <https://doi.org/10.1037/0278-7393.30.1.3>

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, 68(3), 350–386. <https://doi.org/10.3102/00346543068003350>

Khemlani, S., & Johnson-Laird, P. N. (2013). Cognitive changes from explanations. *Journal of Cognitive Psychology*, 25(2), 139–146. <https://doi.org/10.1080/20445911.2012.720968>

Kilpatrick, J. (1985). A retrospective account of the past twenty-five years of research on teaching mathematical problem solving. In E. A. Silver (Hrsg.), *Teaching and Learning Mathematical Problem Solving: Multiple Research Perspectives* (S. 1–15). Hillsdale, NJ: Erlbaum.

Kilpatrick, J., Swafford, J., & Findell, B. (Hrsg.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kindfield, A. C. H. (1993). Biology diagrams: Tools to think with. *The Journal of the Learning Sciences*, 3(1), 1–36.

King, A. (2007). Beyond literal comprehension: A strategy to promote deep understanding of text. In D. S. McNamara (Hrsg.), *Reading comprehension strategies: Theories, interventions, and technologies* (S. 267–290). New York: Psychology Press.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92(1), 109–129. <https://doi.org/10.1037/0033-295X.92.1.109>

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29(2), 133–159. [https://doi.org/10.1016/0749-596X\(90\)90069-C](https://doi.org/10.1016/0749-596X(90)90069-C)

Kirsch, A. (1979). Beispiele für „prämathematische“ Beweise. In W. Dörfler & R. Fischer (Hrsg.), *Beweisen im Mathematikunterricht: Vorträge des 2. Internationalen Symposiums für „Didaktik der Mathematik“ vom 26.09. bis 29.09.1978 in Klagenfurt* (S. 261–274). Wien: Hölder-Pilcher-Tempsky.

Kirsh, D. (2010). Thinking with external representations. *AI & Society*, 25(4), 441–454. <https://doi.org/10.1007/s00146-010-0272-8>

Kittel, A. (2013). Mathematische Interpretation ikonischer Darstellungen. In J. Sprenger, A. Wagner, & M. Zimmermann (Hrsg.), *Mathematik lernen, darstellen, deuten, verstehen. Didaktische Sichtweisen vom Kindergarten bis zur Hochschule*. (S. 109–118). Wiesbaden: Springer.

Klauer, K. J. (2001). Trainingsforschung: Ansätze – Theorien – Ergebnisse. In K. J. Klauer (Hrsg.), *Handbuch kognitives Training* (S. 3–68). Göttingen: Hogrefe.

Kline, S. L. (1998). Influence opportunities and the development of argumentation competencies in childhood. *Argumentation*, 12(3), 367–386. <https://doi.org/10.1023/A:1007780507894>

Klix, F. (1971). *Information und Verhalten: Kybernetische Aspekte der organismischen Informationsverarbeitung* (3. Aufl.). Berlin: VEB Deutscher Verlag der Wissenschaften.

KMK & IQB. (2013). Kompetenzstufenmodell zu den Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4). Abgerufen von https://www.iqb.hu-berlin.de/bista/ksm/KSM_GS_Mathemati_2.pdf (letzter Zugriff am 06.11.2016)

KMK: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2004). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (Jahrgangsstufe 10)*. München: Luchterhand. Abgerufen von http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluessel/2003/2003_12_04-Bildungsstandards-Mathe-Mittleren-SA.pdf (letzter Zugriff am 06.11.2016)

KMK: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2005a). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4)*. München: Luchterhand. Abgerufen von http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluessel/2004/2004_10_15-Bildungsstandards-Deutsch-Primar.pdf (letzter Zugriff am 06.11.2016)

KMK: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2005b). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4)*. München: Luchterhand. Abgerufen von http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluessel/2004/2004_10_15-Bildungsstandards-Mathe-Primar.pdf (zuletzt Zugriff am 06.11.2016)

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1534–1555.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, 29(7), 1000–1009.

Knoblich, G., & Öllinger, M. (2006). Einsicht und Umstrukturierung beim Problemlösen. In J. Funke (Hrsg.), *Enzyklopädie der Psychologie, Denken und Problemlösen* (Bd. 8, S. 1–85). Göttingen: Hogrefe.

Knoblich, G., & Wartenberg, F. (1998). Unbemerkte Lösungshinweise begünstigen Veränderungen der Problemrepräsentation. *Zeitschrift für Psychologie*, 206(3), 207–234.

Koerber, S. (2000). *Der Einfluss externer Repräsentationsformen auf proportionales Denken im Grundschulalter* (Dissertation). Technische Universität, Berlin. Abgerufen von <http://dx.doi.org/10.14279/depositonce-182> (letzter Zugriff am 06.11.2016)

Koffka, K. (1935). *Principles of gestalt psychology*. London: Routledge.

König, H. (1992). Einige für den Mathematikunterricht bedeutsame heuristische Vorgehensweisen. *Der Mathematikunterricht*, 38(3), 24–38.

Kosslyn, S. M. (1994). *Image and brain. The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Krägeloh, N., & Prediger, S. (2015). Der Textaufgabenknacker – Ein Beispiel zur Spezifizierung und Förderung fachspezifischer Lese- und Verstehensstrategien. *Der mathematische und naturwissenschaftliche Unterricht*, 68(3), 138–144.

Kramarski, B., Weisse, I., & Kololshi-Minsker, I. (2010). How can self-regulated learning support the problem solving of third-grade students with mathematics anxiety? *ZDM*, 42(2), 179–193.
<https://doi.org/10.1007/s11858-009-0202-8>

Krauthausen, G. (1994). *Arithmetische Fähigkeiten von Schulanfängern: Eine Computersimulation als Forschungsinstrument und als Baustein eines Softwarekonzeptes für die Grundschule*. Wiesbaden: Deutscher Universitäts-Verlag.

Krauthausen, G. (2001). Wann fängt das Beweisen an? Jedenfalls, ehe es einen Namen hat. In W. Weiser & B. Wollring (Hrsg.), *Beiträge zur Didaktik der Mathematik für die Primarstufe. Festschrift für Siegbert Schmidt* (S. 99–113). Hamburg: Dr. Kovač.

Krauthausen, G., & Scherer, P. (2010). *Umgang mit Heterogenität. Natürliche Differenzierung im Mathematikunterricht der Grundschule*. Handreichungen des Programms SINUS an Grundschulen, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Universität Kiel. Abgerufen von http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Handreichung_Krauthausen-Scherer.pdf (letzter Zugriff am 06.11.2016)

Krauthausen, G., & Scherer, P. (2011). Natürliche Differenzierung durch offene Aufgabenstellungen. Erfahrungen mit den substanzielien Aufgabenformaten Zahlenketten und Rechendreiecke. *Grundschulunterricht Mathematik*, 1, 4–7.

Krauthausen, G., & Scherer, P. (2014). *Einführung in die Mathematikdidaktik* (3. Aufl.). Berlin: Springer.

Krelle, M. (2014). *Mündliches Argumentieren in leistungsorientierter Perspektive: Eine empirische Analyse von Unterrichtsdiskussionen in der neunten Jahrgangsstufe*. Baltmannsweiler: Schneider.

Kremer, J. (2015). *Beeinflusst der Aufgabenkontext die Schwierigkeit von problemhaltigen Textaufgaben? Eine explorative Studie* (unveröffentlichte Bachelorarbeit). Universität Koblenz-Landau, Landau.

Krummheuer, G. (2003a). Argumentationsanalyse in der mathematikdidaktischen Unterrichtsforschung. *Zentralblatt für Didaktik der Mathematik*, 35(6), 247–256.

Krummheuer, G. (2003b). Wie wird Mathematiklernen im Unterricht der Grundschule zu ermöglichen versucht? – Strukturen des Argumentierens in alltäglichen Situationen des Mathematikunterrichts der Grundschule. *Journal für Mathematik-Didaktik*, 24(2), 122–138.

Krummheuer, G., & Brandt, B. (2001). *Paraphrase und Traduktion: partizipationstheoretische Elemente einer Interaktionstheorie des Mathematiklernens in der Grundschule*. Weinheim: Beltz.

Krummheuer, G., & Fetzer, M. (2005). *Der Alltag im Mathematikunterricht: Beobachten – Verstehen – Gestalten*. München: Elsevier.

Krumsdorf, J. (2009). Beweisen am Beispiel – beispielgebundenes Beweisen zwischen induktivem Prüfen und formalem Beweisen. *Praxis der Mathematik in der Schule*, 51(30), 8–13.

Kuckartz, U. (2014a). *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: Springer.

Kuckartz, U. (2014b). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (2. Aufl.). Weinheim: Beltz Juventa.

Kuntze, S. (2013). Vielfältige Darstellungen nutzen im Mathematikunterricht. In J. Sprenger, A. Wagner, & M. Zimmermann (Hrsg.), *Mathematik lernen, darstellen, deuten, verstehen. Didaktische Sichtweisen vom Kindergarten bis zur Hochschule*. (S. 17–37). Wiesbaden: Springer.

Kuntze, S., & Prediger, S. (2005). Ich schreibe, also denk' ich – Über Mathematik schreiben. *Praxis der Mathematik in der Schule*, 47(5).

Kürschner, C., & Schnottz, W. (2007). Konstruktion mentaler Repräsentationen bei der Verarbeitung von Text und Bildern. *Unterrichtswissenschaft*, 35(1), 48–67.

Kürschner, C., & Schnottz, W. (2008). Das Verhältnis gesprochener und geschriebener Sprache bei der Konstruktion mentaler Repräsentationen. *Psychologische Rundschau*, 59(3), 139–149. <https://doi.org/10.1026/0012-1924.59.3.139>

Küspert, P., & Schneider, K. (1998). *Würzburger Leise Leseprobe (WLLP)*. Göttingen: Hogrefe.

Lakatos, I. (1979). *Beweise und Widerlegungen: Die Logik mathematischer Entdeckungen*. Braunschweig: Vieweg.

Lampert, M. (1990). Connecting inventions with conventions. In L. P. Steffe & T. Wood (Hrsg.), *Transforming children's mathematics education: International perspectives* (S. 253–265). Hillsdale: Routledge.

Lange, D. (2014). Kooperationsarten in mathematischen Problemlöseprozessen. *Journal für Mathematik-Didaktik*, 35(2), 173–204. <https://doi.org/10.1007/s13138-014-0063-8>

Lauter, J. (1994). Formen des mündlichen Rechnens und ihre Bedeutung für den Arithmetikunterricht in der Grundschule (1–4). *Mathematische Unterrichtspraxis*, (1), 1–6.

Leiss, D., Schukajlow, S., Blum, W., Messner, R., & Pekrun, R. (2010). The role of the situation model in mathematical modelling – task analyses, student competencies, and teacher interventions. *Journal Für Mathematik-Didaktik*, 31(1), 119–141. <https://doi.org/10.1007/s13138-010-0006-y>

Lengnink, K. (2013). Prozesse beim Mathematiklernen initiieren und begleiten – vom Wert des Intersubjektiven. In M. Rathgeb, M. Helmerich, R. Krömer, K. Lengnink, & G. Nickel (Hrsg.), *Mathematik im Prozess* (S. 211–223). Wiesbaden: Springer.

Lenhard, W., & Schneider, W. (2006). *ELFE 1–6: Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.

Lester, F. K. (1985). Methodological considerations in research on mathematical problem-solving instruction. In E. A. Silver (Hrsg.), *Teaching and learning mathematical problem solving: Multiple research perspectives* (S. 41–69). Hillsdale, NJ: Erlbaum.

Lester, F. K., Garofalo, J., & Kroll, D. L. (1989a). Self-confidence, interest, beliefs, and metacognition: Key influences on problem-solving behavior. In D. B. McLeod & V. M. Adams (Hrsg.), *Affect and Mathematical Problem Solving* (S. 75–88). New York: Springer.

Lester, F. K., Garofalo, J., & Kroll, D. L. (1989b). *The role of metacognition in mathematical problem solving: A study of two grade seven classes. Final report.*

Lester, F. K., & Kroll, D. L. (1990). Assessing student growth in mathematical problem solving. In G. Kulm (Hrsg.), *Assessing higher order thinking in mathematics* (S. 53–70). Washington, DC: AAAS Publication.

Lewis, A. B. (1989). Training students to represent arithmetic word problems. *Journal of Educational Psychology*, 81(4), 521–531. <https://doi.org/10.1037/0022-0663.81.4.521>

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, 79(4), 363–371. <https://doi.org/10.1037/0022-0663.79.4.363>

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22. <https://doi.org/10.2307/2336267>

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.

Lindmeier, A. (2011). *Modeling and measuring knowledge and competencies of teachers: A threefold domain-specific structure model, exemplified for mathematics teachers, operationalized with computer- and video-based methods*. Münster: Waxmann.

Link, F. (2011). *Problemlöseprozesse selbstständigkeitsorientiert begleiten: Kontexte und Bedeutungen strategischer Lehrerinterventionen in der Sekundarstufe I*. Wiesbaden: Springer.

Link, M. (2012). *Grundschulkinder beschreiben operative Zahlenmuster: Entwurf, Erprobung und Überarbeitung von Unterrichtsaktivitäten als ein Beispiel für Entwicklungsforschung*. Wiesbaden: Springer.

Lompscher, J. (1975). Wesen und Strukturen allgemeiner Fähigkeiten. In J. Lompscher (Hrsg.), *Theoretische und experimentelle Untersuchungen zur Entwicklung geistiger Fähigkeiten* (2. Aufl., S. 17–73). Berlin: Volk und Wissen.

Lorenz, J. H. (2004). *Kinder entdecken die Mathematik*. Braunschweig: Westermann.

Lorenz, J. H. (2005). Kommunikation über Rechenwege – Nur mittels Sprache? In J. Engel, R. Vogel, & S. Wessolowski (Hrsg.), *Strukturieren – Modellieren – Kommunizieren: Leitbilder mathematischer und informatischer Aktivitäten; Festschrift für Karl-Dieter Klose, Siegfried Krauter, Herbert Löhne und Heinrich Wölpert* (S. 151–164). Hildesheim: Franzbecker.

Lorenz, J. H. (2006a). Muster erkennen. *Grundschule Mathematik*, 8, 20–21.

Lorenz, J. H. (2006b). Verschiedene Bereiche – gleiche Struktur. *Grundschule Mathematik*, 8, 44–45.

Lorenz, J. H. (2013). Zahlen und Rechenoperationen. Wie sind sie im Kopf des Lernenden? In J. Sprenger, A. Wagner, & M. Zimmermann (Hrsg.), *Mathematik lernen, darstellen, deuten, verstehen. Didaktische Sichtweisen vom Kindergarten bis zur Hochschule*. (S. 3–16). Wiesbaden: Springer.

Lorenz, J. H., & Radatz, H. (1993). *Handbuch des Förderns im Mathematikunterricht*. Hannover: Schroedel.

Lüken, M. M. (2012). *Muster und Strukturen im mathematischen Anfangsunterricht: Grundlegung und empirische Forschung zum Struktursinn von Schulanfängern*. Münster: Waxmann.

Luo, J., Niki, K., & Knoblich, G. (2006). Perceptual contributions to problem solving: Chunk decomposition of Chinese characters. *Brain Research Bulletin*, 70(4-6), 430–443.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention science: The official journal of the society for prevention research*, 1(4), 173–181.

Mac Lane, S. (1981). Mathematical models: A sketch for the philosophy of mathematics. *The American Mathematical Monthly*, 88(7), 462–472.

Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Hrsg.), *Reading comprehension strategies: Theories, interventions, and technologies* (S. 107–136). New York: Psychology Press.

Maier, H. (2000). Schreiben im Mathematikunterricht. *Mathematik lehren*, 99, 10–13.

Malle, G. (2002). Begründen. Eine vernachlässigte Tätigkeit im Mathematikunterricht. *Mathematik lehren*, 110, 4–8.

Mayer, R. E. (1983). *Thinking, problem solving, cognition*. New York: Worth Publishers.

Mayer, R. E. (1995). The search for insight: Grappling with gestalt psychology's unanswered questions. In R. J. Sternberg & J. E. Davidson (Hrsg.), *The Nature of insight* (S. 3–32). Cambridge: MIT Press.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 26(1-2), 49–63. <https://doi.org/10.1023/A:1003088013286>

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Aufl., S. 43–71). New York: Cambridge University Press.

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Hrsg.), *The nature of mathematical thinking* (S. 29–54). Mahwah, NJ: Lawrence Erlbaum.

Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Hrsg.), *Advances in the psychology of human intelligence* (Bd. 2, S. 231–273). Hillsdale, NJ: Erlbaum.

Mayring, P. (2002). *Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken* (5. Aufl.). Weinheim: Beltz.

Mayring, P. (2014). *Qualitative content analysis. Theoretical foundation, basic procedures and software solution*. Klagenfurt. Abgerufen von <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173> (zuletzt abgerufen am 06.11.2016)

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.

Mayring, P. (2016). *Einführung in die qualitative Sozialforschung: eine Anleitung zu qualitativem Denken* (6. Aufl.). Weinheim: Beltz.

Mayring, P., & Fenzl, T. (2014). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 543–556). Wiesbaden: Springer.

McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007). The 4-pronged comprehension strategy framework. In D. S. McNamara (Hrsg.), *Reading comprehension strategies: Theories, interventions, and technologies* (S. 465–496). New York: Psychology Press.

Meissner, H. (1979). Beweisen im Elementarbereich. In W. Dörfler & R. Fischer (Hrsg.), *Beweisen im Mathematikunterricht: Vorträge des 2. Internationalen Symposiums für „Didaktik der Mathematik“ vom 26.09. bis 29.09.1978 in Klagenfurt* (S. 307–314). Wien: Hölder-Pilcher-Tempsky.

Meyer, A. (2015). *Diagnose algebraischen Denkens: Von der Diagnose- zur Förderaufgabe mithilfe von Denkmustern*. Wiesbaden: Springer.

Meyer, A., & Fischer, A. (2013). Wie algebraische Symbolsprache die Möglichkeiten für algebraisches Denken erweitert – Eine Theorie symbolsprachlichen algebraischen Denkens. *Journal für Mathematik-Didaktik*, 34(2), 177–208. <https://doi.org/10.1007/s13138-013-0054-1>

Meyer, M. (2007a). *Entdecken und Begründen im Mathematikunterricht: Von der Abduktion zum Argument*. Hildesheim: Franzbecker.

Meyer, M. (2007b). Entdecken und Begründen im Mathematikunterricht – Zur Rolle der Abduktion und des Arguments. *Journal für Mathematik-Didaktik*, 28(3-4), 286–310. <https://doi.org/10.1007/BF03339350>

Meyer, M., & Prediger, S. (2009). Warum? – Argumentieren, Begründen, Beweisen. *Praxis der Mathematik in der Schule*, 51(30), 1–7.

Meyer, M., & Prediger, S. (2012). Sprachenvielfalt im Mathematikunterricht. Herausforderungen, Chancen und Förderansätze. *Praxis der Mathematik in der Schule*, 54(45), 2–9.

Meyer, M., & Voigt, J. (2009). Beweisen durch Entdecken. *Praxis der Mathematik in der Schule*, 51(30), 14–20.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
<https://doi.org/10.1037/h0043158>

Miller, M. (1986). *Kollektive Lernprozesse: Studien zur Grundlegung einer soziologischen Lerntheorie*. Frankfurt am Main: Suhrkamp.

Moore, R. T. (2012). Multivariate continuous blocking to improve political science experiments. *Political Analysis*, 20(4), 460–479. <https://doi.org/10.1093/pan/mps025>

Moore, R. T. (2015). Overcoming barriers to heterogeneous-group learning in the political science classroom. *PS: Political Science & Politics*, 48(01), 149–156.
<https://doi.org/10.1017/S1049096514001693>

Moore, R. T., & Moore, S. A. (2013). Blocking for sequential political experiments. *Political Analysis*, 507–523. <https://doi.org/10.1093/pan/mpt007>

Moore, R. T., & Schnakenberg, K. (2015). blockTools: Block, assign, and diagnose potential interference in randomized experiments [CRAN Package]. Abgerufen 14. Februar 2016, von <https://cran.r-project.org/web/packages/blockTools/index.html>

Moosbrugger, H. (2012). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 103–117). Berlin: Springer.

Moosbrugger, H., & Schermelleh-Engel, K. (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 325–343). Berlin: Springer.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386. <https://doi.org/10.1177/1094428106291059>

Moser Opitz, E. (2013). *Rechenschwäche/Dyskalkulie: Theoretische Klärungen und empirische Studien an betroffenen Schülerinnen und Schülern* (2. Aufl.). Bern: Haupt.

Muir, T., Beswick, K., & Williamson, J. (2008). “I’m not very good at solving problems”: An exploration of students’ problem solving behaviours. *The Journal of Mathematical Behavior*, 27(3), 228–241. <https://doi.org/10.1016/j.jmathb.2008.04.003>

Murray, M. A., & Byrne, R. M. J. (2013). Cognitive change in insight problem solving: Initial model errors and counterexamples. *Journal of Cognitive Psychology*, 25(2), 210–219.
<https://doi.org/10.1080/20445911.2012.743986>

Myers, A., & Hansen, C. H. (2012). *Experimental Psychology* (7. Aufl.). Belfort, CA: Wadsworth.

Naujoks, N. (2002). Formen von Schülerkooperation aus der Perspektive Interpretativer Unterrichtsforschung. In G. Breidenstein, A. Combe, W. Helsper, & B. Stelmaszyk (Hrsg.), *Forum qualitative Schulforschung 2: Interpretative Unterrichts- und Schulbegleitforschung* (S. 61–80). Opladen: Leske + Budrich.

NCTM: National Council of Teachers of Mathematics (Hrsg.). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Nesher, P., Greeno, J. G., & Riley, M. S. (1982). The development of semantic categories for addition and subtraction. *Educational Studies in Mathematics*, 13(4), 373–394. <https://doi.org/10.1007/BF00366618>

Neumann, A., Beier, F., & Ruwisch, S. (2014). Schriftliches Begründen im Mathematikunterricht. *Zeitschrift für Grundschulforschung*, 7(1), 113–125.

Neumann, A., & Ruwisch, S. (2015). Sprachliche Prozeduren beim fachlichen Lernen: Begründen in Mathematik. In J. Kilian & J. Eckhoff (Hrsg.), *Deutscher Wortschatz – beschreiben, lernen, lehren: Beiträge zur Wortschatzarbeit in Wissenschaft, Sprachunterricht, Gesellschaft* (Bd. 13, S. 313–330). Frankfurt am Main: Peter Lang.

Neumeister, N., & Vogt, R. (2012). Erklären im Unterricht. In M. Becker-Mrotzek (Hrsg.), *Mündliche Kommunikation und Gesprächsdidaktik* (2. Aufl., Bd. 3, S. 562–583). Baltmannsweiler: Schneider.

Newell, A., Shaw, J. C., & Simon, H. A. (1962). The processes of creative thinking. In H. E. Gruber, M. Wertheimer, & G. Terrell (Hrsg.), *Contemporary approaches to creative thinking: A symposium held at the University of Colorado* (S. 63–119). New York: Atherton.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. New Jersey: Prentice-Hall.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Cambridge, MA: Perseus Books.

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak & R. G. Morrison (Hrsg.), *The Cambridge handbook of thinking and reasoning* (S. 321–349). Cambridge: University Press.

Nührenbörger, M., & Schwarzkopf, R. (2010). Mathematische Denkprozesse von Kindern. In C. Böttlinger, K. Bräuning, M. Nührenbörger, R. Schwarzkopf, & E. Söbbeke (Hrsg.), *Mathematik im Denken der Kinder: Anregungen zur mathematikdidaktischen Reflexion* (S. 8–16). Seelze: Klett.

Nussbaumer, M. (1995). *Argumentation und Argumentationstheorie* (Bd. 12). Heidelberg: Groos.

Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait–multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical and Statistical Psychology*, 59(1), 195–213. <https://doi.org/10.1348/000711005X67490>

Oakhill, J., & Cain, K. (2007). Issues of causality in children's reading comprehension. In D. S. McNamara (Hrsg.), *Reading comprehension strategies: Theories, interventions, and technologies* (S. 47–71). New York: Psychology Press.

Ohlsson, S. (1984a). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, 25(2), 117–129. <https://doi.org/10.1111/j.1467-9450.1984.tb01005.x>

Ohlsson, S. (1984b). Restructuring revisited. I. Summary and critique of the gestalt theory of problem solving. *Scandinavian Journal of Psychology*, 25(1), 65–78. <https://doi.org/10.1111/j.1467-9450.1984.tb01001.x>

Ohlsson, S. (1990). The mechanism of restructuring in geometry. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (S. 237–244). Hillsdale, NJ: Lawrence Erlbaum.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. Keane & K. Gilhooly (Hrsg.), *Advances in the psychology of thinking* (S. 1–44). London: Harvester-Wheatsheaf.

Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. New York: Cambridge University Press.

Ohlsson, S. (2012). The problems with problem solving: Reflections on the rise, current status, and possible future of a cognitive research paradigm. *The Journal of Problem Solving*, 5(1). <https://doi.org/10.7771/1932-6246.1144>

Öllinger, M., Jones, G., Faber, A. H., & Knoblich, G. (2013). Cognitive mechanisms of insight: The role of heuristics and representational change in solving the eight-coin problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 931–939. <https://doi.org/10.1037/a0029194>

Öllinger, M., Jones, G., & Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology*, 55(4), 270–282. <https://doi.org/10.1027/1618-3169.55.4.270>

Öllinger, M., Jones, G., & Knoblich, G. (2014). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychological Research*, 78(2), 266–275. <https://doi.org/10.1007/s00426-013-0494-8>

Oswald, H. (1994). Gruppen- und Freundschaftsbeziehungen. *Grundschule*, 26(4), 10–12.

Ott, B. (2014). Kinder zeichnen zu Textaufgaben - Vorstellung eines Instruments zur Analyse graphischer Darstellungen. In J. Ames & J. Roth (Hrsg.), *Beiträge zum Mathematikunterricht 2014: Beiträge zur 48. Jahrestagung der Gesellschaft für Didaktik der Mathematik vom 10.–14. März in Koblenz* (Bd. 2, S. 879–882). Münster: WTM.

Ott, B. (2015). Qualitative Analyse grafischer Darstellungen zu Textaufgaben – eine Untersuchung von Kinderzeichnungen in der Primarstufe. In G. Kadunz (Hrsg.), *Semiotische Perspektiven auf das Lernen von Mathematik* (S. 163–182). Berlin: Springer.

Palmer, S. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Hrsg.), *Cognition and categorization* (S. 259–303). Hillsdale, NJ: Erlbaum.

Pantziara, M., Gagatsis, A., & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics*, 72(1), 39–60. <https://doi.org/10.1007/s10649-009-9181-5>

Papert, S. (1993). *The children's machine: rethinking school in the age of the computer*. New York: BasicBooks.

Pauen, S., Pahnke, J., & Valentiner, I. (2007). *Erfassung kognitiver Kompetenzen im Vorschul- bis Jugendalter: Intelligenz, Sprache und schulische Fertigkeiten. Empfehlungen zum Ausbau des Erhebungsinstrumentariums über Kinder im Sozio-oekonomischen Panel (SOEP); Expertise*. Berlin: DIW.

Pearson, E. S., & Hartley, H. O. (1958). *Biometrika tables for statisticians* (3. Aufl., Bd. 1). New York: Cambridge University Press.

Pehkonen, E. (1995). *Pupils' view of mathematics: Initial report for an international comparison project*. Helsinki: University of Helsinki. Abgerufen von ERIC Database. (ED419712)

Pehkonen, E. (2001). Offene Probleme: Eine Methode zur Entwicklung des Mathematikunterrichts. *Der Mathematikunterricht. Beiträge zu seiner wissenschaftlichen und methodischen Gestaltung*, 47(6), 60–72.

Pehkonen, E., Näveri, L., & Laine, A. (2013). On teaching problem solving in school mathematics. *CEPS Journal*, 3(4), 9–23.

Peirce, C. S. (1906). Prolegomena to an apology for pragmaticism. *The Monist*, 16(4), 492–546. <https://doi.org/10.2307/27899680>

Peirce, C. S. (1932). *Collected papers of Charles Sanders Peirce. Elements of Logic* (Bd. 2). Bristol: Thoemmes Press.

Peirce, C. S. (1933). *Collected papers of Charles Sanders Peirce. The Simplest Mathematics* (Bd. 4). Bristol: Thoemmes Press.

Perelman, C. (1970). *Le champ de l'argumentation*. Brüssel: Presses universitaires de Bruxelles.

Perels, F., Gürtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction*, 15(2), 123–139. <https://doi.org/10.1016/j.learninstruc.2005.04.010>

Perfetti, C. A. (2001). Reading skills. In N. J. Smelser & P. B. Baltes (Hrsg.), *International encyclopedia of the social & behavioral sciences* (S. 12800–12805). Oxford: Pergamon.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>

Petermann, F., & Petermann, U. (2011). *WISC-IV. Wechsler Intelligence Scale for Children* (4. Aufl.). Frankfurt: Pearson.

Peterson, D. (1996). Introduction. In D. Peterson (Hrsg.), *Forms of representation: An interdisciplinary theme for cognitive science* (S. 7–27). Exeter: Intellect.

Peterßen, K. (2012). *Begründungskultur im Mathematikunterricht der Grundschule: Eine Untersuchung der Lehrer zu ihren Vorstellungen vom Begründen und einer begründungsfördernden Unterrichtsgestaltung*. Hildesheim: Franzbecker.

Petko, D., Waldis, M., Pauli, C., & Reusser, K. (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. *Zentralblatt für Didaktik der Mathematik*, 35(6), 265–280. <https://doi.org/10.1007/BF02656691>

Philipp, K. (2013). *Experimentelles Denken: Theoretische und empirische Konkretisierung einer mathematischen Kompetenz*. Wiesbaden: Springer.

Piaget, J. (1972). *Sprechen und Denken des Kindes*. Düsseldorf: Pädagogischer Verlag Schwann.

Pickert, G. (1957). Die Bedeutung der Anschauung für den mathematischen Beweis. *Der Mathematikunterricht. Beiträge zu seiner wissenschaftlichen und methodischen Gestaltung*, 3(4), 49–62.

Plötzner, R., & Fehse, E. (1998). Learning from explanations: Extending one's own knowledge during collaborative problem solving by attempting to understand explanations received from others. *International Journal of Artificial Intelligence in Education*, 9, 193–218.

Pöhler, B. (2014). Umgang mit Prozentaufgaben – Herausforderungen für konzeptuelles Verständnis und Leseverständnis. In J. Ames & J. Roth (Hrsg.), *Beiträge zum Mathematikunterricht 2014: Beiträge zur 48. Jahrestagung der Gesellschaft für Didaktik der Mathematik vom 10.–14. März in Koblenz* (Bd. 2, S. 923–926). Münster: WTM.

Pólya, G. (1962). *Mathematik und plausibles Schliessen: Induktion und Analogie in der Mathematik*. Basel: Birkhäuser.

Pólya, G. (1966). *Vom Lösen mathematischer Aufgaben. Einsicht und Entdeckung, Lernen und Lehren* (Bd. I). Basel: Birkhäuser.

Prediger, S. (2013a). Darstellungen, Register und mentale Konstruktion von Bedeutungen und Beziehungen – mathematikspezifische sprachliche Herausforderungen identifizieren und bearbeiten. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 167–184). Münster: Waxmann.

Prediger, S. (2013b). Sprachmittel für mathematische Verstehensprozesse – Einblicke in Probleme, Vorgehensweisen und Ergebnisse von Entwicklungsforschungsstudien. In A. Pallack (Hrsg.), *Impulse für eine zeitgemäße Mathematiklehrer-Ausbildung: 16. Fachleiterstagung Mathematik; 25. - 27. September 2013, Hessisches Institut für Lehrerbildung, Reinhardswaldschule, Fulda* (S. 26–36). Neuss: Seeberger. Abgerufen von <http://www.mnu.de/images/PDF/fachbereiche/mathematik/16-fachleiterstagung.pdf> (letzter Zugriff am 06.11.2016)

Prediger, S., Freesemann, O., Moser Opitz, E., & Hußmann, S. (2013). Unverzichtbare Verstehensgrundlagen statt kurzfristiger Reparatur – Förderung bei mathematischen Lernschwierigkeiten in Klasse 5. *Praxis der Mathematik in der Schule*, 55(51), 12–17.

Prediger, S., Hußmann, S., Leuders, T., & Barzel, B. (2011). „Erst mal alle auf einen Stand bringen....“ Diagnosegeleitete und individualisierte Aufarbeitung arithmetischen Basiskönnens. *Pädagogik*, 63(5), 20–24.

Prediger, S., Renk, N., Büchter, A., Gürsoy, E., & Benholz, C. (2013). Family background or language disadvantages? Factors for underachievement in high stakes tests. In A. Lindmeier & A. Heinze (Hrsg.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 4, S. 49–56). Kiel: PME.

Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (Hrsg.). (2007). *Pisa 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.

Pressley, M., & Harris, K. R. (1994). Increasing the quality of educational intervention research. *Educational Psychology Review*, 6(3), 191–208. <https://doi.org/10.1007/BF02213181>

Rammstedt, B. (2010). Reliabilität, Validität, Objektivität. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 239–258). Wiesbaden: Springer.

Ramsey, F. L., & Schafer, D. W. (2013). *The statistical sleuth: a course in methods of data analysis* (3. Aufl.). Boston: Brooks/Cole, Cengage Learning.

Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2014). *Quantitative Methoden 2: Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.

Rasch, R. (2001a). Kinder mit unterschiedlicher mathematischer Leistungsfähigkeit lernen in (m)einer Klasse. *Grundschulunterricht*, 48(11), 2–3.

Rasch, R. (2001b). *Zur Arbeit mit problemhaften Textaufgaben im Mathematikunterricht der Grundschule: Eine Studie zu Herangehensweisen von Grundschulkindern an anspruchsvolle Textaufgaben und Schlussfolgerungen für eine Unterrichtsgestaltung, die entsprechende Lösungsfähigkeiten fördert*. Hildesheim: Franzbecker.

Rasch, R. (2005). Anspruchsvolle Textaufgaben im Mathematikunterricht der Grundschule. In J. Engel, R. Vogel, & S. Wessolowski (Hrsg.), *Strukturieren – Modellieren – Kommunizieren: Leitbilder mathematischer und informatischer Aktivitäten; Festschrift für Karl-Dieter Klose, Siegfried Krauter, Herbert Löthe und Heinrich Wölpert* (S. 93–104). Hildesheim: Franzbecker.

Rasch, R. (2008a). *42 Denk- und Sachaufgaben. Wie Kinder mathematische Aufgaben lösen und diskutieren* (3. Aufl.). Seelze-Velber: Kallmeyer.

Rasch, R. (2008b). Offene Aufgaben im Mathematikunterricht der Grundschule. In I. Esslinger-Hinz & H. Hahn (Hrsg.), *Kompetenzen entwickeln. Unterrichtsqualität in der Grundschule steigern: Entwicklungslinien und Forschungsbefunde* (2. Aufl., S. 143–150). Baltmannsweiler: Schneider.

Rasch, R. (2009). Textaufgaben in der Grundschule. Lernvoraussetzungen und Konsequenzen für den Unterricht. *Mathematica didactica*, 32, 67–92.

Rasch, R. (2012a). Herausforderung Textaufgaben. Individuelles Lernen bei Arbeiten mit Textaufgaben begleiten. *Grundschule*, 10, 6–9.

Rasch, R. (2012b). Individuelle Lernprozesse beim Arbeiten mit Sachaufgaben begleiten. In E. C. Wittmann, G. N. Müller, & C. Selter (Hrsg.), *Zahlen, Muster und Strukturen: Spielräume für aktives Lernen und Üben* (S. 130–136). Stuttgart: Klett.

Rasch, R. (2015a). Problemhafte Textaufgaben im Mathematikunterricht der Grundschule. In R. Rink (Hrsg.), *Von Guten Aufgaben bis Skizzen Zeichnen: Zum Sachrechnen im Mathematikunterricht der Grundschule* (S. 203–216). Baltmannsweiler: Schneider.

Rasch, R. (2015b). Von der Routineaufgabe bis zur Fermi-Aufgabe. *Die Grundschulzeitschrift*, 29(283/284), 31–35.

Rasch, R. (2016). *Textaufgaben für Grundschulkinder zum Denken und Knobeln. Mathematische Probleme lösen – Strategien entwickeln*. Seelze: Kallmeyer.

Rasch, R., & Schütte, S. (2008). Zahlen und Operationen. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer, & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule: Mathematik konkret* (S. 66–88). Berlin: Cornelsen Scriptor.

Rathgeb-Schnierer, E. (2005). Kommunikation als zentrales Element im Mathematikunterricht – Kinder artikulieren Entdeckungen und Lösungswege. In J. Engel, R. Vogel, & S. Wessolowski (Hrsg.), *Strukturieren – Modellieren – Kommunizieren: Leitbilder mathematischer und informatischer Aktivitäten; Festschrift für Karl-Dieter Klose, Siegfried Krauter, Herbert Löthe und Heinrich Wölpert* (S. 165–176). Hildesheim: Franzbecker.

Rathgeb-Schnierer, E. (2006). *Kinder auf dem Weg zum flexiblen Rechnen: Eine Untersuchung zur Entwicklung von Rechenwegen bei Grundschulkindern auf der Grundlage offener Lernangebote und eigenständiger Lösungsansätze*. Hildesheim: Franzbecker.

Rehm, M. (1990). Zum „beispiel-“ und „repräsentantengebundenen“ Beweisen im Mathematikunterricht. In M. Glatfeld (Hrsg.), *Finden, Erfinden, Lernen: zum Umgang mit Mathematik unter heuristischem Aspekt* (S. 95–111). Frankfurt am Main: Peter Lang Publishing.

Reisberg, D. (1987). External representations and the advantages of externalizing one's thought. In *The Ninth Annual Conference of the Cognitive Science Society* (S. 281–293). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reiss, K., Hellmich, F., & Thomas, J. (2002). Individuelle und schulische Bedingungsfaktoren für Argumentationen und Beweise im Mathematikunterricht. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 51–64). Weinheim: Beltz.

Reiss, K., & Törner, G. (2007). Problem solving in the mathematics classroom: The German perspective. *ZDM*, 39(5–6), 431–441. <https://doi.org/10.1007/s11858-007-0040-5>

Reiss, K., & Ufer, S. (2009). Was macht mathematisches Arbeiten aus? *Jahresbericht der Deutschen Mathematiker-Vereinigung (DMV)*, 111(4), 155–177.

Rellensmann, J., Schukajlow, S., & Leopold, C. (2015). Gute Skizze – Bessere Lösung? Selbsterstellte Skizzen beim mathematischen Modellieren. In H. Linneweber-Lammerkitten, F. Caluori, & C. Streit (Hrsg.), *Beiträge zum Mathematikunterricht 2015: Beiträge zur 49. Jahrestagung der Gesellschaft für Didaktik der Mathematik vom 09.–13. Februar in Basel* (S. 732–735). Münsster: WTM.

Renkl, A., & Nückles, M. (2006). Lernstrategien der externen Visualisierung. In H. Mandl (Hrsg.), *Handbuch Lernstrategien* (S. 135–147). Göttingen: Hogrefe.

Renkl, A., & Stern, E. (1994). Die Bedeutung von kognitiven Eingangsvoraussetzungen und schulischen Lerngelegenheiten für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für Pädagogische Psychologie*, 8(1), 27–40.

Reusser, K., & Reusser-Weyeneth, M. (1994). *Verstehen. Psychologischer Prozess und didaktische Aufgabe*. Bern: Huber.

Reuter, T. (2016). *Die Rolle externer Repräsentationen für die Konstruktion und Nutzung mentaler Modelle bei der Lösung problemhaltiger Textaufgaben in der Primarstufe* (Dissertation). Universität Koblenz-Landau, Landau. Abgerufen von Deutsche Nationalbibliothek (IDN 1110895682)

Rezat, S., & Rezat, S. (2012). Mathematische Sachaufgaben verstehen und erklären. *Praxis Deutsch*, 39(233), 36–41.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. Ginsburg (Hrsg.), *The development of mathematical thinking* (S. 153–196). New York: Academic Press. Abgerufen von ERIC Database. (ED252410)

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, 101(3), 529–544. <https://doi.org/10.1037/a0014224>

Robertson, I. (2001). *Problem solving*. Hove: Psychology Press.

Rohrbeck, C. A., Ginsburg-Block, M. R., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, 95(2), 240–257.

Roos, S. (2014). Was sind gute Darstellungen? Kinder gestalten aktiv ihren Lernweg. *Grundschule Mathematik*, 41, 36–39.

Roppelt, A., & Reiss, K. (2012). Beschreibung der im Fach Mathematik untersuchten Kompetenzen. In P. Stanat, H. A. Pant, K. Böhme, & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik: Ergebnisse des IQB-Ländervergleichs 2011* (S. 34–48). Münster: Waxmann.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, 2(3), 235–276.

Rösch, H., & Paetsch, J. (2011). Sach- und Textaufgaben im Mathematikunterricht als Herausforderung für mehrsprachige Kinder. In S. Prediger & E. Özil (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit. Stand und Perspektiven der Forschung und Entwicklung in Deutschland* (S. 55–76). Münster: Waxmann.

Rosseel, Y. (2013). The lavaan tutorial. Abgerufen 21. April 2016, von http://dornsifecms.usc.edu/assets/sites/210/docs/GC3/lavaan_tutorial.pdf

Rost, J. (2000). Allgemeine Standards für die Evaluationsforschung. In W. Hager, J.-L. Patry, & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionen. Standards und Kriterien: Ein Handbuch* (S. 129–140). Bern: Hans Huber.

Ruf, U., & Gallin, P. (1998). *Dialogisches Lernen in Sprache und Mathematik. Band II: Spuren legen - Spuren lesen. Unterricht mit Kernideen und Reisetagebüchern*. Seelze-Velber: Kallmeyer.

Ruf, U., & Gallin, P. (2005). *Dialogisches Lernen in Sprache und Mathematik. Band I: Austausch unter Ungleichen. Grundzüge einer interaktiven und fächerübergreifenden Didaktik* (3. Aufl.). Seelze-Velber: Kallmeyer.

Ruwisch, S. (2003). Gute Aufgaben im Mathematikunterricht der Grundschule – Eine Einführung. In S. Ruwisch & A. Peter-Koop (Hrsg.), *Gute Aufgaben im Mathematikunterricht der Grundschule* (2. Aufl., S. 5–14). Offenburg: Mildenberger.

Ruwisch, S. (2014). Herausforderungen im Mathematikunterricht der Grundschule durch veränderte sprachliche Anforderungen. In M. Michalak (Hrsg.), *Sprache als Lernmedium im Fachunterricht. Theorien und Modelle für das sprachbewusste Lehren und Lernen*. (S. 34–52). Baltmannsweiler: Schneider.

Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, 66(2), 201–223.
<https://doi.org/10.1111/j.2044-8317.2012.02049.x>

Scheld, J., & Hahn, K. (2011). „Das sind ja viel zu viele Zahlen!“ Fehlerprovokierende Darstellungsformen von Sachaufgaben helfen den Kindern, geeignete Rechenstrategien zu entwickeln. *Grundschulunterricht Mathematik*, 2, 31–33.

Scherer, P. (2014). Low achievers' understanding of place value – Materials, representations, and consequences for instruction. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen – Using Tools for Learning Mathematics and Statistics* (S. 43–56). Wiesbaden: Springer.

Scherer, P., & Steinbring, H. (2001). Strategien und Begründungen an Veranschaulichungen – statische und dynamische Deutungen. In W. Weiser & B. Wollring (Hrsg.), *Beiträge zur Didaktik der Mathematik für die Primarstufe. Festschrift für Siegbert Schmidt* (S. 188–201). Hamburg: Dr. Kovač.

Schermelleh-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 119–141). Berlin: Springer.

Schipper, W. (2005). Rücksicht auf die Bildungsstandards im Fach Mathematik – Jahrgangsstufe 4. In J. Engel, R. Vogel, & S. Wessolowski (Hrsg.), *Strukturieren – Modellieren – Kommunizieren: Leitbilder mathematischer und informatischer Aktivitäten; Festschrift für Karl-Dieter Klose, Siegfried Krauter, Herbert Löthe und Heinrich Wölpert* (S. 351–360). Hildesheim: Franzbecker.

Schipper, W. (2013). *Handbuch für den Mathematikunterricht an Grundschulen* (3. Aufl.). Braunschweig: Schroedel.

Schnotz, W. (1988). Textverständnis als Aufbau mentaler Modelle. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie* (S. 299–330). München: Psychologie-Verlag-Union.

Schnotz, W. (1994a). *Aufbau von Wissensstrukturen: Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten*. Weinheim: Beltz.

Schnotz, W. (1994b). Wissenserwerb mit logischen Bildern. In B. Weidenmann (Hrsg.), *Wissenserwerb mit Bildern: instruktionale Bilder in Printmedien, Film/Video und Computerprogrammen* (S. 95–147). Bern: Huber.

Schnotz, W. (2002). Wissenserwerb mit Texten, Bildern und Diagrammen. In L. J. Issing (Hrsg.), *Information und Lernen mit Multimedia und Internet: Lehrbuch für Studium und Praxis* (3. Aufl., S. 65–82). Weinheim: Beltz.

Schnotz, W. (2009). *Pädagogische Psychologie kompakt*. Weinheim: Beltz.

Schnotz, W. (2014). The integrated model of text and picture comprehension. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Aufl., S. 72–103). New York: Cambridge University Press.

Schnotz, W., Baadte, C., Müller, A., & Rasch, R. (2011). Kreatives Denken und Problemlösen mit bildlichen und beschreibenden Repräsentationen. In K. Sachs-Hombach & R. Totzke (Hrsg.), *Bilder – Sehen – Denken. Zum Verhältnis von begrifflich-philosophischen und empirisch-psychologischen Ansätzen in der bildwissenschaftlichen Forschung* (S. 204–252). Köln: Hämmel.

Schnotz, W., & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Experimental Psychology*, 46(3), 217–236. <https://doi.org/10.1026/0949-3964.46.3.217>

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)

Schnotz, W., & Dutke, S. (2004). Kognitionspsychologische Grundlagen der Lesekompetenz: Mehrebenenverarbeitung anhand multipler Informationsquellen. In U. Schiefele, C. Artelt, W. Schneider, & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz* (S. 61–99). Wiesbaden: VS.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19(4), 469–508. <https://doi.org/10.1007/s10648-007-9053-4>

Schoenfeld, A. H. (1985). *Mathematical problem solving*. San Diego: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 334–370). New York: Macmillan.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 484–494. <https://doi.org/10.1037/0278-7393.8.5.484>

Schreiber, C. (2014). Beine, Striche, Zahlen. Welche Darstellungen entwerfen Kinder? *Grundschule Mathematik*, 41, 6–9.

Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. Abgerufen 11. Juni 2016, von <http://www.qualitative-research.net/index.php/fqs/rt/prINTERFRIENDLY/2043/3635#gref>

Schukajlow, S. (2011). *Mathematisches Modellieren*. Münster: Waxmann.

Schukajlow, S., & Blum, W. (2011). Zur Rolle von multiplen Lösungen in einem kompetenzorientierten Mathematikunterricht. In K. Eilerts, A. H. Hilligus, G. Kaiser, & P. Bender (Hrsg.), *Kompetenzorientierung in Schule und Lehrerbildung. Perspektiven der bildungspolitischen Diskussion, der Bildungsforschung und der Mathematik-Didaktik. Festschrift für Hans-Dieter Rinckens* (S. 249–267). Berlin: LIT.

Schultz, D. P. (1969). *A history of modern psychology* (3. Aufl.). New York: Academic Press.

Schumacher, S. (2014). Darstellungskompetenz – Ein Schlüssel zum forschenden Lernen?! In J. Ames & J. Roth (Hrsg.), *Beiträge zum Mathematikunterricht 2014: Beiträge zur 48. Jahrestagung der Gesellschaft für Didaktik der Mathematik vom 10.–14. März in Koblenz* (Bd. 2, S. 1123–1126). Münster: WTM.

Schumann, S. (2012). *Repräsentative Umfrage: Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren*. München: Oldenbourg.

Schütte, S. (2002). Das Lernpotenzial mathematischer Gespräche nutzen. *Grundschule*, 34(3), 16–18.

Schütte, S. (2004). Rechenwegnotation und Zahlenblick als Vehikel des Aufbaus flexibler Rechenkompetenzen. *Journal für Mathematik-Didaktik*, 25(2), 130–148.
<https://doi.org/10.1007/BF03338998>

Schütte, S. (2008a). Mathematikunterricht zwischen Offenheit und Lenkung – Zum Verständnis von Konstruktion und Instruktion bei mathematischen Lernprozessen. In I. Esslinger-Hinz & H. Hahn (Hrsg.), *Kompetenzen entwickeln. Unterrichtsqualität in der Grundschule steigern: Entwicklungslinien und Forschungsbefunde* (2. Aufl., S. 135–142). Baltmannsweiler: Schneider.

Schütte, S. (2008b). *Qualität im Mathematikunterricht der Grundschule sichern: Für eine zeitgemäße Unterrichts- und Aufgabenkultur*. München: Oldenbourg.

Schwarzkopf, R. (2000). *Argumentationsprozesse im Mathematikunterricht – Theoretische Grundlagen und Fallstudien*. Hildesheim: Franzbecker.

Schwippert, K., Bos, W., & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther, & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 265–302). Münster: Waxmann.

Segal, E. (2004). Incubation in insight problem solving. *Creativity Research Journal*, 16(1), 141–148.

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Leutner, & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 298–306). Weinheim: Beltz.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499. <https://doi.org/10.3102/0034654307310317>

Selter, C. (1994). *Eigenproduktionen im Arithmetikunterricht der Primarstufe: Grundsätzliche Überlegungen und Realisierungen in einem Unterrichtsversuch zum multiplikativen Rechnen im zweiten Schuljahr*. Wiesbaden: Deutscher Universitätsverlag.

Selter, C. (2005). Bildungsstandards, zentrale Lernstandserhebungen und Unterrichtsqualität. In J. Engel, R. Vogel, & S. Wessolowski (Hrsg.), *Strukturieren – Modellieren – Kommunizieren: Leitbilder mathematischer und informatischer Aktivitäten; Festschrift für Karl-Dieter Klose, Siegfried Krauter, Herbert Löthe und Heinrich Wölpert* (S. 361–371). Hildesheim: Franzbecker.

Semadeni, Z. (1974). *The concept of premathematics as a theoretical background for primary mathematics teaching*. Warschau: Polnische Akademie der mathematischen Wissenschaft.

Semadeni, Z. (1984). Action proofs in primary mathematics teaching and in teacher training. *For the Learning of Mathematics*, 4(1), 32–34.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2. Aufl.). Wadsworth: Cengage Learning.

Simon, H. A. (1996). *The sciences of the artificial* (3. Aufl.). Cambridge, MA: MIT Press.

Slavin, R. E. (1993). Kooperatives Lernen und Leistung: Eine empirisch fundierte Theorie. In G. L. Huber (Hrsg.), *Neue Perspektiven der Kooperation: Ausgewählte Beiträge der Internationalen Konferenz 1992 über Kooperatives Lernen: Veranstaltet von der Internationalen Gesellschaft für Gruppenarbeit in der Erziehung in Utrecht, Holland, vom 1.-4. Juli 1992* (Bd. 6, S. 151–170). Baltmannsweiler: Schneider.

Smith, R. L., Ager, J. W., & Williams, D. L. (1992). Suppressor variables in multiple regression/correlation. *Educational and Psychological Measurement*, 52(1), 17–29. <https://doi.org/10.1177/001316449205200102>

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles: Sage Publications.

Söbbeke, E. (2005). *Zur visuellen Strukturierungsfähigkeit von Grundschulkindern: Epistemologische Grundlagen und empirische Fallstudien zu kindlichen Strukturierungsprozessen mathematischer Anschauungsmittel*. Hildesheim: Franzbecker.

Söbbeke, E. (2009). Welche Faktoren beeinflussen eine strukturorientiert relationale Deutung von Anschauungsmitteln? Ansätze zur Erhebung möglicher Rahmungen bei der Interpretation von Anschauungsmitteln in der Grundschule. In M. Neubrand (Hrsg.), *Beiträge zum Mathematikunterricht 2009 Online. Vorträge auf der 43. Tagung für Didaktik der Mathematik vom 02.03. bis 06.03.2009 in Oldenburg* (S. 115–118). Münster: WTM.

Söbbeke, E., & Steenpaß, A. (2010). Mathematische Deutungsprozesse zu Anschauungsmitteln unterstützen. In C. Böttlinger, K. Bräuning, M. Nührenbörger, R. Schwarzkopf, & E. Söbbeke (Hrsg.), *Mathematik im Denken der Kinder: Anregungen zur mathematikdidaktischen Reflexion* (S. 216–244). Seelze: Klett.

Söbbeke, E., & Steenpaß, A. (2014). Deutungsaufgaben zu Anschauungsmitteln. *Mathematik differenziert*, 5(4), 10–13.

Söbbeke, E., & Steinbring, H. (2007). Anschauung und Sehverstehen – Grundschulkinder lernen im Konkreten das Abstrakte zu sehen und zu verstehen. In J. H. Lorenz & W. Schipper (Hrsg.), *Hendrik Radatz: Impulse für den Mathematikunterricht* (S. 62–68). Braunschweig: Schroedel.

Spiel, C., Lüftenergger, M., Wagner, P., Schober, B., & Finsterwald, M. (2011). Förderung von Lebenslangem Lernen – eine Aufgabe der Schule. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Bildungsforschung* (S. 305–319). Wiesbaden: Springer.

Spreckels, J. (2009). *Erklären im Kontext: Neue Perspektiven aus der Gesprächs- und Unterrichtsforschung*. Hohengehren: Schneider.

Stahl, R. (1975). Die Entwicklung geistiger Fähigkeiten im Mathematikunterricht der Klassen 1 bis 3. In J. Lompscher (Hrsg.), *Theoretische und experimentelle Untersuchungen zur Entwicklung geistiger Fähigkeiten* (2. Aufl., S. 258–302). Berlin: Volk und Wissen.

Stahn, C. (2011). *Evaluation einer Interventionsmaßnahme zur Steigerung der kognitiven Leistungsfähigkeit bei älteren Arbeitnehmern in der Automobilbranche* (Dissertation). Technische Universität, Dortmund. Abgerufen von <https://eldorado.tu-dortmund.de/handle/2003/29183> (letzter Zugriff am 06.11.2016)

Steenpaß, A. (2014). *Grundschulkinder deuten Anschauungsmittel. Eine epistemologische Kontext- und Rahmenanalyse zu den Bedingungen der visuellen Strukturierungskompetenz* (Dissertation). Universität Duisburg-Essen, Essen. Abgerufen von DuEPublico (ID 35866)

Steigert, T., & Schrenk, M. (2012). Fördert eigenständiges Experimentieren die Entwicklung wissenschaftsnaher Vorstellungen zum Pflanzenstoffwechsel? - Teilprojekt 3. In W. Rieß, M. Wirtz, B. Barzel, & A. Schulz (Hrsg.), *Experimentieren im mathematisch-naturwissenschaftlichen Unterricht: Schüler lernen wissenschaftlich denken und arbeiten* (S. 199–212). Münster: Waxmann.

Steinau, B. (2014). In die richtige Reihenfolge bringen. Ein Problem zeichnerisch lösen. *Grundschule Mathematik*, 41, 32–35.

Steinbring, H., & Nührenbörger, M. (2010). Mathematisches Wissen als Gegenstand von Lehr-/Lerninteraktionen. Eigenständige Schülerinteraktionen in Differenz zu Lehrerinterventionen. In U. Dausendschön-Gay, C. Domke, & S. Ohlhus (Hrsg.), *Wissen in (Inter-)Aktion: Verfahren der Wissensgenerierung in unterschiedlichen Praxisfeldern* (S. 161–188). Berlin: Walter de Gruyter.

Steiner, G. (1988). Analoge Repräsentationen. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie* (S. 99–119). München: Psychologie-Verlag-Union.

Steinheider, P. (2014). *Was Schulen für ihre guten Schülerinnen und Schüler tun können: Hochbegabtenförderung als Schulentwicklungsaufgabe*. Wiesbaden: Springer.

Stein, M. (1999). Elementare Bausteine der Problemlösefähigkeit: Logisches Denken und Argumentieren. *Journal für Mathematik-Didaktik*, 20(1), 3–27. <https://doi.org/10.1007/BF03338881>

Steinmetz, H. (2015). *Lineare Strukturgleichungsmodelle: Eine Einführung mit R*. München: Rainer Hampp.

Steinweg, A. S. (2001). *Zur Entwicklung des Zahlenmusterverständnisses bei Kindern: Epistemologisch-pädagogische Grundlegung*. Münster: Lit.

Steinweg, A. S. (2005). Arithmetik ist mehr als Ausrechnen. *Grundschulunterricht Mathematik*, 52(7-8), 15–17.

Steinweg, A. S. (2013). *Algebra in der Grundschule: Muster und Strukturen – Gleichungen – funktionale Beziehungen*. Berlin: Springer.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science*, 26(1-2), 127–140. <https://doi.org/10.1023/A:1003096215103>

Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology*, 85(1), 7–23. <https://doi.org/10.1037/0022-0663.85.1.7>

Stern, E. (1997). Die Entwicklung schulbezogener Kompetenzen: Mathematik. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 95–113). Weinheim: Beltz.

Stern, E. (1998). *Die Entwicklung des mathematischen Verständnisses im Kindesalter*. Lengerich: Pabst Science Publishers.

Stern, E. (2001). Intelligenz, Wissen, Transfer und der Umgang mit Zeichensystemen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 163–204). Lengerich: Pabst Publisher.

Stern, E. (2003). Früh übt sich – Neuere Ergebnisse aus der LOGIK-Studie zum Lösen mathematischer Textaufgaben. In A. Fritz, G. Ricken, & S. Schmidt (Hrsg.), *Handbuch Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen* (S. 116–130). Leverkusen: Leske + Budrich.

Stern, E. (2004). Schubladendenken. Zum Umgang mit unterschiedlichen Lernvoraussetzungen und Lerntypen. *Forum Schule*, 1, 8–11.

Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction*, 13(2), 191–203. [https://doi.org/10.1016/S0959-4752\(02\)00020-8](https://doi.org/10.1016/S0959-4752(02)00020-8)

Stern, E., Felbrich, A., & Schneider, M. (2006). Mathematiklernen. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 461–469). Weinheim: Beltz.

Stevens, J. P. (2007). *Intermediate statistics: A modern approach* (3. Aufl.). New York: Erlbaum.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5. Aufl.). New York: Routledge.

Strauss, A., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2. Aufl.). London: SAGE Publications.

Streefland, L. (1990). Free productions in teaching and learning mathematics. In K. Gravemeijer, M. van den Heuvel, & L. Streefland (Hrsg.), *Contexts free productions tests and geometry in realistic mathematics education* (S. 33–53). Utrecht: OW & OC.

Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: Matching methods for causal inference. In J. W. Osborne (Hrsg.), *Best Practices in Quantitative Methods* (S. 155–176). Los Angeles: SAGE.

Süllwold, F. (1960). Bedingungen und Gesetzmäßigkeiten des Problemlöseverhaltens. In H. Thomae (Hrsg.), *Bericht über den 22. Kongress der Deutschen Gesellschaft für Psychologie: in Heidelberg vom 27. September bis 1. Oktober 1959* (S. 96–115). Göttingen: Hogrefe.

Sundermann, B. (2014). „Wie treffen wir die 1000?“ Lösungswege mit Forschermitteln entwickeln und darstellen. *Grundschule Mathematik*, 41, 22–27.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (S. 19–30). New York: Cambridge University Press.

Sweller, J., & Paas, S. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Aufl., S. 27–42). New York: Cambridge University Press.

Sweller, J., van Marriënboer, J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Belmont, CA: Thomson.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6. Aufl.). Boston: Pearson.

Tarim, K. (2009). The effects of cooperative learning on preschoolers' mathematics problem-solving ability. *Educational Studies in Mathematics*, 72(3), 325–340. <https://doi.org/10.1007/s10649-009-9197-x>

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *The Quarterly Journal of Experimental Psychology Section A*, 58(7), 1311–1323. <https://doi.org/10.1080/02724980443000593>

Thevenot, C., & Oakhill, J. (2006). Representations and strategies for solving dynamic and static arithmetic word problems: The role of working memory capacities. *European Journal of Cognitive Psychology*, 18(5), 756–775. <https://doi.org/10.1080/09541440500412270>

Thevenot, C., & Oakhill, J. (2008). A generalization of the representational change theory from insight to non-insight problems: The case of arithmetic word problems. *Acta Psychologica*, 129(3), 315–324. <https://doi.org/10.1016/j.actpsy.2008.08.008>

Thies, S. (2002). *Zur Bedeutung diskreter Arbeitsweisen im Mathematikunterricht* (Dissertation). Justus-Liebig-Universität, Gießen. Abgerufen von <http://geb.uni-giessen.de/geb/volltexte/2002/854/> (letzter Zugriff am 06.11.2016)

Törner, P. D. G., & Zielinski, S. U. (2013). Problemlösen als integraler Bestandteil des Mathematikunterrichts – Einblicke und Konsequenzen. *Journal für Mathematik-Didaktik*, 13(2-3), 253–270. <https://doi.org/10.1007/BF03338781>

Toulmin, S. E. (1996). *Der Gebrauch von Argumenten* (2. Aufl.). Weinheim: Beltz.

Toulmin, S. E. (2008). *The uses of argument* (reprinted). Cambridge: University Press.

Triandis, H. C. (1975). *Einstellungen und Einstellungsänderungen*. Weinheim: Beltz.

Uesaka, Y., Manalo, E., & Ichikawa, S. (2007). What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? *Learning and Instruction*, 17(3), 322–335. <https://doi.org/10.1016/j.learninstruc.2007.02.006>

Ulm, V. (2008). Einführung: Mit „guten Aufgaben“ arbeiten. In V. Ulm (Hrsg.), *Gute Aufgaben Mathematik: Heterogenität nutzen. 30 gute Aufgaben für die Klassen 1 bis 4* (S. 8–11). Berlin: Cornelsen.

Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from the serial reaction time task. *Memory & Cognition*, 33(2), 213–220. <https://doi.org/10.3758/BF03195310>

van der Waerden, B. L. (1968). *Einfall und Überlegung* (verbesserter Nachdruck). Basel: Birkhäuser.

van Dijk, T. A. (1980). *Textwissenschaft: Eine interdisziplinäre Einführung*. Berlin: Walter de Gruyter.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

van Essen, G., & Hamaker, C. (1990). Using selfgenerated drawings to solve arithmetic word problems. *The Journal of Educational Research*, 83(6), 301–312.

van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, 17(4), 285–325. <https://doi.org/10.1007/s10648-005-8136-3>

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement*, 38(4), 953–958.
<https://doi.org/10.1177/001316447803800415>

Verboom, L. (2014a). Darstellen – eine vernachlässigte Kompetenz. *Grundschule Mathematik*, 41, 40–43.

Verboom, L. (2014b). Mathematische Zeichensprache. Zur Bedeutung nonverbaler Darstellungsmittel im Mathematikunterricht. *Grundschule Mathematik*, 41, 4–5.

Verschaffel, L., de Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology*, 84(1), 85–94. <https://doi.org/10.1037/0022-0663.84.1.85>

Verschaffel, L., Greer, B., & de Corte, E. (2000). *Making sense of word problems*. Lisse: Swets & Zeitlinger.

Vitale, M. R., & Romance, N. R. (2007). A knowledge-based framework for unifying content-area reading comprehension and reading comprehension strategies. In D. S. McNamara (Hrsg.), *Reading comprehension strategies: Theories, interventions, and technologies* (S. 73–104). New York: Psychology Press.

Vollrath, H.-J. (1980). Eine Thematisierung des Argumentierens in der Hauptschule. *Journal für Mathematik-Didaktik*, 1(1-2), 28–41. <https://doi.org/10.1007/BF03338629>

Vollrath, H.-J. (1989). Funktionales Denken. *Journal für Mathematik-Didaktik*, 10(1), 3–37. <https://doi.org/10.1007/BF03338719>

Vollstedt, M., Ufer, S., Heinze, A., & Reiss, K. (2015). Forschungsgegenstände und Forschungsziele. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 567–590). Berlin: Springer.

Vygotskij, L. S. (2002). *Denken und Sprechen: Psychologische Untersuchungen*. (J. Lompscher & G. Rückriem, Hrsg.). Weinheim: Beltz.

Wagenschein, M. (1968). *Verstehen lehren: Genetisch – Sokratisch – Exemplarisch*. Weinheim: Beltz.

Waldis, M., Gautschi, P., Hodel, J., & Reusser, K. (2006). Qualitätsmerkmalen im Geschichtsunterricht. Methodologische Überlegungen am Beispiel der Videostudie „Geschichte und Politik im Unterricht“. In H. Günther-Arndt & M. Sauer (Hrsg.), *Geschichtsdidaktik empirisch: Untersuchungen zum historischen Denken und Lernen* (S. 155–188). Berlin: LIT.

Walsch, W. (2000). Zum Beweisen im Mathematikunterricht. Interview mit Herrn Prof. Werner Walsch. *Mathematik in der Schule*, 38(1), 5–9.

Walther, G. (1984). Action proof vs. illuminating examples? *For the Learning of Mathematics*, 4(3), 10–12.

Walther, G. (2004). *Gute Aufgaben. Handreichungen des Programms SINUS an Grundschulen, Mathematik*. Handreichungen des Programms SINUS an Grundschulen, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Universität Kiel. Abgerufen von http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_STG/Mathematik/Module/Mathe1.pdf (letzter Zugriff am 06.11.2016)

Walther, G., Selter, C., & Neubrand, J. (2008). Die Bildungsstandards Mathematik. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer, & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule: Mathematik konkret* (S. 16–41). Berlin: Cornelsen Scriptor.

Wälti, B. (2009). Sachrechnen – Rechnen und Handeln an der Sache. Tragfähige Vorstellungen zu proportionalen Zuordnungen entwickeln. *Grundschulmagazin*, 7(5), 33–38.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294.
<https://doi.org/10.3102/00346543063003249>

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1997). What helps students learn? Spotlight on student success. *Spotlight on Student Success*, 209, 1–2.

Warne, R. T. (2014). A primer on multivariate analysis of Variance (MANOVA) for behavioral scientists – Add health. *Practical Assessment, Research & Evaluation*, 19(17), 1–10.

Webb, N. M. (1989). Peer interaction, problem-solving, and cognition: Multidisciplinary perspectives. *International Journal of Educational Research*, 13, 1–19.

Webb, N. M. (1995). Testing a theoretical model of student interaction and learning in small groups. In R. Hertz-Lazarowitz & N. Miller (Hrsg.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (S. 102–119). Cambridge: University Press.

Wegerich, M. (2015). Sachrechnen im fächerverbindenden Unterricht. In R. Rink (Hrsg.), *Von Guten Aufgaben bis Skizzen Zeichnen: Zum Sachrechnen im Mathematikunterricht der Grundschule* (S. 261–274). Baltmannsweiler: Schneider.

Weinert, F. E. (1996). Lerntheorien und Instruktionsmodelle. In F. E. Weinert (Hrsg.), *Psychologie des Lernens und der Instruktion* (Bd. 2, S. 1–48). Göttingen: Hogrefe.

Weinert, F. E. (1998). *Entwicklung im Kindesalter*. Weinheim: Beltz.

Weise, G. (1975). *Psychologische Leistungstests: Ein Handbuch für Studium und Praxis* (Bd. 1). Göttingen: Hogrefe.

Wertheimer, M. (1964). *Produktives Denken*. (W. Metzger, Übers.) (2. Aufl.). Frankfurt: Waldemar Kramer.

Wessells, M. G. (1990). *Kognitive Psychologie* (2. Aufl.). München: Ernst Reinhardt.

Wickelgren, W. A. (1974). *How to solve mathematical problems*. San Francisco: Courier Corporation.

Wiley, J., Sanchez, C. A., & Jaeger, A. J. (2014). The individual differences in working memory capacity principle in multimedia learning. In R. E. Mayer (Hrsg.), *The cambridge handbook of multimedia learning* (2. Aufl., S. 598–619). New York: Cambridge University Press.

Winkel, K. (2011). Entwicklungsmechanismen von Metakognition im mathematischen Unterrichtsdis-
kurs der Grundschule. In R. Haug & L. Holzapfel (Hrsg.), *Beiträge zum Mathematikunterricht
2011: Vorträge auf der 45. Tagung für Didaktik der Mathematik vom 21. bis 25.02.2011 in
Freiburg* (Bd. 2, S. 899–902). Münster: WTM.

Winkel, K. (2012). *Entwicklungsmechanismen von Metakognition im mathematischen Unterrichtsdis-
kurs der Grundschule. Ein designbasierter Unterrichtsversuch über vier Schuljahre*. Mün-
chen: Dr. Hut.

Winkel, K. (2013). Darstellungswechsel trainieren von Anfang an. Ein Schlüssel zum Verstehen von
Mathematik. *Grundschulunterricht Mathematik*, 3, 29–31.

Winter, H. (1975). Allgemeine Lernziele für den Mathematikunterricht? *Zentralblatt für Didaktik der
Mathematik*, 3, 106–116.

Winter, H. (1983). Zur Problematik des Beweisbedürfnisses. *Journal für Mathematik-Didaktik*, 4(1),
59–95. <https://doi.org/10.1007/BF03339229>

Winter, H. (1992). Aufgaben, Probleme, Kontexte – Zur grundsätzlichen Problematik des Sachrech-
nens in der Grundschule. *Sachunterricht und Mathematik in der Primarstufe*, 20(8), 350, 363–
369.

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für
Didaktik der Mathematik*, 61, 37–46.

Winter, H. (1997). Problemorientierung des Sachrechnens in der Primarstufe als Möglichkeit entde-
ckendes Lernen zu fördern. In P. Bardy (Hrsg.), *Mathematische und mathematikdidaktische
Ausbildung von Grundschullehrerinnen/-lehrern: Vorträge auf der gleichnamigen Tagung
vom 30. September bis 2. Oktober 1996 an der Universität Halle-Wittenberg (in Köthen)*. (Bd.
6, S. 57–92). Weinheim: Beltz.

Winter, H. (2000). *Sachrechnen in der Grundschule: Problematik des Sachrechnens; Funktionen des
Sachrechnens; Unterrichtsprojekte* (5. Aufl.). Frankfurt am Main: Cornelsen Scriptor.

Winter, H. (2001). Wahrnehmung und Phantasie im Mathematikunterricht der Grundschule. In W.
Weiser & B. Wollring (Hrsg.), *Beiträge zur Didaktik der Mathematik für die Primarstufe.
Festschrift für Siegbert Schmidt* (S. 289–306). Hamburg: Dr. Kovač.

Winter, H. (2003). „Gute Aufgaben“ für das Sachrechnen. In M. Baum & H. Wielpütz (Hrsg.), *Ma-
thematik in der Grundschule: Ein Arbeitsbuch* (S. 177–183). Seelze: Kallmeyer.

Wirtz, M. (2004). Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf
Analyseergebnisse entdeckt und reduziert werden kann. *Die Rehabilitation*, 43(2), 109–115.
<https://doi.org/10.1055/s-2003-814839>

Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur
Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensys-
temen und Ratingskalen*. Göttingen: Hogrefe.

Wittmann, E. C. (1993). „Weniger ist mehr“: Anschauungsmittel im Mathematikunterricht der Grundschule. In *Beiträge zum Mathematikunterricht 1993: Vorträge auf der 27. Bundestagung für Didaktik der Mathematik vom 22. bis 26.3.1993 in Freiburg/Schweiz* (S. 394–397). Hildesheim: Franzbecker.

Wittmann, E. C. (1998). Standard number representations in the teaching of arithmetic. *Journal für Mathematik-Didaktik*, 19(2-3), 149–178. <https://doi.org/10.1007/BF03338866>

Wittmann, E. C. (2003). Was ist Mathematik und welche pädagogische Bedeutung hat das wohlverstandene Fach auch für den Mathematikunterricht der Grundschule. In M. Baum & H. Wielputz (Hrsg.), *Mathematik in der Grundschule: Ein Arbeitsbuch* (S. 18–46). Seelze: Kallmeyer.

Wittmann, E. C. (2014). Operative Beweise in der Schul- und Elementarmathematik. *Mathematica didactica*, 37, 213–232.

Wittmann, E. C., & Müller, G. N. (1988). Wann ist ein Beweis ein Beweis? In P. Bender (Hrsg.), *Mathematikdidaktik: Theorie und Praxis: Festschrift für Heinrich Winter* (S. 237–257). Berlin: Cornelsen.

Wittmann, E. C., & Müller, G. N. (2013a). *Das Zahlenbuch 4. Begleitband*. Stuttgart: Klett.

Wittmann, E. C., & Müller, G. N. (2013b). *Das Zahlenbuch 4. Schülerband*. Stuttgart: Klett.

Wittmann, G. (2014). Beweisen und Argumentieren. In H.-G. Weigand, A. Filler, R. Hölzl, S. Kuntze, M. Ludwig, J. Roth, ... G. Wittmann, *Didaktik der Geometrie für die Sekundarstufe I* (2. Aufl., S. 35–54). Berlin: Springer.

Wöller, M. E. S. (2016). Mathematik und Sprache im Grundschulunterricht am Beispiel des Dialogischen Lernens. In K. Liebers, B. Landwehr, S. Reinhold, S. Riegler, & R. Schmidt (Hrsg.), *Facetten grundschulpädagogischer und -didaktischer Forschung* (Bd. 20, S. 213–218). Wiesbaden: Springer.

Zhe, L. (2012). Survey of primary students' mathematical representation status and study on the teaching model of mathematical representation. *Journal of Mathematics Education*, 5(1), 63–76.