

Die Qualität von Schulbüchern und ihre Bedeutung für die Arithmetikleistung in der Primarstufe

Eine längsschnittliche Analyse anhand von drei
Inhaltsbereichen

Dissertation

zur Erlangung des Doktorgrades
an der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Henning Christoph Sievert

Kiel, August 2020

Erstgutachter: Prof. Dr. Aiso Heinze

Zweitgutachterin: Prof. Dr. Anke M. Lindmeier

Tag der mündlichen Prüfung: 26.10.2020

Danksagung

Diese Arbeit ist das Ergebnis eines längeren Weges, den ich nicht allein beschritten habe, sondern mit der Unterstützung vieler Menschen um mich herum. Für diese bin ich sehr dankbar, denn ohne sie hätte diese Arbeit nicht entstehen können. Allen voran möchte ich meinem Doktorvater Prof. Dr. Aiso Heinze sowie Dr. Ann-Katrin van den Ham danken für die gute Betreuung meiner Arbeit, die stets offenen Ohren für all meine Fragen und die vielen konstruktiven Gespräche. Mein Dank gilt ebenso Prof. Dr. Anke Lindmeier, die einen erheblichen Anteil an meinem Weg in die Wissenschaft hatte und es mir ermöglicht hat, am IPN promovieren zu können.

Für die inhaltliche Unterstützung möchte ich außerdem Dr. Jessica Hoth, Prof. Dr. Christiane Benz sowie der Arbeitsgruppe Arithmetik des Arbeitskreises Grundschule unter der Koordination von Prof. Dr. Charlotte Rechtsteiner danken. Für die Beratung bei methodischen Fragen bedanke ich mich bei Dr. Simon Grund. Für spannende Diskussionen und Impulse aus der Schulpraxis sowie der Bändigung eines äußerst umfangreichen Datensatzes danke ich außerdem Dr. Torben von Seeler, Christiane Meerstein sowie Karin Rogalski vom IQSH.

Die Datengrundlage dieser Studien wäre nicht ohne die tatkräftige Unterstützung ehemaliger Mitarbeiterinnen der „Mathe macht stark“-Evaluation zustande gekommen. Vielen Dank, Prof. Dr. Meike Grüßing, Dr. Anne-Katrin Jordan, Dr. Inga Niedermeyer! Für die sorgfältige und ausdauernde Arbeit danke ich außerdem den studentischen Hilfskräften Jerg Dillmann, Lena Sachler, Pia Kohn, Lara Harbich, Svea Lambers, Maren Sievering sowie meinen jetzigen Kolleginnen Femke Sporn und Kristin Litteck.

Ein gesundes Arbeitsumfeld war eine wertvolle Grundlage bei der Entstehung dieser Arbeit. Ich danke daher den Kolleginnen und Kollegen am IPN, insbesondere der Abteilung Didaktik der Mathematik, für das gemeinsame Forschen, den Austausch und die gegenseitige Unterstützung. Danke, Anje, dass du die letzten dreieinhalb Jahre das Büro mit mir geteilt hast. Ich hätte es nicht besser treffen können! Für die Unterstützung in den letzten Zügen dieser Arbeit gilt mein Dank außerdem Dr. Colin Jeschke sowie Robert von Hering.

Selbstverständlich bin ich ebenso dankbar für die vielen außeruniversitären Bedingungsfaktoren. Ich danke meiner Mutter, meinem Vater und Leni dafür, dass sie mich stets unterstützt haben und mir zu jederzeit ermöglicht haben, den Weg zu gehen, den ich mir selbst ausgesucht habe. Dieses Privileg ist tief in meinem Bewusstsein verankert. Ebenso dankbar bin ich meinen Freundinnen, Freunden und meiner Familie, insbesondere meinen musikalischen Kollegen, die mich während dieser Zeit begleitet haben und mir stets auch einen passenden Ausgleich gegeben haben. Ich danke meinem Opa dafür, dass er mich schon in jungen Jahren für die Mathematik begeistert hat und meiner Oma für das wöchentliche Kaffeetrinken. Manchmal sind es eben die kleinen Dinge, die für uns wesentlich sind. Danke Jessi, dass du stets mein Fels bist, selbst wenn ich die Brandung bin.

Zusammenfassung

Obwohl das Schulbuch im Mathematikunterricht an Grundschulen eine zentrale Rolle spielt, ist dessen Qualität gemessen am Lernerfolg von Schülerinnen und Schülern bislang wenig erforscht. Abgesehen von einer adäquaten Abbildung des Curriculums sind bisher keine weiteren Kriterien definiert, die empirisch gesichert zu einem höheren Lernerfolg bei Schülerinnen und Schülern führen. Dieser Sachverhalt wird in der vorliegenden Arbeit durch drei Studien adressiert, in denen sowohl die Schulbuchqualität für einzelne Inhaltsbereiche definiert und erfasst als auch ihr Zusammenhang mit dem Lernerfolg der Kinder anhand der längsschnittlichen Daten von 1.664 Schülerinnen und Schülern analysiert werden. Die Themenschwerpunkte bilden die Nutzung Operativer Grundaufgaben und das Lösen quantitativer Vergleiche jeweils in der Klassenstufe 1 sowie das adaptive Rechnen bei der Addition und Subtraktion in der Klassenstufe 3. Alle drei Fähigkeitsbereiche stellen wichtige Voraussetzungen für das weitere Mathematiklernen dar. Eine Verbesserung der Schulbuchqualität könnte entsprechend dazu beitragen.

Zur Bestimmung der Schulbuchqualität wurden für jeden Inhaltsbereich Bewertungskriterien aus dem Stand der Forschung abgeleitet und für den jeweiligen Bereich konkretisiert. Für das adaptive Rechnen und die Nutzung operativer Beziehungen wurden diese Kriterien aus einem etablierten psychologischen Modell zur Strategiewahl hergeleitet. Für den Bereich quantitative Vergleiche existiert bisher kein vergleichbares Modell. Die Kriterien dafür wurden entsprechend auf Grundlage des aktuellen Forschungsstands neu entwickelt und in Diskussionen mit Expertinnen und Experten aus der Schulpraxis sowie der Mathematikdidaktik elaboriert. In der untersuchten Stichprobe wurden vier Schulbuchreihen verwendet, die alle das gleiche Arithmetikcurriculum adressieren. Ihre Analyse ergab signifikante Qualitätsunterschiede in den Lerngelegenheiten für alle drei Inhaltsbereiche. Dabei unterschieden sich die Rangfolgen der vier Schulbuchreihen sowohl zwischen den drei Bereichen als auch bezüglich der Kriterien innerhalb der einzelnen Bereiche zum Teil erheblich.

Der Zusammenhang zwischen der bereichsspezifischen Schulbuchqualität und dem entsprechenden Lernerfolg der Schülerinnen und Schüler wurde anschließend anhand von Mehrebenenmodellierungen der Längsschnittdaten untersucht. Dazu wurden die Ergebnisse der Schülerinnen und Schüler für die jeweils bereichsspezifischen Items skaliert und der Einfluss der entsprechenden Schulbuchqualität unter Kontrolle individueller sowie unterrichtlicher Faktoren untersucht. Die Ergebnisse aller drei Analysen weisen signifikante Effekte der Schulbuchqualität auf den Lernerfolg der Schülerinnen und Schüler auf. Dies zeigt, dass sich Schulbucheffekte auf den Lernerfolg der Schülerinnen und Schüler durch die theoriebasierte Analyse der Qualität der Schulbücher erklären lassen und indiziert einen Einfluss der Schulbuchqualität auf die Lernenden. Zudem geben die Ergebnisse Hinweise für die Gestaltung lerneffektiver Schulbücher. Sie bieten darüber hinaus einen Forschungsansatz zur bereichsspezifischen Bestimmung von Schulbuchqualität, welche sowohl theoriebasiert als auch empirisch bestätigt ist. Dies kann einen ersten Schritt zu Beantwortung der übergeordneten Frage nach einer globalen Bewertung der Qualität von Mathematikbüchern darstellen.

Abstract

Despite the central role of the textbook in primary school mathematics teaching, the state of research on its quality as measured by pupils' learning success is limited. Beyond the adequate representation of the curriculum, no further criteria have yet been defined which have been empirically proven to lead to higher learning success for students. This issue is addressed in this thesis by three studies in which both the quality of textbooks is defined and determined for individual subject areas and their relationship to students' learning success is examined using longitudinal data from 1,664 students. The three subject areas studied are the adaptive use of strategies for addition and subtraction in Grade 3, the use of arithmetic principles in Grade 1, and the solving of quantitative comparisons also in Grade 1. All three skill areas addressed provide important prerequisites for further learning in mathematics. A potential improvement in textbook quality could contribute accordingly.

In order to determine textbook quality, evaluation criteria for each subject area were first derived from the current state of research and elaborated for the respective area. For the adaptive use of strategies and the use of arithmetic principles, these criteria were derived from an established psychological model for strategy choice. No comparable model exists so far for quantitative comparisons. Therefore, these criteria were developed based on current research and elaborated in discussions with experts from school practice and mathematics education. Four textbook series were used in the sample studied, which all address the same arithmetic curriculum. Their analysis revealed significant quality differences in learning opportunities for all three subject areas. The ranking of the four textbook series differed both between the three areas and between the categories of the individual areas, in some cases considerably.

We subsequently examined the relationship between area-specific textbook quality and the corresponding learning success of the students using multi-level modelling of the longitudinal data. For this purpose, the students' results of the respective topic-specific items were scaled and the influence of the corresponding textbook quality was examined while controlling for individual and classroom factors. The results of all three analyses showed a significant effect of textbook quality on the students' learning success. This shows that textbook effects on students' learning success can be explained by the theory-based analysis of textbook quality and indicates an influence of textbook quality on the students' learning. In addition, the results provide indications for the design of learning effective textbooks. Furthermore, they offer a research approach for the topic-specific determination of textbook quality, which is both theory-based and empirically confirmed. This can be a first step towards answering the overarching question of a global assessment of textbook quality.

Inhaltsverzeichnis

Danksagung	III
Zusammenfassung	V
Abstract	VII
1 Einleitung	1
2 Theoretischer Hintergrund	3
2.1 <i>Schulbücher als Lernressource</i>	3
2.1.1 Verortung der Untersuchungen im Modell des sozio-didaktischen Tetraeders	3
2.1.2 Die Rolle von Mathematikbüchern	4
2.1.3 Der Inhalt von Mathematikbüchern	4
2.1.4 Die Nutzung von Mathematikbüchern	4
2.1.5 Die Effekte der Schulbuchwahl	5
2.2 <i>Adaptives Rechnen bei der Addition und Subtraktion bis 1.000</i>	7
2.3 <i>Nutzung Operativer Grundaufgaben</i>	13
2.4 <i>Quantitative Vergleiche</i>	17
2.5 <i>Forschungsdiesiderate bereichsspezifischer Schulbuchqualität</i>	20
3 Die drei Studien dieser Arbeit	23
3.1 <i>Studie 1</i>	23
3.1.1 Ziel und Hypothesen	23
3.1.2 Untersuchungsansatz	23
3.2 <i>Studie 2</i>	24
3.2.1 Ziel und Hypothesen	24
3.2.2 Untersuchungsansatz	25
3.3 <i>Studie 3</i>	26
3.3.1 Ziel und Hypothesen	26
3.3.2 Untersuchungsansatz	26
4 Stichproben und Instrumente	28
4.1 <i>Stichproben</i>	28
4.1.1 Evaluationsstichprobe	28
4.1.2 Gesamtstichprobe der IPN-Schulbuchstudie	28
4.1.3 Stichprobe von Studie 1	29
4.1.4 Stichprobe von Studie 2	29
4.1.5 Stichprobe von Studie 3	30
4.2 <i>Instrumente</i>	30
4.2.1 Individuelle und unterrichtliche Kovariaten	30
4.2.2 Schulbuchqualität und Lernerfolg in Studie 1	32
4.2.3 Schulbuchqualität und Lernerfolg in Studie 2	36
4.2.4 Schulbuchqualität und Lernerfolg in Studie 3	43
5 Zusammenfassungen der Studienergebnisse	49
5.1 <i>Zusammenfassung der Ergebnisse von Studie 1</i>	49

5.2	<i>Zusammenfassung der Ergebnisse von Studie 2</i>	51
5.3	<i>Zusammenfassung der Ergebnisse von Studie 3</i>	53
6	Diskussion	56
6.1	<i>Zusammenfassung und Einordnung der Studien in den Forschungsstand</i>	56
6.2	<i>Implikationen für die Forschung</i>	57
6.2.1	Implikationen für die Forschung zum Erwerb bestimmter Fähigkeiten	57
6.2.2	Implikationen für die Forschung zu Mathematikschulbüchern	59
6.3	<i>Implikationen für die Praxis</i>	60
6.4	<i>Limitationen</i>	62
A	Artikel zu Studie 1	i
B	Artikel zu Studie 2	xxxv
C	Artikel zu Studie 3	lxvii
	Abbildungsverzeichnis	lxxxix
	Tabellenverzeichnis	xc
	Literaturverzeichnis	xciii
	Eidesstattliche Erklärung	cv

1 Einleitung

Die wichtigste Lernressource im Mathematikunterricht ist das Schulbuch. International stellt es insbesondere bei Grundschullehrkräften das am häufigsten genutzte Medium dar, sowohl für die Unterrichtsvorbereitung und die Instruktion neuer Inhalte als auch als Aufgabenrepertoire für das eigenständige Üben der Schülerinnen und Schüler (Lepik et al., 2015; Mullis et al., 2012). Inwiefern der Inhalt von Schulbüchern dabei einen Einfluss auf den Lernerfolg von Schülerinnen und Schülern hat, ist trotz der Bedeutung dieser Fragestellung bisher wenig untersucht worden und der bisherige Forschungsstand erweist sich als inkonsistent. Während einige Studien auf einen solchen Zusammenhang hindeuten (z. B. Fagginger Auer et al., 2016; Hadar, 2017; Heinze et al., 2009; Schmidt et al., 2001; Törnroos, 2005; van den Ham & Heinze, 2018), gibt es auch einzelne Untersuchungen, die keinen Zusammenhang zwischen Schulbuch und Lernerfolg der Schülerinnen und Schüler finden konnten (z. B. Blazar et al., 2019; van Steenbrugge et al., 2013).

Die Bedeutung von Schulbüchern im Mathematikunterricht spiegelt sich spätestens seit den 2000er Jahren auch in einem wachsenden Forschungsinteresse wider. So gibt es seit 2014 eine eigene internationale Konferenz zur Forschung und Entwicklung von Mathematikbüchern, welche zunächst in Southampton (UK), das zweite Mal 2017 in Rio de Janeiro (Brasilien) und das dritte Mal 2019 in Paderborn (D) stattgefunden hat (International Conference on Mathematics Textbook Research and Development – ICMT). Trotz des wachsenden Korpus an Studien ist der Wissensstand über die Wirkung von Schulbüchern auf die Lernenden sehr begrenzt. In seinem Überblicksartikel zum Stand der Forschung über Mathematikbücher stellt Fan (2013) dazu fest, dass der Großteil der bisherigen Studien deskriptive Forschungsfragen adressiert – beispielsweise welche soziokulturellen Normen in Schulbüchern transportiert werden oder welche Unterschiede es zwischen Schulbüchern verschiedener Länder gibt. Untersuchungen zu Schulbucheffekten basieren zudem häufig auf querschnittlichen Designs und kleinen Stichprobengrößen (s. a. Fan et al., 2013). Eine weitere Ursache für den ausbaufähigen Stand der Forschung ist systemischer Natur: Zwar gibt es in Nordamerika eine ganze Reihe von Untersuchungen zur Lerneffektivität von Schulbüchern (z. B. Blazar et al., 2019; Koedel et al., 2017; Tarr et al., 2008), da es in den USA allerdings kaum einheitlich verpflichtende Lehrpläne in den Bundesstaaten gibt, sondern sich die Schulen meist autonom für einen Lehrplan entscheiden können, werden Schulbücher und Curricula in diesen Studien jedoch synonym betrachtet. Daher wird in den amerikanischen Untersuchungen hauptsächlich die Lerneffektivität verschiedener Curricula miteinander verglichen, wenn Schulbucheffekte untersucht werden. Curricula und Schulbücher sind in den Studien konfundiert.

Im Gegensatz dazu wird in den Untersuchungen dieser Arbeit eine fokussierte Perspektive auf Schulbücher eingenommen, indem verschiedene Bücher analysiert werden, welche dem gleichen Lehrplan folgen. Sie bauen auf der Studie von van den Ham & Heinze (2018) auf, in der in einem Blackbox-Modell ein Effekt der Schulbuchwahl auf die Arithmetikleistung der Schülerinnen und Schüler in Klassenstufe 1–3 gezeigt werden konnte. Für die Aufklärung, wodurch dieser Effekt zustande kommt, wird in dieser Dissertation das Konstrukt der bereichsspezifischen Schulbuchqualität im Sinne der Lerneffektivität der Bücher entwickelt und eingesetzt.

Um die Bedeutung und einen möglichen Einfluss der Schulbuchqualität auf den Lernerfolg in Mathematik in der Grundschule zu untersuchen, wurden drei Studien durchgeführt, welche in dieser Arbeit berichtet werden. Die drei Studien untersuchen die bereichsspezifische Qualität von vier Schulbuchreihen für drei Inhaltsbereiche¹: Adaptives Rechnen bei der Addition und Subtraktion bis 1.000 (Studie 1), die Nutzung Operativer Grundaufgaben (Studie 2) und quantitative Vergleiche (Studie 3). Für jeden der Bereiche werden zunächst theoriebasiert Kriterien für die Qualität der Schulbücher entwickelt, anhand derer die vier Schulbuchreihen für die Grundschule untersucht und bewertet werden. Auf Grundlage dessen werden in Sekundäranalysen der Längsschnittdaten von etwa 1.600 Schülerinnen und Schülern die Zusammenhänge zwischen bereichsspezifischer Schulbuchqualität und dem Lernerfolg der Kinder im entsprechenden Inhaltsbereich untersucht.

¹ Die Studien werden in der Reihenfolge berichtet, in der sie im Rahmen des Promotionsvorhabens entstanden sind, auch wenn sich Studie 1 auf die Klassenstufen 2 bzw. 3 bezieht und die Studien 2 und 3 auf die Klassenstufe 1.

2 Theoretischer Hintergrund

In diesem Kapitel werden zunächst die Arbeit im Kontext der Forschung zu Mathematikschulbüchern verortet sowie der Stand der Forschung zur Rolle, zum Inhalt, zur Nutzung und zu den Effekten dieser Lernressource dargestellt (Abschnitt 2.1). Die Abschnitte 2.2–2.4 stellen dann die drei analysierten Inhaltsbereiche des Schulcurriculums dar, bevor in Abschnitt 2.5 die Forschungsdesiderate abgeleitet werden.

2.1 Schulbücher als Lernressource

2.1.1 Verortung der Untersuchungen im Modell des sozio-didaktischen Tetraeders

Als von Menschen erstelltes Unterrichtsmaterial stellen Schulbücher Artefakte dar (z. B. Rabardel, 2002). Den komplexen und von vielen Faktoren bedingten Einsatz von solchen Artefakten im Unterricht haben Rezat und Sträßer (2012) in ihrem Modell des *Tetraeders der didaktischen Situation* dargestellt. Darin bilden das Schulbuch (das Artefakt), die Lehrkraft, die Lernenden sowie die Mathematik die vier Eckpunkte des Tetraeders. Aus diesen lassen sich vier Dreiecke ableiten: das didaktische Dreieck (Mathematik – Lehrkraft – Lernende) als Basis des Modells, das Dreieck Mathematik – Schulbuch – Lernende für die durch ein Artefakt vermittelte Aktivität des Mathematiklernens, das Dreieck Schulbuch – Lehrkraft – Lernende für die Rolle der Lehrkraft als Orchestrator und Mediator sowie das Dreieck Mathematik – Schulbuch – Lehrkraft für die Rolle des Schulbuchs als aufbereitete Mathematik und Unterrichtsressource für die Lehrkraft (vgl. Rezat & Sträßer, 2012, S. 645). In einem zweiten Schritt wird dieses Modell durch soziale und kulturelle Aspekte gerahmt und als sozio-didaktisches Tetraeder aufgefasst, welches durch Einbezug weiterer Punkte eine größere Komplexität aufweist. Hinzugefügt werden dabei Konventionen und Normen über Lehrkräfte und über Lernende, eingebundene Personen wie der Peer-Gruppe oder der Familie, die Institution, das gesellschaftliche Bild und die Relevanz von Mathematik sowie die Noosphäre aller weiteren Personen und Institutionen, die sich für das Lehren und Lernen von Mathematik interessieren (vgl. Rezat & Sträßer, 2012, S. 648). Die Komplexität dieses Modells spiegelt zum einen die zahlreichen Mechanismen des Einsatzes von Artefakten wie Schulbüchern sowie die Vielschichtigkeit an möglichen Perspektiven darauf wider. Zum anderen eignet es sich gerade aufgrund dieser Komplexität dazu, die eigene Perspektive zu verorten.

Die Studien dieser Arbeit adressieren den Zusammenhang von Schulbüchern und Lernerfolg der Schülerinnen und Schüler. Dabei spielen Lehrkräfte nicht nur eine vermittelnde Rolle zwischen diesen, sie wählen das Schulbuch in der Regel außerdem (mit) aus, planen den Unterricht, in welchem die Bücher eingesetzt werden und führen ihn durch. Sie entscheiden also nicht nur, welches Schulbuch eingesetzt wird, sondern auch, wie es genutzt wird. Dementsprechend lassen sich die Untersuchungen vor allem in dem Dreieck Schulbuch – Lehrkraft – Lernende einordnen. Die mathematikdidaktische Analyse und Bewertung der Schulbücher bzw. ihrer Qualität ist dagegen im Zusammenspiel von Mathematik, Schulbuch und Lernenden zu verorten, sodass die Untersuchungen auch dieses Dreieck betreffen. Dabei versteht sich, dass

die Einnahme dieser Perspektiven lediglich den Schwerpunkt der Arbeit darstellt und keine isolierte Betrachtung bedeutet. Insbesondere die Institution Schule, welche organisatorische Rahmenbedingungen setzt, spielt daher in allen drei Studien eine wesentliche Rolle.

2.1.2 Die Rolle von Mathematikbüchern

Schulbüchern kommt beim Mathematiklernen in der Grundschule eine zentrale Rolle als Lernressource für Lehrkräfte und Schülerinnen und Schüler zu. Sie übersetzen den abstrakten Lehrplan in konkrete Aufgaben, die von Lehrkräften und Lernenden ausgeführt und bearbeitet werden können, und fungieren so als Mediator zwischen dem offiziellen und dem von Lehrkräften umgesetzten Curriculum (*intendiertes vs. implementiertes Curriculum* nach Valverde et al., 2002). Daher eignen sie sich auch als Instrument zum Monitoring im Bildungssystem. In einigen Ländern bedarf es zudem der Zulassung jedes Schulbuchs durch das Bildungsministerium (z. B. Hongkong, Norwegen). Andere Länder haben keine Genehmigungspflicht (z. B. Niederlande, einige Bundesstaaten der USA) (Mullis et al., 1997). Ähnlich heterogen verhält es sich in Deutschland, wo es sowohl Bundesländer mit als auch solche ohne eine Zulassungspflicht für Schulbücher gibt (Stöber, 2010). In Schleswig-Holstein, wo die Daten der hier vorgestellten Studien erhoben worden sind, gibt es eine solche Genehmigungspflicht seit 2005 nicht mehr. Der Grad eines solchen behördlichen Monitorings kann die Interpretation des Lehrplans im Schulbuch beeinflussen (Schmidt, Raizen et al., 1997).

2.1.3 Der Inhalt von Mathematikbüchern

Mathematikbücher bilden einen konkreten Lehrplan ab. Sie unterscheiden sich dabei sowohl in ihren Inhalten als auch in ihren pädagogischen und didaktischen Ansätzen (Pepin & Haggarty, 2001). Darüber hinaus zeigen Studien auch Unterschiede kultureller Natur (beispielsweise in Bezug auf die Darstellung von Geschlecht, Ethnizität, Gleichberechtigung, Fan et al., 2013). Die Inhalte von Schulbüchern variieren bezüglich ihrer Realisierung des Curriculums (Fan et al., 2013). Dies betrifft beispielsweise die kognitiven Anforderungen, die in den Büchern gestellt werden (z. B. Kolovou et al., 2009) oder die Struktur, in der ein Inhalt dargestellt wird (z. B. Jones et al., 2015). Die Umsetzung eines Curriculums in Form eines Schulbuchs unterscheidet sich also sowohl in Abhängigkeit von kulturellen Normen als auch aufgrund unterschiedlicher pädagogischer und didaktischer Ansätze. Durch diese im Schulbuch dargestellten Inhalte wird ebenfalls das potentielle Angebot an Lerngelegenheiten für Schülerinnen und Schüler im Unterricht mitgestaltet. Wenngleich Lehrkräfte wie auch Schülerinnen und Schüler die im Schulbuch angebotenen Lerngelegenheiten unterschiedlich nutzen, wird davon ausgegangen, dass der Inhalt von Schulbüchern einen Einfluss auf das Lehren und Lernen im Unterricht hat (Valverde et al., 2002).

2.1.4 Die Nutzung von Mathematikbüchern

Weltweit stellen Schulbücher das wichtigste Medium im Mathematikunterricht der Primarstufe dar. So gaben in der "Trends in International Mathematics and Science Study" (TIMSS) 2011 durchschnittlich 75 % der Grundschullehrkräfte an, das Mathematikbuch als Grundlage für ihren Unterricht zu verwenden (Mullis et al., 2012). In Deutschland, wo die Daten der in dieser

Arbeit vorgestellten Studien erhoben worden sind, stimmten 86 % der Lehrkräfte dieser Aussage zu. Lepik et al. (2015) befragten 402 Mathematiklehrkräfte, die in Klassenstufe 7 bis 9 in Estland, Finnland und Norwegen unterrichten. Dabei gaben je nach Land 49–64 % von ihnen an, sich bei der Unterrichtsvorbereitung stark am Schulbuch zu orientieren. 79–92 % der Lehrkräfte gaben sogar an, in mindestens der Hälfte der Unterrichtsstunden das Schulbuch als einzige Aufgabenquelle zu nutzen. Einschränkend ist dabei jedoch hervorzuheben, dass sich Lehrkräfte, welche angeben, das Schulbuch generell zu nutzen, dennoch stark darin unterscheiden können, wie häufig sie es konkret einsetzen (Blazar et al., 2019). Dennoch kann angenommen werden, dass das Schulbuch die am meisten genutzte Lernressource im Mathematikunterricht darstellt, insbesondere in der Grundschule. Dies trifft auch auf die Lehrkräfte der Stichprobe zu, welche in den Studien dieser Arbeit untersucht wurde: Fast 97 % von ihnen gaben an, dass sie für das Unterrichten arithmetischer Inhalte das Schulbuch jeden oder fast jeden Tag nutzen. Über 87 % stimmten der Aussage zu, dem Schulbuch in ihrem Unterricht in Bezug auf Inhalte, Methoden und verwendete Materialien weitestgehend zu folgen (van den Ham & Heinze, 2018).

2.1.5 Die Effekte der Schulbuchwahl

Da Mathematikbücher intensiv genutzt werden, sich aber erheblich voneinander unterscheiden können, ist anzunehmen, dass dies auch einen Effekt auf Unterricht und Lernerfolg haben könnte. Es existieren einige Untersuchungen, welche auf einen Einfluss des Mathematikbuchs auf den Unterricht der Lehrkräfte hinweisen: So konnte Krammer (1985) in einer Studie in den Niederlanden Unterschiede zwischen den Unterrichtsmethoden von 50 Mathematiklehrkräften zeigen, welche eines der drei am weitesten verbreiteten niederländischen Schulbücher genutzt haben. In den Daten der TIMSS Video Studie von 1995 fanden Schmidt et al. (2001) einen positiven Zusammenhang zwischen dem relativen Platz, den ein Thema im Schulbuch einnimmt, und der Unterrichtszeit, die die Lehrkräfte für das Thema verwenden. Ebenso ist es unwahrscheinlich, dass Themen im Unterricht behandelt werden, die nicht im Schulbuch vorkommen (Schmidt, McKnight et al., 1997).

Über diese Ergebnisse zum Einfluss auf den Unterricht hinaus gibt es Hinweise auf einen Einfluss auf den Lernerfolg der Schülerinnen und Schüler. Agodini, et al. (2010) teilten 111 Grundschulen in 10 US-Bundesstaaten vier verschiedene Schulbücher zufällig zu. Schülerinnen und Schüler aus zwei dieser Schulbuchgruppen übertrafen dabei die Leistung einer anderen Schulbuchgruppe in Klassenstufe 2. Bei der Analyse der 1995er-TIMSS-Daten der USA (Klasse 8) konnten Schmidt et al. (2001) auch einen Zusammenhang zwischen dem relativen Anteil eines Themas im Buch und dem entsprechenden Lernzuwachs der Schülerinnen und Schüler aufzeigen. Ähnliche Zusammenhänge fand Törnroos (2005) in den finnischen TIMSS-Daten von 1999. Durch die Analyse von neun Mathematikbuchserien, die in 104 finnischen Schulen (Klasse 5–7) verwendet wurden, konnte er zeigen, dass die Anzahl der Lerngelegenheiten in den Schulbüchern, welche den TIMSS-Inhalten entsprechen, signifikant mit der Leistung der Schülerinnen und Schüler im TIMSS-Test von 1999 korreliert ist. In einer Untersuchung von 4.040 Schülerinnen und Schülern der arabischen Community in Israel stellte Hadar (2017) fest, dass Schülerinnen und Schüler eine höhere Punktzahl in einem staatlichen Abschlusstest erreichten, wenn

das von ihnen genutzte Schulbuch Lerngelegenheiten zur Bearbeitung von Aufgaben enthält, welche ein höheres Maß an Verständnis erfordern.

Diesen Ergebnissen aus querschnittlich angelegten Studien entsprechend konnten van den Ham und Heinze (2018) auch signifikante Effekte der Schulbuchwahl auf den Lernerfolg der Schülerinnen und Schüler anhand von längsschnittlichen Analysen der Daten von 1.664 Schülerinnen und Schülern (Klasse 1–3) aus Deutschland zeigen. Diese basieren auf der gleichen Stichprobe, welche den Studien dieser Arbeit zugrunde liegt, und bilden gleichermaßen ihren Ausgangspunkt. Bereits nach dem ersten Schuljahr zeigen die Schülerinnen und Schüler einer von vier untersuchten Schulbuchgruppen signifikant höhere Lernerfolge im Bereich Arithmetik als eine der anderen drei (unter Kontrolle individueller sowie unterrichtlicher Voraussetzungen). Nach dem zweiten Schuljahr zeigt sich dieser Effekt eines im Verhältnis zur Vergleichsgruppe höheren Lernerfolgs im Bereich Arithmetik auch für die übrigen beiden Schulbuchgruppen der Stichprobe. Bei den Ergebnissen des dritten Schuljahres sind die Effekte ähnlich. Auch hier weisen Schülerinnen und Schüler aus zwei der vier Schulbuchgruppen signifikant höhere Lernzuwächse auf als diejenigen aus der Vergleichsgruppe – mit Effektstärken bis zu einer Größe, welche mit dem durchschnittlichen Lernzuwachs eines halben Schuljahres vergleichbar ist.

Es gibt aber auch Ergebnisse, welche keinen Effekt der Schulbuchwahl auf den Lernerfolg zeigen. So konnten van Steenbrugge et al. (2013) in ihrer querschnittlichen Analyse der Daten von 1.579 Schülerinnen und Schülern aus 89 belgischen Klassen der Stufen 1–6 keine wesentlichen Unterschiede zwischen den Leistungen der Schülerinnen und Schüler gruppiert nach den fünf verwendeten Schulbuchreihen finden. In Übereinstimmung damit fanden Blazar et al. (2019) keine Unterschiede im Leistungswachstum von Schülerinnen und Schülern, welche unterschiedliche Mathematikschulbücher für die Grundschule genutzt haben. Ihre Stichprobe umfasste Klassen mit etwa 1.200 Lehrern aus sechs US-Bundesstaaten, die alle die Common Core State Standards (CCSS) übernommen haben.

Auf der Grundlage öffentlich zugänglicher und hoch aggregierter Leistungsdaten auf Schulebene gibt es des Weiteren in bildungsökonomischen Studien mit Daten aus Kalifornien, Florida und Indiana Hinweise auf Effekte von unterschiedlichen Schulbüchern auf die Leistung der Lernenden (Bhatt & Koedel, 2012; Bhatt et al., 2013; Koedel et al., 2017). Diese Ergebnisse beruhen allerdings auf Blackbox-Modellen, welche keine Erklärungsansätze für die aufgezeigten Effekte bieten. So stellen Koedel et al. (2017) fest: „Our ability to gain inference into the mechanisms underlying the differential curriculum effects that we estimate here, and related estimates elsewhere, is limited by the lack of a larger literature within which our findings can be contextualized. In analytic terms, empirical analyses that would aim to link specific textbook characteristics and/or contextual evaluation factors (such as the assessment used) to efficacy estimates are currently hampered by underidentification — there are too many potential explanatory factors and too few efficacy estimates“ (Koedel et al., 2017, S. 14). Darüber hinaus gibt es in den USA nicht in jedem Bundesstaat landesweit verbindliche Curricula, sodass Schulen zwischen verschiedenen Lehrplänen wählen können, die wiederum durch unterschiedliche Schulbücher repräsentiert werden (Mullis et al., 1997). Entsprechend werden Schulbücher in

diesen Studien als Repräsentanten für verschiedene Curricula betrachtet bzw. die Begriffe Schulbuch und Curriculum synonym verwendet (z. B. Blazar et al., 2019). Im Gegensatz dazu werden Schulbücher in dieser Arbeit als unterschiedliche Realisationen *eines* Curriculums betrachtet und entsprechend die Lernwirksamkeit von Schulbüchern als Wirkung dieser unterschiedlichen Realisationen. Während Effekte durch unterschiedliche Curricula also beispielsweise durch verschiedene verbindliche Inhalte oder Methoden bedingt sein können, lässt sich dies bei Schulbüchern ausschließen, die das gleiche Curriculum adressieren. Dementsprechend sind mögliche Effekte durch das Schulbuch in den Studien dieser Arbeit auch nicht durch Unterschiede im Curriculum konfundiert.

Zusammenfassend lässt sich festhalten, dass es einige Studien gibt, die auf einen möglichen Einfluss des Mathematikbuchs auf den Lernerfolg hinweisen, der Forschungsstand allerdings nicht konsistent ist. Dementsprechend stellte Fan (2013) bereits vor einigen Jahren in seinem Überblicksartikel zur Mathematikbuchforschung fest, dass sich der Großteil der Studien mit deskriptiven Fragen beschäftigt, wie beispielsweise ein spezielles Thema in Schulbüchern behandelt wird, während es zu wenig Studien mit korrelativen und kausalen Fragestellungen gibt. Zudem ist die Aussagekraft vieler Untersuchungen aufgrund kleiner Stichprobengrößen oder querschnittlicher Designs beschränkt (Fan, 2013). Entsprechend fordert er eine Weiterentwicklung der Schulbuchforschung im Bereich Mathematik über die Schulbuchanalyse, den Schulbuchvergleich und die Schulbuchnutzung hinaus hin zu einem Forschungsparadigma, das mehr empirische und experimentelle Methoden einsetzt und Schulbuchforschung strenger als wissenschaftliche Forschung betrachtet. In Übereinstimmung damit ist die Feststellung von Koedel et al. (2017), dass es empirischer Untersuchungen bedarf, welche spezifische Eigenschaften von Schulbüchern mit kontextuellen Faktoren wie dem Lernerfolg von Schülerinnen und Schülern verknüpfen.

Da potentielle Effekte von Schulbüchern auf deren Inhalten basieren, ist für die Studien dieser Arbeit ein Forschungszugang gewählt worden, welcher einzelne curriculare Themen genauer analysiert. Dazu werden die drei Themengebiete Adaptives Rechnen bei der Addition und Subtraktion bis 1.000 (Studie 1), die Nutzung Operativer Grundaufgaben (Studie 2) und quantitative Vergleiche (Studie 3) betrachtet, die in den Abschnitten 2.2–2.4 genauer dargestellt werden.

2.2 Adaptives Rechnen bei der Addition und Subtraktion bis 1.000

Der folgende Abschnitt stellt das Themengebiet des adaptiven Rechnens dar und konkretisiert es für den Bereich der Addition und Subtraktion bis 1.000. Dazu wird zunächst das adaptive Rechnen differenziert beschrieben und für den entsprechenden Bereich operationalisiert. Anschließend wird der Stand der Forschung dargestellt.

Die Idee des adaptiven Rechnens wird in der Literatur auf unterschiedliche Weise interpretiert (z. B. Rathgeb-Schnierer, 2006; Star & Newton, 2009; Threlfall, 2002; Verschaffel et al., 2009). Dieser Arbeit liegt die Interpretation als adaptive Nutzung von Strategien beim Lösen mathematischer Probleme zugrunde, welche Christoph Selter (2009) wie folgt definiert: „Adaptivity

is the ability to creatively develop or to flexibly select and use an appropriate solution strategy in a (un)conscious way on a given mathematical item or problem, for a given individual, in a given sociocultural context.“ (Selter, 2009, S. 624). Diese Definition umfasst mehrere Aspekte, welche bereits von anderen Autorinnen und Autoren genannt worden sind (z. B. Siegler & Lemaire, 1997; Threlfall, 2009; Verschaffel et al., 2009). Adaptivität beschreibt demnach die Fähigkeit, geeignete Lösungsstrategien zu nutzen, welche kreativ entwickelt bzw. flexibel ausgewählt worden sind. Dabei ist die Eignung der Lösungsstrategie – und damit auch der vorhergehende Prozess der Entwicklung bzw. der Auswahl – abhängig von drei Faktoren: (1) dem mathematischen Problem bzw. Aufgabe, (2) dem problemlösenden Individuum sowie (3) dem soziokulturellen Kontext. Mit Fokus auf den Faktor des mathematischen Problems, dargestellt am Beispiel der Aufgabe 401 – 399, ist die Strategie des Ergänzens, d. h. der Fragestellung „Wieviel muss zu 399 hinzugefügt werden, um 401 zu erhalten?“ effizienter und damit geeigneter als der schriftliche Subtraktionsalgorithmus, da erstere weniger Rechenschritte mit einer oftmals geringeren kognitiven Belastung erfordert und damit insgesamt weniger fehleranfällig ist (siehe Abschnitt 4.2.2.2). Aus Perspektive von Lernenden (Faktor Individuum), die die schriftliche Subtraktion sehr gut und weitere Strategien weniger gut beherrschen, wäre die Nutzung des schriftlichen Rechenverfahrens geeignet. Legen es die soziomathematischen Normen im Unterricht (Faktor soziokultureller Kontext) nahe, dass die Schülerinnen und Schüler die Erwartung ihrer Lehrkraft antizipieren, eine bestimmte Strategie zum Lösen einer Aufgabe zu verwenden, ist eine Nutzung dieser als geeignet zu betrachten (z. B. wenn die Aufgabe 401 – 399 beim Üben der schriftlichen Subtraktion gestellt wird).

In dieser Arbeit und damit insbesondere in Studie 1 wird Adaptivität aus normativer Perspektive mit Fokus auf das mathematische Problem betrachtet. Entsprechend wird die Adaptivität der Strategiewahl der Schülerinnen und Schüler in Abhängigkeit der zu lösenden Probleme bewertet. Das ist keineswegs in einer Beurteilung dieser drei Faktoren (Aufgabe – Individuum – soziokultureller Kontext) begründet, sondern einerseits darin, dass die adaptive Nutzung von Strategien zum effizienten Lösen verschiedener Probleme in vielen Ländern ein curriculares Lernziel und damit normativ gesetzt ist (Baroody & Dowker, 2003; Heinze et al., 2009; Mullis et al., 2016; van den Heuvel-Panhuizen, 2001). Andererseits sind es pragmatische Gründe bei der Sekundäranalyse eines bestehenden Datensatzes, in dem weder Informationen zu Eigenschaften und Präferenzen der Schülerinnen und Schüler noch zu soziomathematischen Normen bei der Strategienutzung enthalten sind. Untersuchungen der individuellen Charakteristik der Lernenden bei ihrer Strategienutzung erfordern komplexere Designs. So haben etwa Siegler und Lemaire (1997) das Choice/No-Choice-Design entwickelt, in welchem Schülerinnen und Schüler in einer oder mehreren Bedingungen Probleme mit einer vorgegebenen Strategie lösen sollen und in einer weiteren Bedingung ihren Lösungsweg frei wählen können (z. B. die Studie von Hickendorff, 2020). Zur Erhebung soziomathematischer Normen zur Strategienutzung bedarf es zudem eines weitaus breiteren Instrumentariums (vgl. Cooper et al., 1996; Nunes et al., 1987; Ruthven, 1998; Thompson, 1994) – beispielsweise zu Erfassung von Überzeugungen der Lehrkräfte oder der Instruktion im Unterricht, aber auch wahrgenommener Erwartungen von Lernenden – als das, was für den in dieser Arbeit genutzten Datensatz verwendet wurde.

Aus der oben genannten Definition von Selter (2009) lässt sich des Weiteren die Unterscheidung von flexiblem und adaptivem Rechnen hervorheben. Während einige Autorinnen und Autoren die Begriffe synonym verwenden (z. B. Blöte et al., 2001; Threlfall, 2002), werden sie in dieser Arbeit unterschieden. Die flexible Strategienutzung ist die Fähigkeit, beim Lösen verschiedener Probleme zwischen unterschiedlichen Strategien zu wechseln. Die adaptive Strategienutzung differenziert sich durch das zusätzliche Kriterium, dass mit diesem Wechsel auch eine Eignung der Strategie in Bezug auf einen der drei Faktoren (hier mit Fokus auf die Aufgabe) verbunden ist.

Die Strategien bei der Addition und Subtraktion bis 1.000 werden typischerweise klassifiziert in Kopfrechnen, halbschriftliches Rechnen und schriftliches Rechnen auf Basis der schriftlichen Algorithmen (Krauthausen, 1993; Padberg & Benz, 2011). Krauthausen (1993) führt außerdem den Einsatz des Taschenrechners an, den wir hier nicht weiter betrachten, da er in unserer Untersuchung keine und im Mathematikunterricht der Grundschule keine große Rolle spielt. Zum Kopfrechnen werden dabei alle Verfahren und Strategien gezählt, die gänzlich im Kopf und ohne Notationen erfolgen (ebd.), das heißt alle Zwischenschritte werden mental durchgeführt. Zum halbschriftlichen Rechnen werden alle Strategien gezählt, bei denen Zwischenschritte zwar schriftlich erfolgen, die Rechenwege jedoch nicht einem Standardalgorithmus entsprechen (ebd.). Dies reicht von der Notation einzelner Zahlen oder Zeichen bis hin zum vollständigen Aufschreiben sämtlicher Zwischenschritte bzw. Teilrechnungen der Aufgabengleichung. Bei den schriftlichen Algorithmen werden die Zahlen untereinander geschrieben und stellen- bzw. ziffernweise gerechnet. Daher werden diese in der englischsprachigen Literatur auch als „digit-based strategies“ (dt. Ziffernrechnen, z. B. Torbeyns et al., 2017) bezeichnet. Die halbschriftlichen Strategien, bei denen mit der Zahl als Ganzes bzw. einer Zerlegung dieser gerechnet wird, werden entsprechend als „number-based strategies“ (dt. Zahlenrechnen) abgegrenzt (Kilpatrick et al., 2001). Der Übergang von den halbschriftlichen Strategien zum Kopfrechnen wird dabei als fließend beschrieben. Das heißt, beim Kopfrechnen handelt es sich vornehmlich um internalisierte halbschriftliche Strategien, welche mit zunehmender Erfahrung und Übung immer weniger Notationen benötigen (Padberg & Benz, 2011). Das mentale Ausführen der schriftlichen Algorithmen wird dagegen als aufwändig und fehleranfällig beschrieben (Kilpatrick et al., 2001), falls es für Kinder überhaupt im Kopf durchführbar ist. Daher ist davon auszugehen, dass Lernende, die eine Additions- oder Subtraktionsaufgabe korrekt mental lösen, dazu vornehmlich internalisierte halbschriftliche Strategien verwenden.

In ihrem Modell „Ebenen des Lösungsprozesses“ (Rathgeb-Schnierer, 2011, S. 16) unterscheidet Rathgeb-Schnierer den Strategieeinsatz noch differenzierter in den Ebenen *Formen* (schriftliches Rechnen, halbschriftliches Rechnen und Kopfrechnen), *Referenzen* (Verfahrens- oder Beziehungsorientierung) und *Lösungswerkzeuge* (Zählen, Faktenwissen und strategische Werkzeuge). Aufgrund der Untersuchung schriftlicher Lösungen und Darstellungen von Lösungswegen in Schulbüchern wird der Strategiebegriff in dieser Arbeit etwas breiter verwendet und umfasst alle Vorgehensweisen, die zu einer (korrekten oder nicht korrekten) Lösung führen. Es lässt sich dabei anhand einer halbschriftlichen Rechnung nicht oder nur schwierig beurteilen, ob diese beziehungs- oder verfahrensorientiert eingesetzt wurde und ob sie zählend, mithilfe

von Basisfakten oder mithilfe eines anderen strategischen Werkzeugs gelöst wurde. Aufgrund der Analyse der Effizienz der Strategienutzung lassen sich aber die meisten Strategien, die beim adaptiven Addieren und Subtrahieren bis 1.000 aus normativer Perspektive eine Rolle spielen, beim halbschriftlichen Rechnen oder Kopfrechnen (Form), der Beziehungsorientierung (Referenz) sowie den strategischen Werkzeugen (Lösungswerkzeuge) verorten. Im Gegensatz zum schriftlichen Rechnen, welches einen einheitlichen Lösungsweg vorgibt, bietet das halbschriftliche Rechnen eine große Bandbreite an möglichen Strategien mit unterschiedlicher Effizienz, die wiederum vom gegebenen mathematischen Problem abhängt. Entsprechend sind die halbschriftlichen Strategien für das adaptive Rechnen von großer Bedeutung, da ein Repertoire an unterschiedlichen Strategien die Grundlage für eine adaptive Strategiewahl darstellt. Tabelle 2.1 gibt eine Übersicht an halbschriftlichen Additions- und Subtraktionsstrategien, welche üblicherweise in Schulbüchern und im Unterricht eingeführt bzw. behandelt werden, und zeigt je eine Variante des schriftlichen Algorithmus (vgl. Padberg, 2005; für internationale Vergleiche siehe Threlfall, 2002). Während die schriftlichen Algorithmen und die Strategien "Schrittweise" und "Stellenweise" als universelle Strategien eingeführt werden, die sich auf fast alle Arten von Additions- bzw. Subtraktionsproblemen anwenden lassen, werden die Strategien "Schrittweise alternativ", "Hilfsaufgabe", "Vereinfachen" und "Ergänzen" in der Regel als aufgabenspezifische Strategien betrachtet, die nur für bestimmte Problemtypen geeignet, bei diesen jedoch sehr effizient sind. Aufgrund dieser hohen Effizienz geht man auch davon aus, dass die adaptive Anwendung halbschriftlicher Strategien die mentalen Rechenfähigkeiten besonders bei mehrstelligen Zahlen fördert (Padberg & Benz, 2011).

Neben der Relevanz für informelle Kontexte und die Alltagsnutzung Erwachsener wird die Förderung des Kopfrechnens für die Entwicklung des Verständnisses vom dezimalen Stellenwertsystem und allgemeiner Problemlösungsfähigkeiten sowie für das erfolgreiche Erlernen der nach dem halbschriftlichen Rechnen eingeführten schriftlichen Algorithmen als wichtig erachtet (Threlfall, 2002). Grundschülerinnen und Grundschüler sind zudem kreativ, wenn es darum geht, weitere halbschriftliche Strategien zu erfinden und die zuvor erlernten Strategien zu modifizieren (z. B. Heinze et al., 2018; Selter, 2001; Threlfall, 2002; siehe auch Anhang A).

Bezüglich des Stellenwerts der Algorithmen gibt es in der Forschungsliteratur einerseits eine allgemeine Akzeptanz der Forderung, dass Schülerinnen und Schüler für jede Operation mindestens einen schriftlichen Algorithmus lernen sollten (Verschaffel et al., 2007), andererseits aber auch die Übereinstimmung, dass ein verfrühtes Einführen schriftlicher Algorithmen aufgrund eines fehlenden Verständnisses vermieden werden sollte (ebd.). Auch Selter (2001) führt eine geringe Rolle der schriftlichen Algorithmen für Schülerinnen und Schüler in der Sekundarschule sowie für Alltagssituationen von Kindern und Erwachsenen auf (z. B. Hart, 1981; Nunes et al., 1993; Scribner, 1984). Entsprechend hat sich über die Bedeutung des adaptiven Rechnens in der Arithmetik seit Ende der 1990er-Jahre in vielen Ländern ein breiter Konsens entwickelt (z. B. Baroody & Dowker, 2003; Csíkos, 2016; Kilpatrick et al., 2001; National Council of Teachers of Mathematics, 2003; Selter, 2001; Threlfall, 2009; Torbeyns et al., 2017; Verschaffel et al., 2007). Die Diskussion darüber kam in den Jahren zuvor auf, als mehrere Längsschnittstudien Einblicke in die konzeptionellen Strukturen von Kindern für mehrstellige Zahlen lieferten

Schriftlicher Algorithmus	Halbschriftliche Strategien					
	Schrittweise	Stellenweise	Schrittweise alternativ	Hilfsaufgabe	Vereinfachen	Ergänzen
$\begin{array}{r} 198 \\ + 456 \\ \hline 11 \\ 654 \end{array}$	$\begin{array}{r} 198 + 456 = 654 \\ 198 + 400 = 598 \\ 598 + 50 = 648 \\ 648 + 6 = 654 \end{array}$	$\begin{array}{r} 198 + 456 = 654 \\ 100 + 400 = 500 \\ 90 + 50 = 140 \\ 8 + 6 = 14 \\ 500 + 140 + 14 = \\ 645 \end{array}$	$\begin{array}{r} 198 + 456 = 654 \\ 198 + 2 = 200 \\ 200 + 454 = 654 \end{array}$	$\begin{array}{r} 456 + 198 = 654 \\ 456 + 200 = 656 \\ 656 - 2 = 654 \end{array}$	$\begin{array}{r} 456 + 198 = 654 \\ 454 + 200 = 654 \end{array}$	
$\begin{array}{r} 403 \\ - 398 \\ \hline 11 \\ 005^{**} \end{array}$	$\begin{array}{r} 403 - 398 = 5 \\ 403 - 300 = 103 \\ 103 - 90 = 13 \\ 13 - 8 = 5 \end{array}$	$\begin{array}{r} 345 - 123 = 222 \\ 300 - 100 = 200 \\ 40 - 20 = 20 \\ 5 - 3 = 2^{***} \end{array}$	$\begin{array}{r} 403 - 398 = 5 \\ 403 - 3 = 400 \\ 400 - 395 = 5 \end{array}$	$\begin{array}{r} 403 - 398 = 5 \\ 403 - 400 = 3 \\ 3 + 2 = 5 \end{array}$	$\begin{array}{r} 403 - 398 = 5 \\ 405 - 400 = 5 \\ 398 + 5 = 403 \end{array}$	

*Die Strategie „Ergänzen“ ist nur auf Subtraktionsprobleme anwendbar.
 **Es existieren unterschiedliche Versionen für den schriftlichen Subtraktionsalgorithmus. Diese Version war im Lehrplan unserer Stichprobe vorgeschrieben.
 ***Die Empfehlungen für die Behandlung von Überträgen beim stellenweisen Subtrahieren sind uneinheitlich, da negative Zahlen in der Grundschule nicht behandelt werden. Dies ist ein Beispiel ohne Überträge.

Tabelle 2.1: Gänge im Unterricht eingeführte Additions- und Subtraktionsstrategien

(z. B. Fuson et al., 1997). Es zeigte sich, dass eine hohe Adaptivität der Strategiewahl mit einem breiteren konzeptuellen Wissen, einem tieferen Verständnis des dezimalen Stellenwertsystems oder einer geringeren Fehleranfälligkeit beim Lernen der Algorithmen einhergeht (Carpenter et al., 1997; Hiebert & Wearne, 1996). Strategische Variabilität wird außerdem als Beginn eines mathematischen Denkens angesehen, welches über die Kenntnis von Fakten und Verfahren hinausgeht, und als Voraussetzung für das Lösen von Problemen der realen Welt. Sie wird daher als eine ebenso wichtige Komponente in der Entwicklung mathematischer Kompetenz angesehen wie die Fähigkeit, eine mathematische Aufgabe korrekt zu lösen (Threlfall, 2009; Torbeyns et al., 2017). Dementsprechend wurde die adaptive Strategienutzung in viele Mathematiklehrpläne oder -standards aufgenommen (Baroody & Dowker, 2003; Heinze et al., 2009; van den Heuvel-Panhuizen, 2001), was sich ebenfalls auf die Lehrkräfteausbildung auswirkte: Adaptives Rechnen wurde zu einem Lernziel der Mathematik in der Grundschule sowie zu einem Bestandteil in der Ausbildung von Mathematiklehrkräften.

Die Entwicklung strategischer Kompetenz wird in dem Modell von Lemaire und Siegler (1995) beschrieben. Sie unterscheiden darin vier Dimensionen, für welche sie zeigen konnten, dass eine Veränderung in jeder einzelnen von ihnen die Geschwindigkeit und Genauigkeit der Strategiewahl insgesamt verbessern kann (Lemaire & Siegler, 1995). Diese Dimensionen sind:

- das *Strategierepertoire* als die Kenntnis verschiedener Strategien;
- die *Strategieverteilung* als das Wissen über die relativen Nutzungshäufigkeiten der einzelnen Strategien;
- die *Strategieeffizienz* als die Fähigkeit, die Strategien schnell und korrekt auszuführen;
- die *Strategiewahl* als die Fähigkeit, eine Strategie für ein gegebenes Problem flexibel auszuwählen.

Basierend auf diesem Modell und auf einer Reihe von Studien sowie computergestützten Modellen modellierten Chen und Siegler (2000), wie Schülerinnen und Schüler neue Strategien und deren adaptiven Einsatz lernen. Ihre Ergebnisse deuten darauf hin, dass der Erwerb adaptiver Strategiekompetenz in der Arithmetik auf Grundlage geeigneter Aufgaben erfolgt, die es den Lernenden ermöglichen, Strategien zu entdecken und sie selbst auszuführen. Dies fördert das schnelle und sichere Ausführen und verringert die kognitive Belastung. Zudem bedarf es der Anregung metakognitiver Prozesse wie der Untersuchung der Effizienz einzelner Strategien, welches die Assoziation zwischen bestehenden Strategien und Problemtypen stärkt.

Eine alternative Perspektive zum Strategiewahlmodell nimmt Threlfall (2002, 2009) ein. Er kritisiert, dass das Modell der aktiven Wahl einer Strategie der Komplexität des Problemlösens von Schülerinnen und Schülern nicht gerecht wird. Statt einer aktiven Wahl beschreibt er den Problemlöseprozess als schrittweise fortschreitende Interaktion zwischen dem konzeptuellen Wissen der Lernenden und den Eigenschaften des Problems. Daher stellt er dem Modell der Strategiewahl sein *Modell der Strategieemergenz* entgegen (Threlfall, 2009). Nach diesem ist die explizite Vermittlung von Strategien und deren adaptiver Einsatz ungeeignet für das Erlernen adaptiver Strategienutzung (Heinze et al., 2018). Sie sollte sogar vermieden werden, da sie

nicht kreativ sei und keine Analyse von Problemmerkmalen beinhalte (Threlfall 2009). Im Gegensatz zum Strategiewahlmodell ist dieses Modell jedoch empirisch nicht fundiert. Es betont allerdings die Bedeutung von konzeptuellem Vorwissen und dem Vergleich von Lösungsstrategien, die auch empirische Studien bei der Untersuchung der Strategienutzung bestätigt haben (z. B. Fuson et al., 1997; Rittle-Johnson et al., 2009; Rittle-Johnson & Star, 2007, 2009; Schulz, 2018; Threlfall, 2009).

Entgegen der Vermutung von Threlfall (2009), dass eine explizite Einführung von Strategien und deren adaptivem Einsatz für dessen adaptive Nutzung durch die Lernenden nicht notwendig bzw. sogar ungünstig ist, sind die Effekte unterschiedlicher Unterrichtsansätze zur Strategieinstruktion in einer Reihe von Studien untersucht worden (z. B. Klein et al., 1998; Heinze et al., 2018, 2020; Rittle-Johnson & Star, 2007, 2009; Torbeyns et al., 2009a, 2009b). Als besonders förderlich für eine adaptive Strategienutzung erwiesen sich dabei die frühzeitige Betonung einer flexiblen Strategienutzung (Blöte et al., 2001; Klein et al., 1998) sowie der Vergleich und die Gegenüberstellung verschiedener Lösungsstrategien (Rittle-Johnson & Star, 2007, 2009). Außerdem wird dem arithmetischen Vorwissen eine große Bedeutung beigemessen (z. B. Fuson et al., 1997; Rittle-Johnson et al., 2009; Schulz, 2018; Threlfall, 2009).

Empirische Untersuchungen haben dagegen wiederholt gezeigt, dass die Adaptivität der Strategienutzung bei dem Großteil der Schülerinnen und Schüler wenig ausgeprägt ist (z. B. Csíkos, 2016; Heinze et al., 2009; Torbeyns et al., 2017; Torbeyns & Verschaffel, 2016). Trotz einer feststellbaren Aufwertung der halbschriftlichen Strategien im Vergleich zum traditionellen Unterricht (vgl. Krauthausen, 1993) neigen viele Lernende noch immer dazu, pro Operation eine Strategie zu bevorzugen, die sie auf alle Problemtypen anwenden. Während also aus der Forschung eine Reihe von Erkenntnissen und Angeboten zur Förderung des adaptiven Rechnens geboten wird und die nationalen Curricula und Standards die Vermittlung verschiedener Strategien und deren Nutzung vorsehen, zeigen empirische Ergebnisse zur Adaptivität der Strategienutzung von Schülerinnen und Schülern beim Lösen arithmetischer Probleme wenig Fortschritt. Ein Grund dafür könnte ein Mangel an adäquaten Lerngelegenheiten im Mathematikunterricht oder zumindest eine mangelnde Qualität der Lerngelegenheiten in den Lernressourcen der Schülerinnen und Schüler sein.

2.3 Nutzung Operativer Grundaufgaben

Die Nutzung Operativer Grundaufgaben als Kern der operativen Beziehungen in Klassenstufe 1 stellt das zweite betrachtete Themengebiet dar und wird in diesem Abschnitt zusammenfassend beschrieben. Dazu wird zuerst der Begriff eingeordnet und erörtert, bevor der aktuelle Forschungsstand dargestellt wird.

Operative Grundaufgaben bilden in Klassenstufe 1 den Kern der operativen Beziehungen zur Addition und Subtraktion (Schipper, 2009). Letztere beschreiben Zusammenhänge zwischen Zahlen und Operationen oder mit anderen Worten: das Netzwerk, in welchem Rechenaufgaben miteinander verbunden sind. Beispielhaft ist dies in Abbildung 2.1 für die Aufgabe $3 + 5 = 8$ dargestellt.

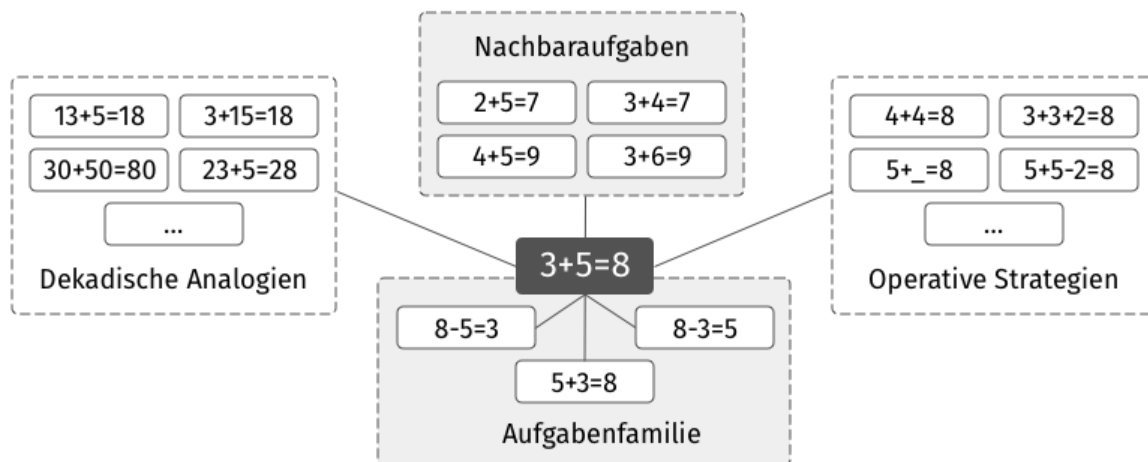


Abbildung 2.1: Operative Beziehungen am Beispiel der Aufgabe $3 + 5 = 8$

Durch Vertauschen der beiden Summanden 3 und 5 erhält man die Tauschaufgabe $5 + 3$, welche aufgrund der Kommutativität der Addition die gleiche Summe ergibt. Durch die Komplementarität von Addition und Subtraktion – letztere ist die Umkehrung der ersteren – ergeben sich entsprechend zwei Umkehraufgaben. Diese vier Additions- bzw. Subtraktionsaufgaben, die sich aus drei positiven ganzen Zahlen bilden lassen, werden unter dem Begriff der *Aufgabenfamilie* zusammengefasst. Erhöht oder vermindert man einen der Summanden der Ausgangsaufgabe um 1, so erhöht bzw. vermindert sich das Ergebnis entsprechend. Die vier sich so ergebenden Aufgaben werden als *Nachbaraufgaben* bezeichnet. Die Aufgabenfamilie und die Nachbaraufgaben fasst Schipper (2009) unter dem Begriff *Operative Grundaufgaben* zusammen. Bereits im Zahlenraum ab 10, vor allem aber über den Zahlenraum bis 20 hinaus, ergeben sich Analogieaufgaben: Aus der Assoziativität der Addition lassen sich dekadische Analogien herleiten wie $13 + 5 = (10 + 3) + 5 = 10 + (3 + 5) = 10 + 8 = 18$, durch die die Aufgabe $13 + 5$ auf die einfachere Aufgabe $3 + 5$ zurückgeführt werden kann. Aus dem Distributivgesetz ergeben sich weitere dekadische Analogien wie $30 + 50 = (10 \times 3) + (10 \times 5) = 10 \times (3 + 5) = 10 \times 8 = 80$. Darüber hinaus können weitere Beziehungen abgeleitet werden, die unter dem Begriff *Operative Strategien* zusammengefasst werden (Schipper, 2009). Dazu gehört beispielsweise das gegensinnige Verändern (z. B. $4 + 4 = 8$), das ebenfalls auf der Assoziativität der Addition beruht und bei dem einer der Summanden erhöht wird, während der andere um den gleichen Betrag vermindert wird, sich dadurch die Summe also nicht verändert. Weitere Beispiele für operative Strategien sind das Ergänzen bei Subtraktionsproblemen ($5 + _ = 8$) oder das Verdoppeln ($4 + 4 = 8$) bzw. Fast-Verdoppeln ($4 + 4 + 1 = 9$). Die nicht ausgefüllten Kästchen in Abbildung 2.1 sollen zum Ausdruck bringen, dass dies nur eine kleine Auswahl an dekadischen Analogien und operativen Strategien darstellt. In Studie 2 dieser Arbeit werden die Operativen Grundaufgaben fokussiert. Dies liegt zum einen an ihrer zentralen Rolle in Bezug auf operative Beziehungen in Klassenstufe 1 und der Trennschärfe, mit der sie sich innerhalb dieser umranden lassen. Zum anderen sind es pragmatische Gegebenheiten des genutzten Datensatzes bzw. der darin enthaltenen Items, welche eine inhaltliche Fokussierung nötig machen.

Das Ablösen von Zählstrategien beim Lösen von Additions- und Subtraktionsproblemen zugunsten von Rechenstrategien und zugunsten von dem Auswendigwissen des kleinen Einspluseins bzw. Einsminuseins ist ein zentrales Ziel des Mathematikunterrichts der ersten Schuljahre (Kilpatrick et al., 2001; Padberg & Benz, 2011; Wright, 2009; Verschaffel et al., 2007). Die meisten Schülerinnen und Schüler können bereits zu Schulbeginn einfache Rechenaufgaben durch Zählstrategien lösen – zumindest mithilfe von Anschauungsmaterial, wie Bausteinen oder Plättchen (Clements & Sarama, 2007; Kilpatrick et al., 2001). Während Zählstrategien auf den Vorkenntnissen der Kinder aufbauen und somit einen Kernbestandteil des mathematischen Anfangsunterrichts darstellen, ist ihre andauernde Anwendung problematisch (Padberg & Benz, 2011). Die Anwendbarkeit und Effizienz von Zählstrategien ist begrenzt, insbesondere wenn die zu lösenden Probleme über den Zahlenraum bis 20 hinausgehen, und sie verursachen verschiedene Schwierigkeiten im weiteren Lernprozess (Cowen, 2003). Der Aufbau von Assoziationen zwischen einem Problem und seiner korrekten Lösung und damit die allmähliche Verinnerlichung dieser Verbindung wird durch den Zählprozess erschwert, vor allem bei Lernenden, die Zählstrategien häufig fehlerhaft anwenden (Siegler et al., 2006). Dieses Hindernis wiederum ist für den weiteren Lernprozess von Nachteil, da sich das Assoziationsnetzwerk der Zahlenkombinationen des kleinen Einspluseins bzw. Einsminuseins kumulativ entwickelt. Folglich sollte die Verwendung von Rechenstrategien die Verwendung von Zählstrategien als Hauptlösungsmethode mit der Zeit ersetzen.

Ein zentrales Ziel des Mathematikunterrichts im ersten Schuljahr ist es also, dass die Kinder die Zahlensätze des kleinen Einspluseins und Einsminuseins auswendig wissen oder mithilfe von Rechenstrategien ableiten können (Padberg & Benz, 2011). Im Modell der „Ebenen des Lösungsprozesses“ (Rathgeb-Schnierer, 2011) ist dies auf der Ebene der *Lösungswerkzeuge* demnach die Entwicklung weg vom Zählen hin zum Lösen mithilfe heuristischer beziehungsweise operativer Strategien (ableitende Strategien) oder dem Abrufen von Fakten. Das Erlernen dieser Grundrechenarten und ihrer Zahlensätze wird dabei als andauernder Prozess beschrieben, da das Wissen über Summen und Differenzen beim Addieren und Subtrahieren relational ist und daher nicht durch reines Auswendiglernen erfolgen kann: „[Children] do not move from knowing nothing about the sums and differences of numbers to having the basic number combinations memorized“ (Kilpatrick et al., 2001, p. 182). Stattdessen handelt es sich um eine längerfristige Entwicklung, in welcher die Lernenden nach und nach elaboriertere und abstraktere Strategien zur Lösung von Additions- und Subtraktionsproblemen entwickeln, indem sie die zugrundeliegenden operativen Beziehungen entdecken und ausnutzen (Rechtsteiner-Merz, 2013). Auch Verschaffel et al. (2007) berichten gut dokumentierte Evidenz für die schrittweise Entwicklung der Beherrschung des kleinen Einspluseins bzw. Einsminuseins durch die Kinder, welche dem Paradigma von Drill und Übung der Zahlensätze widerspricht. „[Lernende] müssen die grundlegende Einsicht erlangen: ‚Schwierige Aufgaben löst man über geeignete leichte Aufgaben‘ unter Ausnutzung von Rechenstrategien.“ (Benz, 2005, S. 60, nach Winter, 1996). Entsprechend kommt der Nutzung der Operativen Grundaufgaben eine zentrale Bedeutung beim Lernen des kleinen Einspluseins und Einsminuseins zu. Durch die Ausnutzung der Beziehungen von Tausch-, Umkehr- und Nachbaraufgaben lassen sich aus einer einzelnen Aufgabe und ihrer

Lösung effizient die Lösungen weiterer Aufgaben ableiten. Im Modell der „Ebenen des Lösungsprozesses“ sind diese Ableitungsstrategien auf der Ebene *Strategische Werkzeuge* einzuordnen (Rathgeb-Schnierer, 2011). Da Studie 2 die Nutzung der operativen Strategien fokussiert und die Untersuchung damit auf das Lösungswerkzeug der Ableitungsstrategien beschränkt ist, wird im Folgenden für das Ableiten aus den Operativen Grundaufgaben der Strategiebegriff verwendet. Er ist aber im Sinne eines strategischen Werkzeugs nach Rathgeb-Schnierer (2011) gemeint. Tabelle 2.2 gibt eine Übersicht der drei Strategien, welche die Tausch-, Umkehr- bzw. Nachbaraufgaben nutzen.

Strategie	Definition	Beispiel
Nutzung der Tauschaufgabe	Ableitung der Lösung aus der Tauschaufgabe	$2 + 9 = \underline{\quad}$
		$\rightarrow 9 + 2 = \underline{11}$
		$\underline{2 + 9 = 11}$
Nutzung der Umkehraufgabe	Ableitung der Lösung aus der Umkehraufgabe	$17 - 15 = \underline{\quad}$
		$\rightarrow 15 + \underline{2} = 17$
		$\underline{17 - 15 = 2}$
Nutzung der Nachbaraufgabe	Ableitung der Lösung aus der Nachbaraufgabe	$8 + 9 = \underline{\quad}$
		$\rightarrow 8 + 8 = 16$
		$\rightarrow 8 + (8 + 1) = 17$
		$\underline{8 + 9 = 17}$

Tabelle 2.2: Übersicht über die drei Strategien, welche die Operativen Grundaufgaben nutzen.

Lernende, die in der Lage sind, alle drei Strategien sicher und schnell anzuwenden, können aus der Lösung einer ihnen bekannten Aufgabe die Lösungen für sieben weitere Additions- und Subtraktionsprobleme ableiten. Dadurch verringert sich ihr Aufwand für das Auswendiglernen der Zahlensätze drastisch.

Trotz der dargestellten Bedeutung sowie der Einigkeit von Vertreterinnen und Vertretern „verschiedener didaktischer Positionen [...], dass die Rechenstrategien des Ableitens auf Grund operativer Beziehungen anzustreben sind“ (Benz, 2005, S. 66), stellen Zählstrategien den Hauptlösungsweg für viele Schülerinnen und Schüler über die erste Klasse hinaus dar (Benz, 2005). Es gibt zudem empirische Befunde, die darauf hindeuten, dass Grundschulkindern wichtige Strategien selten ohne explizite Darstellung selbst entwickeln (De Smedt et al., 2010; Heinze et al., 2018). Weitere Ergebnisse berichten von Vorteilen einer frühzeitigen Betonung der flexiblen Strategienutzung und unterstützen damit auch den vermehrten Umgang mit unterschiedlichen Strategien in frühen Jahrgangsstufen (Blöte et al., 2001; Klein et al., 1998). Darüber hinaus wei-

sen mehrere experimentelle Studien darauf hin, dass das Vergleichen verschiedener Lösungsstrategien durch die Schülerinnen und Schüler die Flexibilität ihrer Strategieauswahl fördert (z. B. Klein et al., 1998; Rittle-Johnson & Star, 2007, 2009). Auf Basis dessen lässt sich vermuten, dass die Rolle der genutzten Lernressourcen auch für eine bessere Förderung und Implementation der Nutzung Operativer Grundaufgaben von Bedeutung ist. Zur Beschreibung einer möglichen Kompetenzentwicklung bei der Nutzung operativer Strategien lässt sich analog das Modell nach Lemaire und Siegler (1995) heranziehen, welches die vier Dimensionen *Strategiepalette*, *Strategieverteilung*, *Strategieeffizienz* und *Strategiewahl* unterscheidet (siehe Abschnitt 2.2).

2.4 Quantitative Vergleiche

Quantitative Vergleiche bilden im Mathematikunterricht der Grundschule einen von vier Situationstypen der Addition und Subtraktion. Schülerinnen und Schüler müssen für den Aufbau einer breiten und tragfähigen Vorstellung der beiden Operationen lernen, verschiedene Typen von Additions- und Subtraktionssituationen mathematisch modellieren zu können. In diesem Abschnitt wird der Situationstyp „quantitative Vergleiche“ eingeführt, der Stand der Forschung berichtet sowie der Lösungsprozess für entsprechende Probleme modelliert.

Ein Schwerpunkt im Mathematikunterricht des ersten Schuljahrs liegt auf dem Aufbau einer breiten und tragfähigen Vorstellung der Addition und Subtraktion (CCSSI, 2010; Education Institute, 2004; KMK, 2005; MoE, 2011). Für die Entwicklung eines solchen Operationsverständnisses müssen Schülerinnen und Schüler mentale Modelle für die verschiedenen Arten von additiven und subtraktiven Situationen aufbauen. Das bedeutet, sie müssen lernen, verschiedene Typen von additiven und subtraktiven Situationen als solche zu erfassen und in eine zu lösende mathematische Aufgabe bzw. Gleichung (das mathematische Modell) zu übersetzen sowie umgekehrt einer gegebenen Additions- oder Subtraktionsaufgabe verschiedene Situationen zuzuordnen (Kintsch & Greeno 1985; Padberg & Benz, 2011). Um zwischen diesen beiden Ebenen – der mathematischen Ebene und der Sachebene – übersetzen zu können, müssen die Schülerinnen und Schüler daher mentale Modelle als verinnerlichte Ideen von Handlungen für die verschiedenen Arten von Situationen aufbauen (Riley et al., 1983).

Der Aufbau solcher mentalen Modelle wird in der deutschsprachigen mathematikdidaktischen Forschung sowohl mit dem Begriff *Operationsverständnis* (z. B. Radatz, 1989; Schäfer, 2005; Royar, 2013) als auch mit dem Begriff der *Grundvorstellungen* (z. B. Blum & vom Hofe, 2003; Kuhnke, 2013) beschrieben (Wessel, 2015). Gemeinsam ist den beiden Begriffen, dass sie die verschiedenen additiven und subtraktiven Situationen klassifizieren. Während unterschiedliche solcher Klassifikationen mit verschiedenen Unterkategorien existieren (z. B. Carpenter & Moser, 1984; Padberg & Benz, 2011; Schipper, 2009; Verschaffel et al., 2007; Van de Walle et al., 2016), ist den meisten davon gemein, dass sie zwischen den Situationen *Verändern* (Hinzufügen zu / Wegnehmen von einer Menge), *Vereinigen* (von zwei Mengen, Teil eines Ganzen), *Ergänzen* (Ausgleichen eines Unterschieds) und *Vergleichen* (Identifizieren eines Unterschieds) unterscheiden. Einige dieser Klassifikationen ordnen die Situationen dabei jeweils der Addition oder Subtraktion zu (z. B. Fuson, 1992; Radatz et al., 1996), während andere auf eine solche

Zuordnung verzichten (z. B. Padberg & Benz, 2011; Schipper, 2009, für eine ausführliche Übersicht siehe Wessel, 2015). Da im Folgenden Situationen zum Vergleichen von Mengen fokussiert werden, welche sowohl additiv als auch subtraktiv gelöst werden können, wird dafür entsprechend keine derartige Operationszuordnung vorgenommen.

Während Veränderungs- und Ausgleichssituationen als dynamisch charakterisiert werden, d. h. eine Handlung ist ihnen bereits inhärent, werden die Situationstypen Vereinigen und Vergleichen als statisch beschrieben, wobei keine Handlung vorgegeben wird (Riley et al., 1983). Innerhalb dieser Kategorisierung nach der semantischen Struktur der Situationen lässt sich außerdem unterscheiden, welche Menge in der jeweiligen Situation gesucht ist, ob die Situation aufsteigend oder absteigend formuliert wurde und ob es sich in der Situation um die Teil-Ganzes-Struktur einer Menge oder um zwei disjunkte Mengen bzw. deren Verhältnis handelt (Gabler & Ufer, 2020).

Von den vier Situationstypen stellen Vergleichsaufgaben länderübergreifend für die meisten Schülerinnen und Schüler in den ersten Schuljahren den schwersten Aufgabentyp dar (Nunes et al., 2016; Riley und Greeno, 1988; Stern, 1993). Ein typisches Beispiel einer Vergleichsaufgabe ist: "Paul hat 3 Murmeln, Mila hat 5 Murmeln, wie viele Murmeln hat Mila mehr als Paul?". Nach der Modellierung von Gabler und Ufer (2020) handelt es sich dabei um eine Aufgabe vom Typ Vergleichen mit zwei disjunkten Mengen (3 Murmeln von Paul, 5 Murmeln von Mila), deren Differenzmenge gesucht ist und die aufsteigend formuliert wurde („wie viele Murmeln hat Mila *mehr* als Paul?“). In dieser Arbeit werden solche Probleme als quantitative Vergleiche bezeichnet, um sie von qualitativen Vergleichen zu unterscheiden ("Paul hat 3 Murmeln, Mila hat 5 Murmeln, wer hat mehr?"). Wie die Problemtypen selbst lassen sich auch die jeweils erforderlichen Fähigkeiten des approximativen und des exakten Vergleichens differenzieren (Obersteiner et al., 2013).

Quantitative Vergleiche sind schwieriger als Probleme anderer Situationstypen, obwohl sie durch äquivalente mathematische Modelle modelliert werden können. Eine Erklärung für diese Schwierigkeit ist der statische Charakter dieser Situationen: Wenn durch das Problem selbst keine Handlung impliziert wird, müssen die Lernenden entweder selbst eine generieren, um ein mentales Modell für das Problem zu entwickeln, oder sie aktivieren ein geeignetes anderes Modell (wie z. B. das Ergänzen), um es zu lösen (vgl. Wessel, 2015). Allerdings sind Aufgaben vom zweiten statischen Situationstyp, dem Vereinigen, für Schülerinnen und Schüler wesentlich einfacher zu lösen. Daher erklärt der statische Charakter allein nicht die Schwierigkeit. Eine zweite Erklärung ist der relationale Charakter des Unterschieds: Im Gegensatz zu den beiden Mengen, die verglichen werden, beschreibt die Differenzmenge und damit die Lösung der Beispielaufgabe (Mila hat 2 Murmeln mehr als Paul) keine konkret vorstellbare Menge, sondern das Verhältnis zwischen zwei anderen Mengen. Stern (1998) beschreibt die Fähigkeit, Zahlen als Verhältnis zwischen zwei Mengen zu verstehen, als relationalen Zahlenbegriff, welchen Schülerinnen und Schüler im Laufe der Zeit entwickeln müssen. Einen dritten Grund für den erhöhten Schwierigkeitsgrad quantitativer Vergleiche können Probleme im Textverständnis darstellen – insbesondere in Klasse 1 – wenn das Problem in schriftlicher Form gegeben ist

(Gabler & Ufer, 2020; Stern, 1993). Ein Beispiel für daraus resultierende Fehlkonzepte sind Signalwortstrategien, bei denen die Operation direkt aus der Formulierung eines Problems abgeleitet wird ("weniger" – Subtraktion, "mehr" – Addition, Stern, 1993). Darüber hinaus kommt Stern (1993) in ihrer Studie über das Lösen von Textaufgaben zu dem Schluss, dass neben Signalwortstrategien auch ein mangelndes Verständnis der sprachlichen Symmetrie quantitativer Vergleiche (A hat zwei mehr als B. \Leftrightarrow B hat zwei weniger als A.) und somit eine fehlende Flexibilität im Umgang mit quantitativen Vergleichen ein möglicher Einflussfaktor für das Verstehen und Lösen von Textaufgaben ist. Ein weiterer Grund, der zwangsläufig mit den vorherigen zusammenhängt, ist ein Mangel an adäquaten Lerngelegenheiten für quantitative Vergleiche im Unterricht. Verschiedene Analysen von Schulbüchern haben eine Unterrepräsentation quantitativer Vergleiche bis hin zu gar keiner Repräsentation in Schulbüchern der ersten Klasse gezeigt und dementsprechend eine starke Betonung der anderen Situationstypen (z. B. Despina & Harikleia, 2014; Tarim, 2017; Wessel, 2015).

Verschiedene Studien haben wiederholt einen langfristigen Nutzen von frühen numerischen Fähigkeiten berichtet (z. B. Reeve et al., 2012; Selter et al., 2012) und verstärken damit die Notwendigkeit, sich im Mathematikunterricht von Beginn an breiter mit den Operationen in all ihren Aspekten auseinanderzusetzen. Dies schließt quantitative Vergleiche ein. Doch obwohl die Schwierigkeit von quantitativen Vergleichen seit mehr als drei Jahrzehnten bekannt ist und untersucht wird, ist die Frage nach effektiven Lernumgebungen noch immer nicht zufriedenstellend beantwortet (Huang et al., 2019). Dabei geben einige explorative Studien Erkenntnisse darüber, wie dieses Problem adressiert werden kann. So misst Stern (1993) der sprachlichen Flexibilität großen Wert bei, insbesondere der Einsicht in die Äquivalenz von "mehr-als-/weniger-als"-Aussagen. Im Einklang damit konnten Schumacher und Fuchs (2012) zeigen, dass Lernende, die an einer Intervention mit Fokus auf Textaufgaben teilgenommen haben, beim Lösen quantitativer Vergleiche sowohl diejenigen Lernenden übertreffen, die an einer Intervention mit Fokus auf das Rechnen teilgenommen haben, als auch diejenigen, die konventionell unterrichtet worden sind. Dieser Interventionseffekt wurde partiell mediiert durch das Verständnis der relationalen Terminologie der Schülerinnen und Schüler. Mwangi und Sweller (1998) berichten zudem, dass das Lernen mithilfe von Beispiellösungen bei quantitativen Vergleichen förderlich für die Lernenden ist, insbesondere wenn diese Beispiele verschiedene Repräsentationen enthalten. Dementsprechend berichten Múñez et al. (2013) ebenfalls einen Mehrwert verschiedener Darstellungen für den Aufbau mentaler Modelle für quantitative Vergleiche. Zu ähnlichen Ergebnissen kommt eine Studie von Huang et al. (2019): Diese indiziert einen förderlichen Effekt von Unterrichtsstunden zum quantitativen Vergleich, welche auf einer Kombination aus einem angenommenen Lernverlauf und der Variationspädagogik basieren. Letztere fokussiert unter anderem einen variationsreichen Gebrauch von Repräsentationen. Während Múñez et al. (2013) Lernende der Sekundarschule untersucht haben, beziehen sich die anderen Studien auf Schülerinnen und Schüler der Klassenstufen 1 bis 3.

Das Lösen eines quantitativen Vergleichs erfordert mehrere Schritte, die im Folgenden am Beispiel der Aufgabe "Paul hat 3 Murmeln, Mila hat 5 Murmeln, wie viele Murmeln hat Mila mehr

als Paul?" idealisiert dargestellt werden (vgl. Kintsch & Greeno, 1985). Zuerst müssen die Lernenden ein Situationsmodell aufbauen. Das bedeutet, sie müssen dem Text das Problem und seinen Kontext entnehmen: Zwei Personen, Paul und Mila, besitzen jeweils eine bestimmte Anzahl an Murmeln, Paul 3 und Mila 5. Der Unterschied zwischen den beiden Anzahlen ist unbekannt und wird gesucht. Auf dieser Grundlage müssen die Schülerinnen und Schüler ein mathematisches Problemmodell erstellen, indem sie von den konkreten Personen und Objekten (Murmeln) abstrahieren und die Mengen 3 und 5 betrachten. Dafür müssen sie die Teilmengenbeziehung zwischen diesen Mengen identifizieren bzw. herstellen und die Bedeutung der Differenzmenge für die Lösung des Problems erkennen. Schließlich müssen sie deren Kardinalität durch eine entsprechende Additions- oder Subtraktionsgleichung modellieren und die Lösung wiederum im Situationsmodell interpretieren. Obwohl diese Beschreibung des Lösungsprozesses sehr kleinschrittig ist und bei den meisten Schülerinnen und Schülern größtenteils implizit ablaufen, gibt sie erste Einblicke in die Fähigkeiten, die zum Lösen quantitativer Vergleiche erworben werden müssen. Dazu zählen die Einsicht in die Teilmengenbeziehung natürlicher Zahlen sowie das Konzept der Differenzmenge. Entsprechend sollten Lernenden für die Entwicklung der Fähigkeit, quantitative Vergleiche zu lösen, nicht nur direkte Lerngelegenheiten für quantitative Vergleiche angeboten werden, sondern auch Lerngelegenheiten für diese Lernvoraussetzungen.

2.5 Forschungsdesiderate bereichsspezifischer Schulbuchqualität

In Abschnitt 2.1.5 wurde bereits herausgearbeitet, dass der empirische Forschungsstand bezüglich der Wirkung der Schulbuchwahl auf den Lernerfolg der Schülerinnen und Schüler inkonsistent ist. In den Abschnitten 2.2 bis 2.4 wurde zudem für die Inhaltsbereiche *adaptives Rechnen bei der Addition und Subtraktion bis 1.000*, *Nutzung Operativer Grundaufgaben* sowie *quantitative Vergleiche* gezeigt, dass es sich um curricular geforderte und damit normativ gesetzte Lernziele handelt, dass der Entwicklung dieser Fähigkeiten eine wichtige Rolle für das weitere Mathematiklernen beigemessen wird und dass bisherige empirische Untersuchungen einen optimierungswürdigen Status quo bei Schülerinnen und Schülern gezeigt haben. Der folgende Abschnitt stellt den Stand der Forschung bezüglich der Rolle von Schulbüchern für die jeweiligen drei Bereiche dar, entwickelt daraus das Konstrukt der bereichsspezifischen Schulbuchqualität und leitet die Forschungsdesiderate für die Studien dieser Arbeit ab.

Mit Bezug auf die Strategienutzung von Schülerinnen und Schülern gibt es Hinweise darauf, dass das verwendete Mathematikbuch einen Einfluss bei der Bearbeitung von Aufgaben hat (vgl. Fagginger Auer et al., 2016; Heinze et al., 2009). Fagginger Auer und Kolleginnen und Kollegen (2016) analysierten die Daten eines 2011 durchgeführten Mathematiktests an niederländischen Grundschulen. Ihre Stichprobe von 1.619 Schülerinnen und Schülern untersuchten sie mithilfe einer Multilevel Latent Class Analysis. Dabei werden verschiedene Profile gebildet – in diesem Fall Profiltypen der Strategienutzung beim Lösen von Arithmetikaufgaben – und unter Berücksichtigung der geschachtelten Datenstruktur von Lernenden in Schulklassen die Wahrscheinlichkeiten berechnet, einem dieser Profile anzugehören. Sie konnten zeigen, dass das verwendete Schulbuch die Wahrscheinlichkeit beeinflusst, dass die Schülerin oder der Schüler

einem bestimmten Profil der Strategienutzung angehört. Heinze et al. (2009) untersuchten 245 Schülerinnen und Schüler der dritten Klassenstufe in Deutschland, wobei sie einen Zusammenhang zwischen der Anzahl der von den Lernenden verwendeten effizienten Strategien und dem instruktionalen Ansatz ihres Schulbuchs fanden. Allerdings handelt es sich in beiden Fällen um Querschnittsstudien, die keine Informationen darüber enthalten, inwieweit sich die Lehrkräfte bei der Thematisierung arithmetischer Strategien an die Schulbücher gehalten haben. Sollten sich diese Hinweise auf einen Einfluss des Schulbuchs auf die Strategienutzung der Schülerinnen und Schüler beim Lösen von Additions- und Subtraktionsproblemen bestätigen, ist dies sowohl für die adaptive Strategiewahl in Klassenstufe 3 also auch für die Nutzung von Strategien, welche auf den Operativen Grundaufgaben in Klassenstufe 1 aufbauen, von Bedeutung.

Auch für den Bereich der quantitativen Vergleiche gibt es Hinweise auf einen Einfluss des genutzten Mathematikbuchs (Despina & Harikleia, 2014; Tarim, 2017; Xin, 2007). Despina und Harikleia (2014) und Tarim (2017) analysierten jeweils die Verteilung verschiedener Typen von Textaufgaben in griechischen bzw. türkischen Grundschulbüchern sowie die Fähigkeit einiger Schülerinnen und Schüler, diese verschiedenen Typen von Textaufgaben zu lösen. Beide Studien kommen zu dem Schluss, dass zum einen quantitative Vergleiche in den untersuchten Schulbüchern unterrepräsentiert sind und zum anderen die Lernenden bei den in Schulbüchern überrepräsentierten Problemtypen besser abschneiden als bei den unterrepräsentierten oder fehlenden Problemtypen. Ähnliches stellt Xin (2007) beim Vergleich jeweils einer Schulbuchreihe aus China und den USA und der Leistungen jeweils einer Gruppe von Schülerinnen und Schüler beider Länder bei multiplikativen Vergleichen fest. Seine Ergebnisse zeigen eine unausgewogene Verteilung der verschiedenen Typen von Textaufgaben im US-Schulbuch im Gegensatz zum chinesischen Schulbuch und eine dementsprechend unausgewogene Leistung bei den verschiedenen Typen von Items für US-Schülerinnen und -Schüler. Der Stand der Forschung basiert jedoch auch hier bisher auf Studien mit querschnittlichen Designs und kleinen Stichprobengrößen.

Für alle drei vorgestellten Inhaltsbereiche gibt es also Hinweise darauf, dass ein Zusammenhang zwischen den Schulbüchern und dem Lernzuwachs der Schülerinnen und Schüler besteht. Bisher wurde dieser Zusammenhang jedoch noch nicht systematisch untersucht. Sollte er sich bestätigen lassen, ist es darüber hinaus insbesondere von Interesse, wodurch ein solcher Zusammenhang in den Schulbüchern bedingt ist – nicht nur aus Perspektive der Forschung, sondern auch aus Sicht von Lehrkräften, Bildungspolitik und Schulbuchautorinnen und -autoren. In den Studien dieser Arbeit wird dafür der Begriff der *bereichsspezifischen Schulbuchqualität* als Merkmal von Schulbüchern im Sinne eines erklärenden Faktors für Zusammenhänge zwischen Schulbüchern und bereichsspezifischer Leistung von Schülerinnen und Schüler verwendet. Der Begriff der bereichsspezifischen Qualität von Schulbüchern wird damit im Sinne der Lerneffektivität wie folgt definiert: Ein Schulbuch hat eine gute bereichsspezifische Qualität, wenn der Lernerfolg der Schülerinnen und Schüler, die es nutzen, im entsprechenden Bereich groß ist. Die Definition als erklärender Faktor für Zusammenhänge zum Lernerfolg der Lernenden erfordert eine inhaltliche Konkretisierung für die Schulbücher und soll die Black-Box-Modelle bisheriger Untersuchungen erweitern. So konnten beispielsweise van den Ham und Heinze (2018)

zeigen, dass es Unterschiede in der Entwicklung der Arithmetikkompetenz der Schülerinnen und Schüler gibt. Die Untersuchungen von Schmidt (2001) und Törnroos (2005) haben einen positiven Zusammenhang zwischen der Quantität der Lerngelegenheiten in den Schulbüchern und dem Lernzuwachs der Schülerinnen und Schüler gezeigt. Schulbücher haben allerdings eine begrenzte Kapazität. Das heißt, eine quantitative Steigerung an Lerngelegenheiten ist kein adäquates Mittel zur Verbesserung des Lernerfolgs. Wie die Qualität der Schulbücher bzw. ihrer Lerngelegenheiten bezogen auf den Lernerfolg inhaltlich charakterisiert werden kann, was also den Inhalt eines Schulbuchs ausmacht, mit dem Schülerinnen und Schüler mehr oder weniger lernen, können diese Untersuchungen nicht beantworten. Angesichts der Komplexität einer globalen und umfassenden, das heißt themenübergreifenden, Bewertung von Schulbuchqualität adressieren die drei Studien dieser Arbeit dieses Desiderat daher im analytischen Sinne anhand dreier ausgewählter Inhaltsbereiche.

3 Die drei Studien dieser Arbeit

In diesem Kapitel werden die drei durchgeführten Studien dieser Arbeit eingeführt. Dabei werden jeweils die Ziele, Forschungshypothesen und Untersuchungsansätze dargestellt.

3.1 Studie 1

3.1.1 Ziel und Hypothesen

Die erste Studie adressiert die Rolle der Schulbuchqualität für die Adaptivität der Strategiewahl der Schülerinnen und Schüler in Klassenstufe 3. Dabei wird die Strategienutzung bei Additions- und Subtraktionsaufgaben im Zahlenraum bis 1.000 untersucht. Da es sowohl Hinweise auf einen Zusammenhang zwischen genutztem Schulbuch und der Strategiewahl der Schülerinnen und Schüler gibt (Fagginger Auer et al., 2016; Heinze et al., 2009) als auch einen Effekt des Schulbuchs auf die Arithmetikleistung der Schülerinnen und Schüler in der untersuchten Gesamtstichprobe (van den Ham & Heinze, 2018), wird angenommen, dass sich die Qualität der Lerngelegenheiten zwischen den vier dort untersuchten Schulbuchreihen unterscheidet. Weiter ist anzunehmen, dass sich diese Qualitätsunterschiede in der Strategiewahl der Schülerinnen und Schüler manifestieren. Die längsschnittliche Datenstruktur ermöglicht darüber hinaus eine Untersuchung des Einflusses des arithmetischen Vorwissens der Schülerinnen und Schüler auf deren Strategienutzung (Torbeyns et al., 2017) sowie dessen möglicher Moderation durch die bereichsspezifische Schulbuchqualität. Entsprechend werden in Studie 1 die folgenden Forschungshypothesen geprüft:

- H 1.1 Mathematikbücher für die Grundschule, welche das gleiche Curriculum abbilden, unterscheiden sich in der Qualität ihrer Lerngelegenheiten für das adaptive Rechnen bei der Addition und Subtraktion bis 1.000.
- H 1.2 Die adaptive Rechenkompetenz der Schülerinnen und Schüler bei der Addition und Subtraktion bis 1.000 am Ende von Klassenstufe 3 hängt mit der Qualität der Mathematikbücher für den Bereich des adaptiven Rechnens zusammen.

Des Weiteren hat sich das arithmetische Vorwissen der Schülerinnen und Schüler als bedeutsam für die Adaptivität ihrer Strategienutzung gezeigt (Torbeyns et al., 2017). Aufbauend auf Hypothese 1.2 wird entsprechend vermutet, dass der Effekt der bereichsspezifischen Schulbuchqualität auf die Strategienutzung der Schülerinnen und Schüler in Abhängigkeit des arithmetischen Vorwissens variiert:

- H 1.3 Eine höhere Ausprägung des arithmetischen Vorwissens geht einher mit einem stärkeren Effekt der Schulbuchqualität auf die adaptive Rechenkompetenz der Schülerinnen und Schüler bei der Addition und Subtraktion bis 1.000.

3.1.2 Untersuchungsansatz

Dieser Studie greift die Ansätze von Törnroos (2005) und Hadar (2017) auf und erweitert sie: Es werden die Lerngelegenheiten zum adaptiven Rechnen in den Schulbüchern analysiert, um

die bereichsspezifische Qualität der Schulbücher entsprechend der Konzeptualisierung in Abschnitt 4.2.2.1 zu bewerten. Die Bewertung der Lerngelegenheiten geht dabei Gegensatz zur Studie von Törnroos (2005) über eine reine Quantifizierung hinaus und schließt auch deren Qualität anhand von theoriebasierten Kriterien mit ein. Diese sind abgeleitet aus dem Modell der Strategiewahl (siehe Abschnitt 2.2, Lemaire & Siegler, 1995) und bilden vier Kategorien der Schulbuchqualität zur Adaptivität der Strategiewahl. Da Additions- und Subtraktionsstrategien für den Zahlenbereich bis 1.000 bereits ab der Erweiterung des Zahlenraums bis 100 in Klassenstufe 2 im Unterricht eine Rolle spielen, werden entsprechend die Schulbücher für das zweite und dritte Schuljahr untersucht. Die Bewertung der Adaptivität der Strategienutzung der Schülerinnen und Schüler erfolgt anhand ihrer Lösungen bei entsprechenden Aufgaben, welche sie am Ende der dritten Klasse bearbeitet haben².

Anhand der Ergebnisse der Schulbuchanalysen erfolgt die Überprüfung von Hypothese 1.1. Weiter werden die Ergebnisse der einzelnen Kategorien gemeinsam skaliert, getrennt nach Klassenstufe 2 und 3, sodass die vier Schulbuchreihen für beide Klassenstufen jeweils auf einer bereichsspezifischen Schulbuchqualitätsskala verortet werden können. Anhand dieser Skalen erfolgt die Untersuchung des Zusammenhangs der Schulbuchqualität mit dem Lernerfolg der Schülerinnen und Schüler (Hypothese 1.2). Dafür werden die Daten einer Substichprobe von $n = 1.404$ Schülerinnen und Schüler in zwei Ebenen (Individual- und Klassenebene) modelliert und der Effekt der Schulbuchqualität unter Kontrolle individueller und unterrichtlicher Kovariaten analysiert.

Um eine mögliche Moderation des arithmetischen Vorwissens durch die bereichsspezifische Schulbuchqualität zu prüfen (H 1.3), wird neben den beiden Haupteffekten auch eine Cross-Level-Interaktion von arithmetischem Vorwissen und bereichsspezifischer Schulbuchqualität ins Modell aufgenommen und untersucht.

3.2 Studie 2

3.2.1 Ziel und Hypothesen

In der zweiten Studie wird die Bedeutung der bereichsspezifischen Schulbuchqualität für die Fähigkeit von Schülerinnen und Schülern am Ende der ersten Klasse untersucht, operative Beziehungen zu nutzen. Der Fokus liegt hierbei auf den Operativen Grundaufgaben nach Schipper (2009), zu denen Tausch-, Umkehr- und Nachbaraufgaben gehören, und den darauf basierenden Rechenstrategien (siehe Abschnitt 2.3). Dementsprechend wird in dieser Untersuchung auf Studie 1 aufgebaut, in der ein Zusammenhang zwischen der bereichsspezifischen Schulbuchqualität und der Strategienutzung gezeigt werden konnte (siehe Abschnitt 5.1). Dies unterstützt vorhandene Hinweise auf einen Effekt des Schulbuchs auf die Strategiewahl der Schülerinnen und Schüler (Fagginger Auer et al., 2016; Heinze et al., 2009). Es wird daher angenommen, dass sich die Qualität der Schulbücher auch für den Bereich der Operativen Grundaufgaben unterscheidet und diese Qualitätsunterschiede für die Nutzung Operativer Grundaufgaben prädiktiv

² Vergleichbare Daten vom Ende der zweiten Klasse liegen in dem untersuchten Datensatz nicht vor.

sind. Da eine höhere Schulbuchqualität bei einem Thema in der Regel mit einer größeren Breite und Tiefe in der Behandlung dieses Themas verbunden ist, wird in dieser Studie darüber hinaus untersucht, ob sich der Zusammenhang zwischen relativem Schulbuchanteil und relativer Unterrichtszeit nach Schmidt et al. (2001) auch in einer Moderation der Effekte von bereichsspezifischer Schulbuchqualität und Unterrichtsquantität auf den Lernerfolg der Schülerinnen und Schüler in diesem Inhaltsbereich zeigt. In Studie 2 werden daher die folgenden Forschungshypothesen geprüft:

- H 2.1 Mathematikbücher für Klassenstufe 1, welche das gleiche Curriculum abbilden, unterscheiden sich in der Qualität ihrer Lerngelegenheiten bezüglich Operativer Grundaufgaben.
- H 2.2 Die Qualität der Mathematikbücher bezüglich Operativer Grundaufgaben hängt mit der Fähigkeit der Schülerinnen und Schüler zusammen, diese am Ende von Klassenstufe 1 zu nutzen.

Um eine mögliche Interaktion von der Qualität der Schulbücher und der Quantität der Schulstunden in Bezug auf die Nutzung Operativer Grundaufgaben durch die Schülerinnen und Schüler zu untersuchen, wird zudem die dritte Forschungshypothese geprüft:

- H 2.3 Der Zusammenhang zwischen der Schulbuchqualität bezüglich Operativer Grundaufgaben und der Fähigkeit der Schülerinnen und Schüler, diese zu nutzen, wird moderiert durch die Anzahl an Unterrichtsstunden, welche die Lehrkräfte für das Thema nutzen.

3.2.2 Untersuchungsansatz

In Studie 2 wird der Ansatz aus Studie 1 basierend auf Törnroos (2005) und Hadar (2017) aufgegriffen und auf den Bereich der Operativen Grundaufgaben übertragen. Die entsprechenden Lerngelegenheiten in den Schulbüchern werden analysiert und darauf aufbauend die bereichsspezifische Qualität der Schulbücher entsprechend der Konzeptualisierung in Abschnitt 4.2.3.1 bewertet. Da auch diese Studie einen Aspekt der Strategiekompetenz der Schülerinnen und Schüler adressiert, werden die Kriterien erneut aus dem Modell der Strategiewahl (siehe Abschnitt 2.2, Lemaire & Siegler, 1995) abgeleitet und in vier Qualitätskategorien zusammengefasst. Der Lernerfolg der Schülerinnen und Schüler wird in dieser Untersuchung anhand der Daten vom Ende der ersten Klasse bewertet. Daher beschränkt sich die Analyse der Schulbücher auf das erste Schuljahr.

Hypothese H 2.1 wird analog zu Studie 1 auf Grundlage der Ergebnisse der Schulbuchanalysen geprüft. Ebenfalls analog zur ersten Studie werden die Ergebnisse der einzelnen Kategorien gemeinsam skaliert mit dem Ergebnis einer gemeinsamen Qualitätsskala der Schulbücher von Klassenstufe 1 für den Bereich Operative Grundaufgaben. Diese Skala wird wiederum in ein Zwei-Ebenen-Modell der Längsschnittdaten integriert. Der Effekt wird unter Kontrolle individueller und unterrichtlicher Kovariaten analysiert, um Hypothese H 2.2 zu prüfen. In dieser Studie besteht die Stichprobe aus $n = 1.614$ Schülerinnen und Schülern. Die Anzahlen zwischen den Untersuchungen unterscheiden sich etwas aufgrund von Stichprobenmortalität im Verlauf

der Erhebung sowie einer leicht variierenden Anzahl teilnehmender Schülerinnen und Schüler bei unterschiedlichen Testterminen (siehe Abschnitt 4.1).

Um eine mögliche Moderation des Effektes der bereichsspezifischen Schulbuchqualität durch die Anzahl an Unterrichtsstunden zu untersuchen und damit Hypothese 2.3 zu prüfen, wird zusätzlich zu den Haupteffekten auch ein Interaktionseffekt der beiden Variablen getestet.

3.3 Studie 3

3.3.1 Ziel und Hypothesen

Ziel der dritten Studie ist die Untersuchung des Zusammenhangs zwischen der Schulbuchqualität und dem Lernerfolg der Schülerinnen und Schüler für den Bereich der quantitativen Vergleiche in Klassenstufe 1. Dabei wird der Lernerfolg an den Lösungen bei quantitativen Vergleichen vom Typ „Unterschied gesucht“ gemessen, das heißt mit fehlender bzw. gesuchter Differenzmenge (Schipper, 2009). Zum einen sind quantitative Vergleiche in Schulbüchern national wie international unterrepräsentiert (Despina & Harikleia, 2014; Tarim, 2017; Wessel, 2015). Zum anderen gibt es auch für diesen Bereich Hinweise auf einen Zusammenhang zwischen dem genutzten Schulbuch und den Fähigkeiten der Schülerinnen und Schüler (Despina & Harikleia, 2014; Tarim, 2017; Xin, 2007). Da quantitative Vergleiche zu den schwierigsten Typen von Textaufgaben in den ersten Schuljahren gehören (Nunes et al., 2016), ist insbesondere von Interesse, ob die bereichsspezifische Qualität des Schulbuchs mit dem Lernerfolg der Kinder im Zusammenhang steht. Auf Basis der ersten beiden Studien sowie der Effekte der Schulbuchwahl auf die Arithmetikleistung der Schülerinnen und Schüler in der untersuchten Gesamtstichprobe (van den Ham & Heinze, 2018) wird auch für diesen Bereich angenommen, dass sich die Qualität der Lerngelegenheiten zwischen den Schulbüchern unterscheidet und mit den Fähigkeiten der Schülerinnen und Schüler zusammenhängt, quantitative Vergleiche zu lösen. Dementsprechend werden in der dritten Studie die folgenden Forschungshypothesen geprüft:

- H 3.1 Mathematikbücher für die erste Klasse, welche das gleiche Curriculum abbilden, unterscheiden sich in der Qualität ihrer Lerngelegenheiten bezüglich quantitativer Vergleiche.
- H 3.2 Die Qualität von Mathematikbüchern bezüglich quantitativer Vergleiche in Klassenstufe 1 ist prädiktiv für die Fähigkeit der Schülerinnen und Schüler, quantitative Vergleiche zu lösen.

3.3.2 Untersuchungsansatz

Der Untersuchungsansatz in Studie 3 ist analog zu denen der beiden vorangegangenen und stellt wie diese eine Erweiterung der Ansätze von Törnroos (2005) und Hadar (2017) dar. Die Lerngelegenheiten hinsichtlich quantitativer Vergleiche in den Schulbüchern werden analysiert und entsprechend der Konzeptualisierung der Schulbuchqualität in Abschnitt 4.2.4.1 bewertet. Im Gegensatz zu den Studien 1 und 2 ist die Frage nach effektiven Lernumgebungen für den Bereich quantitative Vergleiche noch nicht zufriedenstellend beantwortet (Gabler & Ufer,

2020). Es existiert also kein zum Strategiewahl-Modell vergleichbares Modell als Grundlage eines Bewertungssystems. Aufgrund dessen ist zur Bewertung der Schulbuchqualität bezüglich quantitativer Vergleiche ein Kategoriensystem aus dem bisherigen Stand der Forschung (siehe Abschnitt 2.4) abgeleitet und in Diskussionen mit Expertinnen und Experten aus der Schulpraxis sowie aus der Mathematikdidaktik mehrfach überarbeitet, erweitert und elaboriert worden. Die Fähigkeiten der Schülerinnen und Schüler beim Lösen quantitativer Vergleiche werden anhand der entsprechenden Antworten gemessen, welche nach den ersten sieben Monaten von Klassenstufe 1 erhoben worden sind.

In gleicher Weise wie in den Studien 1 und 2 wird Hypothese H 3.1 auf Basis der Ergebnisse der Schulbuchanalysen geprüft. Auch in dieser Untersuchung werden die Ergebnisse der einzelnen Kategorien gemeinsam skaliert und so eine gemeinsame Skala der Schulbuchqualität bezüglich quantitativer Vergleiche in Klassenstufe 1 gebildet, auf welcher sich die vier Schulbuchreihen abbilden lassen. Deren Effekt auf die Fähigkeiten der Schülerinnen und Schüler wird wiederum in einem Zwei-Ebenen-Modell der Längsschnittdaten unter Kontrolle individueller und unterrichtlicher Kovariaten untersucht, um Hypothese H 3.2 zu prüfen. Die Substichprobe in Studie 3 besteht dabei aus $n = 1.513$ Schülerinnen und Schülern.

4 Stichproben und Instrumente

4.1 Stichproben

Die drei Studien dieser Arbeit basieren auf Sekundäranalysen einer Stichprobe aus Schleswig-Holstein, deren Daten zur Evaluation des Förderprogramms „Mathe macht stark – Grundschule“ (IQSH, 2020) erhoben worden sind. Das untersuchte Programm wird in den Klassenstufen 1 und 2 im Bereich Arithmetik eingesetzt und hat die frühzeitige Diagnose und Förderung von leistungsschwachen Schülerinnen und Schülern zum Ziel. Für die Evaluation wurden zwei Interventionsgruppen aus 20 bzw. 10 Schulen, in denen die Lehrkräfte die Materialien des Förderprogramms erhalten haben, mit einer Kontrollgruppe aus 10 Schulen verglichen, in denen der Unterricht konventionell erfolgte. Für die Teilnahme an dem Förderprogramm wurde in den hier präsentierten Analysen mithilfe von Dummy-Variablen kontrolliert.

4.1.1 Evaluationsstichprobe

Für die Evaluation wurden Daten der teilnehmenden Schülerinnen und Schüler zu verschiedenen Inhaltsbereichen der Arithmetik von Beginn bis Ende der Grundschule in den Jahren 2013 bis 2017 erhoben. Zusätzlich wurden Informationen zum Hintergrund der Schülerinnen und Schüler sowie zu unterrichtlichen Rahmenbedingungen durch Eltern- und Lehrkräftefragebögen gesammelt. Die Gesamtstichprobe umfasst 2.330 Schülerinnen und Schüler aus 127 Klassen in 40 Schulen aus städtischen wie auch aus ländlichen Gebieten, was etwa 10 % der Kohorte in Schleswig-Holstein entspricht. In allen Schulen wurde nach dem gleichen, verbindlichen Curriculum unterrichtet, das sowohl Inhalte (wie Zahlaspekte oder Operationsvorstellungen) und zu erwerbende Fähigkeiten (wie Kopfrechnen oder Problemlösen) als auch zu nutzende Repräsentationen (wie den Zahlenstrahl oder die Stellenwerttafel) vorgibt (Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein, 1997).

4.1.2 Gesamtstichprobe der IPN-Schulbuchstudie

In 93 der 127 Klassen ist in den ersten beiden Schuljahren eine der vier Schulbuchreihen „Denken und Rechnen“, „Einstern“, „Flex und Flo“ und „Welt der Zahl“ genutzt worden und das etwa gleichmäßig verteilt (Denken und Rechnen: 18; Einstern: 21; Flex und Flo: 33; Welt der Zahl: 21). Die entspricht knapp 75 % dieser Stichprobe. Diese Verteilung der Schulbücher hat sich zufällig ergeben und ist unabhängig von der Teilnahme am Förderprogramm. Ab der dritten Klasse hat sich diese Verteilung etwas verändert (Denken und Rechnen: 28; Einstern: 6; Flex und Flo: 24; Welt der Zahl: 31). In den restlichen Klassen wurden zum Teil andere, nur vereinzelt auftretende Schulbücher genutzt, zum Teil wurde gar kein Schulbuch genutzt oder die Angabe über das Schulbuch fehlt.

Bei den Reihen „Denken und Rechnen“ und „Welt der Zahl“ handelt es sich um klassische Schulbücher, die alle Themen eines Schuljahres in einem Buch abbilden. „Flex und Flo“ sowie „Einstern“ sind dagegen in bis zu sieben Themenhefte pro Schuljahr gegliedert. „Denken und Rechnen“, „Flex und Flo“ sowie „Welt der Zahl“ sind außerdem vergleichbar in ihrem Ansatz

zum Einsatz des Lehrwerks im Unterricht. Sie lassen Lehrkräften die Entscheidungsfreiheit, welche Beispiele, Aufgaben, Darstellungen, Erklärungen und Übungen sie aus den Büchern auswählen und verwenden wollen. Es werden zwar Lernverläufe empfohlen, aber nicht vorgeschrieben. Bei der Reihe „Einstern“ wird dagegen das Durcharbeiten der Themenhefte in chronologischer Reihenfolge empfohlen, wobei das individuelle und autonome Lernen der Schülerinnen und Schüler betont wird. Die Inhalte sind kleinteilig gegliedert und in der Regel mit jeweils ein bis zwei Seiten Übungsaufgaben versehen.

Die ersten Untersuchungen dieser Substichprobe von $n = 1.664$ Schülerinnen und Schülern (784 weiblich, 47 %) im Rahmen der IPN-Schulbuchstudie durch van den Ham und Heinze (2018) haben gezeigt, dass sich die Arithmetikleistungen der Schülerinnen und Schüler am Ende von Klassenstufe 1, 2 und 3 zwischen den Schulbuchgruppen stark unterscheiden. Über diese Substichprobe verteilen sich also zum einen vier Schulbuchreihen auf statistisch günstige Art und Weise, zum anderen zeigen sich statistische Effekte der Schulbuchwahl auf die Leistungen der Schülerinnen und Schüler im Bereich Arithmetik. Um eine mögliche Erklärung der Schulbucheffekte durch die bereichsspezifische Qualität der Schulbücher zu erhalten und damit Einsicht in die Gestaltung lernerfektiver Schulbücher zu gewinnen, wurden für die drei Studien dieser Arbeit jeweils Substichproben aus der Stichprobe der IPN-Schulbuchstudie gewählt.

4.1.3 Stichprobe von Studie 1

Für die Untersuchung der Schulbuchqualität bezüglich adaptiven Rechnens in Klassenstufe 3 wurden aus der Stichprobe der IPN-Schulbuchstudie 1.404 Schülerinnen und Schüler (676 weiblich, 48 %) aus 82 Klassen gewählt. Die Verteilung der Schulbuchreihen verhält sich ähnlich zur Gesamtstichprobe der IPN-Schulbuchstudie (Denken und Rechnen: 17 (Klassen); Einstern: 20; Flex und Flo: 26; Welt der Zahl: 19). Aufgrund von Stichprobenmortalität bei längsschnittlichen Erhebungen ist die Anzahl der Schülerinnen und Schüler reduziert. Zwar wurden im Gegensatz zu den Studien 2 und 3 auch Schülerinnen und Schüler ausgewählt, die nicht am Abschlusstest der dritten Klasse teilgenommen haben, aus welchem die Items zur Bewertung der Zielvariable von Studie 1 gewählt wurden. Voraussetzung dafür war, dass die vorangegangenen Informationen ausreichen für eine Imputation dieser Daten sowohl durch multiple Imputation (Lüdtke et al., 2007) als auch durch den Full-Information-Maximum-Likelihood-Ansatz (Enders, 2010). Dennoch mussten insgesamt 260 Schülerinnen und Schüler aus methodischen Gründen ausgeschlossen werden, da über diese entweder nicht zu genügend Messzeitpunkten Daten vorlagen oder die Klassenstärke zu klein für eine Mehrebenenmodellierung war ($n < 5$). Letzteres ist neben der Stichprobenmortalität darauf zurückzuführen, dass die Stichprobe jahrgangsübergreifende Klassen enthält, von denen jeweils nur die Daten der Schülerinnen und Schüler aus dem Jahrgang der Evaluationsstichprobe erhoben worden sind.

4.1.4 Stichprobe von Studie 2

Die Stichprobe von Studie 2 zur Analyse der Schulbuchqualität hinsichtlich Operativer Grundaufgaben in Klassenstufe 1 besteht aus 1.614 Schülerinnen und Schülern (756 weiblich, 47 %) aus 86 Klassen. Auch in dieser Substichprobe verteilen sich die vier Schulbuchreihen ähnlich wie in der Gesamtstichprobe der IPN-Schulbuchstudie (Denken und Rechnen: 17 (Klassen);

Einstern: 20; Flex und Flo: 30; Welt der Zahl: 19). Dabei wurden 50 Teilnehmende ausgeschlossen, weil sie nicht am Abschlusstest des ersten Schuljahrs teilgenommen haben, aus welchem die Items zur Bewertung der Nutzung Operativer Grundaufgaben durch die Lernenden gewählt worden sind. In diesen Fällen haben nicht zu ausreichend vielen Messzeitpunkten Daten über sie vorgelegen, um fehlende Werte schätzen zu können. In einigen Klassen hat sich dadurch die Klassenstärke für eine Mehrebenenmodellierung zu stark reduziert ($n < 5$).

4.1.5 Stichprobe von Studie 3

Für die dritte Studie zur Untersuchung der Schulbuchqualität für quantitative Vergleiche besteht die Stichprobe aus 1.513 Schülerinnen und Schülern (720 weiblich, 48 %) aus 84 Klassen. Die Verteilung der Schulbücher ist ebenfalls ähnlich der der Gesamtstichprobe der IPN-Schulbuchstudie (Denken und Rechnen: 17 (Klassen); Einstern: 20; Flex und Flo: 28; Welt der Zahl: 19). Wie in Studie 2 sind dabei insgesamt 151 Schülerinnen und Schüler und 9 Klassen aufgrund von fehlender Teilnahme am Arithmetiktest mit quantitativen Vergleichsaufgaben, nicht ausreichender Datengrundlage oder zu geringer Klassenstärke ausgeschlossen worden.

4.2 Instrumente

Für die Untersuchung der Zusammenhänge zwischen bereichsspezifischer Schulbuchqualität und entsprechendem Lernerfolg der Schülerinnen und Schüler sind jeweils spezifische Instrumente zur Schulbuchanalyse sowie zur Analyse der Fähigkeiten der Lernenden genutzt worden, welche in den Abschnitten 4.2.2–4.2.4 dargestellt werden. Zur Kontrolle individueller und unterrichtlicher Faktoren sind die entsprechenden Instrumente studienübergreifend zum Einsatz gekommen. Diese werden in Abschnitt 4.2.1 vorgestellt.

4.2.1 Individuelle und unterrichtliche Kovariaten

Um den Einfluss der individuellen Voraussetzungen der Kinder zu berücksichtigen, wird in allen Studien für die Schuleingangsvoraussetzungen kontrolliert. Diese haben langfristig Einfluss auf die Entwicklung mathematischer Kompetenz (z. B. Schneider et al., 2013) und sind mithilfe standardisierter Tests mit ausreichender Reliabilität in den ersten Wochen von Klassenstufe 1 erhoben worden. Zur Erfassung der kognitiven Grundvoraussetzungen wurde der CFT 1-R eingesetzt (Weiß & Osterland, 2013, Cronbach's $\alpha = 0,91$). Dieser stellt eine Adaption des „Culture Fair Intelligence Test 1“ (CFT 1: Cattell et al., 1997) dar und besteht aus 45 Items, bei denen die Teilnehmenden formal-logische Probleme nonverbal lösen müssen, wie beispielsweise das Fortsetzen von Reihen oder das Lösen von Labyrinthen. Die sprachlichen Lernvoraussetzungen zu Schulbeginn wurden mit dem Münsteraner Screening (MÜSC: Mannhaupt, 2013, Cronbach's $\alpha = 0,72$) erhoben. Dabei müssen die Teilnehmenden 40 Items bearbeiten, welche das phonologische Bewusstsein als Voraussetzung für den Erwerb der Schriftsprache testen. Zur Erhebung der mathematischen Lernvoraussetzungen wurde der Hamburger Rechentest (HaRet 1: Lorenz, 2007, Cronbach's $\alpha = 0,74$) genutzt. Dieser beinhaltet 43 Items zu den Arithmetik-bezogenen Vorläuferfähigkeiten der Schülerinnen und Schüler zu Schulbeginn. Dazu zählen beispielsweise

Eins-zu-Eins-Zuordnungen, der (approximative) Größenvergleich von Zahlen oder die Zuordnung von Zahlsymbolen und Mengen.

In allen Studien wurden außerdem aggregierte Mittelwerte einiger dieser Variablen auf Klassenebene in die Modelle aufgenommen, um zusätzlich für die Klassenkomposition zu kontrollieren. Aufgrund von Multikollinearität ist jeweils nur eine davon enthalten. In Studie 1 wird anstelle der mittleren mathematischen und sprachlichen Lernvoraussetzungen der Klassen für das mittlere arithmetische Vorwissen am Ende von Klassenstufe 1 kontrolliert. Dieses wurde anhand eines von Lipowsky et al. (2011) adaptierten Tests erhoben, welcher das Arithmetik-Curriculum des ersten Schuljahres umfasst. Die Ergebnisse sind IRT-skaliert und zeigen eine gute WLE-Reliabilität von 0,94. Eine simultane Aufnahme aller vier Variablen zu aggregierten Lernvoraussetzungen auf Klassenebene ins Modell ist aufgrund von Multikollinearität nicht möglich.

Auf Klassenebene wird darüber hinaus in allen Studien mithilfe von Dummy-Variablen für die Teilnahme am Förderprogramm „Mathe macht stark“ kontrolliert sowie für die formale Qualifikation der Lehrkräfte. Da nicht alle Grundschullehrkräfte, die in Deutschland Mathematik unterrichten, auch eine fachbezogene Qualifikation haben, das heißt, ein abgeschlossenes Studium in Mathematik, und in verschiedenen Studien mit Grundschullehrkräften aus Deutschland gezeigt werden konnte, dass deren fachbezogene Qualifikation stark mit ihrem mathematikspezifischen Professionswissen zusammenhängt (z. B. Knievel et al., 2015; Tatto et al., 2012), ist ein Einfluss auch auf den Lernerfolg anzunehmen. In Studie 1 wurde die Skala der Lehrkraftqualifikation um einen Punkt erweitert:

- 0 Punkte: Lehrkraft hat nicht Mathematik studiert
- 1 Punkt: Lehrkraft hat Mathematik studiert und das Studium vor 2005 abgeschlossen
- 2 Punkte: Lehrkraft hat Mathematik studiert und das Studium 2005 oder später abgeschlossen

Seit dem Ende der 1990er Jahre hat das flexible Rechnen und damit einhergehend die Rolle halbschriftlicher Strategien deutlich an Bedeutung gewonnen und damit auch Einzug in die Ausbildungspläne von Lehrkräften erhalten (Padberg & Benz, 2011, siehe Abschnitt 2.2). Daher sollten Lehrkräfte, die 2005 oder später ihren universitären Abschluss gemacht haben, in ihrer Ausbildung mehr dafür sensibilisiert worden sein als ihre dienstälteren Kolleginnen und Kollegen. Dies könnte entsprechend auch einen Effekt auf die Strategienutzung ihrer Schülerinnen und Schüler haben. Da Schülerinnen und Schüler individuell ein Instrumentarium von Strategien aufbauen, wurde zudem für den Grad an konstruktivistischen Überzeugungen der Lehrkraft kontrolliert. Dafür wurde ein erprobter Fragebogen eingesetzt, welcher von Staub und Stern (2002) entwickelt wurde.

In Studie 2 wird außerdem die Anzahl der Unterrichtsstunden, in denen Operative Grundaufgaben behandelt wurden, als Indikator für entsprechende Unterrichtsaktivitäten berücksichtigt. Diese sind in einem Lehrkräftefragebogen am Ende der ersten Klasse mithilfe einer 4-stufigen Likert-Skala erhoben worden (nie, in 1–2 Unterrichtsstunden, in 3–4 Unterrichtsstunden, in mehr als 4 Unterrichtsstunden).

4.2.2 Schulbuchqualität und Lernerfolg in Studie 1

4.2.2.1 Schulbuchqualität für den Bereich adaptives Rechnen in Klassenstufe 2 und 3

Zur Bewertung der Schulbuchqualität für das adaptive Rechnen sind vier Kategorien aus den vier Dimensionen des Strategiewahlmodells (Abschnitt 2.2) abgeleitet und für den Bereich Addition und Subtraktion bis 1.000 konkretisiert worden. Eine Übersicht ist in Tabelle 4.1 gegeben. Die Kategorien und deren Bewertung werden im Folgenden dargestellt.

Kategorie	Strategie-repertoire	Strategie-Verteilung	Strategie-effizienz	Strategie-auswahl
Kriterium	Instruktionsqualität der Strategien	Häufigkeit der Darstellung der Strategien	Anzahl an Übungsaufgaben	Strategievergleiche
Punkte	0–3 Punkte/ Strategie 0–12 Punkte/ Schulbuch	0–2 Punkte/ Strategie 0–10 Punkte/ Schulbuch	Anzahl an Übungsaufgaben	0–2 Punkte/ Schulbuch

Tabelle 4.1: Bewertungskategorien für die Schulbuchqualität für den Bereich adaptives Rechnen

Für die Dimension *Strategierepertoire*, welche die Kenntnis verschiedener Strategien beschreibt, wird die Instruktionsqualität der verschiedenen halbschriftlichen Strategien in den Schulbüchern bewertet. Die Übersicht halbschriftlicher Strategien ist Tabelle 2.1 zu entnehmen. Dabei wird jeder Schulbuchreihe pro Strategie, jeweils für Addition und Subtraktion, die folgende Punktzahl zugewiesen:

- 0 Punkte, wenn die Strategie nicht enthalten ist
- 1 Punkt, wenn die Strategie enthalten, aber nicht als regulärer Lerninhalt hervorgehoben wird (d. h. nicht im instruktionalen Teil auf einer Schulbuchseite, sondern beispielsweise nur als „Rechentrick“ am Rand einer Übungsseite)
- 2 Punkte, wenn die Strategie explizit für ein Additions- oder Subtraktionsproblem eingeführt wird (im instruktionalen Teil auf einer Schulbuchseite)
- 3 Punkte, wenn die Strategie explizit eingeführt und zusätzlich in einer weiteren Repräsentationsform dargestellt wird (z. B. in symbolischer Form und zusätzlich in einer graphischen Illustration am Zahlenstrahl)

Beispiele für die verschiedenen Punktzahlen sind in Abbildung 4.1 zu sehen.

a

3

a) $34 + 15$	b) $67 + 32$	c) $56 + 33$	d) $46 + 27$	e) $67 + 24$
$25 + 31$	$58 + 24$	$45 + 54$	$51 + 48$	$34 + 45$
$42 + 26$	$34 + 59$	$37 + 43$	$44 + 33$	$28 + 41$

49 56 68 69 73 77 79 80 82 87 89 91 93 99 99 99

4

a) $23 + 29$	b) $44 + 19$	c) $35 + 29$
$36 + 19$	$38 + 39$	$35 + 28$
$45 + 29$	$66 + 29$	$46 + 28$
d) $59 + 37$	e) $58 + 29$	f) $64 + 19$
$29 + 54$	$19 + 19$	$64 + 18$
$49 + 26$	$48 + 48$	$57 + 18$

38 48 52 55 63 63 64 74 74 75 75 77 82 83 83 87 95 96 96

b

Rechenwege – Addieren

$47 + 35$

$40 + 30 = 70$
 $7 + 5 = 12$
 $70 + 12 = 82$
 Tom

$47 + 30 + 5 = 82$
 Lena

c

1 Schreibe auf, wie die Kinder rechnen.

Die grünen Pfeile zeigen meinen Rechenweg.

Die roten Pfeile zeigen, wie ich rechne.

a $24 + 32 =$

Seite 21	Aufgabe 1	
a)	$24 + 32 = 56$	
Lena	$24 + 2 + 30 = 56$	
Tom	$24 + 30 + 2 = 56$	
b)	...	

Abbildung 4.1: Beispiele für 1 Punkt- (a, Rinkens, 2010), 2 Punkte- (b, Eidt, 2010) und 3 Punktebewertungen (c, Bauer & Maurach, 2011b)

Für die Dimension *Strategieverteilung*, dem Wissen über die relative Häufigkeit der Anwendung einer Strategie, wird die Relevanz bewertet, die jeder Strategie im Schulbuch zugeschrieben wird. Dafür wurde neben der Bandbreite der Strategien in einem Buch (siehe Strategierepertoire) auch analysiert, ob eine Strategie an mehr als nur einem Beispielproblem dargestellt wird, und folgendermaßen gewertet:

- 0 Punkte, wenn die Strategie gar nicht oder nur ein einziges Mal abgebildet ist (letzteres ist in der vorherigen Kategorie bereits abgebildet)

- 1 Punkt, wenn die Strategie eingeführt und zusätzlich anhand eines zweiten Problems auf einer anderen Schulbuchseite dargestellt wird
- 2 Punkte, wenn die Strategie eingeführt und zusätzlich anhand von mehr als zwei Problemen auf verschiedenen anderen Schulbuchseiten dargestellt wird.

Beispiele für die Einführung einer Strategie entsprechen den Beispielen (b) und (c) in Abbildung 4.1. Das Scoring in dieser Kategorie unterscheidet dabei, ob und wie sich entsprechende Darstellungen pro Strategie über das jeweilige Schulbuch verteilen.

Die Dimension *Strategieeffizienz* beschreibt die Fähigkeit, Strategien schnell und sicher ausführen zu können. Diese entwickelt und verbessert sich mit zunehmender Erfahrung und Übung. Daher haben wir als Kriterium für diese Kategorie die Anzahl an Additions- und Subtraktionsaufgaben gewählt. Diese wurden im Gegensatz zu den anderen drei Kategorien nicht bewertet, sondern gezählt und als direkte Indikatoren genutzt.

Für die letzte Dimension, die *Strategieauswahl*, das heißt, die Fähigkeit, eine Strategie für ein bestimmtes Problem flexibel auszuwählen, ist aus der Forschung bekannt, dass ein expliziter Vergleich von Lösungsstrategien förderlich für dessen Entwicklung ist (vgl. Abschnitt 2.2). Daher wurde bewertet, ob die Schulbücher Lerngelegenheiten systematisch oder vereinzelt enthalten, die explizit einen Strategievergleich fordern. Systematisch bedeutet in diesem Zusammenhang, dass solche Lerngelegenheiten ein wiederkehrendes Element des Schulbuchs darstellen und damit Teil des instruktionalen Ansatzes bei der Einführung neuer Strategien sind: Mit der Einführung neuer Strategien oder im Anschluss daran werden diese mit bereits bekannten Strategien in einer entsprechenden Aufgabe verglichen. Vereinzelt heißt, dass das Schulbuch solche Lerngelegenheiten enthält, diese aber nicht systematisch enthalten sind. Da die Art und Weise des Auftretens von Strategievergleichen (systematisch, gelegentlich oder gar nicht) für die Addition und Subtraktion in allen vier Schulbuchreihen analog ist, wurde dabei nicht zwischen den Operationen unterschieden. Die Schulbücher wurden entsprechend bewertet:

- 0 Punkte, wenn das Schulbuch keine Lerngelegenheiten zum Strategievergleich enthält
- 1 Punkt, wenn das Schulbuch vereinzelt Lerngelegenheiten zum Strategievergleich enthält
- 2 Punkte, wenn das Schulbuch systematisch Lerngelegenheiten zum Strategievergleich enthält

Ein Beispiel für eine solche Lerngelegenheit zeigt Abbildung 4.2.

Rechenwege – Addieren

1

Rechenkonferenz

$47 + 35$

$47 + 30 + 5 = 82$ Lena

$40 + 30 = 70$
 $7 + 5 = 12$
 $70 + 12 = 82$ Tom

$47 + 3 + 30 + 2 = 82$ Felix

2 Wie rechnest du? Vergleicht miteinander.

a) $37 + 16$ b) $56 + 38$ c) $25 + 36$ d) $65 + 27$ e) $27 + 35$

Abbildung 4.2: Beispiel für eine Lerngelegenheit für den Vergleich von Strategien (Eidt, 2010)

Die Schulbücher für Klassenstufe 2 und 3 der vier Schulbuchreihen wurden von drei geschulten und unabhängigen Ratern in den Kategorien *Strategierepertoire*, *Strategieverteilung* und *Strategieauswahl* bewertet. Die Übungsaufgaben (*Strategieeffizienz*) wurden von einem der Rater ermittelt. Die Übereinstimmungswerte zwischen den Urteilen sind akzeptabel bis sehr gut (Fleiss' κ : 0,61 (*Repertoire*), 0,82 (*Verteilung*), 1,00 (*Auswahl*); ICC: 0,93 (*Repertoire*), 0,98 (*Verteilung*), 1,00 (*Auswahl*)). Die Schulbücher wurden in jeder Kategorie jeweils für Klassenstufe 2 und 3 in eine Rangfolge gebracht und durch die Berechnung des mittleren Rangs auf jeweils eine Skala für Klassenstufe 2 und 3 gebracht, bei der alle Kategorien gleich gewichtet sind³.

4.2.2.2 Adaptives Rechnen bei der Addition und Subtraktion bis 1.000

Die Adaptivität der Strategiewahl der Schülerinnen und Schüler wurde anhand ihrer Lösungen bei vier Items aus dem Arithmetiktest in Klassenstufe 3 bewertet. Dieser wurde am Ende des Schuljahres durchgeführt und enthielt verschiedene Arten von Aufgaben aus dem Bereich der Arithmetik. Bei vier Items aus dem Bereich Addition und Subtraktion bis 1.000 (482+218, 473+398, 381-99, 702-698), bietet sich aus aufgabenorientierter Sicht (vgl. Abschnitt 2.2) der Einsatz bestimmter Strategien als effizienter Lösungsweg an. Die geringe Anzahl an Items ergibt sich aus der Sekundäranalyse vorhandener Daten, die mit dem Fokus auf allgemeine Rechenkompetenzen erhoben worden sind. Der Test enthielt zwar noch weitere Items zur Addition und Subtraktion, allerdings war bei diesen eine klare Unterscheidung zwischen effizienten und weniger effizienten Strategien nicht möglich (z. B. 172+424, 824-247). Aus aufgabenorientierter Perspektive bedeutet eine effiziente Lösung eine korrekte Lösung, welche mit einer Strategie gefunden wurde, die sowohl die kognitive Belastung als auch die Anzahl der Lösungsschritte minimiert und daher schnell und am wenigsten fehleranfällig ist. Für jedes dieser vier Items wurde normativ und aufgabenspezifisch festgelegt, welche Strategien effizient (2 Punkte), teilweise effizient (1 Punkt) oder nicht effizient (0 Punkte) sind. Ein Beispiel für diese

³ Die Ergebnisse unterscheiden sich nur marginal, wenn die Punkte der einzelnen Kategorien z-standardisiert und zu einem Gesamtscore aufsummiert werden. Dementsprechend wurde ab Studie 2 vorgegangen. Um konsistent mit den Artikeln (Anhang A–C) zu bleiben, erfolgt die Darstellung in dieser Arbeit wie in den Artikeln.

Zuordnung der Strategien für die Aufgabe 702 – 698 ist in Tabelle 4.2 gegeben, die entsprechenden Kategorisierungen der anderen Items sind Anhang A zu entnehmen.

Effizienz Score = 2			
Hilfsaufgabe	Vereinfachen	Ergänzen	Kopfrechnen
$702 - 700 = 2$ $2 + 2 = 4^*$	$704 - 700 = 4$	$698 + 4 = 702$	<i>keine schriftliche Notation</i>
Effizienz Score = 1			
Stellenweise verkürzt	Schrittweise verkürzt	Indirekte Subtraktion	
$700 - 600 = 100$ $100 - 98 = 2$ $2 + 2 = 4^*$	$702 - 600 = 102$ $102 - 98 = 4^*$	$702 - 2 = 700^*$ $700 - 2 = 698$ $2 + 2 = 4^*$	
Effizienz Score = 0			
Stellenweise	Schrittweise	Schriftlicher Algorithmus	
$702 - 600 = 102$ $102 - 90 = 12$ $12 - 8 = 4$	$700 - 600 = 100$ $100 - 90 = 10$ $10 - 8 = 2$ $2 + 2 = 4$	$\begin{array}{r} 702 \\ - 698 \\ \hline 11 \\ 4 \end{array}$	

* Die Notierung dieser Lösungsschritte variiert zwischen den Schülerinnen und Schülern.

Tabelle 4.2: Zuordnung der Strategien zu den Effizienz kategorien am Beispiel der Aufgabe 702 – 698

Anhand dieser Zuordnung wurde die Effizienz der Lösungen der Schülerinnen und Schüler für jedes der vier Items bewertet. Die daraus resultierende Skala von 0 bis 8 Punkten weist eine zufriedenstellende Reliabilität auf (Cronbach's $\alpha = 0,71$) und wird in Studie 1 als Maß für die Adaptivität der Strategiewahl genutzt.

4.2.3 Schulbuchqualität und Lernerfolg in Studie 2

4.2.3.1 Schulbuchqualität für den Bereich Operative Grundaufgaben in Klassenstufe 1

Wie in Studie 1 basiert auch die Bewertung der Schulbuchqualität für die Nutzung Operativer Grundaufgaben in der ersten Klasse in Studie 2 auf den vier Dimensionen des Strategiewahl-

modells (Abschnitt 2.2). Diese wurden für das erste Schuljahr für die drei Strategien ausgearbeitet, welche die Operativen Grundaufgaben (Tausch-, Nachbar- und Umkehraufgabe) nutzen. Eine Übersicht ist in Tabelle 4.3 zu sehen.

Kategorie	Strategie-repertoire	Strategie-Verteilung	Strategie-effizienz	Strategie-auswahl
Kriterium	Instruktionsqualität der Strategien	Häufigkeit der Darstellung der Strategien	Anzahl an Übungsaufgaben	Strategievergleiche, 1+1- bzw. 1-1-Tabelle, Metakognition
Punkte	0–5 Punkte/ Strategie 0–15 Punkte/ Schulbuch	0–2 Punkte/ Strategie 0–6 Punkte/ Schulbuch	Anzahl an Übungsaufgaben	Vergleiche: 0–6 Punkte/ Schulbuch Tabellen: 0–5 Punkte/ Schulbuch Metakognition: 0–4 Punkte/ Schulbuch

Tabelle 4.3: Bewertungskategorien für die Schulbuchqualität für den Bereich Operative Grundaufgaben

Auch in Studie 2 wird für die Dimension *Strategierepertoire* die Instruktionsqualität der verschiedenen Strategien bewertet. Dabei wird im Vergleich zur ersten Studie der intermodale Transfer stärker gewichtet (Bruner, 1974). Entsprechend erhält jedes Schulbuch pro Strategie jeweils einen Punkt, wenn

- die Strategie ikonisch dargestellt wird
- die Strategie sowohl anhand von didaktischen Materialien als auch in einem für Kinder realitätsnahen Kontext dargestellt wird
- der Name der Strategie eingeführt wird
- die Ausführung der Strategie sprachlich beschrieben wird
- die Strategie anhand von Arbeitsmaterialien dargestellt wird, welche typischerweise im Klassenzimmer vorhanden sind

Beispiele für alle Kriterien sind in Abbildung 4.3 zu sehen.

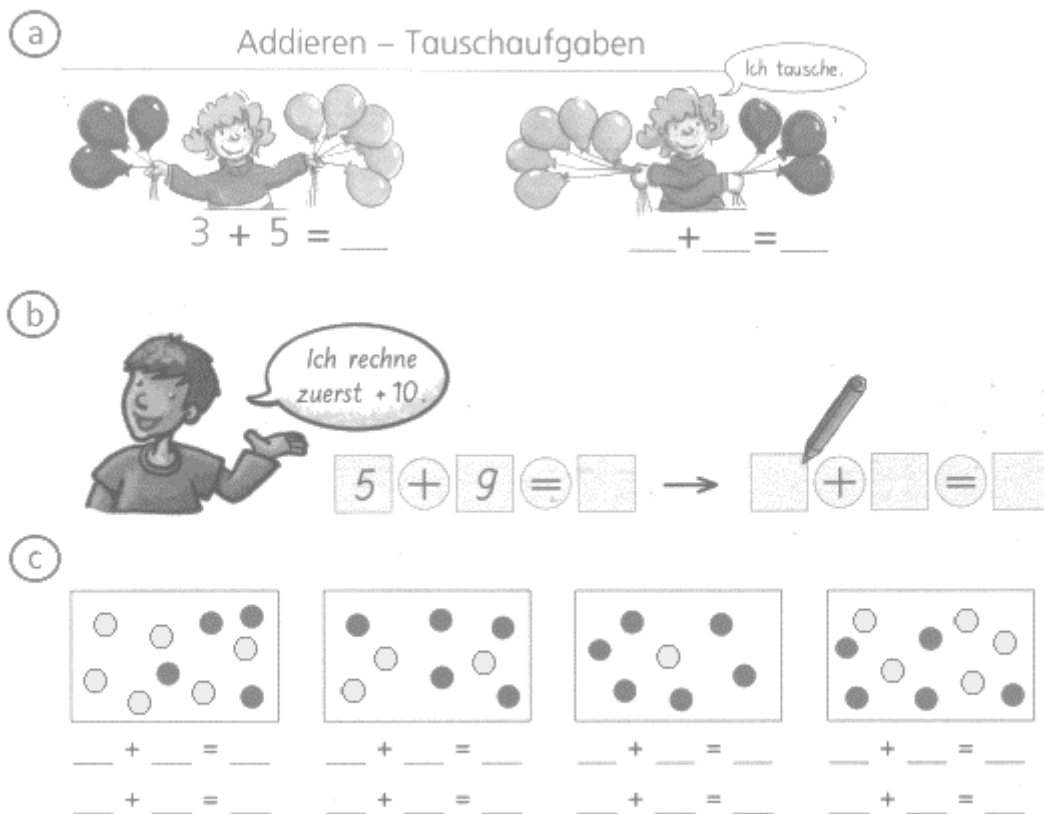




Abbildung 4.3: Beispiele für ikonische Darstellungen (a & c), einen realitätsnahen Kontext (a), Einführung der Namen (a), sprachliche Beschreibung (b) und Darstellung anhand typischerweise vorhandener Arbeitsmaterialien (Wendepüttchen, c); (a): Tauschaufgabe, Buschmeier, 2011, (b): Nachbaraufgabe, Bauer & Maurach, 2011a, (c): Tauschaufgabe, Brall, 2010

Da die wiederholte Darstellung einer Strategie die ihr beigemessene Bedeutung und damit die Häufigkeit ihrer Anwendung verstärkt, wurde auch in dieser Studie die Anzahl der Darstellungen einer Strategie innerhalb eines Schulbuchs als Kriterium für die Dimension *Strategieverteilung* gewählt. Sie wurde für jede der drei Strategien wie folgt bewertet:

- 0 Punkte, wenn die Strategie gar nicht oder nur ein einziges Mal dargestellt wird (letzteres ist in der vorherigen Kategorie bereits abgebildet)
- 1 Punkt, wenn die Strategie eingeführt und zusätzlich anhand eines zweiten Problems auf einer anderen Schulbuchseite dargestellt wird
- 2 Punkte, wenn die Strategie eingeführt und zusätzlich anhand von mehr als zwei Problemen auf verschiedenen anderen Schulbuchseiten dargestellt wird.

Wie schon in Studie 1 wurde die Anzahl an Übungsaufgaben als Indikator für die Dimension *Strategieeffizienz* gewählt. Im Gegensatz zu den halbschriftlichen Strategien beim adaptiven Rechnen lassen sich die Übungsaufgaben bei den Operativen Grundaufgaben den einzelnen Strategietypen Tausch-, Umkehr- und Nachbaraufgaben zuordnen und entsprechend differenziert analysieren. Ein Beispiel für alle drei Typen zeigt Abbildung 4.4.

(a)  $4 + 5 = \underline{\quad}$ $5 + 4 = \underline{\quad}$

 $2 + 7 = \underline{\quad}$ $7 + 2 = \underline{\quad}$

(b) $2 + 4 = 6$ $7 + 3 = \underline{\quad}$ $8 + 1 = \underline{\quad}$
 $6 - 4 = \underline{\quad}$ $\underline{\quad} - \underline{\quad} = \underline{\quad}$ $\underline{\quad} - \underline{\quad} = \underline{\quad}$
 $8 - 1 = 7$ $8 - 5 = \underline{\quad}$ $10 - 7 = \underline{\quad}$
 $7 + \underline{\quad} = \underline{\quad}$ $\underline{\quad} + \underline{\quad} = \underline{\quad}$ $\underline{\quad} + \underline{\quad} = \underline{\quad}$

(c) $5 + 5 = 10$
 $5 + 6 = 11$
 $7 + 7 = \square$
 $7 + 8 = \square$
 $6 + 6 = \square$
 $6 + 7 = \square$

Abbildung 4.4: Übungsaufgaben für die Strategien Tauschaufgabe (a, 2 Aufgaben: Buschmeier, 2011), Umkehraufgabe (b, 6 Aufgaben: Brall, 2010), und Nachbaraufgaben (c, 2 Aufgaben: Bauer & Maurach, 2011a)

Die Dimension *Strategieauswahl* wurde schließlich nach drei Kriterien bewertet: Als erstes wurden wie in Studie 1 Lerngelegenheiten zum Strategievergleich ermittelt. Da es in diesem Bereich eine geringere Anzahl verschiedener Strategien gibt, wurde anstelle einer Kategorisierung der Häufigkeiten die Anzahl dieser pro Schulbuch als Indikator gewählt. Ein Beispiel dafür zeigt Abbildung 4.5.

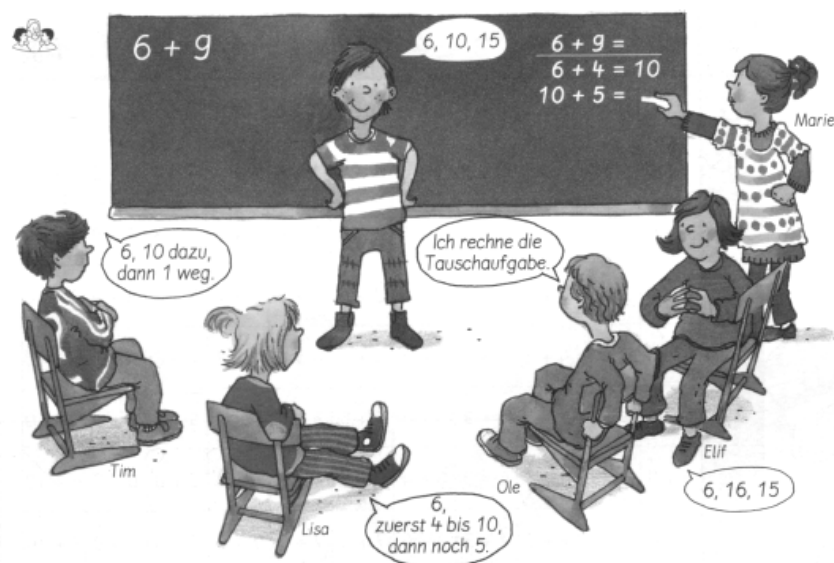


Abbildung 4.5: Beispiel für einen Strategievergleich (Rinkens et al., 2011)

Das zweite Kriterium für die Strategiewahl sind strukturierte Darstellungen der Aufgaben des kleinen Einspluseins bzw. Einsminuseins in entsprechenden Tabellen. Diese visualisieren zahlreiche Zusammenhänge und können die Nutzung operativer Beziehungen fördern. Entsprechend haben wir die Anzahl solcher Lerngelegenheiten in jedem Schulbuch ermittelt. Ein Beispiel ist in Abbildung 4.6 dargestellt.

Das dritte Kriterium für diese Kategorie sind Begründungen für die Anwendung einer Strategie. Im Gegensatz zur Beschreibung der Ausführung, welche in der Kategorie *Strategierepertoire* berücksichtigt wurde, verdeutlicht die Argumentation für eine bestimmte Strategie, *warum* eine Strategie verwendet werden sollte, und spricht damit die metakognitiven Fähigkeiten für eine flexible Auswahl von Strategien an. Entsprechend wurde die Anzahl solcher Erklärungen sowohl in der Instruktion für die Schülerinnen und Schüler als auch in den didaktischen Anmerkungen für die Lehrkraft ermittelt. Abbildung 4.7 zeigt ein Beispiel dafür.

10-10
 11-10 11-11
 12-10 12-11 12-12
 13-10 13-11 13-12 13-13
 14-10 14-11 14-12 14-13 14-14
 15-10 15-11 15-12 15-13 15-14 15-15
 16-10 16-11 16-12 16-13 16-14 16-15 16-16
 17-10 17-11 17-12 17-13 17-14 17-15 17-16 17-17
 18-10 18-11 18-12 18-13 18-14 18-15 18-16 18-17 18-18
 19-10 19-11 19-12 19-13 19-14 19-15 19-16 19-17 19-18 19-19
 20-10 20-11 20-12 20-13 20-14 20-15 20-16 20-17 20-18 20-19 20-20

5 $15 - 12 = \underline{\quad}$
 $\quad - 12 = \underline{\quad}$
 $\quad - 12 = \underline{\quad}$

6 $14 - 12 = \underline{\quad}$ $14 - 13 = \underline{\quad}$ $14 - 14 = \underline{\quad}$

7 $16 - 14 = \underline{\quad}$ $16 - 15 = \underline{\quad}$ $16 - 16 = \underline{\quad}$

8 $\quad - 17 = \underline{\quad}$ $17 - 16 = \underline{\quad}$ $17 - 17 = \underline{\quad}$

Beginne mit der letzten Aufgabe.

Abbildung 4.6: Beispiel einer Lerngelegenheit mit einer 1-1-Tabelle (Rinkens et al., 2011)

1 Finde die Tauschaufgabe und rechne.

$4 + 7 = 11$	\rightarrow	$7 + 4 = 11$
$5 + 9 = \square$	\rightarrow	$\square + \square = \square$
$3 + 8 = \square$	\rightarrow	$\square + \square = \square$
$4 + 9 = \square$	\rightarrow	$\square + \square = \square$
$5 + 7 = \square$	\rightarrow	$\square + \square = \square$

Wenn die größere Zahl vorne ist, ist es für mich leichter.




Abbildung 4.7: Beispiel für eine Begründung für die Anwendung einer Strategie (Nutzen der Tauschaufgabe: Bauer & Maurach, 2011a)

Die Erstklassbücher der vier Schulbuchreihen wurden von drei unabhängigen und geschulten Personen auf der Grundlage dieser vier Kategorien (Strategierepertoire, Strategieverteilung, Strategieeffizienz, Strategiewahl) bewertet. Die Ergebnisse zeigten eine zufriedenstellende Interraterreliabilität (Fleiss' κ : 0,77, ICC: 0,92)⁴. Anschließend wurde durch die Konsensmethode eine einheitliche Bewertung erzielt, mit Ausnahme der Kategorie *Strategieeffizienz*. Die Anzahl der Aufgaben, die von den drei Ratern für diese Kategorie ermittelt wurden, schwankte aufgrund der großen Menge an Übungen innerhalb einer Spanne von 14 % um den Mittelwert. Daher wurde der Mittelwert für die weitere Analyse gewählt. Um alle Kategorien bzw. Subkategorien gleich zu gewichten, wurden die Ergebnisse jeweils z-standardisiert und zu einem Categoriescore bzw. Gesamtscore des jeweiligen Schulbuchs aufsummiert.

4.2.3.2 Nutzung Operativer Grundaufgaben

Die Fähigkeiten der Schülerinnen und Schüler, beim Lösen von Aufgaben des kleinen Einspluseins bzw. Einsminuseins die auf den Operativen Grundaufgaben basierenden Strategien zu nutzen, wurden anhand ihrer Lösungen bei zwei Item-Typen eines Arithmetiktests am Ende der ersten Klasse erhoben. Der erste Item-Typ adressiert die sogenannte Aufgabenfamilie, welche auf der Kommutativität der Addition sowie der Komplementarität von Addition und Subtraktion beruht. Dabei sind drei Zahlen gegeben. Die Schülerinnen und Schüler sollen die vier Aufgabengleichungen der Aufgabenfamilie notieren, das heißt, je zwei Additions- und Subtraktionsgleichungen, welche die drei vorgegebenen Zahlen enthalten. Bei der Aufgabenfamilie handelt es sich, ausgehend von einer der beiden Additionsgleichungen, um die Kombination mit der Tausch- und den beiden Umkehraufgaben. Entsprechend erfordern diese Items die Nutzung

⁴ Die Interraterreliabilität der einzelnen Kategorie zeigte ebenfalls zufriedenstellende Werte. Aus Gründen der Konsistenz werden die Ergebnisse entsprechend der Berichterstattung in den Artikeln (Anhang A–C) dargestellt. Daher werden für Studie 2 und 3 globale Kennwerte angegeben.

der darauf aufbauenden Strategien. Der Test umfasste drei Items dieser Art mit einer Bearbeitungszeit von jeweils 60 Sekunden. Jede richtige Gleichung wurde mit einem Punkt bewertet. Ein Beispiel-Item ist in Abbildung 4.8 (linke Seite) dargestellt.

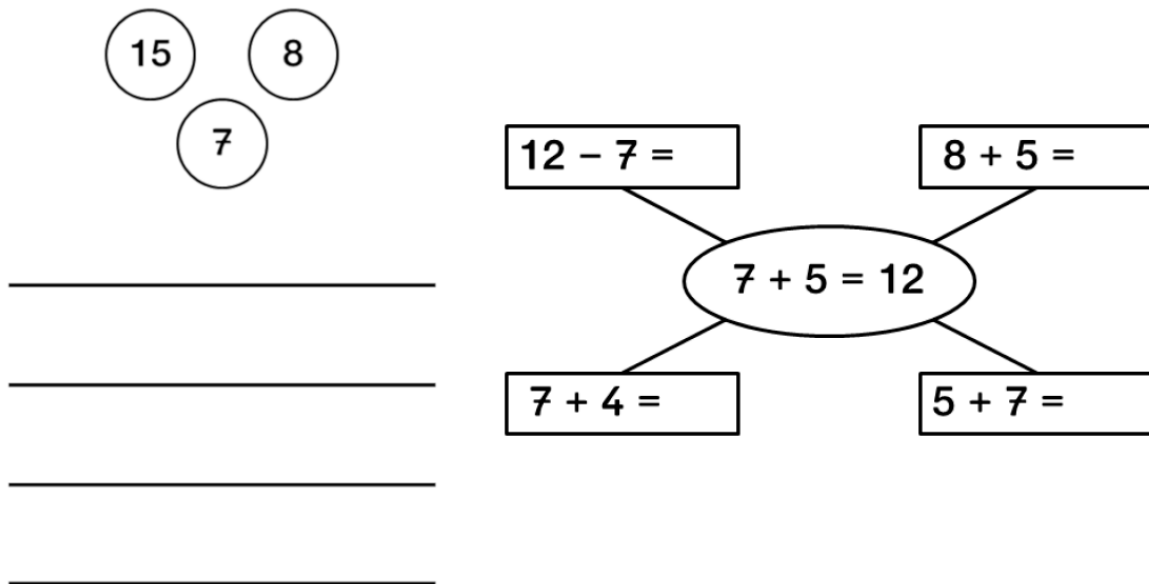


Abbildung 4.8: Beispiele für die beiden Item-Typen zur Nutzung Operativer Grundaufgaben unter Speedbedingungen

Der zweite Item-Typ (Abbildung 4.8, rechte Seite) enthält operative Beziehungen vom Typ Tausch-, Umkehr- und Nachbarschaftsaufgaben. Ein bereits gelöstes Additions- oder Subtraktionsproblem ist in der Mitte zu sehen, umgeben von vier ungelösten Problemen, welche mit dem gelösten in der Mitte in Beziehung stehen. Es handelt sich dabei jeweils um die Tauschaufgabe oder eine der Umkehr- oder Nachbaraufgaben. Entsprechend erfordert dieser Item-Typ die Nutzung aller Strategien, welche auf den Operativen Grundaufgaben basieren. Um sicherzustellen, dass die Bewertung tatsächlich die Verwendung dieser Strategien abbildet, gab es auch bei diesen Items eine restriktive Speed-Komponente: Die Schülerinnen und Schüler hatten 30 Sekunden Zeit, um alle vier Gleichungen zu lösen. Darüber hinaus wurden die Lösungen per Partial Credit Scoring bewertet: Bei vier richtigen Lösungen wurden zwei Punkte gegeben, bei drei richtigen Lösungen ein Punkt und bei zwei oder weniger null Punkte. Die Speed-Komponente und die Bewertung per Partial Credit sollen sicherstellen, dass die Lernenden nur Punkte erhalten, wenn sie die Aufgaben mithilfe der Aufgabenbeziehungen lösen, also durch Ableitungsstrategien ausgehend von der Aufgabe in der Mitte, und nicht durch Ausrechnen der einzelnen Aufgaben. Aufgrund der sehr geringen Zeit ist davon auszugehen, dass Schülerinnen und Schüler am Ende von Klassenstufe 1 kaum mehr als zwei Aufgaben rechnerisch lösen. Um diese Annahme zu validieren, wurde zusätzlich eine Interview-Studie mit 18 Schülerinnen und Schülern (7 weiblich, 39 %) in Klassenstufe 1 und 2 durchgeführt. Die Lernenden erhielten dazu die gleichen Items unter identischen Bedingungen wie in der großen Stichprobe. Im Anschluss wurden sie gefragt, wie sie die Aufgaben gelöst haben und ob sie dazu die Gleichungen in der Mitte verwendet haben. Mithilfe eines Chi-Quadrat-Tests wurden die Verteilungen der Partial Credit Scores der Schülerinnen und Schüler und ihrer Nutzung der Operativen Grundaufgaben

(dichotom) miteinander vergleichen. Die Ergebnisse zeigen einen signifikanten Zusammenhang zwischen ihrer Punktzahl und der Verwendung von Strategien, welche die Operativen Grundaufgaben nutzen, $\chi^2(2) = 6,41$, $p = 0,04$, $V = 0,34$. Das heißt, Schülerinnen und Schüler mit einer höheren Punktzahl verwenden häufiger diese Strategien.

Der Test umfasste von diesem Item-Typ ebenfalls drei Items. Die Punktzahlen der Lernenden für die sechs Items der beiden Typen wurden auf der Grundlage der Item-Response-Theorie unter Verwendung einer Erweiterung des Rasch-Modells geschätzt (Partial Credit Model, Masters, 1982). Die sich daraus ergebende Skala weist eine gute WLE-Reliabilität von 0,82 auf und wird in Studie 2 als Maß für die Fähigkeit der Schülerinnen und Schüler verwendet, auf Operativen Grundaufgaben basierende Strategien zu nutzen.

4.2.4 Schulbuchqualität und Lernerfolg in Studie 3

4.2.4.1 Schulbuchqualität für den Bereich quantitative Vergleiche in Klassenstufe 1

Im Gegensatz zu den ersten beiden Studien konnte bei der Untersuchung der Schulbuchqualität für diesen Bereich nicht auf einem etablierten Modell aufgebaut werden, da die Frage nach effektiven Lernmöglichkeiten noch nicht ausreichend beantwortet ist (Gabler & Ufer, 2020; Huang et al., 2019). Daher wurden zunächst Kriterien aus allgemeineren Theorien und dem aktuellen Stand der Forschung entwickelt und das resultierende Kategoriensystem mit Expertinnen und Experten sowohl aus der Schulpraxis als auch aus der Mathematikdidaktik diskutiert und überarbeitet. Beide Gruppen haben dabei wertvolle Anregungen und Rückmeldungen gegeben, die geholfen haben, die Kategorien und Kriterien mehrfach zu überarbeiten und zu erweitern. Das resultierende Kategoriensystem zur Bewertung der Schulbücher besteht aus fünf Kategorien: *Zahlzerlegungen mit fehlender Teilmenge*, *Komplementarität von Addition und Subtraktion*, *Subtraktion als Unterschied*, *Modellierung quantitativer Vergleiche* und *Impulse für quantitative Vergleiche im Lehrkräftemanual*. Die Kriterien und ihre Bewertung werden im Folgenden näher erläutert. Eine Übersicht gibt Tabelle 4.4.

Wie im modellierten Lösungsprozess (siehe Ende von Abschnitt 2.4) gezeigt wird, ist die Einsicht in die Teilmengenbeziehung natürlicher Zahlen und damit die Vorstellung, dass Zahlen in andere Zahlen zerlegt werden können, entscheidend für das Lösen quantitativer Vergleiche. Zudem haben *Zahlzerlegungen mit einer fehlenden Teilmenge* wie quantitative Vergleiche auch in der Regel einen statischen Charakter (siehe Abb. 4.9). Zahlzerlegungen mit fehlender Gesamtmenge stimulieren eher das mentale Modell des Vereinigens und wurden daher von der Analyse ausgeschlossen.

Kategorie	Zahlzerlegungen	Komplementarität +/-	Subtraktion als Unterschied	Quantitative Vergleiche	Lehrkräfte-Manual
Subkategorien	Einführung – Eigenständiges Üben – Wiederholung				keine
Kriterium	In jeder Subkategorie: 1 Punkt für			In jeder Subkategorie:	0 Punkte: keine Impulse
	<ul style="list-style-type: none"> - enaktive Lerngelegenheiten - ikonische Darstellungen - symbolische Darstellungen - sprachliche Darstellungen oder Aufforderung zum Sprechen - Verknüpfung zwischen den Repräsentationsebenen 			<ul style="list-style-type: none"> 1 Punkt, falls quantitative Vergleiche vorhanden sind 	<ul style="list-style-type: none"> 1 Punkt: Impulse für leistungsstarke Lernende 2 Punkte: Impulse für alle Lernenden
Punkte	0–5 Punkte/ Subkategorie	0–5 Punkte/ Subkategorie	0–5 Punkte/ Subkategorie	0–1 Punkte/ Subkategorie	0–2 Punkte/ Schulbuch
	0–15 Punkte/ Schulbuch	0–15 Punkte/ Schulbuch	0–15 Punkte/ Schulbuch	0–3 Punkte/ Schulbuch	

Tabella 4.4: Bewertungskategorien für die Schulbuchqualität für den Bereich quantitative Vergleiche



Abbildung 4.9: Beispiel für eine Lerngelegenheit in der Kategorie Zahlzerlegungen (Buschmeier, 2011)

Da einerseits Inhalten mehr Bedeutung beigemessen wird, die mehr als einmal dargestellt werden, aber andererseits eine je-mehr-desto-besser-Logik in diesem Fall kein überzeugender Qualitätsindikator ist, wurden die jeweiligen Schulbuchseiten in die Kategorien *Einführung*, *eigenständiges Üben* und *Wiederholung* eingeteilt. So konnten für jede dieser Subkategorien die folgenden Punktzahlen vergeben werden unter Berücksichtigung mehrfacher Darstellungen, ohne dabei sehr große Anzahlen zu überschätzen. Für jede Seite, die Lerngelegenheiten zu Zahlzerlegungen mit einer fehlenden Teilmenge beinhaltet, wurde zunächst die Subkategorie der Seite (*Einführung*, *eigenständiges Üben* und *Wiederholung*) bestimmt und in einem zweiten Schritt jeweils einen Punkt in dieser Subkategorie für jede der folgenden Eigenschaften vergeben:

- eine Lerngelegenheit, die ein Handlung erfordert
- eine ikonische Darstellung
- eine symbolische Darstellung
- eine Aufforderung, mit anderen darüber zu sprechen (z. B. zu erklären, sich mit anderen auszutauschen)
- eine sinnvolle Verknüpfung zwischen diesen Repräsentationsebenen

Diese Kriterien berücksichtigen Ergebnisse früherer Untersuchungen zur Bedeutung der sprachlichen Flexibilität (Schumacher & Fuchs, 2012; Stern, 1993) sowie der Vielfalt an Repräsentation und deren Integration (Huang et al., 2019; Múñez et al., 2013; Mwagi & Sweller, 1998).

Die Einsicht in die *Komplementarität von Addition und Subtraktion* eröffnet zwei Möglichkeiten, eine Differenz zu bestimmen (siehe Abschnitt 2.4). Darüber hinaus stellt sie die Äquivalenz von "mehr als / weniger als" in einem mathematischen Modell dar, dessen Verständnis sich als wichtig zum Lösen quantitativer Vergleiche erwiesen hat (Stern, 1993). Entsprechend wurde diese Kategorie in das Bewertungssystem aufgenommen. Beispiele für entsprechende Lernmöglichkeiten sind die Umkehrbarkeit der Operationen (Abb. 4.10) oder ikonisch dargestellte Rechengeschichten mit zwei möglichen Interpretationsrichtungen. Die Bewertung erfolgt analog zur vorherigen Kategorie (drei Subkategorien mit jeweils 0-5 Punkten, mögliche Bandbreite von 0-15 Punkten), da das Verständnis solcher Zusammenhänge durch die Darstellung in verschiedenen Repräsentationsformen sowie deren Verknüpfung gefördert werden sollte.



Abbildung 4.10: Beispiel für eine Lerngelegenheit der Kategorie Komplementarität (Buschmeier, 2011)

Die Vorstellung der *Subtraktion als Differenz*, das heißt, das Ergebnis einer Subtraktion kann nicht nur den Rest in einer Wegnahme-Situation, sondern auch den Unterschied zwischen zwei betrachteten Mengen darstellen, ist für einen Lösungsansatz bei quantitativen Vergleichen grundlegend (siehe Abschnitt 2.4). Darüber hinaus ist das Konzept der Differenzmenge als Beziehung zwischen zwei Mengen zentral im Lösungsprozess quantitativer Vergleiche (siehe Abschnitt 2.4). Daher wird in dieser Kategorie bewertet, ob die Idee der Subtraktion als Differenz in den Schulbüchern adressiert wird. Auch in dieser Kategorie entspricht das Bewertungsschema dem der beiden vorangegangenen. Ein Beispiel ist in Abbildung 4.11 dargestellt.

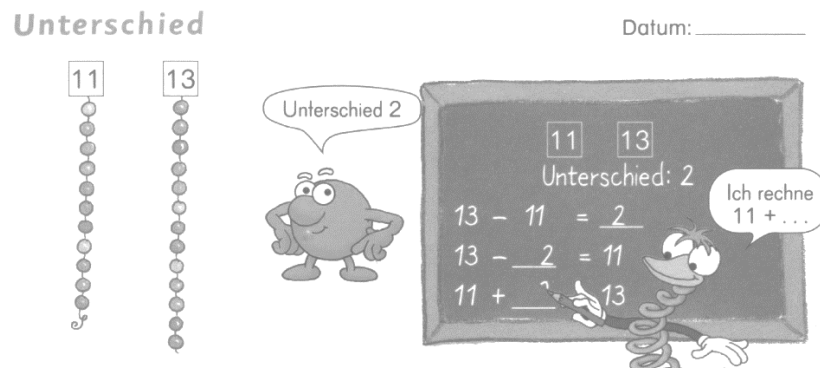


Abbildung 4.11: Beispiel für eine Lerngelegenheit zur Subtraktion als Unterschied (Brall, 2010)

Quantitative Vergleiche sind natürlich selbst eine wichtige Lerngelegenheit für diesen Bereich (Schumacher & Fuchs, 2012). In Schulbüchern werden sie typischerweise schriftlich oder ikonisch als Rechengeschichte präsentiert. Es wurde daher in der Kategorie *Modellierung quantitativer Vergleiche* bewertet, ob die Bücher solche Lernmöglichkeiten enthalten. Da diese stets in einer Repräsentationsform dargestellt werden, welche modelliert, das heißt, übersetzt und in eine symbolische Form gebracht werden muss, wurden die Verknüpfungen in dieser Kategorie nicht berücksichtigt. Stattdessen wurden wieder die gleichen drei Unterkategorien verwendet und für jede davon bewertet, ob quantitative Vergleiche vorhanden sind oder nicht. Dies führte zu einer möglichen Spanne von 0 bis 3 Punkten pro Schulbuch in dieser Kategorie.

Bei ersten Analysen der Schulbücher ist aufgefallen, dass aufgrund der Tatsache, dass die meisten Probleme in Klassenstufe 1 nonverbal dargestellt sind, bei einer Reihe von Bildern potenziell eine vergleichende Interpretation möglich ist. In der Regel regt jedoch der Kontext oder die Aufgabe zu anderen, meist dynamischen Interpretationen an. Darüber hinaus gibt es Beispiele wie in Abbildung 4.12: In einer Multiple-Choice-Beispielaufgabe, bei der die richtige Gleichung zugewiesen werden soll, wird die potenzielle und korrekte vergleichende Interpretation ($6 - 2 = \underline{\quad}$) als falsch dargestellt. Zur Reduzierung von Interpretationen wurde deshalb entschieden, das Lehrkräftemanual zusätzlich auf mögliche Impulse für quantitative Vergleiche zu untersuchen und diese in der Kategorie *Impulse für quantitative Vergleiche im Lehrkräftemanual* zu bewerten. Dies geschah per Partial Credit Scoring:

- 0 Punkte, falls keine derartigen Impulse gegeben werden
- 1 Punkte, falls Impulse zum quantitativen Vergleich als Differenzierungsaufgabe für leistungsstarke Schülerinnen und Schüler gegeben werden
- 2 Punkte, wenn Impulse zum quantitativen Vergleich für alle Schülerinnen und Schüler gegeben werden

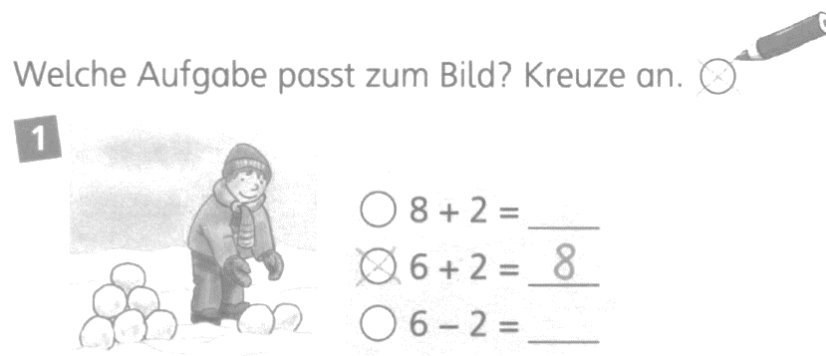


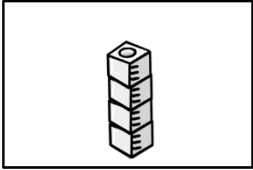
Abbildung 4.12: Vergleichende Interpretation als falsche Antwort (Buschmeier, 2011)

Die vier Schulbücher für die erste Klasse wurden wiederum von drei unabhängigen und geschulten Personen auf Basis der vier Schulbuchkategorien bewertet (*Zahlzerlegungen mit fehlender Teilmenge, Komplementarität von Addition und Subtraktion, Subtraktion als Unterschied, Modellierung quantitativer Vergleiche*). Die Manuale für die Lehrkräfte wurden von einer der drei Personen untersucht, da entsprechende Impulse, sofern vorhanden, expliziert sind und nicht mittels Interpretation bewertet wurden. Die Ergebnisse weisen eine substantielle Übereinstimmung zwischen den Ratern auf (Fleiss' κ : 0,74, ICC: 0,90). Durch die Konsensmethode wurde anschließend ein einheitliches Bewertungsergebnis erzielt, mit Ausnahme der Kategorie *Strategieeffizienz*. Um alle Kategorien bzw. Subkategorien gleich zu gewichten, wurden die Ergebnisse jeweils z-standardisiert und zu einem Kategorie- bzw. Gesamtscore des jeweiligen Schulbuchs aufsummiert.

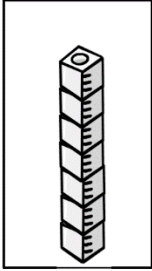
4.2.4.2 Lösen quantitativer Vergleiche

Die Bewertung der Fähigkeit der Schülerinnen und Schüler, quantitative Vergleiche zu lösen, basiert auf zwei Item-Typen aus einem Arithmetiktest, welcher nach den ersten sieben Monaten in Klassenstufe 1 eingesetzt worden ist. Der Test enthält verschiedene Typen von Items aus dem Arithmetikcurriculum für das erste Schuljahr. Die beiden hier genutzten Item-Typen enthalten jeweils einen quantitativen Vergleich in Textform und sind identisch strukturiert mit dem Unterschied, dass der zweite Typ zusätzlich eine ikonische Darstellung der Situation enthält. Abbildung 4.13 zeigt ein Beispiel-Item.

Tom hat 4 Steckwürfel. Anna hat 7 Steckwürfel.



Tom



Anna

Wie viele Steckwürfel hat Anna mehr als Tom?

Schreibe die Aufgabe dazu auf.

Abbildung 4.13: Beispiel-Item mit ikonischer Darstellung

Bei jedem Item wird ein statischer, quantitativer Vergleich mit unbekanntem Unterschied präsentiert („Vergleich: Unterschied gesucht“, Schipper, 2009). Die Schülerinnen und Schüler werden nach dem Unterschied gefragt und sollen anschließend eine entsprechende Aufgabe dazu aufzuschreiben. Die Frage enthält stets die Formulierung, wer "mehr" als der andere hat. Der Test enthält drei Items mit und drei ohne ikonische Darstellung. Bei jedem Item wurde 1 Punkt für eine korrekte Differenz und 1 Punkt für eine geeignete Modellierung gegeben. Als geeignete Modellierung wurde dabei jede vollständige oder partielle Gleichung bzw. Aufgabe akzeptiert, die zur Bestimmung der korrekten Differenz verwendet werden kann, das heißt, Lösungen wie bspw. " $7 - 4$ ", " $4 + _ = 7$ ", " $7 - ? = 4$ " oder " $4 + 3 = 7$ ". Die Punktzahlen der Schülerinnen und Schüler für die sechs Items wurden auf Grundlage der Item-Response-Theorie skaliert. Die resultierende Skala weist eine zufriedenstellende WLE-Reliabilität von 0,76 auf und wird als Maß für die Fähigkeit verwendet, quantitative Vergleiche zu lösen.

5 Zusammenfassungen der Studienergebnisse

In diesem Kapitel werden die Ergebnisse der drei im Rahmen dieser Arbeit durchgeführten Studien zusammengefasst und vorgestellt. Detaillierte Beschreibungen der Untersuchungen sind in den Anhängen A bis C in Form von Zeitschriftenartikeln gegeben. Die Diskussion und Einordnung der Ergebnisse erfolgt in Kapitel 6.

5.1 Zusammenfassung der Ergebnisse von Studie 1

In der ersten Studie wird die Rolle des Schulbuchs für das adaptive Rechnen bei der Addition und Subtraktion in Klassenstufe 3 untersucht. Dafür wurden zunächst die Zweit- und Drittklassausgaben der vier Schulbuchreihen aus der IPN-Studie anhand der in Abschnitt 4.2.2.1 dargestellten Kriterien analysiert und so die Schulbuchqualität für den Bereich adaptives Rechnen bei der Addition und Subtraktion bestimmt. Im Anschluss wurden die Längsschnittdaten (Klasse 1 bis 3) von 1.404 Schülerinnen und Schülern der in Abschnitt 4.1.3 beschriebenen Substichprobe in einem Zwei-Ebenen-Modell reanalysiert. Dadurch konnte der Zusammenhang zwischen der Strategienutzung der Schülerinnen und Schüler bei der Addition und Subtraktion bis 1.000 und der entsprechenden Schulbuchqualität unter Kontrolle individueller sowie unterrichtlicher Bedingungsfaktoren untersucht werden. Neben dem Haupteffekt der Schulbuchqualität auf die Adaptivität der Strategienutzung wurde zudem eine Moderation des arithmetischen Vorwissens durch die Schulbuchqualität untersucht.

Klasse	Strategie-repertoire		Strategie-Verteilung		Strategie-effizienz		Strategie-auswahl		Mittlerer Rang	
	2	3	2	3	2	3	2	3	2	3
Punktzahl										
Denken und Rechnen	21	15	8	7	1146	815	2	2		
Einstern	17	10	7	4	548	727	0	0		
Flex und Flo	22	24	6	8	645	1371	0	2		
Welt der Zahl	22	23	8	10	1392	1769	2	2		
Rang										
Denken und Rechnen	2	3	1	3	2	3	1	1	1.50	2.50
Einstern	3	4	2	4	4	4	2	2	2.75	3.50
Flex und Flo	1	1	3	2	3	2	2	1	2.25	1.50
Welt der Zahl	1	2	1	1	1	1	1	1	1.00	1.25

Tabelle 5.1: Ergebnisse der Schulbuchanalyse in Studie 1 (absolute Punktzahlen und Rangfolge)

Die Ergebnisse der Schulbuchanalyse zeigen deutliche Unterschiede zwischen den vier Schulbuchreihen in der Qualität für den Bereich adaptives Rechnen für beide Schuljahre (Tabelle 5.1). Bei den Rangfolgen in den einzelnen Kategorien ist die Reihe „Welt der Zahl“ in allen (Klasse 2) bzw. fast allen (Klasse 3) Kategorien auf dem ersten Platz, während andererseits die Bücher der Reihe „Einstern“ in allen (Klasse 3) bzw. fast allen (Klasse 2) Kategorien

den letzten Platz belegen. Die anderen beiden liegen in unterschiedlicher Verteilung in den meisten Kategorien zwischen diesen beiden. Für Klasse 2 ergeben sich mittlere Ränge von 1,00 (Welt der Zahl), 1,50 (Denken und Rechnen), 2,25 (Flex und Flo) und 2,75 (Einstern). Für Klasse 3 liegen sie bei 1,25 (Welt der Zahl), 1,50 (Flex und Flo), 2,50 (Denken und Rechnen) und 3,50 (Einstern). Schulbücher mit höheren Rängen enthalten mehr Strategien bzw. variationsreichere und explizite Darstellungen dieser, eine größere Verteilung verschiedener Strategien sowie systematisch verteilt Lerngelegenheiten zum Vergleich von Strategien. Auf Grundlage dieser Ergebnisse lässt sich Forschungshypothese H 1.1 bestätigen. Zur leichteren Interpretation wurden die Skalen für die weiteren Analysen umgedreht (d. h. 3 ist die höchste und 1 die geringste bereichsspezifische Schulbuchqualität), sodass positive Koeffizienten in der Mehrebenenanalyse positive Effekte auf den Lernerfolg der Schülerinnen und Schüler bedeuten.

	Modell 1	Modell 2	Modell 3	Modell 4
Ebene 1 (Schülerinnen und Schüler)				
Kognitive Grundfähigkeiten	.02 (.03)	.02 (.03)	.02 (.03)	.02 (.04)
Arithmetisches Vorwissen	.32** (.04)	.32** (.04)	.32** (.04)	-.07 (.14)
Ebene 2 (Klasse)				
Kognitive Grundfähigkeiten (aggregiert)	.09** (.02)	.08 (.06)	.06 (.04)	.06 (.04)
Arithmetisches Vorwissen (aggregiert)	.02 (.04)	.04 (.05)	.02 (.05)	.02 (.05)
Förderprogramm Gruppe 1		.02 (.12)	.09 (.12)	.10 (.11)
Förderprogramm Gruppe 2		-.10 (.13)	.04 (.13)	.02 (.13)
Lehrkraftqualifikation (Klasse 2)		-.13 (.07)	-.06 (.07)	-.06 (.07)
Lehrkraftqualifikation (Klasse 3)		.00 (.09)	.00 (.08)	.01 (.08)
Grad an konstruktivistischen Überzeugungen der Lehrkraft (Klasse 2 ^a)		-.01 (.05)	-.02 (.05)	-.01 (.05)
Schulbuchqualität (Klasse 2)			.19** (.07)	.18** (.07)
Schulbuchqualität (Klasse 3)			-.07 (.07)	-.05 (.07)
Cross-level Interaktion				
Arithmetisches Vorwissen x Schulbuchqualität (Klasse 2)				.17** (.06)
Arithmetisches Vorwissen x Schulbuchqualität (Klasse 3)				.00 (.05)
<i>Intercept</i>	<i>-0.03 (0.04)</i>	<i>0.08 (0.14)</i>	<i>-0.30 (0.21)</i>	<i>-0.33 (0.21)</i>
Erklärte Varianz innerhalb der Klassen	11.6 %	11.4 %	11.4 %	12.2 %
Erklärte Varianz zwischen den Klassen	10.1 %	18.5 %	31.5 %	28.9 %

* $p < .05$, ** $p < .01$, ^a Daten zu den Überzeugungen der Lehrkräfte in Klasse 3 waren im Datensatz nicht vorhanden.

Tabelle 5.2: Ergebnisse der Mehrebenenanalyse in Studie 1 (Standardfehler in Klammern, alle Variablen sind standardisiert mit Ausnahme der Teilnahme am Förderprogramm (0,1), der Lehrkraftqualifikation in Klasse 2 und 3 (0,1,2) und der Schulbuchqualität in Klasse 2 und 3 (mittlerer Rang))

In den Analysen der Zwei-Ebenen-Modellierung zeigen sich auf individueller Ebene keine Effekte der kognitiven Grundfähigkeiten auf die Adaptivität der Strategienutzung der Schülerinnen und Schüler, aber signifikante Effekte des arithmetischen Vorwissens am Ende von Klassenstufe 1 (Tabelle 5.2). Diese bestätigen vorherige Befunde zur Bedeutung des Vorwissens für die flexible Strategienutzung (Fuson et al., 1997; Rittle-Johnson et al., 2009; Schulz, 2018; Threlfall, 2009). Auf Klassenebene zeigen sich keine signifikanten Effekte der Klassenkomposition bezüglich des arithmetischen Vorwissens, der Teilnahme am Förderprogramm oder der Lehrkraftqualifikation und -überzeugungen. Einen kleinen, aber nicht robusten signifikanten Effekt weist die Klassenkomposition bezüglich kognitiver Grundfähigkeiten in einem der vier Modelle auf. Die Schulbuchqualitätsskala für das zweite Schuljahr hat einen signifikanten Effekt auf die Strategienutzung der Schülerinnen und Schüler ($B = 0,19$, $p < 0,01$). Das bedeutet, dass sich Lernende in Klassen, die ein Schulbuch mit der höchsten Qualität nutzen, und Lernende in Klassen, die ein Schulbuch mit der niedrigsten Qualität nutzen, im Mittel um 0,34 Standardabweichungen bei der Adaptivität ihrer Strategiewahl unterscheiden. Die Schulbuchqualität in Klassenstufe 3 zeigt keinen zusätzlichen Effekt, was frühere Studienergebnisse zur Bedeutung einer frühen Betonung der flexiblen Strategienutzung bestätigt (Blöte et al., 2001; Klein et al., 1998). Durch die Aufnahme der Schulbuchqualität ins Modell steigt der Anteil an aufgeklärter Varianz zwischen Klassen um 13 Prozentpunkte. Im Hinblick auf Forschungshypothese H 1.2 wurde ein signifikanter statistischer Effekt der Schulbuchqualität auf die Adaptivität der Strategienutzung der Schülerinnen und Schüler gefunden und können diese bestätigen.

Auch die Cross-Level-Interaktion der beiden signifikanten Haupteffekte, das heißt, des arithmetischen Vorwissens am Ende von Klassenstufe 1 auf Individualebene und der Schulbuchqualität in Klassenstufe 2 auf Klassenebene, zeigt einen signifikanten Effekt. Damit profitieren Lernende mit einem höheren Maß an arithmetischem Vorwissen noch stärker von Schulbüchern mit hoher Qualität. Die Qualität der Schulbücher moderiert in diesem Bereich also den Effekt des arithmetischen Vorwissens auf die Adaptivität der Strategienutzung der Schülerinnen und Schüler. Damit kann auch Forschungshypothese H 1.3 auf Basis dieser Ergebnisse bestätigt werden.

5.2 Zusammenfassung der Ergebnisse von Studie 2

Studie 2 adressiert die Bedeutung der bereichsspezifischen Schulbuchqualität für den Lernerfolg der Schülerinnen und Schüler bei der Nutzung Operativer Grundaufgaben in Klassenstufe 1. Analog zur ersten Studie wurden die Ausgaben der vier Schulbuchreihen für das erste Schuljahr auf Grundlage der Kriterien aus Abschnitt 4.2.3.1 analysiert und deren Rolle für den Lernerfolg der Kinder in diesem Bereich anhand der Längsschnittdaten (Anfang bis Ende Klasse 1) einer Substichprobe von 1.614 Schülerinnen und Schülern untersucht (siehe Abschnitt 4.1.4). Durch die Modellierung in zwei Ebenen wurde dabei wiederum die geschachtelte Datenstruktur berücksichtigt und entsprechend für individuelle und unterrichtliche Kovariaten kontrolliert. In Studie 2 wurde zudem die Anzahl an Unterrichtsstunden zum Thema Operative Grundaufgaben als Maß für Unterrichtsaktivität ins Modell aufgenommen und eine Interaktion von Schulbuchqualität und Unterrichtsquantität in diesem Bereich getestet.

	Strategie- repertoire	Strategie- verteilung	Strategie- effizienz	Strategie- auswahl	Summen- score
Denken und Rechnen	-1.26	-1.73	-1.58	-0.45	-1.71
Einstern	1.26	0.58	0.19	0.34	0.81
Flex und Flo	0.63	0.58	1.20	-1.30	0.38
Welt der Zahl	-0.63	0.58	0.19	1.41	0.53

Tabelle 5.3: Ergebnisse der Schulbuchanalyse in Studie 2 (z-Scores der vier Kategorien und z-standardisierter Summenscore)

Auch für den Bereich Operative Grundaufgaben zeigen sich erhebliche Unterschiede zwischen den vier Schulbuchreihen in der Qualität ihrer Lerngelegenheiten, wobei sich die Rangfolgen im Vergleich zu Studie 1 stark unterscheiden (Tabelle 5.3). Das Schulbuch „Denken und Rechnen“ weist in diesem Fall die geringste Qualität in drei der vier Kategorien auf und damit auch in der Gesamtpunktzahl, während die Bewertungen der anderen drei Schulbuchreihen zwischen den Kategorien stark variieren. Jede von ihnen erreicht in jeweils einer Kategorie den ersten Rang. Im Gegensatz zur ersten Studie sind die Punktzahlen in den Subkategorie standardisiert und zu einer standardisierten Gesamtpunktzahl aufsummiert worden, welche als Skala in die Mehrebenenanalyse eingegangen ist. Auf dieser sind die Schulbücher wie folgt eingeordnet: Denken und Rechnen (-1,71), Flex und Flo (0,38), Welt der Zahl (0,53) und Einstern (0,81). Eine höhere Punktzahl steht dabei für eine höhere Schulbuchqualität. Entsprechend können wir die Forschungshypothese H 2.1 bestätigen.

Ebenfalls zeigen sich zudem in der Mehrebenenanalyse von Studie 2 signifikante Effekte des individuellen Vorwissens – sowohl des mathematischen als auch des sprachlichen Vorwissens und der kognitiven Grundfähigkeiten zu Schulbeginn (Tabelle 5.4). Auf Ebene der Klasse weisen auch hier weder die Klassenkomposition noch die Teilnahme am Förderprogramm oder die Lehrkraftqualifikation statistisch bedeutsame Effekte auf. Signifikante Prädiktoren für die Nutzung Operativer Grundaufgaben sind dagegen die Qualität der Schulbücher sowie die Quantität der Unterrichtsstunden. Beide haben einen positiven Effekt auf den Lernerfolg. Der Anteil an aufgeklärter Varianz zwischen Klassen erhöht sich um 23,3 bzw. 20,1 Prozentpunkte (Schulbuchqualität bzw. Unterrichtsstunden) bei Hinzunahme zum Modell. Der Unterschied zwischen Klassen mit der höchsten und Klassen mit der niedrigsten Schulbuchqualität beträgt im Mittel 0,35–0,40 Standardabweichungen. Demnach lässt sich auch Forschungshypothese H 2.2 bestätigen.

In Bezug auf die dritte Forschungshypothese H 2.3, einer Interaktion zwischen dem Schulbuch als Lernressource und der Anzahl an Unterrichtsstunden als Indikator für Unterrichtsaktivitäten, weisen die Daten keinen solchen Effekt auf. Daher müssen wir Forschungshypothese H 2.3 verwerfen. Obwohl es keinen solchen Interaktionseffekt auf die Strategienutzung der Schülerinnen und Schüler gibt, zeigen die Skalen eine schwache bis moderate Korrelation ($r = 0,24$, $p < 0,05$).

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
Ebene 1 (Schülerinnen und Schüler)					
Kognitive Grundfähigkeiten	.21** (.04)	.21** (.04)	.21** (.04)	.21** (.04)	.21** (.04)
Sprachliche Lernvoraussetzungen	.15** (.02)	.14** (.02)	.15** (.02)	.15** (.02)	.15** (.02)
Mathematische Lernvoraussetzungen	.27** (.04)	.27** (.04)	.27** (.04)	.27** (.04)	.27** (.04)
Ebene 2 (Klasse)					
Kognitive Grundfähigkeiten (aggregiert)	.06 (.04)	.04 (.04)	.02 (.04)	.02 (.03)	.02 (.04)
Förderprogramm Gruppe 1	-.02 (.09)	.03 (.08)	.01 (.08)	.04 (.07)	.04 (.08)
Förderprogramm Gruppe 2	-.03 (.10)	-.09 (.09)	-.03 (.09)	-.07 (.09)	-.08 (.09)
Lehrkraftqualifikation	.08 (.07)	.09 (.07)	.01 (.06)	.00 (.07)	.01 (.07)
Schulbuchqualität		.16** (.03)		.13** (.03)	.14** (.04)
Quantität der Unterrichtsstunden			.15** (.04)	.11** (.04)	.12** (.04)
Schulbuchqualität x Quantität der Unterrichtsstunden					.02 (.03)
<i>Intercept</i>	<i>-0.01 (0.08)</i>	<i>-0.03 (0.07)</i>	<i>0.03 (0.07)</i>	<i>0.00 (0.07)</i>	<i>-0.01 (0.07)</i>
Erklärte Varianz innerhalb der Klassen	31.4%	31.2%	31.2%	31.1%	31.1%
Erklärte Varianz zwischen den Klassen	6.4%	29.7%	26.5%	41.4%	41.9%

* $p < .05$, ** $p < .01$

Tabelle 5.4: Ergebnisse der Mehrebenenanalyse in Studie 2 (Standardfehler in Klammern, alle Variablen sind standardisiert mit Ausnahme der Teilnahme am Förderprogramm (0,1) und der Lehrkraftqualifikation (0,1))

5.3 Zusammenfassung der Ergebnisse von Studie 3

In der dritten Studie wurde untersucht, welche Rolle das Schulbuch für die Fähigkeit der Lernenden spielt, quantitative Vergleiche in Klassenstufe 1 zu lösen. Wie in den ersten beiden Studien wurde die bereichsspezifische Qualität der vier Schulbuchreihen anhand der Kriterien aus Abschnitt 4.2.4.1 bewertet und deren Effekt auf die entsprechenden Fähigkeiten der Schülerinnen und Schüler in einer Mehrebenenmodellierung der Längsschnittdaten einer Substichprobe von 1.513 Lernenden (siehe Abschnitt 4.1.5) unter Kontrolle individueller und unterrichtlicher Bedingungsfaktoren analysiert.

	Zahl- zerlegungen	Komplemen- tarität +/-	Subtraktion als Unter- schied	Quantitative Vergleiche	Lehrkräfte- Manual	Summen- score
Denken und Rechnen	0.82	-0.12	-0.93	0	-0.30	-0.18
Einstern	-1.63	0.35	-0.93	0	-1.51	-1.28
Flex und Flo	0.00	-1.50	0.42	0	0.90	-0.06
Welt der Zahl	0.82	1.27	1.44	0	0.90	1.52

Tabelle 5.5: Ergebnisse der Schulbuchanalyse in Studie 3 (z-Scores der fünf Kategorien und z-standardisierter Summenscore)

Substantielle Unterschiede bei der Qualität der vier Schulbuchreihen zeigten sich auch in Hinblick auf quantitative Vergleiche in Klassenstufe 1 (Tabelle 5.5). Zwar enthält keines der untersuchten Bücher Lerngelegenheiten zu quantitativen Vergleichen, was bereits Wessel (2015) bei der Analyse von Schulbüchern festgestellt hat. Im Hinblick auf die weiteren Kriterien bezüglich der Fähigkeiten, welche für das Lösen quantitativer Vergleiche identifiziert worden sind (siehe Abschnitt 4.2.4.1), unterscheiden sich die vier Reihen jedoch erheblich. Abgesehen von der Kategorie quantitative Vergleiche erreicht die Schulbuchreihe „Welt der Zahl“ in allen Kategorien die höchste Punktzahl, während die Bücher der Reihe „Einstern“ in fast allen Kategorien die geringste Punktzahl erreicht. Die anderen beiden Reihen erreichen jeweils in einer Kategorie ebenfalls die höchste Punktzahl sowie in einer Kategorie die niedrigste Punktzahl. Entsprechendes spiegelt sich in der standardisierten Skala bestehend aus der Summe der standardisierten Punktzahlen der Subkategorien wider: Einstern (-1,28), Denken und Rechnen (-0,18), Flex und Flo (-0,06) und Welt der Zahl (1,52). Damit kann Forschungshypothese H 3.1 bestätigt werden.

Bei der Analyse der Mehrebenenmodelle zeigen sich für die Fähigkeit, quantitative Vergleiche zu lösen, auf individueller Ebene ebenfalls signifikante Effekte aller drei Lernvoraussetzungen zu Schulbeginn (sprachliche und mathematische Vorläuferfähigkeiten sowie kognitive Grundfähigkeiten, Tabelle 5.6). Wie in den beiden vorangegangenen Studien zeigen sich auf Klassenebene keine Effekte der Klassenkomposition, der Teilnahme am Förderprogramm oder der Lehrkraftqualifikation. Einen signifikanten Effekt auf den Lernerfolg der Schülerinnen und Schüler beim Lösen quantitativer Vergleiche zeigt wiederum die bereichsspezifische Qualität der Schulbücher. Lernende in Klassen mit geringer Schulbuchqualität und in Klassen mit hoher Schulbuchqualität unterscheiden sich in diesem Bereich im Mittel um 0,28 Standardabweichungen. Die Aufnahme der Schulbuchqualität ins Modell führt zu einer Zunahme der erklärten Varianz zwischen Klassen von $\Delta R^2 = 8,1$ Prozentpunkten. Anhand dieser Ergebnisse lässt sich auch Forschungshypothese H 3.2 bestätigen.

	Modell 1	Modell 2	Modell 3
Ebene 1 (Schülerinnen und Schüler)			
Kognitive Grundfähigkeiten	.25** (.03)	.25** (.03)	.25** (.03)
Sprachliche Lernvoraussetzungen	.15** (.02)	.14** (.02)	.14** (.02)
Mathematische Lernvoraussetzungen	.11** (.03)	.11** (.03)	.11** (.03)
Ebene 2 (Klasse)			
Mathematische Lernvoraussetzungen (aggregiert)	.08 (.05)	.08 (.05)	.08 (.05)
Förderprogramm Gruppe 1		.05 (.11)	.09 (.10)
Förderprogramm Gruppe 2		.14 (.11)	.18 (.10)
Lehrkraftqualifikation		-.11 (.08)	-.13 (.07)
Schulbuchqualität			.09** (.03)
<i>Intercept</i>	<i>0.01 (0.04)</i>	<i>-0.01 (0.09)</i>	<i>-0.35 (0.09)</i>
Erklärte Varianz innerhalb der Klassen	20.3%	20.3%	20.2%
Erklärte Varianz zwischen den Klassen	8.3%	13.2%	21.3%

* $p < .05$, ** $p < .01$

Tabelle 5.6: Ergebnisse der Mehrebenenanalyse in Studie 3 (Standardfehler in Klammern, alle Variablen sind standardisiert mit Ausnahme der Teilnahme am Förderprogramm (0,1) und der Lehrkraftqualifikation (0,1))

Da sowohl in Studie 2 als auch in Studie 3 die Schulbücher für das erste Schuljahr untersucht wurden, wurde zudem eine Kreuzvalidierung der entsprechenden Qualitätsskalen durchgeführt. Dazu wurden die entsprechenden Skalen für Schulbuchqualität bezüglich Operativer Grundaufgaben bzw. quantitativer Vergleiche in das jeweilige andere Modell aufgenommen, um zu kontrollieren, dass die entsprechenden Konstrukte nicht eine allgemeine Schulbuchqualität abbilden, die bereichsübergreifend prädiktiv ist. Beide Skalen zeigen für den jeweils anderen Bereich jedoch keine signifikanten Effekte, sind also nicht prädiktiv für den anderen Bereich. Dies entspricht den Erwartungen insofern, als zum einen den Analysen unterschiedliche, bereichsspezifische Kriterien zugrunde lagen und entsprechend unterschiedliche Lerngelegenheiten untersucht worden sind. Zum anderen zeigen die resultierenden Skalen stark variierende Ergebnisse. So hat beispielsweise die Serie „Einstern“ den höchsten Wert auf der Skala zu Operativen Grundaufgaben, aber den niedrigsten auf der Skala zu quantitativen Vergleichen.

6 Diskussion

6.1 Zusammenfassung und Einordnung der Studien in den Forschungsstand

Die drei Studien dieser Arbeit adressieren die Rolle von Schulbüchern als Lernressource für den Lernerfolg in drei ausgewählten Inhaltsbereichen des Arithmetikunterrichts in der Grundschule: dem adaptiven Rechnen bei der Addition und Subtraktion bis 1.000, der Nutzung Operativer Grundaufgaben und dem Lösen quantitativer Vergleiche. Die Definition der bereichsspezifischen Schulbuchqualität für jeden der drei Bereiche erfolgte auf Basis von theoriegeleiteten Analysen der Lerngelegenheiten in vier Schulbuchreihen, welche in der untersuchten Stichprobe genutzt worden sind. Die Anwendung dieser Qualitätsindikatoren auf die vier Schulbuchreihen zeigte für alle drei Bereiche Unterschiede in der bereichsspezifischen Qualität der Schulbücher. Die darauf aufbauende Sekundäranalyse von Daten einer dreijährigen Längsschnittstudie zeigte signifikante Zusammenhänge der Schulbuchqualität mit dem Lernerfolg der Schülerinnen und Schüler beim adaptiven Rechnen in Klasse 3 sowie bei der Nutzung Operativer Grundaufgaben und dem Lösen quantitativer Vergleiche in Klasse 1.

Für das adaptive Rechnen deutet zudem eine Cross-level-Interaktion darauf hin, dass die Qualität der Schulbücher den Einfluss des arithmetischen Vorwissens auf das adaptive Rechnen der Schülerinnen und Schüler moderiert. Bei der Nutzung Operativer Grundaufgaben zeigte die Quantität an Unterrichtsstunden zwar ebenfalls einen signifikanten Haupteffekt, jedoch keine Interaktion mit der Schulbuchqualität und damit keine Moderation der Schulbuchqualität durch die Anzahl an darauf verwendeten Unterrichtsstunden.

Die drei Untersuchungen bauen auf den Arbeiten zu Effekten der Schulbuchwahl auf den Lernerfolg der Schülerinnen und Schüler auf (Fagginger Auer et al., 2016; Hadar, 2017; Heinze et al., 2009; Schmidt et al., 2001; Törnroos, 2005; van den Ham & Heinze, 2018). Insbesondere erweitern sie die Erkenntnisse von Bhatt und Koedel (2012) Bhatt et al. (2013), Koedel et al. (2017) und van den Ham & Heinze (2018), indem sie den Black-Box-Modellen der Schulbucheffekte einen Erklärungsansatz durch die theoriegeleitete Analyse der Lerngelegenheiten in den genutzten Schulbüchern liefern. Dabei werden die Ansätze von Törnroos (2005) und Hadar (2017) genutzt und elaboriert, indem die Bücher ebenfalls auf Ebene der Lerngelegenheiten analysiert wurden und darauf aufbauend die Qualität der Schulbücher bereichsspezifisch bestimmt wurde. Die Ergebnisse liefern Einblicke in die Zusammenhänge der Qualität von Schulbüchern für einzelne Inhaltsbereiche und dem entsprechenden Lernerfolg der Schülerinnen und Schüler. Sie geben damit wichtige Hinweise für die Lerneffektivität von Schulbüchern und deren Bestimmung. Durch die längsschnittliche Struktur der untersuchten Daten ist es in diesen Untersuchungen zusätzlich zur Modellierung in zwei Ebenen möglich, für individuelle Unterschiede vor der Nutzung der Schulbücher zu kontrollieren. Sie folgen damit den Forderungen von Fan (2013) sowie Koedel et al. (2017) nach einer Weiterentwicklung der Forschung zu Mathematikbüchern zu mehr relationalen und kausalen Untersuchungen. Für eine umfassende

Bewertung der Qualität von Schulbüchern sind weitere Studien zu anderen Inhaltsbereichen nötig. Insbesondere für eine empirische Fundierung einer Kausalität von Schulbucheffekten bedarf es entsprechender experimenteller bzw. quasi-experimenteller Studiendesigns. Die damit verbundenen Schwierigkeiten und Anforderungen werden in Abschnitt 6.4 diskutiert.

6.2 Implikationen für die Forschung

Die in dieser Arbeit präsentierten Studien tragen sowohl zur Forschung über den Erwerb bestimmter Fähigkeiten durch die Lernenden – dem adaptiven Rechnen bei der Addition und Subtraktion bis 1.000, der Nutzung Operativer Grundaufgaben sowie dem Lösen quantitativer Vergleiche – als auch zur Forschung über die Rolle von Schulbüchern bei. Mit Blick auf die Unterrichtsforschung legen die Ergebnisse der drei Studien nahe, die Qualität von Schulbüchern als wichtigen Prädiktor bei der Untersuchung des Mathematikunterrichts zu berücksichtigen. Gemäß früherer Schulbuchforschung (siehe Abschnitt 2.1) sowie den Angaben der Lehrkräfte aus der Stichprobe dieser Untersuchung (siehe Abschnitt 2.1.4) liegt die Vermutung nahe, dass die meisten arithmetischen Aufgaben und Probleme im Mathematikunterricht aus dem Schulbuch stammen. Die Lehrerinnen und Lehrer sind für die Unterrichtsvorbereitung und den Unterricht verantwortlich. Die Ergebnisse dieser Arbeit deuten darauf hin, dass sowohl die Inhalte als auch der instruktionale Ansatz der Schulbücher einen wesentlichen Einfluss auf die Lernumgebungen haben könnten, welche Lehrkräfte in ihrem Unterricht implementieren. Dies gilt selbst und insbesondere dann, wenn Schulbücher und Lehrkräfte demselben Lehrplan folgen. Daher sollten Schulbücher bei der Untersuchung der Effektivität des Unterrichts berücksichtigt werden.

6.2.1 Implikationen für die Forschung zum Erwerb bestimmter Fähigkeiten

In Bezug auf die Entwicklung der Strategiekompetenz bei Schülerinnen und Schülern wurden in Abschnitt 2.2 die wiederholt berichteten Ergebnisse dargestellt, dass die Strategienutzung von Grundschülerinnen und Grundschülern bei der Addition und Subtraktion meist einen geringen Grad an Adaptivität aufweist (z. B. Csikos, 2016; Heinze et al., 2009; Torbeyns et al., 2017; Torbeyns & Verschaffel, 2016). Zudem wurde in Abschnitt 2.3 berichtet, dass viele Schülerinnen und Schüler über die erste Klasse hinaus Schwierigkeiten haben, die zählenden Rechenstrategien durch effizientere und breiter anwendbare Strategien zu ersetzen. Die Analyse der vier Schulbuchreihen in Studie 1 und 2 ergab, dass sich die Schulbücher in der Anzahl der enthaltenen Additions- und Subtraktionsprobleme unterscheiden. Gleichmaßen unterschieden sie sich auch bei den enthaltenen Lerngelegenheiten für die verschiedenen Additions- und Subtraktionsstrategien und deren adaptive Nutzung (Studie 1) sowie für die Operativen Grundaufgaben (Studie 2). Wie in den Abschnitten 5.1 und 5.2 ausgeführt, haben diese spezifischen Schulbucheigenschaften, abgebildet in jeweils einer Schulbuchqualitätsskala bezüglich des adaptiven Rechnens und der Operativen Grundaufgaben, einen statistisch signifikanten Einfluss auf das adaptive Rechnen der Schülerinnen und Schüler am Ende von Klassenstufe 3 bzw. die Nutzung Operativer Grundaufgaben am Ende von Klassenstufe 1. Die empirische Evidenz für diese Zusammenhänge wird durch das Längsschnittdesign, die große Stichprobengröße und die Einbeziehung relevanter Kovariaten auf Individual- und Klassenebene fundiert. Dass die explizite Thematisierung und Darstellung von Strategien sowie deren Vergleich effektive instruktionale Ansätze sind, wurde in kontrollierten experimentellen Studien gezeigt (z. B. Klein et al., 1998; Rittle-Johnson & Star, 2007, 2009; Torbeyns et al., 2009a, 2009b). Darüber hinaus legen

empirische Befunde von De Smedt et al. (2010) und Heinze et al. (2018, 2020) nahe, dass wichtige Strategien - wie das Ergänzen oder Vereinfachen - von Grundschulkindern selten ohne explizite Repräsentation erfunden werden. Die Ergebnisse der Studien 1 und 2, welche auf Daten aus dem regulären Mathematikunterricht basieren, unterstreichen die Effektivität einer Erklärung des Strategielernens anhand der Dimensionen des Siegler-Modells und unterstützen die Hypothese, dass Schülerinnen und Schüler im Mittel von der Instruktion der Strategien und deren explizitem Vergleich profitieren. In Übereinstimmung mit früheren Untersuchungen zum Einfluss von Schulbüchern auf den Strategiegebrauch der Schülerinnen und Schüler zeigen beide Untersuchungen einen solchen an (Fagginger Auer et al., 2016; Heinze et al., 2009). Entsprechend legen die Ergebnisse beider Untersuchungen nahe, das Schulbuch als relevanten Faktor für die Entwicklung der Strategiekompetenz der Schülerinnen und Schüler zu berücksichtigen.

Bezüglich des Aufbaus eines Operationsverständnisses unterstreichen die Ergebnisse von Studie 3 die Bedeutung einer breiten Erarbeitung der Operationen anhand des gesamten Spektrums unterschiedlicher Situationstypen. Obwohl die Behandlung quantitativer Vergleiche in Klasse 1 sowohl curricular als auch aus Sicht der Forschung gefordert wird (CCSSI, 2010; KMK, 2005; MoE, 2011; Reeve et al., 2012), scheinen sie beim Lernen der Addition und Subtraktion im Unterricht noch immer unterrepräsentiert zu sein (Despina & Harikleia, 2014; Tarim, 2017; Wessel, 2015). Ausgehend von Daten aus dem regulären Mathematikunterricht zeigen die Ergebnisse von Studie 3 einen Mangel an geeigneten Lerngelegenheiten zu quantitativen Vergleichen in mehreren der untersuchten Schulbuchreihen für die erste Klasse und gleichzeitig einen Zusammenhang zwischen der Qualität der Schulbücher für den Bereich quantitative Vergleiche und den Leistungen der Schülerinnen und Schüler beim Lösen solcher Probleme. In Übereinstimmung mit früheren Forschungsergebnissen weisen diese Ergebnisse also darauf hin, dass die niedrigeren Lösungsraten bei quantitativen Vergleichen nicht nur auf die den Problemen inhärente Schwierigkeit zurückzuführen sind, sondern auch auf fehlende Lerngelegenheiten in Schulbüchern, welche die im Unterricht angebotenen Lerngelegenheiten einschränken (Schmidt, McKnight et al., 1997).

Die Cross-Level-Interaktion von arithmetischem Vorwissen und Schulbuchqualität in Studie 1 deutet darüber hinaus darauf hin, dass der Einfluss der individuellen Unterschiede der Schülerinnen und Schüler auf den individuellen Lernfortschritt im Bereich des adaptiven Rechnens durch die Qualität der Schulbücher moderiert wird. Dies unterstützt die Hypothese, dass die Wirkung der Lerngelegenheiten verstärkt wird, wenn die Schülerinnen und Schüler über ein hohes Vorwissen verfügen, was dem *Matthäus-Effekt* entspricht (Bodovski & Farkas, 2007). Der mehrfach angeführte Einfluss des arithmetischen Vorwissens der Schülerinnen und Schüler auf die Entwicklung ihrer adaptiven Rechenkompetenz (z. B. Fuson et al., 1997; Rittle-Johnson et al., 2009; Schulz, 2018; Threlfall, 2009) wird demnach durch die Qualität der Lerngelegenheiten im Schulbuch moderiert. Das würde bedeuten, dass auch und insbesondere Lernende mit hohem arithmetischem Vorwissen bei der Entwicklung adaptiver Rechenkompetenz von einer hohen Qualität der Schulbücher bezüglich des adaptiven Rechnens profitieren. Dies wiederum würde bedeuten, dass auch die Schülerinnen und Schüler die in den Schulbüchern dargebotenen Lerngelegenheiten zum adaptiven Rechnen je nach ihren mathematischen Leistungen unterschiedlich nutzen. Dies gilt insbesondere für die Schulbuchreihen der genutzten Stichprobe, welche demselben Lehrplan folgen. Für die weitere Bildungsforschung ist daher von Interesse,

ob diese Differenzierung der Schulbuchnutzung auf Seiten der Lernenden auch für andere Bereiche der Mathematik gezeigt werden kann und von Bedeutung ist.

6.2.2 Implikationen für die Forschung zu Mathematikschulbüchern

Die Studien tragen außerdem zur bisherigen Forschung über die Bedeutung von Mathematikbüchern für den Lernerfolg der Schülerinnen und Schüler bei und entwickeln diese weiter. In Übereinstimmung mit den Ergebnissen von Schmidt et al. (2001), Hadar (2017) und Törnroos (2005) für die Sekundarstufe zeigen sich Hinweise darauf, dass die Wahl eines bestimmten Schulbuchs auch in der Primarstufe einen Einfluss auf die Lernergebnisse der Schülerinnen und Schüler hat. Sie unterscheiden sich von den Ergebnissen der Untersuchungen von van Steenbrugge et al. (2013) oder auch Blazar et al. (2019), welche keine bzw. kaum Unterschiede in der Mathematikleistung von Lernenden verschiedener Schulbuchgruppen berichten. Unterschiede zur Studie von van Steenbrugge et al. (2013) können unter anderem am unterschiedlichen Design liegen. Während die Stichprobengröße zwar ähnlich ist ($n = 1.579$ Schülerinnen und Schüler), handelt es sich zum einen um eine querschnittliche Untersuchung, zum anderen verteilen sich die Lernenden über sechs Klassenstufen und fünf Schulbücher. Die Ergebnisse der Schulbuchanalysen haben gezeigt, dass sich die Qualität der Schulbücher für die verschiedenen Inhaltsbereiche deutlich unterscheidet und dieselben Ausgaben einer Klassenstufe für verschiedene Inhaltsbereiche sehr unterschiedliche Qualitätsverteilungen aufweisen. Entsprechend könnten sich bereichsspezifische Effekte innerhalb eines Jahres gegenseitig reduzieren. Ebenso könnten einige Klassen in verschiedenen Schuljahren auch unterschiedliche Schulbuchreihen genutzt haben, was eine Konfundierung der Effekte bedeuten könnte. Diskrepanzen zu den Ergebnissen von Blazar et al. (2019) könnten auch auf Unterschiede im methodischen Ansatz zurückzuführen sein, wie z. B. die verwendeten aggregierten Daten auf Schulebene oder die breitere Datenerhebung in sechs Bundesstaaten. Es ist auch möglich, dass die Unterschiede aus Qualitätsunterschieden innerhalb der Schulbücher resultieren, je nach untersuchtem Thema, da die Studien dieser Arbeit Beziehungen für bestimmte Inhaltsbereiche der Schulbücher untersucht haben. Auch kann der Grad an curricularen Vorgaben, die ein Schulbuch erfüllen muss, zwischen den verschiedenen Studien unterschiedlich sein (d. h. ein Schulbuch in den USA kann ein eigenes Curriculum darstellen, während ein Schulbuch in Deutschland ein verbindliches Curriculum abbilden muss).

In den drei Studien dieser Arbeit wurden spezifische Kriterien für die Schulbuchqualität im Hinblick auf die drei Inhaltsbereiche, adaptives Rechnen bei der Addition und Subtraktion bis 1.000, Nutzung Operativer Grundaufgaben und quantitative Vergleiche, entwickelt. Die Entwicklung dieser Kriterien basierte auf dem jeweiligen Stand der Forschung. Für den Bereich quantitative Vergleiche wurden die Kriterien darüber hinaus mit Expertinnen und Experten aus der Schulpraxis und der Mathematikdidaktik diskutiert. Diese Qualitätskriterien entsprechen den empirischen Befunden zum Lernerfolg der Schülerinnen und Schüler im jeweiligen Inhaltsbereich und liefern damit eine Erklärung für die beobachteten Schulbucheffekte. Die Ergebnisse zeigen, dass eine feinkörnige und theoriebasierte Inhaltsanalyse von Mathematikbüchern ein produktiver Ansatz zur Untersuchung der Qualität und der Wirkung dieser Lernressource ist. Damit leisten die Studien dieser Arbeit einen Beitrag zur Entwicklung der Schulbuchforschung, wie sie von Fan (2013) und auch von Koedel et al. (2017) gefordert wird. Obwohl eine

globale Messung sowie ein globaler Vergleich von Schulbuchqualität ein komplexes und sicherlich weit entferntes Unterfangen ist, zeigen diese Untersuchungen, dass es für verschiedene Teilbereiche in der Arithmetik möglich ist, Beziehungen zwischen der Qualität der in Schulbüchern enthaltenen Lerngelegenheiten und dem Lernerfolg der Schülerinnen und Schüler zu identifizieren. Dies stellt einen vielversprechenden Ausgangspunkt für weitere Forschungen zur Qualität und Wirkung von Mathematikbüchern dar. Es ist denkbar, dass dieser Ansatz auf weitere Themen der Mathematik übertragen werden kann, um noch mehr über die Stärken und Schwächen von Schulbüchern zu erfahren.

Interessanterweise zeigte die Qualifikation der Lehrkraft in allen drei Studien keine signifikanten Effekte auf den Lernerfolg der Schülerinnen und Schüler. Für das adaptive Rechnen in Studie 1 wurde zudem für den Grad an konstruktivistischen Überzeugungen kontrolliert, wobei sich ebenfalls kein Effekt zeigte. Obwohl die Lehrkraft den zentralen Faktor des Mathematikunterrichts darstellt, welche verantwortlich für dessen Vorbereitung und Durchführung ist, legen die Ergebnisse der drei Studien nahe, dass die Qualität der Schulbücher einen nicht zu vernachlässigenden und wesentlichen Einfluss auf die Lernumgebungen hat, welche Lehrkräfte in ihrem Unterricht implementieren. Der Cross-Level-Effekt in Studie 1 deutet zudem darauf hin, dass der Einfluss der individuellen Unterschiede der Schülerinnen und Schüler auf den individuellen Lernfortschritt im Bereich des adaptiven Rechnens durch die Qualität der Schulbücher moderiert wird. Dies gilt insbesondere für die Schulbuchreihen der genutzten Stichprobe, welche dem gleichen Lehrplan folgen.

Obwohl in Studie 2 sowohl die Quantität der Unterrichtsstunden, die Lehrkräfte für Operative Grundaufgaben verwenden, als auch die Qualität der Schulbücher in diesem Bereich mit der Nutzung Operativer Grundaufgaben durch die Schülerinnen und Schüler zusammenhängen, weisen die Daten nicht auf eine Interaktion dieser Prädiktoren hin (H 2.3). Ein möglicher Grund dafür ist, dass die Skala der Unterrichtsaktivitäten nur deren Quantität abdeckt, nicht aber deren Qualität. Außerdem konnten, wie oben erwähnt, nur Informationen bezüglich eines der beiden Item-Typen in die Analyse einbezogen werden, was kein zufriedenstellendes Maß darstellt. Dennoch sind Schulbuchqualität und Unterrichtsquantität korreliert, was die Ergebnisse von Schmidt et al. (2001) unterstützt. Daher sollte die weitere Forschung auch das Verhältnis von Schulbuchqualität und Unterrichtsaktivitäten, insbesondere die Qualität der von der Lehrkraft angebotenen Lerngelegenheiten, berücksichtigen.

6.3 Implikationen für die Praxis

Die vier untersuchten Schulbuchreihen adressieren alle denselben Lehrplan. Die Analysen dieser Schulbuchreihen in den drei vorgestellten Studien haben substantielle Unterschiede in der Qualität der enthaltenen Lerngelegenheiten bezüglich aller drei untersuchten Inhaltsbereiche gezeigt. Damit konnte gezeigt werden, dass der Lernerfolg nicht nur eine Frage des institutionell vorgeschriebenen und verwendeten Curriculums ist, sondern auch eine Frage der konkreten Realisierung des Curriculums in Form des Schulbuchs. In Anbetracht der Tatsache, dass Schulbücher eine den Unterricht mitbestimmende Determinante darstellen, welche im Vergleich zu strukturellen Rahmenbedingungen oder Inhalten in der Lehramtsausbildung relativ kurzfristig und zu minimalen Kosten austauschbar ist, erscheint die potenzielle Verbesserung durch die Wahl eines anderen Schulbuchs vielversprechend. Aus steuerungspolitischer Sicht

können Schulbücher damit auch als ein wirksames Instrument der Qualitätssicherung im Bildungswesen betrachtet werden. Um den Schülerinnen und Schülern qualitativ hochwertige Schulbücher zur Verfügung zu stellen und damit ihre Lernprozesse bestmöglich zu fördern, liefern die vorliegenden Studien erste Qualitätskriterien für lerneffektive Mathematikbücher, welche Lehrkräfte und Lehrkräfteauszubildende nutzen können. Diese Kriterien sind theoretisch fundiert und empirisch belegt.

Nach Mullis et al. (2012) sowie den Ergebnissen der Lehrkräfteangaben der in dieser Arbeit genutzten Stichprobe nutzen Mathematiklehrerinnen und -lehrer in Grundschulen Schulbücher häufig und folgen im Allgemeinen ihren Konzepten. Im Hinblick auf die Ausbildung von Mathematiklehrkräften legen die Ergebnisse deshalb nahe, dass Lehrkräfte sich der Qualitätsunterschiede zwischen den Schulbüchern einerseits bewusst sein und diese insbesondere einschätzen können sollten. Andererseits sollten sie dazu in der Lage sein, eine geringe Qualität innerhalb des von ihnen verwendeten Buchs zu kompensieren (z. B. durch die Verwendung zusätzlicher Materialien). Da Schulbücher sich in der Qualität ihrer Lerngelegenheiten unterscheiden, sollte das fachdidaktische Wissen der Lehrerinnen und Lehrer auch die Identifizierung von Schwächen eines Schulbuchs umfassen, damit sie den Schülerinnen und Schülern alternative Lerngelegenheiten anbieten können. Obwohl es viele Hinweise darauf gibt, dass die Schulbücher intensiv genutzt werden, ist der Wissenstand darüber begrenzt, wie Lehrkräfte sie auswählen, zur Unterrichtsvorbereitung verwenden und ihren Inhalt umsetzen. Dennoch hat sich in den drei Studien dieser Arbeit ein starker Zusammenhang zwischen der Schulbuchqualität und dem Lernerfolg der Schülerinnen und Schüler gezeigt. Daher könnten das Bewusstsein für sowie die kritische Bewertung von Schulbuchqualität wichtige Themen für die Aus- und Weiterbildung von Lehrkräften sein. Entsprechend bedarf es auch hier weiterer Forschung zur Schulbuchwahl und -nutzung von Lehrkräften sowie zum Zusammenhang mit deren Professionswissen. Zudem sollte neben der Aus- und Weiterbildung der Lehrkräfte auch ein kontinuierlicher, niedrigschwelliger Transfer bisheriger Ergebnisse in die Praxis erfolgen, wie kompakte Handreichungen für Lehrerinnen und Lehrer. Ein Beispiel dafür, Lehrkräften dies entsprechend zur Verfügung zu stellen, ist für den Inhaltsbereich quantitative Vergleiche im demnächst erscheinenden Praxisbeitrag von Sievert et al. (2020) gegeben:

„Zahlzerlegungen mit fehlender Teilmenge: Lerngelegenheiten, die die Zerlegung einer Zahl mit einer fehlenden Teilmenge darstellen, sollten sowohl handelnde (enaktive) Arbeitsaufträge beinhalten als auch bildliche (ikonische) und symbolische Repräsentationen. Die drei Ebenen sollten miteinander verknüpft sein, das heißt innerhalb einer Lerngelegenheit ineinander überführt werden. Außerdem sollten die Lernenden dabei aufgefordert werden, über ihre Lösungsansätze zu sprechen. Diese fünf Kriterien (E-I-S-Ebenen, deren Verknüpfung, sprachlicher Austausch) sollten an mehreren Stellen im Schulbuch aufgegriffen werden, also sowohl im erklärenden Teil, als auch auf Seiten zum Üben sowie beim späteren Wiederholen.

Zusammenhang von Addition und Subtraktion: Wie bei den Zahlzerlegungen sollte auch der Zusammenhang der beiden Operationen anhand der drei E-I-S-Ebenen dargestellt, diese miteinander verknüpft und die Schülerinnen und Schüler zum sprachlichen Austausch angeregt werden. Auch hier ist eine Verteilung der Kriterien über das Schulbuch in Form von Erklär-, Übungs- und Wiederholungsteil sinnvoll.

Subtraktion als Unterschied: Die Vorstellung der Subtraktion als Unterschied sollte explizit behandelt werden und die häufig nur dynamischen Subtraktionssituationen ergänzen. Sinnvoll sind statische Darstellungen, die eine Eins-zu-Eins-Zuordnung ermöglichen, und so die Differenzmenge zugänglich machen. Auch hier sollten zudem enaktive (z. B. das Spiel „Hamstern“, Verboom, 2010) und symbolische Darstellungen enthalten sowie miteinander verknüpft sein und Anregungen zum sprachlichen Austausch gegeben werden. Wie in den vorherigen Kategorien sollte sich dies über das Schulbuch verteilen.

Lehrkräfteband: Das Handbuch für Lehrkräfte sollte Impulse zum quantitativen Vergleich beinhalten. Diese sollten für *alle* Lernenden gedacht sein und nicht nur leistungsstarke Kinder adressieren.

Quantitative Vergleiche: Obwohl die Schulbücher unserer Untersuchung keine direkten Lerngelegenheiten zum quantitativen Vergleich enthalten, deuten die Schülerleistungen in unserer Studie darauf hin, dass die Behandlung in Klasse 1 möglich ist. Wir gehen davon aus, dass solche Lerngelegenheiten in Schulbüchern förderlich für den Lernerfolg der Kinder sind.“

In diesem Beitrag werden entsprechend weitere Forschungsergebnisse kompakt dargestellt.

Schließlich dienen die entwickelten Qualitätskriterien auch der (Weiter-)Entwicklung von Schulbüchern und damit den Autorinnen und Autoren als Gestaltungshinweise für lernwirksame Inhalte. Entsprechend sollten weitere Studien das Repertoire an untersuchten Inhaltsbereichen erweitern. In Anbetracht der technischen Entwicklung wird dabei noch zu prüfen sein, inwieweit diese Kriterien auch für die Lernwirksamkeit digitaler Lernressourcen gelten.

6.4 Limitationen

Die Interpretation der Ergebnisse dieser Arbeit unterliegt Limitationen, die im Folgenden diskutiert werden. Einige davon sind durch die Sekundäranalyse eines bestehenden Datensatzes begründet und waren damit nicht beeinflussbar.

Zunächst lassen sich aufgrund des nicht-experimentellen Designs im Hinblick auf die Schulbuchverteilung zwar signifikante Zusammenhänge, jedoch keine Kausalität aus den Untersuchungen ableiten. Das bedeutet insbesondere, dass bei Effekten der Schulbuchwahl oder -qualität statistische Effekte des Modells, nicht aber kausale Effekte gemeint sind. Trotz dieser erheblichen Einschränkung der Ergebnisse erweitern sie den bisherigen Stand der Forschung durch die Kombination aus einer großen Stichprobe und dem längsschnittlichem Design, welches die Analysen unter Kontrolle von individuellem Vorwissen und Vorläuferfähigkeiten vor der Nutzung der Schulbücher sowie unterrichtlicher Rahmenbedingungen ermöglicht. Erstere ist nicht nur der Aussagekraft der Untersuchung dienlich, sondern zeigt auch, dass sich trotz der großen Varianz in der Art und Häufigkeit der Nutzung und des Einsatzes der Bücher im Unterricht, die vermutlich auch in dieser Stichprobe existiert, ein Zusammenhang zwischen der Schulbuchqualität und dem Lernerfolg der Schülerinnen und Schüler in den Daten nachweisen lässt. Allerdings muss auch dabei eingeschränkt werden, dass aufgrund der Fokussierung auf die vier am

häufigsten genutzten Schulbuchreihen etwa ein Drittel der Gesamtstichprobe von den Untersuchungen ausgeschlossen werden musste. Ein streng experimentelles bzw. quasi-experimentelles Design bezüglich der Schulbuchverteilung, welches den hier berichteten Studien nicht zugrunde liegt, bringt zudem eine Reihe anderer Herausforderungen mit sich. Beispielsweise bedeutet der Wechsel zu einem anderen Schulbuch für Lehrkräfte eine Umstellung bzw. Umgewöhnung, sodass bei einer zufälligen Verteilung von Lehrwerken an Klassen für einen ökologisch validen Vergleich zusätzlich zum Erhebungszeitraum eine Einarbeitungszeit von mindestens ein oder zwei Schuljahren eingeplant werden sollte. Zudem kann die Akquise von teilnehmenden Lehrkräften bei einem streng experimentellen bzw. quasi-experimentellen Design zu einer selektiven Stichprobe von Lehrkräften führen, die damit auch die Ergebnisse entsprechend verzerren. So berichten Blazar et al. (2019) von zwei solchen Untersuchungen in den USA, bei denen nur etwa 2,5 % der kontaktierten Distrikte bzw. 0,15 % der kontaktierten Lehrkräfte teilgenommen haben, und vermuten eine Unzufriedenheit bei den Teilnehmenden mit ihrem bisherigen Lehrwerk (vgl. Agodini et al., 2010; Eddy et al., 2014).

Weiter konnte in den vorgestellten Studien auch nicht kontrolliert werden, warum diese Schulbücher in den Klassen der Stichprobe verwendet wurden. Spezifische Merkmale der Lehrkräfte könnten mit bestimmten Eigenschaften der Schulbücher interagieren. In Schleswig-Holstein, wo die Daten der Stichprobe erhoben worden sind, werden Schulbücher auf Schulebene ausgewählt. Das heißt, sie werden in der Regel von der Schulkonferenz oder der Gruppe der Lehrkräfte ausgesucht, die Mathematik unterrichten. Daher ist davon auszugehen, dass der Einfluss einer einzelnen Lehrkraft auf die Schulbuchwahl gering ist. Allerdings könnten Variablen auf Schulebene wie der Standort, das Einzugsgebiet oder die Schulkultur möglicherweise einen Einfluss auf die Schulbuchauswahl haben. Vereinzelt gibt es auch in dieser Stichprobe Schulen, in denen verschiedene Schulbuchreihen genutzt wurden. Dies beschränkt sich vor allem auf größere Grundschulen. Bislang ist insgesamt wenig darüber bekannt, nach welchen Kriterien Schulen die Bücher auswählen. In den Vereinigten Staaten wurde der Auswahlprozess auf Distriktebene als sehr eigenwillig und marktgesteuert beschrieben (National Research Council, 2002). Für Deutschland vermutet Hartung (2014), dass die Schulen bei der Wahl von Schulbüchern aus Mangel an Zeit leichtfertige Entscheidungen treffen, welche durch Werbung der Verlage oder eine enge Beziehung zwischen den Vertriebsmitarbeitenden der Verlage und den Schulen beeinflusst werden. Es ist aber noch unklar, ob bestimmte Merkmale der Schule mit deren Kriterien für die Schulbuchwahl zusammenhängen. Aufgrund der Sekundäranalyse vorhandener, pseudonymisierter Daten war es nicht möglich, zusätzliche Informationen über Schulmerkmale oder die Schulbuchauswahl zu sammeln. Darüber hinaus ist die Anzahl der Klassen (Level-2-Einheiten) pro Schule nicht ausreichend für eine 3-Ebenen-Modellierung, welche die Schachtelung von Klassen in Schulen zusätzlich berücksichtigt. Zukünftige Studien sollten den Zusammenhang zwischen Schulmerkmalen bzw. deren Auswahlkriterien für Schulbücher und ihrer Schulbuchauswahl untersuchen, um die Möglichkeit einer Konfundierung durch diese zu berücksichtigen.

In den drei Studien wurde für die Qualifikation und zum Teil für den Grad ihrer konstruktivistischen Überzeugungen kontrolliert, welche in keiner der drei Untersuchungen signifikante Effekte gezeigt haben. Zwar ist die Unterscheidung fachfremder und nicht fachfremder Lehrkräfte ein erster Indikator für ihre Qualifikation, allerdings ist die Kategorie fachfremder Lehrkräfte aufgrund der heterogenen Situation der Ausbildung in Abhängigkeit vom Bundesland sowie des Zeitpunkts dieser Ausbildung sehr uneinheitlich. So gibt es Lehrkräfte, die keinerlei mathematische oder mathematikdidaktische Veranstaltung in ihrem Studium besucht haben, während andere verpflichtende Seminare und Module zur Mathematikdidaktik oder zum mathematischen Anfangsunterricht belegen mussten. In einigen Bundesländern ist ein mathematisches Grundlagenstudium wiederum verpflichtend für alle Grundschullehrkräfte. Im untersuchten Datensatz waren weiterführende Informationen zum Professionswissen der Lehrkräfte nicht verfügbar. Die Ergebnisse verschiedener Studien mit Grundschullehrkräften aus Deutschland zeigen, dass die formale Qualifikation der Lehrkräfte stark mit ihrem mathematikspezifischen Professionswissen zusammenhängt (z. B. Blömeke et al., 2010; Knievel et al., 2015; Tatto et al., 2012). Entsprechend ist ein differenzierteres Maß für das Professionswissen der Lehrkräfte für zukünftige Studien zum Einsatz von Schulbüchern wünschenswert.

Weitere Kovariaten, die mit der Qualität der Schulbücher und ihren Zusammenhang mit dem Lernerfolg interagieren könnten oder möglicherweise konfundiert sind, aber im untersuchten Datensatz nicht miterhoben wurden, sind Informationen über den tatsächlichen Schulbucheinsatz der Lehrkräfte im Unterricht. Auch der sozioökonomische Status der Schülerinnen und Schüler, der die Auswirkungen von Schulbüchern beeinflussen könnte (Hadar, 2017), konnte aufgrund des großen Anteils fehlender Daten nicht in die Analysen einbezogen werden.

Eine weitere Einschränkung ergibt sich aus der Wahl der jeweiligen Items, die zur Messung der verschiedenen Fähigkeiten genutzt worden sind. Die Adaptivität der Strategienutzung wurde anhand von vier Items aus dem Drittklasstest bewertet, die Fähigkeiten, Operative Grundaufgaben zu nutzen und quantitative Vergleiche zu nutzen, jeweils anhand von sechs Items aus verschiedenen Tests in Klassenstufe 1. Dies ist jeweils eine eher geringe Anzahl von Aufgaben für ein umfassendes Bild der untersuchten Fähigkeit. Dennoch haben die entsprechenden Skalen in allen drei Studien zufriedenstellende bis gute Reliabilitätswerte gezeigt. Zukünftige Studien zum Zusammenhang von Schulbüchern bzw. Schulbuchqualität und bestimmten Fähigkeiten der Schülerinnen und Schüler sollten entsprechend breitere und umfassendere Instrumentarien dafür einsetzen. Auch diese Einschränkung basiert auf dem Design der Sekundäranalyse.

Die Interview-Studie zur Validierung der Items zur Nutzung Operativer Grundaufgaben (Abschnitt 4.2.3.2) liefert Hinweise für einen Zusammenhang zwischen höherer Punktzahl und Gebrauch der zugrundeliegenden Strukturen, ist in ihrer Aussagekraft jedoch eingeschränkt. Beispielsweise ist eine reine Bearbeitung durch Faktenabruf oder sehr schnelles Kopfrechnen bei diesen Items nicht auszuschließen, wenn auch sehr unwahrscheinlich. Eine Untersuchung von Gasteiger (2019) mit vergleichbaren Aufgaben bei Kindern am Ende von Klassenstufe 2 – also mit Lernenden, die ein Jahr älter sind als die Teilnehmenden dieser Untersuchung – zeigte eine durchschnittliche Bearbeitungszeit von 11 Sekunden pro Aufgabe. Die Messung der tatsächlichen Strategienutzung von Schülerinnen und Schüler ist im Allgemeinen ein schwieriges und

vielschichtiges Unterfangen. Selbst wenn die Schülerinnen und Schüler unmittelbar nach dem Lösen eines Problems befragt werden, ist nicht klar, ob die von ihnen angegebene Strategie tatsächlich die von ihnen verwendete Strategie ist (Verschaffel et. al., 2007). Obwohl die zusätzliche Interview-Studie auf einen Zusammenhang zwischen der Verwendung von Strategien auf Basis der Operativen Grundaufgaben und der Anzahl korrekter Lösungen hinweist, ist sie in ihrer Aussagekraft eingeschränkt. Daher ist es möglich, wenn auch unwahrscheinlich, dass einige Erstklässler drei oder sogar vier der Gleichungen der entsprechenden Items durch ungewöhnlich schnelle Einzelberechnungen oder Faktenabruf lösen. Der Abruf von Fakten würde aber wiederum für eine vorangegangene Beziehungsorientierung beim Aufbau eines umfassenden Faktenwissens und damit für die Nutzung Operativer Grundaufgaben sprechen (Kilpatrick et al., 2001; Rechtsteiner-Merz, 2013).

Die Studien 1 und 2 zum adaptiven Rechnen und der Nutzung operativer Beziehungen zeigen zwar jeweils Zusammenhänge zwischen Schulbuchqualität und Lernerfolg der Schülerinnen und Schüler. In beiden Studien war es jedoch aufgrund von Multikollinearität nicht möglich, die Bedeutung der einzelnen Kategorien und damit der verschiedenen Dimensionen des Siegler-Modells zur Entwicklung von Strategiekompetenz zu untersuchen, indem die entsprechenden Skalen einzeln ins Modell aufgenommen wurden. Beispielsweise lässt sich anhand der Daten nicht die alternative Erklärung ausschließen, dass eine Zunahme der Adaptivität der Strategienutzung der Lernenden in Studie 1 lediglich auf den Umfang an Übungsaufgaben zurückgehen könnte. Eine weitergehende Analyse in Studie 2, bei welcher durch das Entfernen jeweils einer der Kategorien aus der Berechnung der Skala die gleichen Modelle mit alternativen Skalen genutzt wurden, zeigte jedoch, dass der statistische Effekt der bereichsspezifischen Schulbuchqualität dennoch für jede dieser Skalen signifikant bleibt (siehe Anhang B). Außerdem wird Erfahrung bzw. Routine auch im Strategiewahl-Modell als notwendige, aber nicht hinreichende Voraussetzung für die Entwicklung der adaptiven Strategienutzung betrachtet. Zudem betonen sowohl das Strategiewahl-Modell, welches empirisch fundiert ist, als auch frühere Studien, die in Abschnitt 2.2 vorgestellt wurden, die förderliche Wirkung der Thematisierung verschiedener Strategien und ihrer Verwendung auf die Adaptivität der Strategienutzung der Schülerinnen und Schüler (z. B. Blöte et al., 2001; Klein et al., 1998; Rittle-Johnson & Star, 2007, 2009; Torbeyns et al., 2009a, 2009b). Analog gilt dies für die fünf Kategorien zur Schulbuchqualität bezüglich quantitativer Vergleiche in Klasse 1. Obwohl die Rangfolgen in den einzelnen Kategorien teilweise variieren, ist es nicht möglich, alle fünf Skalen in einem Modell zu testen. Daher lassen sich keine Schlussfolgerungen für die relativen Bedeutungen der einzelnen Kategorien ziehen.

Abschließend ist anzumerken, dass in den drei Studien die Ausgaben der vier Schulbuchreihen untersucht wurden, die von den Klassen in unserem Datensatz verwendet wurden. Da die Daten in den Jahren 2013 bis 2016 erhoben wurden, sind die Ergebnisse möglicherweise nicht aussagekräftig für überarbeitete Ausgaben der jeweiligen Schulbuchreihe. Dies ändert allerdings nichts an den generellen Aussagen zum Zusammenhang von Schulbuchqualität und Lernerfolg der Schülerinnen und Schülern in den drei Inhaltsbereichen.

A Artikel zu Studie 1

Dieser Artikel wurde am 11.06.2018 erstmals eingereicht und in der untenstehenden Fassung am 26.07.2019 im Journal **Learning and Individual Differences** zunächst online und anschließend im August 2019 in Volume 74, 101716, publiziert (DOI: 10.1016/j.lindif.2019.02.006).

Effects of mathematics textbooks on the development of primary school children's adaptive expertise in arithmetic

Henning Sievert, Ann-Katrin van den Ham, Inga Niedermeyer, Aiso Heinze

Abstract: Over the past two decades educational research has become increasingly interested in the development of children's adaptive expertise in arithmetic. Even though there are several studies investigating the impact of different teaching approaches, knowledge concerning effects of learning resources is relatively low. Using data of a three-year longitudinal study with 1404 students from 82 classes, we examined learning opportunities presented by textbooks regarding adaptive expertise in multi-digit addition and subtraction, and in a subsequent multilevel analysis, the effect of the resulting textbook quality on third-graders' adaptive expertise. The results show discrepancies in the textbooks' quality concerning learning opportunities for adaptive expertise and a substantial effect of textbook quality on students' adaptive expertise after three years of schooling. An additional cross-level interaction suggests that the effect of textbook quality increases with children's prior arithmetic knowledge.

Keywords: Adaptive expertise; Textbook effects; Primary school; Mathematics

1. Introduction

Since the end of the 1990's, there has been a development in many countries towards a broad consensus about the importance of adaptive expertise in arithmetic, that is, an adaptive or flexible use of strategies in arithmetic (e.g., Baroody & Dowker, 2003; Csíkos, 2016; Kilpatrick, Swafford, & Findell, 2001; National Council of Teachers of Mathematics, 2003; Selter, 2001; Threlfall, 2009; Torbeyns, Hickendorff, & Verschaffel, 2017; Verschaffel, Greer, & De Corte, 2007). Though adaptive expertise is considered as a learning goal in primary school mathematics, empirical studies have repeatedly revealed an unsatisfying extent of adaptivity in students' actual strategy use (e.g., Csíkos, 2016; Heinze, Marschick, & Lipowsky, 2009; Selter, 2001; Torbeyns et al., 2017; Torbeyns & Verschaffel, 2016). Consequently, researchers and educators have dealt with this phenomenon and examined instructional approaches which aim to improve the development and understanding of the students' strategy use in favor of a more adaptive manner (e.g., Blöte, Klein, & Beishuizen, 2000; Heinze, Arend, Gruessing, & Lipowsky, 2018; Rittle-Johnson, Star, & Durkin, 2012; Star & Seifert, 2006; Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009a). From an instructional perspective, teaching and learning material is of significant interest in this context because it frames the teaching activities in the mathematics class and thus the opportunities to learn an adaptive use of strategies. According

to primary school mathematics teachers the most important learning resource for their mathematics instruction is the textbook (Mullis, Martin, Foy, & Arora, 2012). Consequently, the opportunities to learn regarding adaptive expertise presented in textbooks affect the teacher's instruction of adaptive expertise on the one hand, and the student's use of those on the other. It is therefore not surprising that some empirical studies indicate a possible relation between mathematics textbooks and students' strategy use (cf. Fagginger Auer, Hickendorff, van Putten, Béguin, & Heiser, 2016; Heinze et al., 2009). In addition, research shows that mathematics textbooks differ in their mathematical content as well as in certain pedagogical, cultural and sociological aspects (e.g., related to gender, ethnicity, equity, Fan, Zhu, & Miao, 2013). However, to our knowledge clear evidence for textbook effects on students' adaptive expertise is still missing.

Among others, Bodovski and Farkas (2007) could show a positive relation between prior knowledge and gains, the so called Matthew Effect, in early mathematics classes. Also for the field of adaptive expertise prior knowledge is ascribed to be an important predictor (Schulz, 2018). Thus, if textbooks indeed affect this development, it is yet to clarify, whether this effect differs among different achievement groups of students. According to the Matthew Effect, students of different achievement groups might differ in their use of learning opportunities. In other words, the textbooks moderate the effect of prior knowledge on student achievement by the opportunities to learn, which they present. Hence, it is of interest, whether this interaction can be shown for the development of adaptive expertise.

In accordance with the claim of Fan (2013) who argues for more relational and causal textbook research, we investigated the relationship between the textbooks used in class and the actual strategy use of students at the end of Grade 3. The basis of our study was an existing longitudinal data set covering 1404 students from 82 school classes during their first three years in primary school. We first conducted a theory-based textbook analysis of the mathematics textbook series used by the sample to assess the learning opportunities presented by textbooks with respect to adaptive expertise in the domain of multi-digit addition and subtraction. Subsequently, we analyzed the effect of this textbook quality measure on the students' strategy use, using a multilevel analysis, and controlling for a variety of individual and class-level covariates.

Before introducing the present study, we elaborate on the main concepts as well as on previous research on children's adaptive expertise in arithmetic and on the role of textbooks for mathematics teaching and learning.

1.1. Adaptive expertise

The notion of adaptive expertise is interpreted in different ways in the research literature. We follow the interpretation in the sense of adaptivity in mathematical strategy use, which is defined by Selter (2009) as: "Adaptivity is the ability to creatively develop or to flexibly select and use an appropriate solution strategy in a (un)conscious way on a given mathematical item or

problem, for a given individual, in a given sociocultural context” (Selter, 2009, p. 624). This definition comprises several aspects mentioned by other authors (e.g. Siegler & Lemaire, 1997; Threlfall, 2009; Verschaffel, Luwel, Torbeyns, & van Dooren, 2009). It particularly includes three influencing factors: problem or task characteristics, which might suggest certain strategies for solving; characteristics of the subject, like preferences and skills, which might induce the use of certain strategies; and the sociocultural context, which explicitly or implicitly sets norms in favor of certain strategies. In this study, we take a normative, problem-based perspective, focusing on task characteristics, motivated by the fact that a flexible or adaptive use of different strategies to generate correct solutions is a learning goal in curricula across several countries (Baroody & Dowker, 2003; Heinze et al., 2009; Mullis, Martin, Goh, & Cotter, 2016; van den Heuvel-Panhuizen, 2001). Thus, we will also assess the efficiency of strategy use task-based rather than individual- or context-based, although we consider each of the perspectives as crucial. While some authors use the terms “adaptive” and “flexible” synonymously in the context of strategy use (e.g. Blöte, van der Burg, & Klein, 2001; Threlfall, 2002), we follow Selter (2009) and Verschaffel et al. (2009) and also require the appropriateness of a solution strategy to call it adaptive. Throughout this article we will use the term “adaptive expertise”, which originates from Hatano (Hatano, 1982; Hatano & Inagaki, 1986), for the ability specified by Selter.

An essential aspect of examining a subject's adaptive expertise is the description of relevant strategies in the domain of interest. In our target domain of three-digit addition and subtraction we first distinguish between digit-based strategies, namely the written algorithms, and the (whole-)number-based strategies, which might be carried out either mentally or by means of jottings or part calculations (e.g., Selter, 2001). For number-based strategies, the transition from written to mental execution is fluid (Padberg, 2005), whereas the mental use of digit-based algorithms is described as “clumsy and error-prone” (Kilpatrick et al., 2001, p. 215). Hence, it is highly reasonable that students who solve a three-digit addition and subtraction task correctly by mental calculation do use internalized number-based strategies. In their review chapter, Verschaffel et al. (2007), on the one hand, report a general acceptance on the claim that students should learn at least one digit-based written algorithm for each operation; on the other hand, they also determine an agreement in the research literature that a premature teaching of written algorithms should be avoided due to a consequential lack of understanding. Further, they see consent in the need for a prolonged work with concrete and mental arithmetic, starting from the student's knowledge on operations, as a condition before introducing written algorithms. Also, Selter (2001) lists empirical evidence for a minor role of written algorithms for students in secondary school or in students' or adults' daily life situations.

While students usually learn one digit-based written algorithm per operation which defines a unitary approach for any kind of problem, number-based strategies allow a variety of solution approaches with varying efficiencies, depending on the particular problem given. Hence, from a problem-based perspective the number-based strategies are of special interest, since they provide a series of possible approaches for students in contrast to the unitary approach of a digit-based algorithm. For this reason we will further distinguish the number-based strategies

into the most relevant categories according to the literature. This categorization applies to both written and mental execution of number-based strategies. It will form a basis for our analyses.

Table 1 is contrasting the digit-based written algorithms and the different number-based multi-digit addition and subtraction strategies which are commonly addressed in mathematics textbooks and by teachers in the mathematics classroom (cf. Padberg, 2005; for international comparisons see Threlfall, 2002). The presentation of the strategies might vary in different countries and cultures; Table 1 shows all strategies in the symbolic representation as they were typically taught in the mathematics classes of our study. The digit-based written algorithms and the strategies “jump” and “split” are taught as universal strategies which can be applied to almost all kinds of problem types, whereas the strategies jump alternative, compensation, simplifying, and indirect addition are usually considered as task-specific strategies which are highly efficient for specific problem types only. Due to this high efficiency, it is assumed that the adaptive use of number-based strategies fosters mental calculation abilities especially in case of multi-digit numbers, because it emphasizes the efficient use of strategies so that students can gradually execute them mentally. Besides its relevance for informal contexts and adult life, fostering mental calculation is considered important for the development of insight into the place value system and general problem-solving skills, as well as for the successful learning of the digit-based written algorithms which are introduced after the number-based strategies (Threlfall, 2002). It should be mentioned that primary school students are creative in inventing additional number-based strategies and in modifying number-based strategies they learned before (e.g., Heinze et al., 2018; Selter, 2001; Threlfall, 2002). As presented below, in our study we found, for example, strategies which can be considered as shortcut-versions of the jump and the split strategy (i.e., the jump or split strategy where two solution steps were merged). These additional strategies which we empirically found in the data of our sample, are not listed in Table 1 but will be described in the method section (Section 2.2.2).

The discussion about the value and importance of number-based strategies and an adaptive use of strategies came up in the 1990's when several longitudinal studies provided insights into the conceptual structures that children built up for multi-digit numbers and their relations to and consequences for children's methods for multi-digit addition and subtraction (e.g., Fuson et al., 1997). Researchers began revealing purposes of adaptive expertise. For instance, students in reform-oriented classrooms with a high level of adaptivity also demonstrated a broader conceptual knowledge, a deeper understanding of base-ten number conceptions and participants of those classes developed less buggy-procedures of algorithms later on (Carpenter, Franke, Jacobs, Fennema, & Empson, 1997; Hiebert & Wearne, 1996). Strategic variability is further seen as the beginning of a mathematical thinking that goes beyond the knowledge of facts and procedures and as a condition for solving real-world problems, it is thus considered to be an equally important component of the development of mathematical expertise as the pure ability to answer a mathematical task accurately (Threlfall, 2009; Torbeyns et al., 2017). Accordingly, an efficient or adaptive use of strategies was adopted in many mathematics curricula or standards (Baroody & Dowker, 2003; Heinze et al., 2009; van den Heuvel-Panhuizen,

Digit-based strategies		Number-based strategies					
Written algorithm		Jump	Split	Jump alternative	Compensation	Simplifying	Indirect Addition*
1 9 8		$\underline{198 + 456 = 654}$	$\underline{198 + 456 = 654}$	$\underline{198 + 456 = 654}$	$\underline{456 + 198 = 654}$	$\underline{456 + 198 = 654}$	
+ 4 5 6		$198 + 400 = 598$	$100 + 400 = 500$	$100 + 400 = 500$	$456 + 200 = 656$	$454 + 200 = 654$	
$\begin{array}{r} 1\ 1 \\ \hline \end{array}$		$598 + 50 = 648$	$90 + 50 = 140$	$90 + 50 = 140$	$656 - 2 = 654$		
6 5 4		$648 + 6 = 654$	$8 + 6 = 14$	$8 + 6 = 14$			
			$500 + 140 + 14 = 645$	$500 + 140 + 14 = 645$			
4 0 3		$\underline{403 - 398 = 5}$	$\underline{345 - 123 = 222}$	$\underline{403 - 398 = 5}$	$\underline{403 - 398 = 5}$	$\underline{403 - 398 = 5}$	$\underline{403 - 398 = 5}$
- 3 9 8		$403 - 300 = 103$	$300 - 100 = 200$	$403 - 3 = 400$	$403 - 400 = 3$	$405 - 400 = 5$	$398 + 5 = 403$
$\begin{array}{r} 1\ 1 \\ \hline \end{array}$		$103 - 90 = 13$	$40 - 20 = 20$	$400 - 395 = 5$	$3 + 2 = 5$		
0 0 5**		$13 - 8 = 5$	$5 - 3 = 2$ ***				

*The strategy indirect addition exists only for subtraction problems.

**There are different ways to conduct the written subtraction algorithms. This version was prescribed in the curriculum for our sample.

***The recommendations for the handling of carry-overs in the split strategy are inconsistent, because negative numbers are not available in the lower grades. This is an example without carry-overs.

2001). This shift of focus also affected the education of teachers as well so that the weight of number-based strategies and their adaptive use changed significantly since the end of the 1990's. The adaptive use of strategies became a learning goal of primary school mathematics as well as a topic in the education of mathematics teachers.

A theoretical framework by Lemaire and Siegler (1995) describes determinants of adaptive expertise. They distinguish four dimensions, for which they showed that changes in any of them can improve speed and accuracy of strategy choice overall (Lemaire & Siegler, 1995): *strategy repertoire* as the knowledge of different types of strategies; *strategy distribution* as the knowledge about the relative frequencies of use of these strategies; *strategy efficiency* as the ability to perform strategies quickly and accurately; and *strategy selection* as the ability to flexibly select a strategy for a given problem. Following this model and based on a series of studies as well as computer-based models, Chen and Siegler (2000) described how students learn new strategies and their adaptive use. These results suggest that the acquisition of adaptive expertise in arithmetic takes place on the basis of suitable tasks allowing students to discover strategies, to experience strategy execution which increases strategy fluency and reduces cognitive load, and finally to perform metacognitive processes like examining strategy efficiency which might result in strengthening the association between existing strategies and problem types.

Though the previously mentioned results suggest that an explicit teaching of strategies and its adaptive use is not necessary, a number of authors have considered the effects of different instruction approaches (e.g., Klein, Beishuizen, & Treffers, 1998; Rittle-Johnson and Star, 2007, Rittle-Johnson and Star, 2009; Torbeyns et al., 2009a; Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009b). Among other advantages and effects of these approaches, an early emphasis on the flexible use of strategies was identified to result in a more eclectic use of solution procedures rather than emphasizing the acquisition of standard procedures (Blöte et al., 2001; Klein et al., 1998). Moreover, Rittle-Johnson and Star, 2007, Rittle-Johnson and Star, 2009 showed that comparing and contrasting alternative solution methods leads to greater gains in the adaptive use of strategies than isolated teaching of single methods. Concerning the role of individual prerequisites, there is a consensus among researchers that conceptual and procedural knowledge is an important predictor for students' development of adaptive expertise (e.g., Fuson et al., 1997; Rittle-Johnson, Star, & Durkin, 2009; Schulz, 2018; Threlfall, 2009).

Despite the progress in research on adaptive expertise thus far, empirical investigations repeatedly reported a lack of adaptivity for most of the students (e.g., Csíkos, 2016; Heinze et al., 2009; Torbeyns et al., 2017; Torbeyns & Verschaffel, 2016). Although a shift towards the weight of number-based strategies is perceptible, many students still tend to favor one strategy per operation which is applied to almost each type of problem. Hence, while research provides a set of learning support for fostering adaptive expertise and while national curricula and standards include the teaching of adaptive or flexible use of strategies, there is little progress regarding students' adaptive expertise in arithmetic. One reason might be lacking opportunities to learn in the mathematics class or at least a lack of quality in the learning opportunities in the students' learning resources.

1.2. Textbooks as learning resources

As elaborated in this section, learning resources like textbooks play an important role for mathematics instruction and, thus, for students' opportunities to learn in mathematics lessons. Accordingly, textbooks might also affect students' development of adaptive expertise.

Valverde, Bianchi, Wolfe, Schmidt, and Houang (2002) conceptualize the textbook as mediator between the official curriculum and the implemented curriculum by the teachers. Textbooks translate the abstract curriculum into concrete operations teachers and students can carry out. According to TIMSS 2011 data (Mullis et al., 2012), textbooks are the most important learning resource in mathematics education for primary school teachers. In Germany, where our data were collected, 86% of the primary school teachers reported that they use the textbook as a basis for their instruction (Mullis et al., 2012). In the last decades, several empirical studies provided evidence that mathematics textbooks used by teachers influence their mathematics instruction. For example, Krammer (1985) found differences in the implemented teaching practices of 50 mathematics teachers in the Netherlands using one of the three best known Dutch mathematics textbooks. Schmidt et al. (2001) analyzed the data from the TIMSS 1995 video study and found a positive correlation between the relative space of a topic in a textbook and the time spent on that topic during the videotaped mathematics lessons. Topics not covered in a textbook were unlikely to be presented in the classroom (Schmidt, McKnight, Valverde, Houang, & Wiley, 1997). In addition to these findings there are some indications that textbooks used by mathematics teachers have an effect on the achievement of their students. Analyzing the United States TIMSS 1995 eighth grade data, Schmidt et al. (2001) found a relationship between the space a topic covers in a textbook and the size of achievement gain on that topic. Similar findings regarding the interaction between mathematics textbooks and learning outcomes were reported by Törnroos (2005) using data from TIMSS 1999 in Finland. The analysis of nine mathematics textbook series used in 104 Finnish schools (grade 5–7) showed that the number of learning opportunities a textbook provided for the content of TIMSS items was significantly correlated with students' performance in the TIMSS 1999 test. In contrast, van Steenbrugge, Valcke, and Desoete (2013) examined 1579 students (Grade 1–6) from 89 classes in Flanders and did not find substantial effects of the implemented five mathematics textbook series on the performances of students.

All in all, there are some indications that the quality of learning opportunities presented by textbooks affects students' learning. With respect to the adaptive use of strategies, there are indications that the mathematics textbook used also affects primary school students' strategy choice when working on arithmetic problems (cf. Fagginger Auer et al., 2016; Heinze et al., 2009). Fagginger Auer et al. (2016) used the data from the 2011 assessment of Dutch primary schools' mathematics, including 1619 students, and conducted a multilevel latent class analysis. They showed that the textbook used by the students affects the students' probability to belong to a certain profile of strategy use. Heinze et al., (2009) examined 245 German third graders and found a relationship between the number of efficient strategies used by the students and the instructional approach of their textbook. However, in both cases the data stemmed from

cross-sectional studies and there was no information to what extent the teachers followed the textbooks when teaching arithmetic computation strategies.

1.3. Present study

The present study aims to contribute to the current research, investigating the role of the mathematics textbook as a learning resource for students' strategy acquisition and, thus, students' adaptive expertise. In contrast to most of the previously presented research, our data were collected in a longitudinal study with a sample of 1404 students from 82 classes. The data cover the first three grades of primary school in which all classes followed the same mathematics curriculum. Following the approach of Törnroos (2005), we analyzed the opportunities to learn in the four different textbook series used in our sample. For this purpose we conceptualized textbook quality with respect to adaptive expertise following research results presented in Section 1.1. This measure of textbook quality allowed us to supplement and advance the research of Fagginger Auer et al. (2016) and Heinze et al. (2009) on textbook effects on students' strategy use. In addition, considering the impact of prior knowledge on the development of adaptive expertise (Torbeys et al., 2017), the longitudinal data enable the examination of an interaction of the textbook quality and the students' prior knowledge in arithmetic.

Based on the reported state of research, this study aims to test the following Research Hypotheses:

Textbook quality

- (1) Primary school mathematics textbooks, which intend to follow the same arithmetic curriculum, vary in the quality of their opportunities to learn for adaptive expertise in multi-digit addition and subtraction.

Effects of the textbook quality

- (2) Since research shows a connection between learning opportunities in mathematics textbooks and learning outcomes (Schmidt et al., 2001; Törnroos, 2005), we expect that the mathematics textbooks' quality concerning adaptive expertise in multi-digit addition and subtraction is related to the development of primary school students' adaptive expertise in this field.
- (3) Provided Hypothesis 2 proves true: A higher prior arithmetic knowledge accompanies an increased effect of textbook quality on students' adaptive expertise in multi-digit addition and subtraction.

2. Method

2.1. Participants and design

We used an existing dataset of a three-year longitudinal evaluation of a mathematics support program for weak students in the regular mathematics classroom to test the Research Hypotheses. In the evaluation, two groups in which teachers received additional teaching material for weak students in Grade 1 and 2, were compared with a control group. The teaching material addressed aspects of basic arithmetic (e.g., concept of cardinal numbers, place value system, concepts of the basic arithmetic operations) to support children struggling with arithmetic at the beginning of primary school. The support program did not address advanced aspects like strategy efficiency or adaptive expertise and the participation was independent of the textbooks the teachers used. We controlled for the participation in the support program by dummy coded variables in the analyses (see Section 2.4) and it turned out that it had no effect on students' adaptive expertise (see Section 3.2).

The dataset comprises longitudinal data from the first to the third grade of 2330 children of 40 schools and 127 classes in Schleswig-Holstein, Germany (about 10% of the cohort in this federal state). Schools were chosen from urban and rural areas from all over the federal state. All teachers in the sample had to follow the same state-wide curriculum which prescribes content (e.g., concepts, algorithms), skills (e.g., mental computation, estimation), and representations (e.g., number line, place value chart) to be addressed in the mathematics classroom for each grade. For our analysis we used the subsample of 82 classes with 1404 children (48.1% female) taught using one of four popular mathematics textbooks.⁵ The distribution of the classes over the four textbooks is relatively even and provides a sufficient number of classes for each textbook. In Grade 1 and 2, the textbook “Denken und Rechnen” was used by 326 students, the textbook “Einstern” by 320 students, the textbook “Flex und Flo” by 433 students and the textbook “Welt der Zahl” by 325 students. In Grade 3 this distribution changed to 475 (“Denken und Rechnen”), 74 (“Einstern”), 301 (“Flex und Flo”) and 425 (“Welt der Zahl”), while the teachers of 129 third-graders changed to another or no textbook. For each of the four textbook series, almost all teachers (90.8–100.0%) stated they had used the textbook for their arithmetic instruction at least once a week and 81.1–94.1% of the teachers reported that they were oriented towards the content, methods and chronology of the textbook. Although according to the official regulation in Schleswig-Holstein, textbooks should be chosen on school level by the group of teachers, there is some variance between classes of the same school in our sample. In Grade 2, in 18 schools the same textbook series was used in all classes, while in 16 schools two different textbooks were used in different Grade 2 classes. In Grade 3, in only one school the

⁵ In 14 classes a variety of other textbooks were used, 4 classes did not use any book and for 16 classes we had no information concerning the textbook used. 11 classes had to be excluded for methodological reasons, since the number of students was less than 5.

classes differed in the textbook they used. To check whether teacher characteristics and textbook choice are confounded, we tested whether the textbooks are distributed differently between the teacher qualification groups (which was not the case, see Section 3.2 for details).

All data were collected by an authority subordinated to the Ministry of Education of Schleswig-Holstein, assuring the official human subject related approval. We received all data in a full pseudonymized condition. The compliance of the human subject related guidelines were monitored by the Data-Protection Supervisor of Schleswig-Holstein.

2.2. Measures

2.2.1. Textbooks quality regarding adaptive expertise in arithmetic

In order to test Research Hypothesis 1, we developed a measure to assess the quality of the learning opportunities presented by the textbooks regarding adaptive expertise. Indicators for textbook quality were derived from the four dimensions of adaptive expertise by Lemaire and Siegler (1995): strategy repertoire, strategy distribution, strategy efficiency, and strategy selection (see Section 1.1). Since our study addresses third-graders' adaptive expertise for multi-digit addition and subtraction, we restricted the textbook quality assessment to the chapters for addition and subtraction in textbooks for the Grades 2 and 3.

Opportunities to learn with respect to the dimension *strategy repertoire*, that is knowledge of different types of strategies, were assessed by analyzing the number-based strategies introduced in the textbook. The strategies were classified into the six types of number-based strategies described in Table 1 (see Section 1.1). We rated how many different strategies are introduced in each textbook as well as the quality of the instruction. We used a partial credit scoring for each strategy, separately for both operations (addition and subtraction), and added them up to a sum score for the respective textbook (Grade 2 or 3). For each strategy and operation we assigned

- 0 points, if the strategy was not contained;
- 1 point, if the strategy was contained in the textbook but not emphasized as regular learning content (e.g., not explained in the instruction part of a page but only presented as a hint on the margin of an exercise page);
- 2 points, if the strategy was explicitly introduced for an addition or subtraction problem (within an instruction part of a page);
- 3 points, if the strategy was explicitly introduced for an addition or subtraction problem and additionally illustrated in another representation (e.g., a solution is presented in a symbolic notation and in a graphical representation using a number line).

Regarding learning opportunities for the dimension *strategy distribution*, which is the knowledge about the relative frequency of use of a strategy, we considered the relevance the textbook ascribes to each strategy. Hence, in addition to the variety of strategies introduced in

a book (see strategy repertoire), we analyzed whether a textbook presented a strategy for more than only one example problem. This leads to the following partial credit scoring:

- 0 points, if the strategy is not mentioned at all or introduced only once (the latter case is already scored in the preceding category strategy repertoire);
- 1 point, if the strategy is introduced and additionally presented by means of a second problem;
- 2 points, if the strategy is introduced and additionally presented more than once by means of different problems.

Again, the scores of the strategies are added up to an overall score for the respective textbook (Grade 2 or 3).

Strategy efficiency is the ability to perform strategies fluently and accurately. This evolves and enhances by growing experience and practice. Therefore, our criterion for a textbook is the number of multi-digit addition and subtraction problems and tasks, to foster this ability. In contrast to the other three categories, these numbers were not scored but simply counted and directly used as indicators for the textbooks.

Concerning the last dimension, *strategy selection*, which is the ability to flexibly select a strategy for a given problem, we know from former research, that an explicit comparison of solution strategies is conducive for the development of adaptive expertise (cf. Section 1.1). Thus, we assessed whether the textbooks systematically or occasionally provide exercises which explicitly ask for a strategy comparison. Systematically in this context means that this type of exercise is recurring within the textbook as part of an instructional approach when introducing new strategies: After or with the introduction of new strategies, new strategies are contrasted to known strategies in a strategy comparison exercise. Occasionally means that such comparison exercises occur only sporadically, that is, the textbooks do contain a few of them, but these exercises are not integrated as recurring elements into an instructional approach for the introduction of new strategies. The scoring was partial credit: 2 points for textbooks with such exercises as systematic component, 1 point for textbooks with occasionally provided exercises for strategy comparisons, and 0 points for textbooks without. Since the way of occurrence of comparison exercises (systematically as part of the instructional approach, occasionally, or not at all) is analog for addition and subtraction in each of the four textbook series, we did not distinguish for the two operations here.

Overall, we got four scores for each textbook as a quality measure for opportunities to learn with respect to adaptive expertise.

2.2.2. Students' adaptive expertise in arithmetic

A measure for the adaptivity of students' actual strategy use was created from the data of four addition and subtraction tasks from the Grade 3 arithmetic test. This test was administered at

the end of the school year and contained different types of arithmetic items. There were four items ($482 + 218$, $473 + 398$, $381 - 99$, $702 - 698$) suggesting the use of specific strategies as efficient solutions. Following a task-based perspective an efficient solution means a correct solution in combination with a strategy that minimizes both cognitive load and the number of solution steps and thus is quick and least error-prone. As described below, for each of these items it was determined normatively and task-specifically, which strategies are more efficient than others. The small number of items results from our using existing data of a study which addressed general arithmetic competences and not specifically adaptive expertise. There were some more addition and subtraction problems in the test but for these items a clear distinction between efficient and less efficient strategies was not possible (e.g., $172 + 424$, $824 - 247$).

Regarding the scoring, in a first step the calculations and notes of the students' were coded by trained research assistants. We distinguished digit-based written algorithms and number-based strategies with an explicit presentation of the solution. The latter were specified corresponding to the classification Table 1 presented in Section 1.1. Besides the above mentioned shortcut versions of split and jump, where one or two solution steps of the according strategy were merged, we empirically found another strategy students applied for the item $702 - 698$, which we denoted as "Indirect Subtraction" (see Table 3). In addition, we coded student solutions where only the result was presented (i.e. without any computation steps) as purely mental computation. Table 2 shows the list of codes. About 22% of the solutions were double-coded, providing a sufficient interrater agreement for each of the four items (Cohen's κ : .83–.90). Not distinctly assignable solutions were coded as 28 ("Others"). This category primarily comprises different mixed versions of jump and split strategies (an example is displayed in Appendix D) as well as single inconclusive approaches, which both appeared only sporadically (there are 43 solutions in the category "Others", i.e., 0.77% of all solutions; 24 of the 43 solutions lead to incorrect results).

Code	Strategy
0	Mental computation (no computation-steps are written down)
1	Written algorithm
21	Split
219	Split shortcut
22	Jump
229	Jump shortcut
23	Jump alternative
24	Compensation
25	Simplifying
26	Indirect Addition
27	Indirect Subtraction
28	Others

Table 2: List of strategy codes

Efficiency Score = 2			
Compensation	Simplifying	Indirect Addition	Mental Computation
$702 - 700 = 2$ $2 + 2 = 4^*$	$704 - 700 = 4$	$698 + 4 = 702$	No written strategy
Efficiency Score = 1			
Split shortcut	Jump shortcut	Indirect Subtraction	
$700 - 600 = 100$ $100 - 98 = 2$ $2 + 2 = 4^*$	$702 - 600 = 102$ $102 - 98 = 4^*$	$702 - 2 = 700^*$ $700 - 2 = 698$ $2 + 2 = 4^*$	
Efficiency Score = 0			
Jump	Split	Written Algorithm	
$702 - 600 = 102$ $102 - 90 = 12$ $12 - 8 = 4$	$700 - 600 = 100$ $100 - 90 = 10$ $10 - 8 = 2$ $2 + 2 = 4$	$\begin{array}{r} 702 \\ - 698 \\ \hline 11 \\ 4 \end{array}$	

* The presentation of these solution steps varied across students.

Table 3: Efficiency Coding of the task 702–698

In a second step, each solution was scored with respect to an adaptive use of strategies. For each item we specifically determined efficient, partly efficient and inefficient solutions from a normative, task-based perspective so that the strategy classification differs for the four items. As relevant characteristics we took into account the correctness of the solution, the number of solution steps, and the demand on cognitive load induced by carry-overs in case of digit-based algorithms and by crossing the tens or hundreds boundary in case of number-based strategies. First all incorrect solutions were classified as inefficient. Strategies used to obtain correct solutions were classified either as efficient or as partly efficient when they reduce the problems to easier ones which are demanding less cognitive load. These are strategies which either needed only one or two far easier solutions steps that do not cross the tens or hundreds boundary (Simplifying, Compensation for all items, Jump shortcut for $482 + 218$, and Jump alternative for $473 + 398$), or only a single solution step with crossing the hundreds or tens boundary (Indirect Addition for $702-698$). Strategies demanding either three solution steps without crossing the hundreds or tens boundary or two solution steps and exactly one crossing of the tens or hundreds boundary were classified as partly efficient. Strategies which led to more complex calculations than the previously described solutions were classified as inefficient. These were in particular, the written algorithms as well as the jump and split strategy for all four items, since they

represent considerably longer and more cumbersome solutions with a higher demand on cognitive load for some of the solution steps. All four items include specific characteristics which are considered as error-prone when using the digit-based written algorithm, like carry-overs to a 0 (381–99), carry-overs to a 9 (473 + 398, 702–99) subtracting from 0 (702–698), adding up to 10/0 as a column value (482 + 218), or two carry-overs in a row (all items, see Table 3 and Appendix A Efficiency coding of the task 482 + 218, Appendix B Efficiency coding of the task 473 + 398, Appendix C Efficiency coding of the task 381 – 99). Also the number-based jump and split strategies, which are decomposing one or both numbers and calculating all the way down with each decomposed element without avoiding crossing the tens or hundreds boundary, were classified as inefficient for all four items. Moreover, for the problem 473 + 398 the shortcut version of the split strategy was classified as inefficient because it requires three solution steps as well as crossing the tens and the hundreds boundary (see Appendix B) and for 381–99 the indirect addition strategy was considered as inefficient due to four solution steps (Appendix C).

Finally, all correct solutions from purely mental calculations were classified as efficient. This was motivated firstly by the assumption that mental calculation strategies are mentally conducted number-based strategies (cf. Section 1.1, Kilpatrick et al., 2001; Padberg, 2005). Secondly, efficient number-based strategies have a reduced cognitive load in comparison to partial efficient and inefficient strategies. Therefore we assume that third-graders are able to conduct efficient strategies on three-digit addition and subtraction mentally, correctly and without any jottings. Solutions of the category “Others” were categorized as inefficient, since they all constitute cumbersome approaches or lead to incorrect results. An example is given in Appendix D. The coding scheme is exemplified for the task 702–698 by student solutions in Table 3 (the coding schemes for the other three items are given in Appendix A Efficiency coding of the task 482 + 218, Appendix B Efficiency coding of the task 473 + 398, Appendix C Efficiency coding of the task 381 – 99). Students' solution examples for each scoring are given in Fig. 1.

Following these guidelines, we conducted a partial-credit scoring on each item assigning 0 points for inefficient strategies, 1 point for partly efficient strategies, and 2 points for efficient strategies. The resulting scale from 0 to 8 points showed an acceptable reliability (Cronbach's $\alpha = .71$) and was used as a measure for students' adaptive expertise in arithmetic.

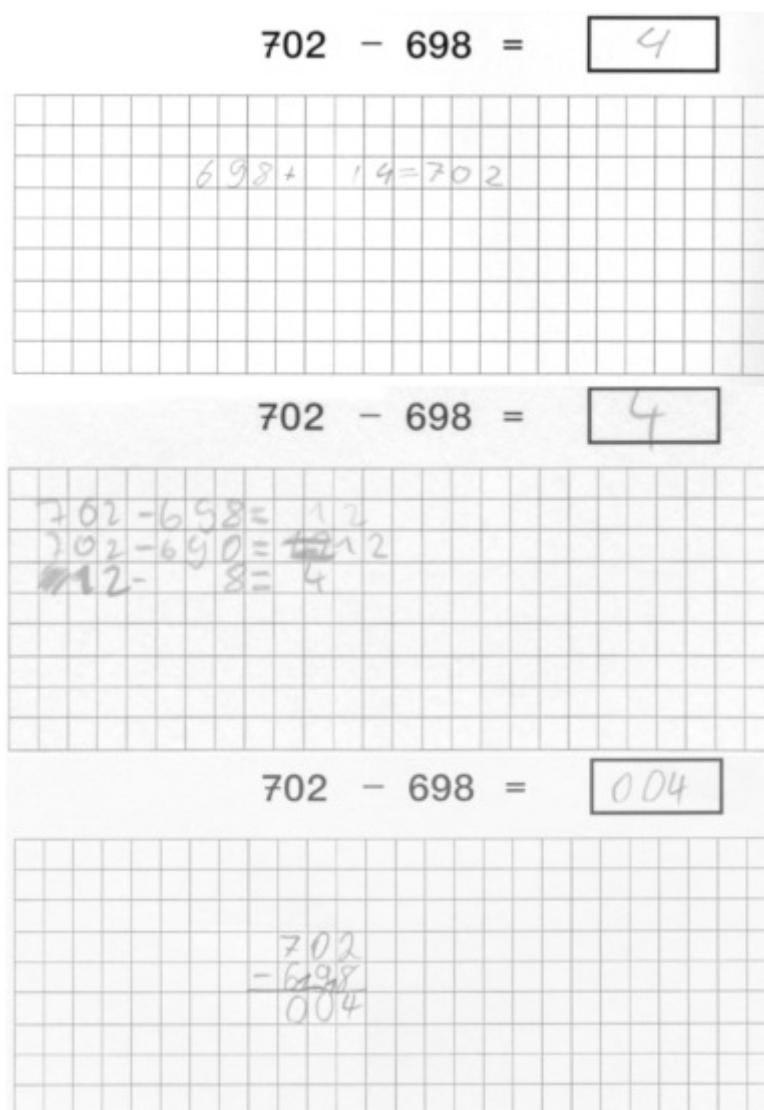


Figure 1: Examples for a 2-Point (Indirect Addition), a 1-Point (Jump shortcut), and a 0-Point (Written algorithm) efficiency score.

2.2.3. Individual and classroom covariates

To examine the unique effect of textbook quality on students' adaptive expertise, we included several individual and classroom-related covariates in our analysis. In order to control for individual differences between students, data for the basic cognitive abilities were collected upon school entrance in Grade 1. An approved standardized test with a sufficient reliability was administered (CFT 1-R, Weiß & Osterland, 2013, Cronbach's $\alpha = .94$). Furthermore we used data collected by an arithmetic test which covered the content of the Grade 1 arithmetic curriculum (i.e. basic arithmetic with numbers 0–20) and which was adapted from Lipowsky, Faust, and Karst (2011) to control for students' arithmetic knowledge at the end of Grade 1. We decided to include this variable because in Germany the teaching of multi-digit addition and subtraction strategies starts in Grade 2 after the extension of the number space to 0–100 and because arithmetic knowledge is considered a predictor for adaptive expertise in arithmetic (see Section 1.1). The students' arithmetic scores were estimated based on Item Response Theory and the unidimensional scale showed a good reliability (WLE Person separation reliability: .94).

We included the participation of the classes in one of the three groups of the arithmetic support program (see Section 2.1) as dummy variable to control classroom characteristics. Moreover, we controlled for class composition by including the aggregated variables for the basic cognitive abilities at school entrance and the prior arithmetic knowledge at the end of Grade 1. As primary school teachers in Germany are generalists, that is, all teachers teach mathematics but only some teachers have a formal qualification to teach mathematics, there is a need to control for the teacher qualification (studied mathematics or not). We combined data about teacher qualification with information about the time when the teachers went to university. As mentioned in the introduction, students' adaptive expertise in arithmetic as a learning goal became of particular interest in mathematics education in the late 1990's. It was included in the five-year mathematics teacher education programs in Germany since the beginning of the 2000's. It is reasonable to assume that teachers who finished their mathematics teacher education in the year 2005 or later have been taught about and sensitized to this topic, which might possibly affect their teaching. Hence, we constructed a teacher qualification scale for the teachers in Grade 2 and Grade 3, ranging from 0 to 2 points, whereby 1 point is given for a formal qualification in mathematics and 1 additional point if the teacher completed the teacher education program in mathematics after 2005. Finally, we used data on Grade-2-teachers' beliefs to control for the extent of constructivist orientation. These data were collected using an approved questionnaire developed by Staub and Stern (2002). Empirical research has repeatedly shown relations between teacher's beliefs and the development of their students' mathematics achievement (e.g., Kunter et al., 2013; Staub & Stern, 2002). Furthermore, the adaptive use of strategies depends on students' personal preferences and tendencies (see Section 1.1), meaning students could have an advantage in developing adaptive expertise if their teachers have rather constructivist beliefs.

2.3. Procedure

2.3.1. Textbook analysis

Three independent and trained persons rated both grade levels of the four textbook series on the basis of the three categories (1) strategy repertoire, (2) strategy distribution and (4) strategy selection following the criteria presented in Section 2.2.1. The number of tasks (category (3): strategy efficiency) was determined by only one of them, since this was not a rating procedure. The raters' results showed a solid agreement (Fleiss' κ : .61 (1), .82 (2), 1.00 (4); ICC: .93 (1), .98 (2), 1.00 (4)). Since different strategies for addition and subtraction are introduced in Grade 2 and Grade 3, we analyzed the textbooks for each grade. A uniform scoring was derived by consensus method. A ranking was set up in each category for the Grade 2 and the Grade 3 textbooks. For each grade we compiled a relative overall ranking scale of the books' quality for both years by computing the mean value of the four category ranks for each textbook so that each category was weighted equally.

2.3.2. Longitudinal study

The longitudinal study started in the school year 2013/2014. The basic cognitive abilities were assessed upon starting school. The arithmetic tests were administered at the end of the first and the third school year (testing time 30 min and 40 min, respectively). All data were collected by trained test administrators in accordance with the respective test manuals. All instructions were read out aloud and for each new item type the administrator worked an example with the children. For the Grade 3 items we used to measure adaptive expertise, the students were asked to solve the problems in a way which they find fastest and most accurate. They were instructed to decide on their own whether to use the space given below the tasks for written calculation or jottings, or calculate mentally and just write down the result.

At the end of Grade 1 teachers were given a questionnaire to collect data about teacher qualification, teacher experience, teacher beliefs and textbook use. After Grade 3 the school principals were asked about the textbooks the classes used in Grade 3 as well as the teachers' formal qualification and experience.

2.4. Data analyses

Differences in textbook quality with respect to adaptive expertise (Research Hypothesis 1) were analyzed for the four categories strategy repertoire, strategy distribution, strategy efficiency, and strategy selection by comparing and interpreting the scores for the Grade 2 and Grade 3 volumes of the textbook series.

To analyze textbook effects (Research Hypotheses 2 and 3) we conducted multilevel random intercept models with the software Mplus 7.0 (Muthén & Muthén, 1998–2012). Using multilevel modeling the nested levels (students nested within classes) are treated as meaningful. Students in the same classes are likely to be more similar to each other than with students in other classes and residuals are likely to be correlated. Multilevel modeling allows for calculating correct standard errors (Hox, 2010). Significance testing was undertaken at the 5% level. A step by step procedure was used. The Model 0 without predictors (null model) was run to estimate the partition of variance between and within classes. In Model 1 we included individual characteristics (at individual level) and information about the class composition (at class level). The basic cognitive abilities at school entrance and the arithmetic knowledge at the end of Grade 1 were included on individual level to account for individual differences between children before multi-digit addition and subtraction strategies were taught. In addition, we controlled for the class composition which in previous research turned out to be influential for the learning conditions and thus students' achievement development (e.g., Dumont, Neumann, Maaz, & Trautwein, 2013). Therefore, the values of the individual characteristics were aggregated to class means and included on class level in the multilevel models.

In Model 2 of our analysis the support programs (see Section 2.1), the teacher qualification scale and the teachers' beliefs about constructivism were included, to control for possible effects concerning instructional quality. In Model 3 the quality of the Grade 2 and 3 textbooks with respect to adaptive expertise were included. Based on this model we can analyze the effect of textbook quality on the third-graders adaptive expertise in arithmetic (Research Hypothesis 2). Finally, in Model 4, we additionally examined cross level interaction effects between individual prior arithmetic knowledge at the end of Grade 1 and the textbook quality scales with respect to adaptive expertise (Research Hypothesis 3). Model 4 allows the analysis whether the textbook quality moderates the influence of students' prior arithmetic knowledge on the acquisition of adaptive expertise. It is plausible, that students need arithmetic knowledge to develop an adaptive use of strategies in addition and subtraction.

Scores for the learning prerequisites as well as the mathematics achievement scores were standardized so that the corresponding β -coefficients can be interpreted as effect sizes similar to Cohen's d (cf. Tymms, 2004). For missing data on the independent variables we applied a full-information-maximum-likelihood (FIML) approach in Mplus 7.0. FIML combines missing data and parameter estimation in a single step and uses all the available information and therefore is very useful (Enders, 2010).⁶ Due to sample selection there was no missing data for textbooks of Grade 2 and support program.

3. Results

3.1. Textbook quality of opportunities to learn regarding adaptive expertise

Table 4 shows the results of the consensus rating regarding textbook quality. The single rater results are given in Appendix E. Due to different scoring procedures, the magnitudes of the scores differ between the categories (i.e., the scores of the different categories are not comparable). Higher scores in the category strategy repertoire represent a larger scope of strategies and/or a higher quality of their introduction. For example, the textbooks "Welt der Zahl" in Grade 2 and 3 include a greater number of strategies, which are mostly more valuable in terms of quality of their instruction, than the textbooks "Einstern". In the category strategy distribution, textbooks with higher scores are characterized by a higher number of strategies which are emphasized through a more frequent appearance of examples within the book. Again, the textbooks "Welt der Zahl" emphasize more strategies in Grade 2 and 3 than "Einstern". The scores for the category strategy selection show whether a textbook systematically or occasionally contains exercises for strategy comparison. While the textbooks for Grade 2 and Grade 3 of the series "Welt der Zahl" and "Denken und Rechnen" systematically contain such exercises, only the Grade 3 textbook of "Flex und Flo" does, and none of the "Einstern" textbooks include such exercises at all. Overall, the scores indicate that the textbook series "Welt

⁶ We controlled for differences to Multiple Imputation using the R-package "mitml". We did not find substantial differences in the estimated coefficients. There are no differences in the significance levels.

der Zahl” shows the highest quality whereas the series “Einstern” shows substantial lower scores in most categories. This means that teachers, who narrowly follow the textbook series “Einstern” in Grade 2 and 3 will provide fewer opportunities to learn strategy acquisition and the adaptive use of strategies. Regarding Research Hypothesis 1 we have found substantial quality differences between the textbook series with respect to adaptive expertise.

In each of the categories we ranked the textbooks in accordance with their score to weight the different categories equally and to get a relative assessment in each of the categories (see the lower section of Table 4). Subsequently, we compiled a relative overall ranking of the textbooks for Grade 2 and Grade 3, representing the relative quality of the textbooks' opportunities to learn regarding adaptive expertise. All in all “Welt der Zahl” (in Grade 2 and 3) has the highest ranks on the textbook quality scale. “Einstern” has the lowest ranking in both grades. For a better interpretation of the effects, we reversed the ranking in the following multilevel analysis, so that positive regression coefficients represent positive effects of textbook quality on students' adaptive expertise.

	Strategy Repertoire		Strategy Distribution		Strategy Efficiency		Strategy Selection		Mean Rank	
	2	3	2	3	2	3	2	3	2	3
Grade	2	3	2	3	2	3	2	3	2	3
Score										
Denken und Rechnen	21	15	8	7	1146	815	2	2		
Einstern	17	10	7	4	548	727	0	0		
Flex und Flo	22	24	6	8	645	1371	0	2		
Welt der Zahl	22	23	8	10	1392	1769	2	2		
Rank										
Denken und Rechnen	2	3	1	3	2	3	1	1	1.50	2.50
Einstern	3	4	2	4	4	4	2	2	2.75	3.50
Flex und Flo	1	1	3	2	3	2	2	1	2.25	1.50
Welt der Zahl	1	2	1	1	1	1	1	1	1.00	1.25

Table 4: Results of the textbook analysis: absolute scores and ranking

3.2. Effects of the textbooks' quality on students' adaptive expertise

Table 5 provides the means and standard deviations of students' basic cognitive abilities, prior arithmetic knowledge at the end of Grade 1 and adaptive expertise at the end of Grade 3 for

the textbook groups. The deviations in basic cognitive abilities confirm former results of a substantial heterogeneity in the learning conditions at school entrance between as well as within classes in Germany (e.g. Padberg, 2005). Differences in prior arithmetic knowledge, tested at the end of Grade 1, are related to those of basic cognitive abilities and are additionally influenced by the first year of schooling. To control for effects concerning these predictors, we included them in the multilevel analysis on both levels.

We tested whether the textbooks are distributed differently between the teacher qualification groups or between intervention and control group of the support program. The distribution did neither differ significantly between the teacher qualification groups (Grade 2: $H(2) = 0.966$, $p = .62$; Grade 3: $H(2) = 4.307$, $p = .12$) nor between participant and non-participants of the support program (Grade 2: $H(1) = 2.568$, $p = .11$; Grade 3: $H(1) = 2.672$, $p = .10$).⁷

	Basic cognitive abilities (0–45)	Prior arithmetic knowledge (WLE)	Adaptive expertise (0–8)
Grade 2			
Denken und Rechnen	27.60 (7.66)	0.19 (1.70)	2.00 (2.56)
Einstern	26.33 (8.10)	-0.04 (1.66)	1.03 (1.72)
Flex und Flo	28.27 (8.48)	0.37 (1.46)	1.78 (2.29)
Welt der Zahl	29.38 (8.48)	0.66 (1.57)	2.33 (2.66)
Grade 3			
Denken und Rechnen	26.86 (8.17)	0.09 (1.68)	1.84 (2.47)
Einstern	26.77 (7.89)	-0.34 (1.59)	1.16 (1.92)
Flex und Flo	28.45 (8.36)	0.33 (1.44)	1.81 (2.25)
Welt der Zahl	29.00 (8.24)	0.67 (1.53)	1.95 (2.50)

Table 5: Means and standard deviations of individual characteristics for the different textbook groups

For the analysis of the textbook effects, we first computed the null model to determine the share of between-class variance in students' adaptive expertise. It turned out that 11.1% of the variance lay between classes indicating that the adaptive expertise of the third-graders differs depending on the class they attended.

Table 6 shows the results of the Models 1–4 (as described in Section 2.4).

⁷ The Kruskal-Wallis test was used because the expected frequencies in 33–50% of the categories were smaller than five.

	Model 1	Model 2	Model 3	Model 4
Level 1 (students)				
Basic cognitive abilities	.02 (.03)	.02 (.03)	.02 (.03)	.02 (.04)
Prior arithmetic knowledge	.32** (.04)	.32** (.04)	.32** (.04)	-.07 (.14)
Level 2 (class)				
Basic cognitive abilities (aggregated)	.09** (.02)	.08 (.06)	.06 (.04)	.06 (.04)
Prior arithmetic knowledge (aggregated)	.02 (.04)	.04 (.05)	.02 (.05)	.02 (.05)
Support program Group 1		.02 (.12)	.09 (.12)	.10 (.11)
Support program Group 2		-.10 (.13)	.04 (.13)	.02 (.13)
Teacher qualification (Grade 2)		-.13 (.07)	-.06 (.07)	-.06 (.07)
Teacher qualification (Grade 3)		.00 (.09)	.00 (.08)	.01 (.08)
Teacher beliefs (Grade 2 ^a)		-.01 (.05)	-.02 (.05)	-.01 (.05)
Textbook quality (Grade 2)			.19** (.07)	.18** (.07)
Textbook quality (Grade 3)			-.07 (.07)	-.05 (.07)
Cross-level interaction				
Prior arithmetic knowledge x Textbook quality (Grade 2)				.17** (.06)
Prior arithmetic knowledge x Textbook quality (Grade 3)				.00 (.05)
<i>Intercept</i>	<i>-0.03 (0.04)</i>	<i>0.08 (0.14)</i>	<i>-0.30 (0.21)</i>	<i>-0.33 (0.21)</i>
Explained within class variance	11.6 %	11.4 %	11.4 %	12.2 %
Explained between class variance	10.1 %	18.5 %	31.5 %	28.9 %

* $p < .05$, ** $p < .01$, ^a in the dataset there were no data on teacher beliefs for Grade 3 teachers.

Table 6: Results of the multilevel analysis of students' adaptive expertise score (standard errors in brackets, all variables are standardized except Support program Group 1, 2 (0,1), Teacher qualification Grade 2, 3 (0,1,2) and Textbook quality Grade 2, 3 (rank))

First of all it should be noted that there is no effect of the basic cognitive abilities on the individual level in any of the models. On Level 2 there is a positive class composition effect in Model 1, which is no more significant and decreases further in the remaining models. In contrast, the prior arithmetic knowledge has a strong significant effect on the individual level (Models 1–3) or as cross-level interaction (Model 4), but no significant composition effect on class level. Overall, these effects are plausible since (i) a sound individual arithmetic knowledge base is indispensable for adaptive expertise, (ii) the school entrance basic cognitive abilities do no longer have an additional influence on the individual level after three years of schooling, and (iii) the cognitive potential of a class (indicated by the class mean of cognitive abilities) seems to have at least a slight impact on the learning opportunities for challenging learning content (like adaptive expertise) chosen by the teachers.

Regarding the covariates for the learning environment and teacher characteristics, we did not find any effects. The mathematics support programs in Grade 1 and 2 did not affect students' adaptive expertise at the end of Grade 3 which seems surprising at first glance (though it is advantageous for our analysis). However, these support programs specifically addressed very weak students in Grade 1 and 2 with problems in the basic understanding of numbers and operations, and did not address advanced aspects like different arithmetic strategies. Since we know from other studies that prior knowledge is highly relevant for adaptive expertise (e.g., Rittle-Johnson et al., 2009), the non-existing effect on students' adaptive expertise at the end of Grade 3 is less surprising. In addition, the intervention ended in Grade 2. We did expect significant effects for the teacher variables but neither the teacher qualification of Grade 2 or 3 nor the teacher beliefs of Grade 2 had an effect on the students' adaptive expertise. Nevertheless, the inclusion of the learning environment variables into Model 2 led to an increase of the explained between class variance by $\Delta R^2 = 8.4$ percentage points.

A substantial effect on students' adaptive expertise on class level was caused by the textbook quality scale for Grade 2 textbooks ($B = 0.19$ in Model 3). Thus, a shift of 1 point on the textbook quality scale leads to an increase of 0.19 standard deviations of the average adaptive expertise score in a class. In other words: The change from textbook with the lowest (1.25) to the textbook with the highest (3.00) quality score leads to an increase of 0.34 standard deviations of the average adaptive expertise score in a class, which we consider as substantial. The textbook quality in Grade 3 had no additional effect on the adaptive expertise of students at the end of Grade 3. This result is in line with and underlines the importance of an early emphasis of a flexible use of strategies (Blöte et al., 2001; Klein et al., 1998). The explained between class variance increased by $\Delta R^2 = 13.0$ percentage points, when including the textbook quality. With regard to Research Hypothesis 2 we have found a significant effect of textbook quality on adaptive expertise.

Moreover, we found a significant cross-level interaction effect between the individual prior arithmetic knowledge at the end of Grade 1 and the quality of the textbook used in Grade 2 (Model 4). Similar to the textbook main effect, there is no interaction between the prior knowledge and the textbook quality of Grade 3. When including the interaction effects, the regression coefficient of the textbook quality in Grade 2 on class level slightly decreased ($B = 0.18$), but still remained highly significant, whereas the regression coefficient of the prior arithmetic knowledge on individual level was no longer significant. Hence, the effect of the individual prior arithmetic knowledge on students' adaptive expertise in Model 3 is moderated by the textbook quality of Grade 2 and we can confirm Research Hypothesis 3. Even for high achieving students it is not irrelevant which specific learning opportunities the textbook of Grade 2 provides with respect to an adaptive strategy use.

4. Discussion

This study addresses the role of textbooks as learning resources for students' development of adaptive expertise in the arithmetic classroom. A theory-based analysis of the opportunities to

learn with respect to adaptive expertise in common mathematics textbooks allowed the definition of textbook quality for this specific topic. The application of this quality indicator to four mathematics textbooks yielded differences in the textbooks' quality. Moreover, data of a three-year longitudinal study with 82 classes revealed substantial effects of the textbook quality on students' adaptive expertise after three years of schooling. A cross-level interaction indicated that textbook quality can be considered as a moderator for the influence of prior arithmetic knowledge on students' adaptive expertise in arithmetic.

4.1. Impact for educational research

Our study contributes to research on students' acquisition of adaptive expertise as well as to research on textbooks effects. Concerning students' acquisition of adaptive expertise we mentioned the frequently replicated results that primary school students show a low adaptive expertise in multi-digit addition and subtraction (e.g., Csíkos, 2016; Heinze et al., 2009; Torbeyns et al., 2017; Torbeyns & Verschaffel, 2016) in Section 1.1. Based on the findings of textbook research (see Section 2.2) supported by the teacher questionnaire data of our study, we can assume that the arithmetic problems in the mathematics lessons mainly stem from the mathematics textbooks. The analysis of the four textbooks in our study revealed that the textbooks differ in the number of multi-digit addition and subtraction problems they offered to the teachers and their students. In accordance with that, they differed in the explicitly presented opportunities to learn for different addition and subtraction strategies as well as their adaptive use. As elaborated in Section 3.2, these specific textbook characteristics, which we compiled to a textbook quality score with respect to adaptive expertise, had a substantial effect on students' adaptive expertise. The empirical evidence for this effect is substantiated by the three-year longitudinal design, the large sample size and the inclusion of relevant covariates on the individual and the class level. The fact that an explicit presentation of strategies and opportunities for strategy comparison is an effective instructional approach was shown in controlled experimental studies (e.g., Klein et al., 1998; Rittle-Johnson and Star, 2007, Rittle-Johnson and Star, 2009; Torbeyns et al., 2009a, Torbeyns et al., 2009b). Moreover, empirical findings of De Smedt, Torbeyns, Strassens, Ghesquière, and Verschaffel (2010) and Heinze et al. (2018) suggest that important strategies – like the indirect addition or simplifying strategy – are rarely invented by primary school children without explicit presentation. Our results based on data from the regular mathematics classroom support findings that – on average – students benefit from the explication of strategies and inducing strategy comparison. In line with former research on textbook effects, our study suggests an impact of textbooks on students' strategy use (Fagginger Auer et al., 2016; Heinze et al., 2009). Based on previous findings (see section 1.2), we assume that the textbooks affect the teachers and the learning opportunities for adaptive expertise they provide in classroom. The cross-level interaction (see Model 4 in Table 6 in Section 3.2) supports the hypothesis that the impact of explicit opportunities to learn is strengthened when students possess high prior knowledge, which is in line with the “Matthew Effect” (Bodovski & Farkas, 2007). Hence, the well-known influence of students' prior arithmetic knowledge on the development of their adaptive expertise in arithmetic (e.g., Fuson et al., 1997; Rittle-Johnson

et al., 2009; Schulz, 2018; Threlfall, 2009) is moderated by explicit opportunities to learn which in our study are induced by the textbook. It seems that even for students with high prior knowledge the development of adaptive expertise depends on the variety of strategies and related tasks presented in the learning resources. This indicates a difference in the students' use of the learning opportunities for adaptive expertise presented in textbooks, depending on their mathematical achievement. It is of interest for further educational research whether this differentiation can be shown for other fields of mathematics.

Our study also contributes to previous research on mathematics textbook effects on students' achievement. In line with the findings of Schmidt et al. (2001) and Törnroos (2005) for secondary students, we found evidence that the choice of a certain textbook has an effect on students' learning outcomes. In our study, we were able to develop specific criteria for textbook quality with respect to the acquisition of adaptive expertise. These quality criteria correspond to the empirical findings on students' development in adaptive expertise and, thus, provide an explanation for the observed textbook effects. Interestingly, teacher qualification and beliefs did not play a role in students' adaptive expertise. Our results suggest that textbook quality with respect to adaptive expertise should be taken into account as an important variable when examining students' development of adaptive expertise in the mathematics classroom context. Though the teacher is the responsible person for preparing and teaching the mathematics lesson, our findings suggest that the content and the instructional approaches presented in the textbooks have a substantial influence on the learning environments teachers implement in their lessons which cannot be ignored. The cross-level effect suggests that the influence of students' individual differences on the individual learning progress is moderated by the textbooks' quality. This holds true even in the case of textbooks and teachers aiming for the same curriculum.

The results show that a fine-grained and theory-based content analysis of mathematics textbooks is a fruitful approach to investigate the impact and effects of these learning resources. Hence, our study contributes to textbook research as required by Fan (2013). Though a global measurement and comparison of textbook quality might be a complex and far off endeavor, our investigation reveals that for the subdomain of adaptive expertise in arithmetic, it is possible to identify relations between opportunities to learn presented by textbooks and students' learning achievements. This might be a good starting point for further research into the quality and impact of mathematics textbooks by expanding this method to other mathematical domains and identifying strengths and weaknesses using this method.

4.2. Impact for educational practice

While the four textbook series all address the same curriculum, the textbook analysis unveiled crucial differences in the quality of their learning opportunities with respect to adaptive expertise. Hence, we generated evidence, that the learning process is not only a matter of the curricula used but also a matter of the concrete realization of a curriculum by the textbooks. From a political level textbooks can be considered an instrument of quality assurance in education.

Thus, in order to provide students with high quality textbooks and to foster their learning processes, this study provides evidence for an introduction or a reinforcement of a regulation for textbook approval by means of clear quality criteria. These criteria should be theory-based and empirically proven.

With a view to teachers, who use the textbooks and – according to Mullis et al. (2012) and our own findings (cf. Section 2.1) – follow their concepts in general, they should be aware of textbook quality discrepancies. In particular, teachers should be sufficiently strengthened in assessing the quality of the textbook they use. Though we do not know much about the cognitive processes when teachers use the textbooks for lesson preparation and implement their content, we have seen a strong effect of textbooks' quality. Thus, awareness and critical evaluation of textbook quality might be enriching issues in teacher education and training.

4.3. Limitations

As with any study, our results have limitations. Since we did a secondary analysis with existing data, we did not have an experimental setting to test textbook effects. Hence, due to the design we cannot strictly conclude causality for the effect of textbook quality on students' learning. In particular, we were not able to control why these textbooks were used in the considered classes. Specific teacher characteristics may be interacting with textbook features. In the federal state our data stemmed from the schools decide which textbook they use. Hence, we think that the influence of a single teacher on textbook choice is low, though there were some schools using different series. Consequently, when textbooks are chosen at school level, variables on school level could potentially be confounded with the textbook. So far, little is known about the criteria schools use to select textbooks and to what extent variables on school level are related to textbook choice. The selection process in the United States has been described at district level as highly idiosyncratic and market-driven (National Research Council, 2002). Hartung (2014) suspects for Germany that in the selection of textbooks, due to a lack of time, the schools make light-hearted decisions which are influenced by advertising of the publishers or a close relationship between the sales representatives of the publishers and the schools. But it is still unclear whether certain school characteristics (e.g. the social milieu of the school or the school culture) are related to certain criteria for selecting textbooks. Because we reanalyzed existing data it was not possible for us to collect additional information on school characteristics or textbook choice. Furthermore, the number of classes (level 2 units) per school is not sufficient for a possible 3-level modeling. Future studies should explore the relationship of school characteristic, textbook selection criteria and textbook choice to exclude the possibility of confounding school level variables. Moreover, we controlled for teacher qualification and beliefs (which did not have a significant effect). However, data on the teachers' professional knowledge was not available. From the Teacher Education and Development Study (TEDS, Blömeke, Kaiser, & Lehmann, 2010) we know that teachers' mathematics-specific knowledge is strongly related to their qualification (qualified as mathematics teachers vs. out-of-field teacher). Nevertheless, a more differentiated measure for teachers' professional knowledge would be desirable.

A second limitation is due to the test items we could use for measuring adaptive expertise. Although we got a reliable scale for adaptive expertise with respect to multi-digit addition and subtraction and although there were noteworthy results on the textbook quality effects, it must be mentioned that our test was only based on four items. Both limitations are due to the fact that we used existing data. For future studies in this field, experimental designs with more elaborated tests on students' adaptive expertise and control data with respect to teachers' professional knowledge would be beneficial.

A third limitation is that our data cannot exclude the alternative explanation that the increase of students' adaptive expertise in our study may simply be an effect of amount of computation practice. To test this hypothesis we need to compare the effects of the four categories of the Siegler model – in particular the effects of strategy repertoire, strategy distribution, and strategy selection with the effect of strategy efficiency – which leads to multicollinearity in the regression model. Nevertheless, following the Siegler model we consider experience or routine as a necessary but not sufficient prerequisite for the development of an adaptive use of strategies. Still, based on this model as well as on earlier studies, which we present in Section 1.1 (e.g. Blöte et al., 2001; Klein et al., 1998; Rittle-Johnson and Star, 2007, Rittle-Johnson and Star, 2009; Torbeyns et al., 2009a, Torbeyns et al., 2009b), making different strategies and strategy use a subject of discussion is conducive for students' adaptive expertise.

5. Conclusion

The presented study has shown a substantial effect of mathematics textbooks' quality on the students' adaptive expertise with respect to multi-digit addition and subtraction. The findings of a large longitudinal study with students learning in regular mathematics classroom support the results of previous controlled experimental studies that emphasizing different strategies and a strategy comparison are beneficial for the development of students' adaptive expertise. The findings particularly suggest that students with high prior knowledge benefit more from this kind of explicit opportunities to learn than students with low prior knowledge. In addition, the study contributes to textbook research by offering an idea of theory-driven quality criteria for textbooks and providing evidence for their relevance.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Efficiency Coding of the task $482 + 218$

Efficiency Score = 2				
Simplifying	Simplifying	Compensation	Jump shortcut	Mental Computation
$500 + 200 = 700$	$480 + 220 = 700$	$480 + 218 = 698$	$482 + 210 = 692$	No written strategy
		$698 + 2 = 700$	$692 + 8 = 700$	
Efficiency Score = 1	Efficiency Score = 0			
Split shortcut	Jump	Split	Written Algorithm	
$82 + 18 = 100$	$482 + 200 = 682$	$400 + 200 = 600$		482
$400 + 200 = 600$	$682 + 10 = 692$	$80 + 10 = 90$		$+ 218$
$600 + 100 = 700$	$692 + 8 = 700$	$2 + 8 = 10$		$\frac{11}{700}$
		$600 + 90 + 10 = 700$		

Appendix B. Efficiency Coding of the task 473 + 398

Efficiency Score = 2				
Simplifying	Compensation	Jump alternative	Mental Computation	
471 + 400 = 871	473 + 400 = 873 873 - 2 = 871	473 + 27 = 500 500 + 371 = 871	No written strategy	
Efficiency Score = 1	Efficiency Score = 0			
Jump shortcut	Split shortcut	Jump	Split	Written Algorithm
473 + 300 = 773 773 + 98 = 871*	470 + 390 = 860 3 + 8 = 11 860 + 11 = 871	473 + 300 = 773 773 + 90 = 863 863 + 8 = 871	400 + 300 = 700 70 + 90 = 160 3 + 8 = 11 700 + 160 + 11 = 871	$\begin{array}{r} 473 \\ + 398 \\ \hline 11 \\ \hline 871 \end{array}$

Appendix E. Single rater results of the textbook analysis

Rater	Strategy Repertoire			Strategy Distribution			Strategy Selection		
	1	2	3	1	2	3	1	2	3
Grade 2									
Denken und Rechnen	21	20	24	8	8	8	2	2	2
Einstern	24	16	18	7	7	7	0	0	0
Flex und Flo	22	17	24	6	5	5	0	0	0
Welt der Zahl	18	21	26	8	7	9	2	2	2
Grade 3									
Denken und Rechnen	15	14	16	7	7	7	2	2	2
Einstern	17	9	14	4	3	3	0	0	0
Flex und Flo	21	20	27	8	6	8	2	2	2
Welt der Zahl	20	21	27	10	10	11	2	2	2

References

- Baroody, A.J., & Dowker, A. (Eds.). (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Blömeke, S., Kaiser, G., Lehmann, R. (Eds.). (2010). *TEDS-M 2008 - Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich [TEDS-M 2008- international comparison of future primary school teachers' professional competence and opportunities to learn]*. Münster: Waxmann.
- Blöte, A.W., Klein, A.S., Beishuizen, M. (2000). Mental computation and conceptual understanding. *Learning and Instruction*, 10(3), 221–248.
- Blöte, A. W., van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving twodigit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, 93, 627–638.
- Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal*, 108(2), 115–130.
- Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1997). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 29(1), 3–20.
- Chen, Z., & Siegler, R. S. (2000). Across the great divide: Bridging the gap between understanding of Toddlers' and older Children's thinking. *Monographs of the Society for Research in Child Development*, 65 (2, serial no. 261).
- Csikos, C. (2016). Strategies and performance in elementary students' three-digit mental addition. *Educational Studies in Mathematics*, 91(1), 123–139.

- De Smedt, B., Torbeyns, J., Strassens, N., Ghesquiere, P., & Verschaffel, L. (2010). Frequency, efficiency and flexibility of indirect addition in two learning environments. *Learning and Instruction, 20*, 205–215.
- Dumont, H., Neumann, M., Maaz, K., & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen: Internationale und nationale Befunde [the composition of students as an influence factor for school performance: International and national findings]. *Psychologie in Erziehung und Unterricht, 60*, 163–183.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Publications.
- Fagginger Auer, M. F., Hickendorff, M., van Putten, C. M., Beguin, A. A., & Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data. Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education, 29*(2), 144–159. <https://doi.org/10.1080/08957347.2016.1138959>.
- Fan, L. (2013). Textbook research as scientific research. Towards a common ground on issues and methods of research on mathematics textbooks. *ZDM—The International Journal on Mathematics Education, 45*(5), 765–777. <https://doi.org/10.1007/s11858-013-0530-6>.
- Fan, L., Zhu, Y., & Miao, Z. (2013). Textbook research in mathematics education: Development status and directions. *Zentralblatt für Didaktik der Mathematik, 45*(5), 633–646. <https://doi.org/10.1007/s11858-013-0539-x>.
- Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., ... Fennema, E. (1997). Children's conceptual structures for multidigit numbers and methods of multidigit addition and subtraction. *Journal for Research in Mathematics Education, 28*(2), 130–162.
- Hartung, T. (2014). Schulbuchauswahl und Lernmittelfreiheit in den deutschen Bundesländern im Kontext von Schülerpartizipation [textbook choice and the free supply of educational aids in the German federal states in the context of student participation]. Eckert. Working Papers, 11, 1–15. Retrieved from <http://www.edumeres.net/> urn/urn:nbn:de:0220-2014-00346.
- Hatano, G. (1982). Cognitive consequences of practice in culture specific procedural skills. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 4*, 15–18.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. W. Stevenson, H. Azuma, & K. Hakuta (Eds.). A series of books in psychology. *Child development and education in Japan* (pp. 262–272). New York: W H Freeman/Times Books/Henry Holt & Co.
- Heinze, A., Arend, J., Gruessing, M., & Lipowsky, F. (2018). Instructional approaches to foster third graders' adaptive use of strategies: An experimental study on the effects of two learning environments on multi-digit addition and subtraction. *Instructional Science. Advance online publication*. <https://doi.org/10.1007/s11251-018-9457-1>.
- Heinze, A., Marschick, F., & Lipowsky, F. (2009). Addition and subtraction of three-digit numbers. Adaptive strategy use and the influence of instruction in German third grade. *ZDM—The International Journal on Mathematics Education, 41*(5), 591–604. <https://doi.org/10.1007/s11858-009-0205-5>.
- Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction, 14*(3), 251–284.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications (2nd ed.)*. Quantitative methodology series. New York: Routledge.

- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up*. Washington, D.C.: National Academies Press. <https://doi.org/10.17226/9822>.
- Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades. Realistic versus gradual program design. *Journal for Research in Mathematics Education*, 29(4), 443–464.
- Krammer, H. P. M. (1985). The textbook as classroom context variable. *Teaching and Teacher Education*, 1(4), 273–278. [https://doi.org/10.1016/0742-051X\(85\)90015-0](https://doi.org/10.1016/0742-051X(85)90015-0).
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers-results from the COACTIV project*. New York: Springer.
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change. Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124(1), 83–96.
- Lipowsky, F., Faust, G., & Karst, K. (Eds.). (2011). *Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschulkindern" (PERLE) – Teil 2. Materialien zur Bildungsforschung [Documentation of the test instrument of the project "personal and learning development of primary school students" (PERLE) – Part 2. Materials for educational research]*. Frankfurt am Main, Germany: GFPF.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, Mass: TIMSS & PIRLS International Study Center.
- Mullis, I. V. S., Martin, M. O., Goh, S., & Cotter, K. (2016). TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science. Retrieved from <http://timssandpirls.bc.edu/timss2015/encyclopedia/>
- Muthen, L. K., & Muthen, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthen & Muthen.
- National Council of Teachers of Mathematics (2003). Standards for school mathematics: number and operations. Retrieved from <http://standards.nctm.org/document/chapter3/numb.htm>
- National Research Council (2002). Investigating the influence of standards: A framework for research in mathematics, science, and technology education. In I. R. Weiss, M. S. Knapp, K. S. Hollweg, & G. Burrill (Eds.). *Committee on understanding the influence of standards in K-12 science, mathematics, and technology education, Center for Education, division of behavioral and social sciences and education*. Washington, DC: National Academy Press.
- Padberg, F. (2005). *Didaktik der Arithmetik [arithmetic education]*. Heidelberg: Spektrum Akademischer Verlag.
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561–573.
- 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), 529–543.
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836–852. <https://doi.org/10.1037/a0016026>.
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology*, 82, 436–455. <https://doi.org/10.1111/j.2044-8279.2011.02037.x>.

- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.
- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics. Vol. 1*. Dordrecht: Kluwer.
- Schulz, A. (2018). Relational reasoning about numbers and operations—Foundation for calculation strategy use in multi-digit multiplication and division. *Mathematical Thinking and Learning*, 20(2), 108–141. <https://doi.org/10.1080/10986065.2018.1442641>.
- Selter, C. (2001). Addition and subtraction of three-digit numbers. German elementary children's success. Methods and Strategies. *Educational Studies in Mathematics*, 47(2), 145–174.
- Selter, C. (2009). Creativity, flexibility, adaptivity, and strategy use in mathematics. *ZDM—The International Journal on Mathematics Education*, 41(5), 619–625. <https://doi.org/10.1007/s11858-009-0203-7>.
- Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication. Testing predictions of ASCM using the choice-no-choice method. *Journal of Experimental Psychology: General*, 126(1), 71–92.
- Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, 31, 280–300.
- Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for Students' achievement gains. Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94(2), 344–355.
- Threlfall, J. (2002). Flexible mental calculation. *Educational Studies in Mathematics*, 50, 29–47.
- Threlfall, J. (2009). Strategies and flexibility in mental calculation. *ZDM—The International Journal on Mathematics Education*, 41(5), 541–556.
- Torbeyns, J., De Smedt, B., Ghesquiere, P., & Verschaffel, L. (2009a). Jump or compensate? Strategy flexibility in the number domain up to 100. *ZDM—The International Journal on Mathematics Education*, 41(5), 581–590.
- Torbeyns, J., De Smedt, B., Ghesquiere, P., & Verschaffel, L. (2009b). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics*, 71(1), 1–18.
- Torbeyns, J., Hickendorff, M., & Verschaffel, L. (2017). The use of number-based versus digit-based strategies on multi-digit subtraction: 9–12-year-olds' strategy use profiles and task performance. *Learning and Individual Differences*, 58, 64–74.
- Torbeyns, J., & Verschaffel, L. (2016). Mental computation or standard algorithm? Children's strategy choices on multi-digit subtractions. *European Journal of Psychology of Education*, 31(2), 99–116. <https://doi.org/10.1007/s10212-015-0255-8>.
- Tornroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, 31(4), 315–327.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen, & K. Elliot (Eds.). *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London: National Foundation of Educational Research.
- Valverde, G., Bianchi, L. J., Wolfe, R., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht: Kluwer Academic Publishers.

- van den Heuvel-Panhuizen, M. (2001). Realistic mathematics education in the Netherlands. In J. Anghileri (Ed.), *Principles and practices in arithmetic teaching* (pp. 49–64). Buckingham: Open University Press.
- van Steenbrugge, H., Valcke, M., & Desoete, A. (2013). Teachers views of mathematics textbook series in Flanders. Does it (not) matter which mathematics textbook series schools choose? *Journal of Curriculum Studies*, 45(3), 322–353.
- Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. Lester (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 557–628) (2nd ed.). New York: MacMillan.
- Verschaffel, L., Luwel, K., Torbeyns, J., & van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24(3), 335–359.
- Weis, R. H., & Osterland, J. (2013). *CFT 1-R–Grundintelligenztest Skala 1 Revision*. Göttingen, Germany: Hogrefe.

B Artikel zu Studie 2

Dieser Artikel wurde am 26.08.2019 erstmals eingereicht beim Journal **Learning and Instruction** und nach Begutachtung sowie ‚invitation to resubmit after major revision‘ in der vorliegenden Fassung am 11.08.2020 wieder eingereicht. Er wurde dort in einer überarbeiteten Version zunächst online und anschließend im Februar 2021 in Volume 71, 101401, publiziert (DOI: 10.1016/j.learninstruc.2020.101401).

Are First Graders' Arithmetic Skills Related to the Quality of Mathematics Textbooks? A Study on Students' Use of Arithmetic Principles

Henning Sievert, Ann-Katrin van den Ham, Aiso Heinze

Abstract: A major goal of the first years of schooling is students' development from using counting strategies to using calculation strategies – or even recall – to solve addition or subtraction problems. It requires the perception and usage of arithmetic principles. Although the continued use of counting strategies is problematic for further learning progress, they are frequently used by many students beyond Grade 1. While learning resources possibly play a role in students' development in this field, our knowledge about their impact is limited. In this study we investigated the presented learning opportunities of four different textbooks regarding arithmetic principles in Grade 1, and subsequently the relation to first graders' strategy use by analyzing a one-year dataset of 1614 students from 86 classes. The analyses show differences in the textbooks' quality concerning arithmetic principles as well as a considerable connection to students' strategies use at the end of Grade 1.

Keywords: primary school; arithmetic principles; mathematics textbooks; textbook quality

1. Introduction

The development of efficient solution methods – namely the use of strategies or recall – for solving addition and subtraction problems is a key objective of initial mathematics instruction (Kilpatrick, Swafford, & Findell, 2001; Wright, 2009; Verschaffel, Greer, & DeCorte, 2007). Many children develop efficient counting strategies for addition and subtraction when dealing with numbers up to 20. However, counting strategies will become inefficient, error-prone and a barrier for the further learning process when the number range is expanded to 100 and subsequently to 1000. For the learning of basic arithmetic, Kilpatrick and colleagues (2001) emphasize that the knowledge about sums and differences in addition and subtraction is relational and does not require plain memorization. Children “do not move from knowing nothing about the sums and differences of numbers to having the basic number combinations memorized” (Kilpatrick et al., 2001, p. 182), but go through a prolonged progress, in which they gradually develop more advanced and abstract solution methods for addition and subtraction problems

by perceiving and utilizing the underlying arithmetic principles. Verschaffel and colleagues (2007) report well-documented evidence for the gradual development in children's mastery of single-digit computation, which is opposite to the drill-and-practice paradigm of number facts.

From the instruction perspective teaching and learning material is of crucial interest in this context. It frames the teaching activities in the mathematics classroom and thus the opportunities to learn the use of strategies. In primary schools, textbooks are the most important learning resource used for mathematics instruction (Mullis, Martin, Foy, & Arora, 2012). Hence, learning opportunities regarding calculation strategies presented in textbooks might affect the teacher's instruction, and hence – indirectly – also the student's learning and strategy use. Stein, Remillard and Smith (2007) conceptualize this mediation as enacted curriculum, which refers to what actually happens in the classroom. Accordingly, some studies indicate a relation between mathematics textbooks and students' strategy use for advanced arithmetic strategies (cf. Fagginger Auer, Hickendorff, van Putten, Béguin, & Heiser, 2016; Heinze, Marschick, & Lipowsky, 2009, Sievert, van den Ham, Niedermeyer, & Heinze, 2019). Earlier research also shows that mathematics textbooks differ in their mathematical content as well as in certain pedagogical, cultural and sociological aspects (e.g., related to gender, ethnicity, equity, Fan, Zhu, & Miao, 2013). To our knowledge no studies exist providing evidence regarding textbook effects on first grade students' strategy use. Since the first year of schooling plays a significant role for students' further arithmetic development, especially for strategy change, a relation of students' learning to the textbook used can be considered an important factor for improving arithmetic education.

In his review on textbook research, Fan (2013) concludes a need for more relational and causal studies. Pursuant to this, we studied the relation of textbooks and students' strategy use at the end of Grade 1. This study is based on a secondary analysis of a data set including 1614 students from 86 school classes from the complete first year of school. First, we carried out a theory-based textbook analysis of the four textbooks used in this sample to determine the textbooks' quality with respect to opportunities to learn arithmetic principles. Second, we conducted a multilevel analysis to examine the relation of the resulting textbook quality scale to the students' use of arithmetic principles, while controlling for relevant individual and classroom covariates.

1.1. Basic Arithmetic Principles

Most students can solve simple arithmetic problems using counting strategies – at least by means of concrete objects on their first day of school (Clements & Sarama, 2007; Kilpatrick et al., 2001). Although counting strategies build on children's prior knowledge and thus constitute a core component of their initial learning at school, the continued use is problematic. The usability is limited, in particular beyond the number 20 and it causes several problems in the further learning process (Cowen, 2003). Creating associations between a problem and its correct solution and thus gradually internalizing those associations for recall is difficult especially for students continuously using error-prone counting strategies (Siegler, Deloache, & Eisenberg,

2006). This is disadvantageous for further learning as the associational network of basic number combinations develops cumulatively. Consequently, the use of advanced and more efficient strategies should replace the use of counting strategies as major solution methods.

First strategies students use in this development, which go beyond simple counting, are based on arithmetic principles. Prather and Alibali (2009) describe mathematical principles as “fundamental properties about the functioning of the system of mathematics” (p. 222). Unlike concepts, which principles are related to, they are not constructed by learners, but inherent in a certain domain. Thus, in arithmetic, certain laws or regularities of relationships between numbers or operations can be considered as principles of this domain (e.g., the laws of commutativity, associativity and distribution). Four basic tasks generate the so called *family of facts*, the core of arithmetic principles which occur in the Grade 1 curriculum: a basic task $a + b = c$, its commutative equivalent $b + a = c$, and the two inversions $c - a = b$, $c - b = a$ (c.f. Baroody, 2016). They are based on the commutative law of addition and the complementarity of addition and subtraction. Students with the ability to fluently utilize the family of facts can derive one other task or solution out of a basic one by using the commutative equivalent (swap strategy) and two other tasks or solutions by using the inversions (reverse strategy). Furthermore, for each addition and subtraction task four new tasks (neighbor tasks) arise by decreasing or increasing one addend, the subtrahend or the minuend by 1 (e.g., $a + b = c \Rightarrow (a + 1) + b = c + 1$). We use the term neighbor strategy for deriving a solution of a task by utilizing the relation between neighbor tasks. We refer to the family of facts and the neighbor tasks as *basic arithmetic principles (BAP)*, and to the strategies utilizing these principles as BAP-based strategies (c.f. Kilpatrick et al., 2001). Table 1 illustrates the three typical BAP-based strategies for simple addition and subtraction problems.

Strategy	Definition	Example
Swap	Deriving the solution of a task using the commutative equivalent	$2 + 9 = \underline{\quad}$ $\rightarrow 9 + 2 = \underline{11}$ $\underline{2 + 9 = 11}$
Reverse	Deriving the solution of a task by using the inversion	$17 - 15 = \underline{\quad}$ $\rightarrow 15 + \underline{2} = 17$ $\underline{17 - 15 = 2}$
Neighbor	Deriving the solution of a task by using one of the neighbor tasks	$8 + 9 = \underline{\quad}$ $\rightarrow 8 + 8 = 16$ $\rightarrow 8 + (8 + 1) = 17$ $\underline{8 + 9 = 17}$

Table 1: Strategies utilizing the BAP

Students capable of fluently using the three strategies based on BAP can easily derive solutions for seven simple addition and subtraction problems from the solution of only one basic task which they know by heart. Hence, their effort to memorize number facts drastically decreases.

Research literature provides several models describing children’s acquisition and efficient use of strategies. In their theoretical framework of strategic competence, Lemaire and Siegler (1995) distinguish four dimensions for which they showed that changes in any of them can

improve speed and accuracy of flexible use of strategies overall: *strategy repertoire* as the knowledge of different types of strategies; *strategy distribution* as the knowledge about the relative frequencies of use of these strategies; *strategy efficiency* as the ability to perform strategies quickly and accurately; and *strategy selection* as the ability to flexibly select a strategy for a given problem. Following this model and based on a series of studies as well as computer-based models, Chen and Siegler (2000) described how students learn new strategies and their efficient use. These results suggest that the acquisition of an advantageous use of strategies takes place on the basis of suitable tasks allowing students

- to discover strategies,
- to experience strategy execution which increases strategy fluency and reduces cognitive load, and
- to perform metacognitive processes like examining strategy efficiency which might result in strengthening the association between existing strategies and problem types.

There are further empirical findings which suggest that primary school children seldom invent important strategies without explicit presentation (De Smedt, Torbeyns, Strassens, Ghesquière, & Verschaffel, 2010; Heinze, Arend, Gruessing, & Lipowsky, 2018). Other results report benefits of an early emphasis on a flexible use of strategies, and thus also support the frequent handling of different strategies in early grades (Blöte, van der Burg, & Klein, 2001; Klein, Beishuizen, & Treffers, 1998). Moreover, several experimental studies indicate that an explicit comparison of solution strategies improves the flexibility of students' strategy selection (e.g., Klein et al., 1998; Rittle-Johnson & Star, 2007, 2009).

1.2 Textbooks as learning resources

Textbooks play a crucial role as a learning resource for teachers and students in school. They translate the abstract curriculum into concrete operations, which teachers and students can perform. Thus, they act as a mediator between the official curriculum and the curriculum implemented by the teachers and are referred to as *potentially implemented curriculum* (Valverde, Bianchi, Wolfe, Schmidt, & Houang, 2002). In this manuscript we use the term curriculum to refer to the level of obligatory learning objectives, which is the most concrete and binding. In some countries like in the US or in Germany there exist national standards as a framework for the whole country which are adapted by the federal states to different state-specific curricula. In contrast, we consider the term textbook as a possible interpretation and specification of a given curriculum. If a textbook is used in one federal state, it generally has to follow the statewide curriculum.

According to TIMSS 2011, which collected specific data on textbook use in schools, textbooks appeared to be the most important learning resource in mathematics education for primary school teachers (Mullis et al., 2012). This particularly applies to Germany where the data of our study presented below were collected: 86 % of the primary school teachers use the textbook

as a basis for their instruction (Mullis et al., 2012)¹. Furthermore, several empirical studies indicate mathematics textbooks influence teachers' mathematics instruction. Krammer (1985), for instance, has shown differences in the implemented teaching practices of 50 mathematics teachers in the Netherlands using one of the three best known Dutch mathematics textbooks. By analyzing data of the TIMSS 1995 video study, Schmidt et al. (2001) found a positive correlation between the relative space of a topic within a textbook and the time teachers spent on that topic during lessons. Also, topics not covered in a textbook were unlikely to be covered in class (Schmidt, McKnight, Valverde, Houang, & Wiley, 1997).

In addition to these findings pertaining to the impact of mathematics textbooks on teachers' instruction, there are indications regarding the impact on student achievement as well. Agodini, Harris, Thomas, Murphy, & Gallagher (2010) randomly assigned four textbooks to 111 primary schools in 10 US states. Students of two textbook groups outperformed those of another one in Grade 2. Other correlational studies underline the connection between textbook and student achievement. For example, by analyzing the US TIMSS 1995 data of Grade 8, Schmidt and colleagues (2001) found a relationship between the portion a topic takes up in a textbook and the students' achievement gain in that topic. Törnroos (2005) reports similar results from the Finnish TIMSS 1999 data. The investigation of nine mathematics textbook series used in 104 Finnish schools (Grade 5–7) showed that the number of learning opportunities a textbook provided for the content of TIMSS items was significantly correlated with students' performance in the TIMSS 1999 test. Using publicly available and highly aggregated school-level achievement data, studies from economics of education provide evidence for textbook effects in California, Florida, and Indiana (Bhatt & Koedel, 2012; Bhatt, Koedel, & Lehmann, 2013; Koedel, Li, Polikoff, Hardaway, & Wrabel, 2017). In addition, by analyzing data of a three-year study (Grade 1–3) with 1664 German primary school students from 93 classes van den Ham & Heinze (2018) found a substantial effect of the textbook choice on the students' achievement in arithmetic. Their results also indicate that the effect is cumulative when teachers use textbooks of the same textbook series from Grade 1 to Grade 3. However, there are conflicting results as well. Van Steenbrugge, Valcke and Desoete (2013) analyzed cross-sectional data of 1579 students (Grade 1–6) from 89 classes in Belgium. Their results showed no substantial effects of the five implemented mathematics textbook series on the students' achievement. In line with that, Blazar and colleagues (2019) did not find differences in students' achievement growth between learners using varying primary school mathematics textbook. Their sample included classes of about 1200 teachers from six US states, all of which adopted the Common Core State Standards (CCSS).

Although the majority of the previously mentioned studies are of correlational nature, there are some indications that the choice of a mathematics textbooks is related to students' learning. Summarizing the results from studies in the TIMSS context, the space a topic covers in a textbook correlates with the time it is dealt with in class as well as with the students' achievement gain. It is also plausible to assume that a higher quality of topic-specific learning

¹ Nevertheless, in a more recent study, Blazar and colleagues (2019) showed that teachers' general textbook use can still mean quite a variance in frequencies.

opportunities in a textbook means a broader treatment in class covering more aspects of the topic. Hence, similar correlations for the textbook's quality on a topic might be expected as well where the lesson quantity of a topic plays the role of a moderator of the relation between textbook quality and student learning.

A relation between textbooks and student achievement is not only possible for the specific textbooks that are used, but also for different topics within one textbook (Bhatt et al., 2013). Regarding children's use of strategies in primary schools, there are also indications of the used mathematics textbook's influence when working on arithmetic problems (cf. Fagginger Auer et al., 2016; Heinze et al., 2009; Sievert et al., 2019). Fagginger Auer and colleagues (2016) analyzed data from the 2011 assessment of Dutch primary schools' mathematics, comprising 1619 students. Using a multilevel latent class analysis, they could show that textbook choice is related to the students' probability of fitting a certain profile of strategy use. In line with this, Heinze et al. (2009) found a relationship between the number of efficient strategies used by 245 German third graders and the instructional approach of the used textbook. Both studies are based on cross-sectional data and do not contain information regarding the teachers' use of or agreement with the textbook when teaching arithmetic strategy use. Sievert et al. (2019) examined data of 1404 students in Germany and their adaptive use of addition and subtraction strategies at the end of Grade 3. They also found that students' strategy use differed depending on the quality of opportunities to learn presented by the textbooks their teachers used.

1.3 Present Study

The purpose of the present study is to analyze the role of the mathematics textbook as a learning resource for first graders' use of BAP-based strategies. The analysis enriches the current research since they are based on large-scale data of a one-year study with 1614 students from 86 classes over the first school year. Since all textbooks in this study follow the same statewide curriculum, correlations are not confounded by different curricula. Following the approach of Törnroos (2005), we analyzed the learning opportunities in the four different textbook series used in our sample. We used the research results presented in Section 1.1 to conceptualize textbook quality concerning BAP-based strategies. The resulting textbook quality scale facilitates the supplementation and further development of the research of Fagginger Auer et al. (2016), Heinze et al. (2009) and Sievert et al. (2019) on the relation between textbooks and students' use of arithmetic strategies. Specifically, it provides initial insights into the role of textbooks and teachers' emphasis of a topic in class. We pursue this objective by addressing the following research hypotheses:

Hypothesis on textbook quality

- (1) Grade 1 textbooks which follow the same arithmetic curriculum differ in the quality of learning opportunities concerning basic arithmetic principles.

Hypothesis on the relation between textbook quality and students' strategy use

- (2) The quality of learning opportunities in basic arithmetic principles presented in Grade 1 textbooks is related to students' use of basic arithmetic principles at the end of Grade 1.
- (3) The relation between the quality of learning opportunities for arithmetic principles presented in Grade 1 textbooks and the students' use of basic arithmetic principles at the end of Grade 1 is moderated by the lesson quantity teachers spent on this topic in class.

2. Method

2.1. Participants and design

Our study is a secondary analysis of a dataset stemming from the evaluation of a mathematics support program for low-achieving students which was conducted within the regular mathematics classroom². In the evaluation, two groups, in which teachers received additional teaching material for those students, are compared to a control group. The teaching material addressed aspects of basic arithmetic (e.g., concept of cardinal numbers, place-value system, concepts of the basic arithmetic operations) to identify and support low-achievers at the beginning of primary school. Thus, it also included material dealing with the family of facts (cf. Section 1.1). However, only low-achieving students were supported with this material. Dealing with the family of facts was one of 23 content components of the program and therefore it was a subject at only one point in time during the first grade. Furthermore, participation was independent of the textbook the teachers used. We controlled for possible effects of participation in the support program using dummy coded variables in the analyses (see Section 2.4), but it did not show any effect on the students' use of BAP-based strategies (see Section 3.2).

The entire dataset involves data of 2462 students from 40 schools and 127 classes in [federal state in Germany] (about 10 % of the cohort in this German federal state) from the start to the end of Grade 1. Schools were selected from rural as well as from urban areas all over the federal state. A state-wide obligatory curriculum prescribes content (e.g., number concept, concept and strategies of addition and subtraction), skills (e.g., switch between representation, split numbers), and forms of number representations (concrete, iconic, symbolic) to be addressed in the mathematics classroom for all teachers. Within this dataset, 86 classes (1664 students) used one of the four textbook series "Denken und Rechnen", "Einstern", "Flex und Flo" and "Welt der Zahl". In 15 classes teachers used a variety of other textbooks, no textbooks were used in 4 classes, for 18 classes we had no information about the textbook used, and 4 classes were excluded due to an insufficient number of students ($n < 5$). We excluded 50 students, because they did not participate in the test at the end of Grade 1. Thus, to ensure a robust comparison, we chose the subsample of 1614 students (756 female, 47 %) with a sufficient

² The studies of van den Ham & Heinze (2018) as well as Sievert et al. (2019) also used data from this data set. Their subsamples slightly differ from the sample used in the present study and their studies focused different aspects of textbook effects and children in a different grade.

number of classes and students for each of the four textbooks for our analyses (Denken und Rechnen: 371 students; Einstern: 351 students, Flex und Flo: 536 students; Welt der Zahl: 356 students).

The four textbook series are similar in their content structure and follow the same curriculum. While the series “Denken und Rechnen” and “Welt der Zahl” include all topics of the school year within one book, “Flex and Flo” is divided into three issues, “Einstern” into six. All textbooks comprise the same topics for Grade 1 (e.g., beginning with numbers, operations, geometry, patterns). For each textbook the vast majority of teachers (90.8–100 %) stated that they had used the textbook for arithmetic instruction at least once a week. Moreover, 81.1–94.1 % reported a general orientation towards the content, methods and chronology of the textbook. All textbooks were published by commercial publishing houses. In contrast to other federal states in Germany, in [federal state of Germany] the state authority waived the procedure for a formal textbook approval at the time the study was conducted and referred to the textbook approval in neighboring states with similar curricula. The textbook choice is made on the school level, usually by the group of mathematics teachers and not by individual teachers, while little is known about the criteria (Hartung, 2014). However, there is one school in the sample with different textbooks being used between classes, while in all other schools the same textbook series was used in all classes of these schools. We tested whether the textbooks are distributed differently between teachers who studied mathematics and out-of-field teachers in order to check whether teacher qualification and textbook choice are confounded (which was not the case, see Section 3.2 for details). The dataset did not include school-level variables to control for.

All data were collected by an authority subordinated to the Ministry of Education of [federal state of Germany], assuring the official approval for human subject related research in schools. We received all data in a full pseudonymized condition and had no contact to the students or teachers. The compliance of the human subject related guidelines were monitored by the Data-Protection Supervisor of [federal state of Germany].

2.2. Measures

2.2.1. Textbook quality regarding BAP-based strategies

We developed a measure for the quality of opportunities to learn presented by the textbooks to analyze textbook quality regarding BAP-based strategies and to test Research Hypothesis 1. For this purpose we derived indicators for textbook quality from the four dimensions of the model for strategy competence development by Lemaire and Siegler (1995): strategy repertoire, strategy distribution, strategy efficiency, and strategy selection (see Section 1.1). These categories also reflect relevant cognitive domains which describe the addressed level of cognitive processes (Grønmo, Lindquist, Arora, & Mullis, 2015). The indicators were applied to the textbooks, in order to determine the quality of learning opportunities for BAP-based strategies presented by the textbooks. In the following we describe the criteria and the scoring for the four strategy categories in detail together with illustrative examples.

In order to assess the learning opportunities regarding different types of strategies (strategy repertoire), we rated the quality of instruction for the three strategies swap, reverse and neighbor. We scored one point each, if:

- the strategy is introduced iconically
- the strategy is presented both by educational materials and in a lifelike context for students
- a name of the strategy is introduced
- the execution of the strategy is also described linguistically
- the strategy is presented by manipulatives typically available for students in the classroom.

Lifelike contexts may take up students' everyday life and are thus meaningful. Likewise, giving a solution method a name gives it meaning and relevance. Since representations are not self-explanatory, a linguistic description might give some students additional learning support. Finally, the active use of physical objects is particularly important in the first year of schooling for students to create mental models of arithmetic operations. The scores for the three strategies were totaled for each textbook resulting in a maximum of 15 points per textbook in the category "strategy repertoire".

Since a repeated appearance of a strategy strengthens the importance ascribed to a strategy as well as the frequency of its use, we chose the quantity of presentations of a strategy within a textbook as criterion for the dimension "strategy distribution". We did a partial credit scoring for each of the three strategies as follows:

- 0 point for no or only one presentation of the strategy (the latter case is already scored in the preceding category strategy repertoire);
- 1 point if the strategy is introduced and also presented by means of another problem on another page of the textbook
- 2 points if the strategy is introduced and also presented more than once by means of different problems on several pages of the textbook.

The scores of the strategies were again totaled to an overall score (0–6 points) for the respective textbook.

The dimension "strategy efficiency" covers the ability to perform strategies fluently and accurately. Its development requires experience and practice in the use of the relevant strategy. Thus, the criterion for our textbook analysis in this domain was the number of problems or tasks specifically determined for each strategy. Since the strategies swap and reverse demand specific tasks formats (e.g., three numbers given, find both addition tasks), we counted the number of tasks for these strategies. For the strategy neighbor, we determined the quantity of packages, which consist of neighbor tasks, since only several tasks in combination can form neighbors. Figure 1 shows examples for each strategy. In contrast to the other three categories,

these numbers were not scored but simply counted and the quantity was directly used as an indicator for the textbook.

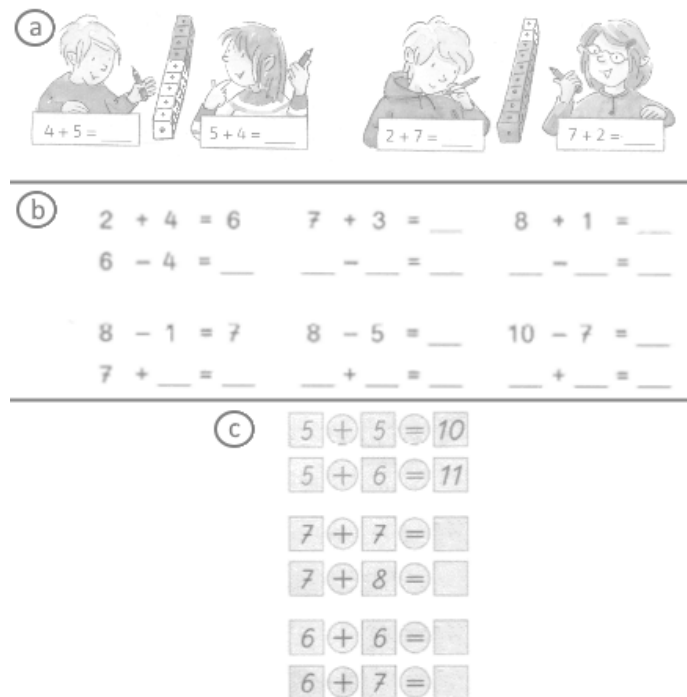


Figure 1: Example tasks from textbooks for the strategies swap (a, 2 tasks: Buschmeier, 2011), reverse (b, 6 tasks: Brall, 2010), and neighbor (c, 2 tasks: Bauer & Maurach, 2011)

Regarding the dimension “strategy selection”, which is the ability to flexibly select a strategy for a given problem, we know from earlier research that an explicit comparison of solution strategies is conducive for the development of a flexible selection (cf. Section 1.1). Thus, we identified the number of learning opportunities which demand the comparison of one of the strategies swap, reverse and neighbor to any other solution strategy. An example is given in Figure 2.

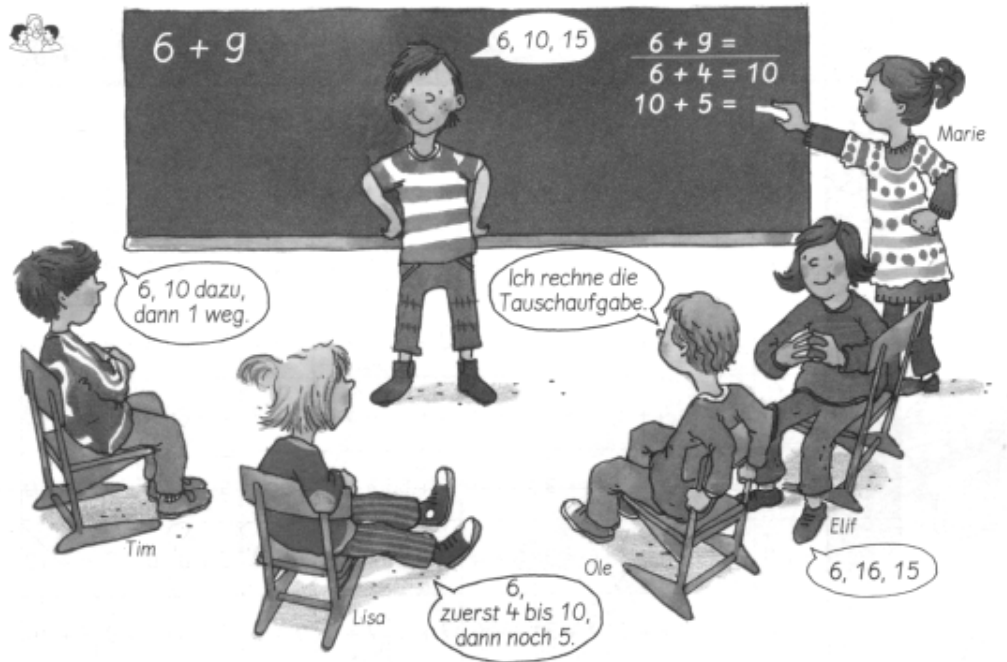


Figure 2: Example for a learning opportunity for a strategy comparison (speech bubbles from the left: “6, add 10, then take away 1”; “6, first 4 to the 10, then another 5”; “I use the swap task”: Rinkens, Hönisch, & Träger, 2011)

Further, we followed the assumption that a structured overview in the form of an addition table or subtraction table, in connection with tasks to discover the interconnectedness of addition and subtraction problems (cf. Section 1.1), fosters students’ skills to make use of the resulting arithmetic principles as strategies in a flexible way. Consequently, we determined the number of learning opportunities using an addition or subtraction table. An example is given in Figure 3.

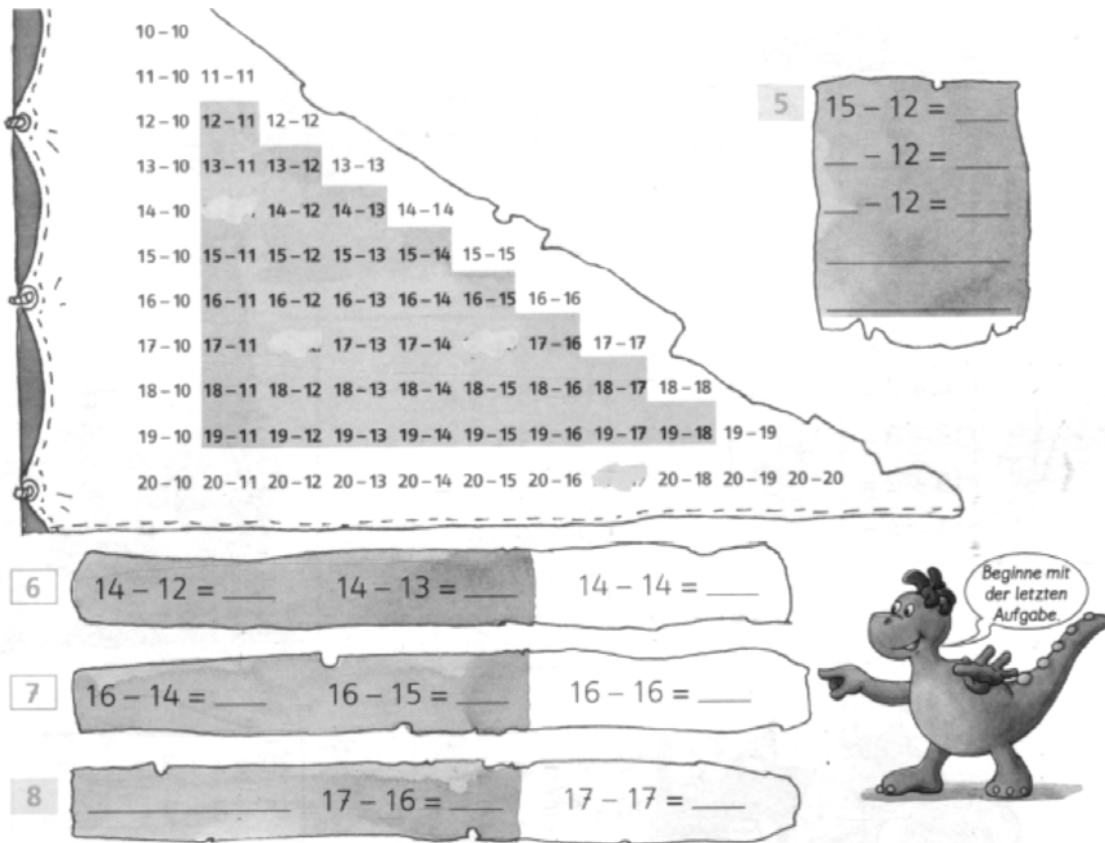


Figure 3: Example for a learning opportunity using a subtraction table (“Start with the last problem”: Rinkens et al., 2011)

A third criterion for the category “strategy selection” is the occurrence of written explanations for the use of a strategy. Unlike the simple description of strategy execution, which we accounted for in the category “strategy repertoire”, reasoning in favor of a certain strategy clarifies *why* to use a strategy, and thus addresses metacognitive abilities of a flexible selection of strategies. Hence, we counted the number of opportunities to learn which include such explanations either in the instruction for students or in the footnotes with instructional comments meant for the teacher. Figure 4 gives an example for such a learning opportunity.

1 Finde die Tauschaufgabe und rechne.

$4 + 7 = 11$	\rightarrow	$7 + 4 = 11$
$5 + 9 = \square$	\rightarrow	$\square + \square = \square$
$3 + 8 = \square$	\rightarrow	$\square + \square = \square$
$4 + 9 = \square$	\rightarrow	$\square + \square = \square$
$5 + 7 = \square$	\rightarrow	$\square + \square = \square$

Figure 4: Example for a learning opportunity with an explanation for the use of the swap strategy (instruction: "Find the swap task and calculate"; speech bubble: "If the higher number comes first, it is easier for me.": Bauer & Maurach, 2011)

Table 2 summarizes the categories, criteria and the scoring. Scores in the categories "strategy repertoire" and "strategy distribution" were totaled to a category score. Scores for the different criteria in the categories "strategy efficiency" and "strategy selection" were standardized and then totaled to a category score. All category scores were standardized and totaled to a standardized mean score, reflecting the assessed quality for the textbook's instruction of strategies based on BAP.

Category	Strategy repertoire	Strategy distribution	Strategy efficiency	Strategy selection
Cognitive domain	Knowing	Knowing	Applying	Reasoning
Criteria	1 point each for: - iconic introduction - presentation by educational materials <u>and</u> in a lifelike context - introduction of the name - strategy execution is described linguistically - strategy is presented by manipulatives that are typically available in the classroom	0 points: 0–1 presentations of the strategy 1 point: 2 presentations 2 points: 3 or more presentations of the strategy	Number of specific problems for each strategy	Number of strategy comparisons Number of tasks using an addition or subtraction table Number of metacognitive elements regarding a strategy choice
Scoring	0–5 points/ strategy 0–15 points/ textbook	0–2 points/ strategy 0–6 points/ textbook	<i>Means of the raters</i> <u>Swap</u> : 34.67–74.0/ textbook <u>Reverse</u> : 14.33–92.67/ textbook <u>Neighbor</u> : 68.33–166.33/ textbook	<u>Comparisons</u> : 1–6 / textbook <u>+/- Tables</u> : 0–5 / textbook <u>Metacognition</u> : 2–4 / textbook

Table 2: Categories, criteria and scoring of the textbook assessment

2.2.2. Students' use of BAP-based strategies

The students' skills to use BAP-based strategies were assessed using two item types from the Grade 1 arithmetic test. This test was administered at the end of the school year and contained different types of arithmetic items. The first item type addressed the family of facts: Three numbers were given (e.g., 11, 7 and 4) and the first graders were asked to write down the four equations of the associated family of facts. These items required the use of the strategies swap and reverse. The test included three items of this type and the students were given 60 seconds for each item. Every correct equation was scored with one point. An example item is given in Figure 5 (left-hand side).

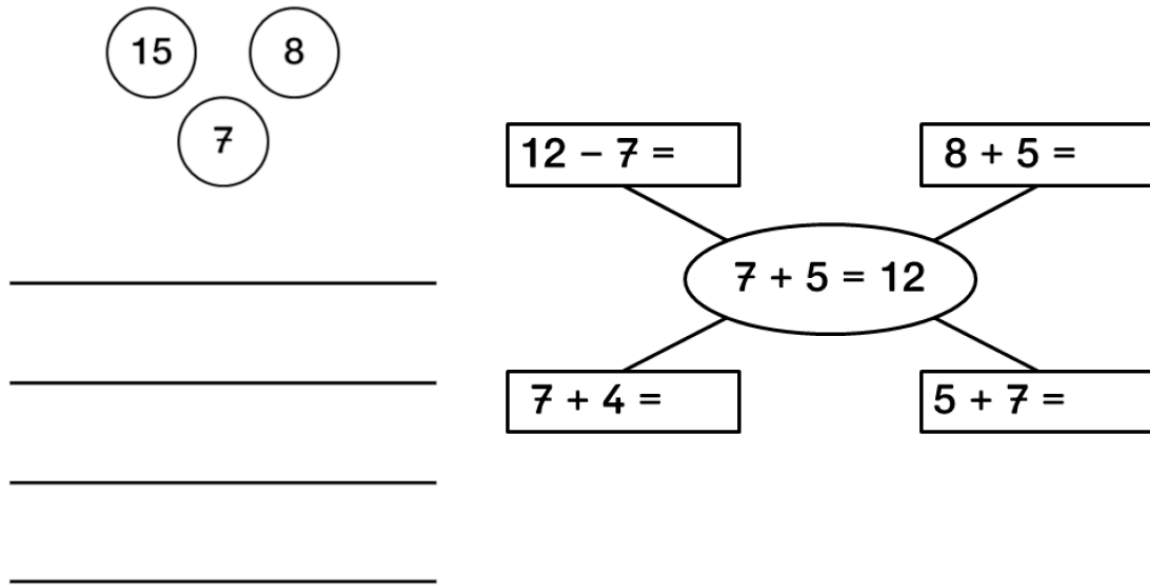


Figure 5: Examples of the item types 1 & 2

The second item type addressed the family of facts and the neighbor tasks. One already solved addition or subtraction problem was given in the center surrounded by four unsolved, related problems. These problems comprised the commutative equivalents, inversions or neighbor tasks. Hence, these items required the use of all BAP-based strategies. There was a speed component for these items to ensure that the scoring actually represents the use of BAP-based strategies and the relations between the problems: The students were given 30 seconds to solve all four equations. In addition, we used a partial credit scoring, giving two points for four correct solutions, one point for three correct solutions and zero points for two, one or zero correct solutions. The speed component and the partial credit scoring should ensure that students, who started calculating the problems separately one by one get no credit, since it is unlikely that they solve more than two problems under these conditions (though it is not impossible). The test included three items of this item type as well; an example is given in Figure 5 (right-hand side). We conducted an additional interview study with 18 students from Grade 1 and 2 (39% female) to confirm the validity of this item type. The students were given the same items under the same conditions as in the large sample. Afterwards they were asked how they solved the items and whether they used the equations in the center. A Chi-square test was used to compare students' partial credit score and their use of BAP-based strategies (dichotomous). The results showed a significant connection between score and use of BAP-based strategies, $\chi^2(2) = 6.41$, $p = .04$, $V = 0.34$. Hence, students with a higher score showed a more frequent use of BAP-based strategies and vice versa.

The students' scores for the six items of both types were estimated based on Item Response Theory using an extension of the Rasch model (partial credit model; Masters, 1982). The resulting scale showed a good reliability (WLE Person separation reliability = .82) and was used as a measure for students' ability to use BAP-based strategies.

2.2.3. Individual and classroom covariates

We included a number of available individual and classroom-related covariates in our analysis, in order to explore the specific connection between textbook quality and the students' use of BAP-based strategies. Several studies provide evidence that students' abilities at school entry influence the development of their mathematics skills in primary school (e.g., Schneider, Küspert, & Krajewski, 2013). Accordingly, we controlled for students' basic cognitive abilities, basic numerical skills and German language skills. For this purpose approved standardized tests with sufficient reliabilities were administered at school entry (basic cognitive abilities: CFT 1-R, Weiß & Osterland, 2013, Cronbach's $\alpha = .91$; basic numerical skills: HaReT 1, Lorenz, 2007, Cronbach's $\alpha = .74$; German language skills: Münsteraner Screening, MÜSC, Mannhaupt, 2013, Cronbach's $\alpha = .72$). For CFT 1-R (45 items), students had to solve problems by means of formal logic, like continuing series, finding substitutions or identify classifications non-verbally. Concerning German language skills, MÜSC (40 items) was implemented to tests the students' phonological awareness as requirement for the acquisition of written language. HaRet 1 (43 items) tested students' arithmetic-related precursory skills at school entry, like one-to-one mappings, comparing sets or numbers, linking number symbols and to sets. BAP-based strategies are not part of the HaRet 1 test. Although preschoolers bring a variety of mathematical proficiency to school (e.g., counting up to ten or solving simple addition or subtraction word problems), their conceptual understanding is usually limited to small numbers, counting strategies, the use of concrete materials and familiar word problems (Kilpatrick et al., 2001). Sophisticated thinking strategies, like those based on BAP, are gradually developed in the first years of schooling based on the curriculum for Grade 1 and 2. The standardized school entry tests were scored according to the test manuals and the results for the different textbook groups are given in Section 3.1.

On the class level we included the participation of the classes in one of the three groups of the arithmetic support program (see Section 2.1) as dummy variables to control classroom characteristics. Since differences between classes could also be caused by different levels of students' ability on average, we controlled for class composition effects by including the aggregated variable for basic cognitive abilities (CFT 1-R) at school entry. Aggregated numerical and German language skills were excluded by reason of multicollinearity. The latter results in distorted estimates. Further, we used the mathematics-specific teacher qualification as an indicator for this covariate (as dummy variable indicating whether the teacher studied mathematics or not) since the dataset did not provide detailed information on the teachers' professional knowledge. In Germany, primary school teachers are generalists, i.e., all teachers teach mathematics but only some of them have a formal qualification to teach mathematics. Findings from different studies with German primary school teachers revealed that the formal teacher qualification is strongly related to mathematics-specific professional knowledge (e.g., Knievel, Lindmeier, & Heinze, 2015; Tatto et al., 2012).

Finally, the learning opportunities presented in textbooks might affect the teaching activities in the classroom and these teaching activities might affect the students' learning. Hence, we included data on the quantity of lessons in which teachers dealt with BAP-based strategies in the

mathematics classroom as indicator for the teaching activities. The available data in the dataset were collected by a teacher questionnaire simultaneously administered to the student test at the end of Grade 1. The teachers were shown an example of both BAP item types, an example like in Figure 5 supplemented by the item instruction, and asked how often they have presented the corresponding mathematical topics in the mathematics lessons of their class (never, 1-2 lessons, 3-4 lessons, more than 4 lessons). Unfortunately, the question for the second item type did not include the essential speed component, which caused invalid and biased data so that we had to exclude the second questionnaire item from our analyses. Figure 6 gives an overview about all variables included in the analyses.

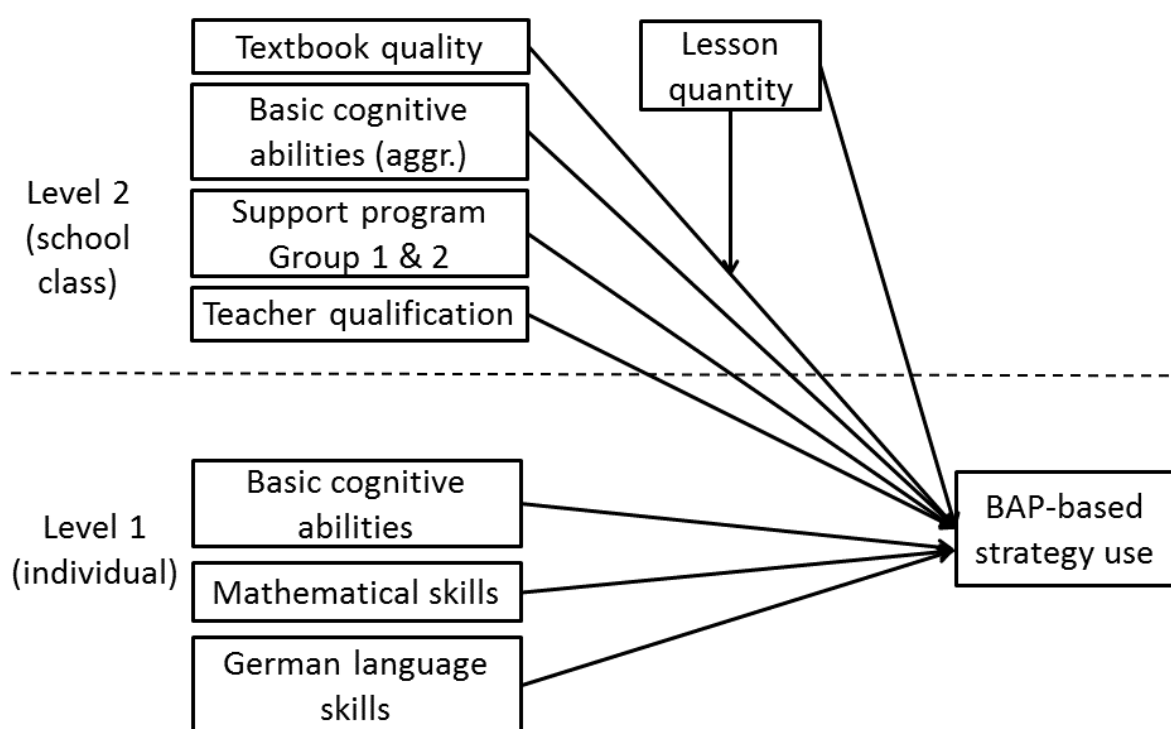


Figure 6: Final model of the multilevel analysis (basic cognitive abilities, mathematical skills, and German language skills at school entry)

2.3. Procedure

2.3.1. Textbook analysis

The Grade 1 textbooks of the four series were rated by three independent and trained persons on the basis of the four categories (strategy repertoire, strategy distribution, strategy efficiency, strategy selection), following the criteria presented in Section 2.2.1. Their results indicated a solid agreement (Fleiss' κ : .77, ICC: .92). Subsequently, the consensus method was applied to obtain a unique scoring for all but the category "strategy efficiency". The number of problems determined by the three raters for strategy efficiency varied within a range of 14 %

around the mean due to the large number of exercises. Hence, the mean was chosen for the further analysis. In order to weight all four categories equally, the resulting scores of the categories were standardized and aggregated to a single standardized quality scale for the textbooks' learning opportunities with respect to BAP-based strategies. The subscales of the single categories could not be included separately due to multicollinearity.

2.3.2. Data collection timeline

The data collection started at the beginning of the school year 2013/2014 when the students entered primary school. Basic cognitive abilities, basic numerical skills and German language skills were assessed at the beginning of Grade 1. The BAP items were administered at the end of Grade 1. All data were collected by trained test administrators in accordance with the respective test manuals. All instructions were read out aloud and the administrator worked an example with the children for the BAP items. At the end of Grade 1 teachers were given two questionnaires to collect data about their lesson quantity with respect to arithmetic principles in the classroom, their qualification and textbook use.

2.4. Data analysis

Differences in textbook quality with respect to BAP-based strategies (Research Hypothesis 1) were analyzed on the basis of the four categories (strategy repertoire, strategy distribution, strategy efficiency, strategy selection) by comparing and interpreting the appropriate scores of the four textbook series.

We applied multilevel random intercept models with the software Mplus 7.0 (Muthén, & Muthén, 1998–2012) to analyze the relation of textbook quality to students' strategy use (Research Hypotheses 2 & 3) while accounting for the nested data structure (Hox, 2010). The Model 0 without predictors (null model) was run to estimate the partition of variance between and within classes. In Model 1, we included the individual and classroom covariates: (1) students' basic cognitive abilities, basic numerical skills and German language skills at individual level to account for individual characteristics at school entry and students' basic cognitive abilities aggregated to a class mean at class level to control for the class composition, all of these variables were standardized to facilitate interpretation; (2) the support programs as dummy variable (see Section 2.1), and (3) the teacher qualification. The textbook quality scale regarding BAP-based strategies was included in Model 2. Model 3 included the lesson quantity teachers spent on BAP-based strategies instead of the textbook quality scale. We chose this model to separately control the individual meanings of both lesson quantity in classroom and textbook quality. In Model 4, both of these variables were included simultaneously. Here, we were able to analyze the unique relation between textbook quality and the students' use of BAP-based strategies at the end of Grade 1. Finally, Model 5 also included the interaction of textbook quality and the lesson quantity regarding BAP-based strategies, in order to address Research Hypothesis 3.

Scores for the learning prerequisites, the aggregated class mean of basic cognitive abilities, the lesson quantity, the textbook quality scale and the dependent variable for students' use of BAP-

based strategies were standardized so that the corresponding β -coefficients could be interpreted as effect sizes similar to Cohen's d (cf. Tymms, 2004). For missing data on the independent variables we applied a full-information-maximum-likelihood (FIML) approach in Mplus 7.0. FIML combines missing data and parameter estimation in a single step and uses all the available information and is therefore very useful (Enders, 2010). Due to sample selection there was no missing data for textbooks, support programs, aggregated cognitive abilities, and students' use of BAP-based strategies. The share of missing data for the covariates are given in Table 4 in Section 3.2.

3. Results

3.1. Textbook quality of opportunities to learn regarding BAP-based strategies

The results for the textbook quality with respect to BAP-based strategies are presented in Table 3.

	Strategy Repertoire	Strategy Distribution	Strategy Efficiency	Strategy Selection	Overall Score
Denken und Rechnen	-1.26	-1.73	-1.58	-0.45	-1.71
Einstern	1.26	0.58	0.19	0.34	0.81
Flex und Flo	0.63	0.58	1.20	-1.30	0.38
Welt der Zahl	-0.63	0.58	0.19	1.41	0.53

Table 3: Results of the textbook analysis: z-scores for the four strategy categories and overall z-score (mean)

Table 3 reveals that the four textbooks show specific strengths and weaknesses. The textbook "Denken und Rechnen" always shows the lowest quality concerning opportunities to learn for BAP-based strategies, except for the category "strategy selection" where it is rated the second lowest. Each of the other three textbooks represents the highest quality in one category. Flex & Flo shows the highest score for the category "strategy efficiency", which is assigned to the cognitive domain of applying and the lowest score for the category "strategy selection", which is assigned to the cognitive domain of reasoning. The overall scores indicate a gap of quality regarding the instruction of BAP-based strategies between the textbook "Denken und Rechnen" and the other three. Hence, teachers who closely comply with this textbook will provide fewer learning opportunities dealing with BAP-based strategies. We find crucial quality differences between the textbook series in terms of opportunities to learn regarding BAP-based strategies, and thus we can confirm Research Hypothesis 1. In contrast to the black box model of overall effects of textbook choice on students' arithmetic achievement in Grade 1 (van den Ham & Heinze, 2018), the ranking for the specific topic BAP-based strategies is entirely different. According to their results, students using „Denken und Rechnen" and "Welt der Zahl" showed significantly higher scores in their overall arithmetic achievements than students using "Einstern". Hence, these results indicate that the textbook quality for the specific topic BAP-based strategies is not simply determined by an "overall" textbook quality.

3.2 The relation between textbook quality and students' use of BAP-based strategies

Table 4 provides the means and standard deviations of all covariates for the four textbook groups. The analyses of variance and subsequent post-hoc tests showed significant differences for a few covariates and textbook groups with negligible effect sizes of $\eta^2 \leq .015$.

	Denken und Rechnen	Einstern	Flex und Flo	Welt der Zahl
Basic cognitive abilities				
(0–45/ 0–44/ 6.8%)	26.19 ^a (8.56)	25.43 ^{b,c} (8.54)	27.39 ^b (9.22)	28.47 ^{a,c} (9.14)
Basic numerical skills				
(0–43/ 7–43/ 8.1%)	30.96 (7.37)	31.83 (6.64)	31.57 (7.36)	32.49 (6.63)
German language skills				
(0–40/ 4–40/ 8.7%)	32.61 ^a (6.17)	33.24 (5.92)	33.02 ^b (6.44)	34.52 ^{a,b} (5.44)
Basic cognitive abilities (aggregated)*				
(0–45/ 16.20–36.06)	26.14 (2.54)	25.48 (2.79)	27.29 (2.99)	28.47 (2.95)
Support program				
Group 1				
(0–1/ 0–1)	0.77	0.38	0.68	0.33
Support program				
Group 2				
(0–1/ 0–1)	0.06	0.38	0.28	0.29
Teacher qualification				
(0–1/ 0–1/ 1.4%)	0.44	0.45	0.33	0.56
Lesson quantity*				
(1–4/ 1–4/ 3.2%)	2.02 (1.06)	2.41 (0.78)	2.61 (1.04)	2.95 (0.72)

Table 4: Means and standard deviations of the covariates for the different textbook groups (possible range/actual range/share of missing data below the variables name, ^{a,b,c} significant differences between groups with same reference mark, *all group differences are significant)

The null model of our analysis indicates a between-class variance in students' use of BAP-based strategies of 12.0 %. Hence, depending on the class the students attended, their use of BAP-

based strategies differed at the end of Grade 1. Table 5 shows the results of the Models 1–5 (as described in Section 2.4).

First of all, there are significant main effects of all aspects of students' abilities on individual level. They explain a substantial amount of almost one third of the within-class variance. These effects confirm earlier research on the meaning of students' abilities at school entry for the development of mathematics competence (e.g., Schneider et al., 2013). Our results show that they are also highly predictive for the students' use of BAP-based strategies at the end of Grade 1.

	Model 1	Model 2	Model 3	Model 4	Model 5
Level 1 (students)					
Basic cognitive abilities	.21** (.04)	.21** (.04)	.21** (.04)	.21** (.04)	.21** (.04)
German language skills	.15** (.02)	.14** (.02)	.15** (.02)	.15** (.02)	.15** (.02)
Basic numerical skills	.27** (.04)	.27** (.04)	.27** (.04)	.27** (.04)	.27** (.04)
Level 2 (class)					
Basic cognitive abilities (aggregated)	.06 (.04)	.04 (.04)	.02 (.04)	.02 (.03)	.02 (.04)
Support program Group 1	-.02 (.09)	.03 (.08)	.01 (.08)	.04 (.07)	.04 (.08)
Support program Group 2	-.03 (.10)	-.09 (.09)	-.03 (.09)	-.07 (.09)	-.08 (.09)
Teacher qualification	.08 (.07)	.09 (.07)	.01 (.06)	.00 (.07)	.01 (.07)
Textbook quality		.16** (.03)		.13** (.03)	.14** (.04)
Lesson quantity			.15** (.04)	.11** (.04)	.12** (.04)
Textbook quality x lesson quantity					.02 (.03)
<i>Intercept</i>	<i>-0.01 (0.08)</i>	<i>-0.03 (0.07)</i>	<i>0.03 (0.07)</i>	<i>0.00 (0.07)</i>	<i>-0.01 (0.07)</i>
Explained within class variance	31.4%	31.2%	31.2%	31.1%	31.1%
Explained between class variance	6.4%	29.7%	26.5%	41.4%	41.9%

Table 5: Results of the multilevel analysis of students' use of BAP-based strategies (standard errors in brackets, all variables are standardized except Support program Group 1, 2, and Teacher qualification (0,1)), * $p < .05$, ** $p < .01$

On the class level, there is no effect of the classroom-aggregated scores of basic cognitive abilities. Participation in the support program does not show any effect on students' use of BAP-based strategies so that our specific analysis of textbook correlation is not influenced by this intervention. In addition, there are no effects of teacher qualification. The lesson quantity spent on BAP-based strategies appears to be predictive for the students' ability to utilize these strategies.

The relation of textbook quality to the students' use of BAP-based strategies is significant in all models (2, 4 & 5). To interpret the coefficients ($\beta = .16$ in Model 2, $\beta = .13$ in Model 4, $\beta = .14$ in Model 5), we refer to the results of Table 3 which show the textbook quality range from -1.71 to 0.81 (z-scores). Students using the textbook of the lowest quality on this scale – on average – show a lower score for the use of BAP-based strategies of about .35–.40 standard

deviation than students using the textbook of the highest quality. The textbook quality and the teacher activities appear predictive in our models. Both of them lead to a considerable increase of explained between class variance when added to the model ($\Delta R^2 = 23.3$ & 20.1 percentage points, respectively). The corresponding coefficients of the standardized variables are similar in size. In direct comparison, including the textbook quality scale appears to have a higher amount of explained variance than including the lesson quantity. When including both variables (Model 4) the explained between-class variance increases up to 41.4% . With respect to Research Hypothesis 2 we find a significant relation of textbook quality to the students' use of BAP-based strategies. However, there is no moderation (Model 5) although an interaction of the textbook as learning resource and the lesson quantity as indicator for classroom activities was expected. Hence, based on these data we have to reject Research Hypothesis 3. Despite no interaction effect on students' strategy use, the scales show a weak to moderate correlation ($r = .24, p < .05$).

4. Discussion

In this study we considered the role of the textbook as a learning resource for the students' learning in terms of their ability to use strategies based on basic arithmetic principles (BAP) in Grade 1. We established a definition of textbook quality for this topic by conducting a theory-based analysis of the respective learning opportunities presented in commonly used mathematics textbooks. We found differences in their quality concerning BAP-based strategies by applying these quality indicators to four mathematics textbooks. A subsequent analysis of this quality measure on the basis of a one-year dataset of 86 classes showed a significant relation of the textbook quality with the students' use of BAP-based strategies after the first year of schooling. The further analysis revealed that the relation of textbook quality with students' learning outcomes is not moderated by the lesson quantity.

4.1. Impact for educational research

With respect to classroom research, our results suggest taking textbook quality into account as an important predictor when studying the mathematics classroom. According to earlier textbook research (see Section 1.2) as well as data of our teacher questionnaires (see Section 2.1) it is reasonable to hypothesize that most arithmetic problems in the mathematics classroom originate from the textbook. Teachers are responsible for lesson preparation and teaching. Our findings indicate one should consider the possibility that both textbook content and instructional approaches might have a substantial effect on the learning environments teachers implement in their lessons. This holds true even when textbooks and teachers follow the same curriculum. Hence, one should take textbooks into account when investigating teaching effectiveness, particularly as an additional factor for the quantity of classroom activities. Moreover, although both the quantity of lessons teachers spent on BAP-based strategies and the quality of textbooks concerning BAP-based strategies are related to the students' use of BAP-based strategies, an interaction of these predictors is not indicated by our data. One reason for this

might be that the scale on the classroom activities only covers the provided quantity of classroom activities for BAP-based strategies but not their quality. Moreover, as mentioned above, we could only include teaching information on one of the two item types in the analysis, which is not a satisfactory measure. Nevertheless, textbook quality and lesson quantity are correlated which supports the findings of Schmidt et al. (2001). Hence, further research should also consider the relation of textbook quality and classroom activities, in particular the quality of learning opportunities provided by teachers.

With respect to mathematics teacher education, our results suggest that teachers should be aware of quality differences between mathematics textbooks on the one hand, and be able to compensate poor quality within the textbook they use, on the other hand (e.g., by using additional materials). Since textbooks differ in the quality of their learning opportunities, teachers' pedagogical content knowledge should include identifying a textbook's weaknesses and providing students with alternative learning opportunities. More research needs to be conducted on the relation of teachers' professional knowledge and their textbook choice and use.

For the field of strategy learning in arithmetic, our studies underline the meaning of an explicitation of strategy learning by means of the dimensions of the Siegler model. Hence, the identified types and quantity of learning opportunities targeting at principle-based understanding appear to have an impact on the acquisition of students' conceptual understanding. The varied and repeated representation of principle-based strategies and the possibility to execute them play a role as well as metacognitive processes for their evaluation. The effectiveness of such instructional approaches is also supported by controlled experimental studies (DeSmedt et al., 2010; Heinze et al., 2018; Klein et al., 1998; Rittle-Johnson & Star, 2007, 2009). Our results, which stem from data of the regular mathematics classroom, support findings that the instruction and presentation of strategies is related to students' later strategy use. In line with earlier research on the relation of textbooks and students' strategy learning, this study suggests that the textbook should be taken into account as a relevant factor (Fagginger Auer et al., 2016; Heinze et al., 2009; Sievert et al., 2019).

Finally, our studies contribute to research regarding mathematics textbooks' relation to students' achievement. Supporting findings of Schmidt et al. (2001) and Törnroos (2005) for secondary students, we found further evidence that the textbook choice is also related to students' learning outcomes at the beginning of primary school (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Koedel et al., 2017). However, our results are not in line with the findings of van Steenbrugge et al. (2013) and Blazar et al. (2019). Although the sample of van Steenbrugge and colleagues was comparable in size, it was distributed over Grades 1-6, and data collection was cross-sectional. Hence, differences in textbook quality of distinct grades within one series might be confounded, as well as the change of the textbook series used between school years. Discrepancies to the results of Blazar et al (2019) might also stem from differences in the methodological approach, like the school-level aggregated data used or the broader data collection in six federal states. It is also possible that the differences result from quality discrepancies within the textbooks, depending on the topic investigated, since this study

examined the relationship for a specific topic of the textbooks. Also worth mentioning, the degree of curricular specifications a textbook must comply might differ between the different studies (i.e., textbook as own curriculum following national standards vs. textbook following a curriculum based on national standards). As mentioned in Section 1.2, there is a lack of experimental study designs in research about the relation between mathematics textbooks and students' achievement. For generalizable and causal interpretations, textbook research will have to grow up from its "infancy" (Blazar et al., 2019, p. 29) and develop new and more sophisticated methods, as also requested by Fan (2013). Blazar and colleagues (2019) discuss possible stumbling blocks along this way. In ecologically valid settings, however, correlational evidence is a step forward, though not satisfying. Our study's contributions to this development are the data collection timeline (allowing to consider a baseline of student characteristics before textbook use), the two-level-modeling (considering individual and classroom covariates), and the definition of a topic-specific quality assessment – which can explain the relation between textbook and topic-specific student achievement.

Although this study is based on the same four-year dataset as Sievert et al. (2019) and the results coincide with respect to an overall significant relation between textbook quality and students' strategy use, they differ both methodologically and in their resulting textbook quality scales. While this study analyzed data from the beginning to the end of Grade 1, the time interval of Sievert et al. (2019) range from the beginning of Grade 1 to the end of Grade 3. Accordingly, the students' strategy use was assessed based on different instruments at the end of the respective data collection timeline. Similarly, different instruments were used for relevant covariates in the two studies. But the most important difference lies in the topic-specific textbook quality. For instance, while the series "Einstern" in Grade 1 reaches the highest score for BAP-based strategies, it turns out to have the lowest score for the adaptive use of addition and subtraction strategies in Grade 2 (see Appendix A). A cross-validation of both quality scales, each for the other area, showed no significant relation. Correspondingly, each observed textbook quality seems to be topic-specific and the same textbook varies crucially in its quality among different topics which it comprises. In our study, we accomplished the development of specific criteria for textbook quality regarding the development of students' use of BAP-based strategies. These quality criteria are in accordance with the empirical research on students' use of BAP-based strategies and, thus, explain the observed relation of textbooks to students' learning outcomes. The results show that a fine-grained and theory-based content analysis of mathematics textbooks is an auspicious approach to study the possible effects of these learning resources. Therefore, our study contributes to textbook research as required by Fan (2013) and extends earlier research by Sievert et al. (2019). In a broader sense, it also joins a growing body of research aiming at investigating the characteristics of learning opportunities and their importance in the learning process of children (e.g. for mathematics: Braithwaite, Pyke, & Siegler, 2017). Although a global assessment and comparison of textbook quality might be a complex and future endeavor, our investigation reveals that one can identify relations between learning opportunities presented by textbooks and students' achievements for the subdomain of BAP. This might be a fruitful contribution to research regarding the quality and impact of

mathematics textbooks. One could conceivably expand this method to other mathematical topics and also identify further strengths and weaknesses of textbooks.

4.2 Impact for educational practice

With respect to BAP-based strategies, the textbook analysis brought up substantial differences in the quality of learning opportunities presented in four textbooks following the same curriculum. Consequently, this study indicates that students' achievement not only depends on the prescribed curricula but also on the specific realization of a curriculum in the form of the textbook used. Hence, textbook content that appear to be less effective should be compensated by more effective material in classes. Considering that textbooks represent a classroom determinant very easily interchangeable at minimal cost, the potential improvement by textbook choice appears not only remarkable but also cost-effective. Hence, textbooks can be regarded as relevant instruments of quality assurance in education. These studies provide an approach for a criteria-based assessment of different subdomains of a textbook, which teachers and educators can utilize since students and teachers should receive high quality textbooks, and students' learning processes should be fostered as well as possible. Such criteria require both theoretical basis and empirical proof.

According to Mullis et al. (2012) and our findings, most mathematics teachers in primary schools frequently use the textbooks and generally follow their conceptions. Thus, they should be sensitized for these quality differences, and especially trained in the assessment of textbook quality. We have limited knowledge about how teachers use a textbook for the preparation of lessons or the implementation of its content. However, the study has shown a significant relation between textbook quality and students' learning of a medium effect size. Therefore, awareness and critical evaluation of textbook quality might be important topics for teacher education and training.

Concerning the acquisition of students' conceptual knowledge in basic arithmetic, this study makes the case for an instruction of strategies that goes beyond routine expertise of basic number combination. As proposed by Kilpatrick et al. (2001), the results underline the meaning of a prolonged learning process, which is enriched by repeated and varied representations of principles and metacognitive learning opportunities.

4.3 Limitations

There are several limitations of this study which might influence the results and their interpretation. The reanalysis of existing data, which is not based on an experimental setting concerning textbook distribution, provides correlational relations, but does not allow a causal interpretation in the strict sense. The use of an existing dataset also affects the sample used in our study. Since we could only include the four widely distributed textbook series to ensure adequate sample sizes per textbook, about 1/3 of the classes with either other or no textbooks or with missing textbook information had to be excluded. A third restriction due to the analysis of

existing data was that we could not include relevant covariates which might interact or be confounded with textbook quality. This comprises, for instance, information about the textbook choice in the primary schools of the sample, observational data about teachers' actual textbook use in the lessons (instead of teacher reports), or data on teachers' professional knowledge which is more detailed than the formal teacher qualification we used. Students' socioeconomic status, which might influence effects of textbooks (Hadar, 2017) could not be included in the analysis due to a large portion of missing data. All these variables require consideration in future investigations to explore their relevance. Especially, the selection of textbooks is crucial for this field of research and little is known about the criteria schools or teachers apply. This process is described as highly idiosyncratic and market-driven for the United States (National Research Council, 2002). For Germany, Hartung (2014) assumes light-hearted decisions, influenced by publishers' advertising or a close relationship between the publisher's sales representatives and the schools. However, the used dataset did not provide additional information about school characteristics that might influence the textbook choice and the number of classes per school is not sufficient for a 3-level modeling to account for the school level. Future studies should explore the relationship of specific school characteristics, textbook selection criteria and textbook choice to exclude the possibility of confounding school level variables.

A second limitation is that our data does not provide information about the weighting of the four categories of the Siegler model (repertoire, distribution, efficiency, flexible selection). Although the rankings vary among these categories, testing all four scales in a single model is not possible because of multicollinearity. Testing Model 2 with alternative scales by leaving out one of the four categories results in slightly lower β -coefficients (.135–.15) of similar size. The same applies to the share of explained variance between classes (25.1–28.9 %), whereas leaving out the category "strategy efficiency" results in the lowest values. Thus, despite this indication for a slightly higher importance of specific problems or tasks within the textbook, we should be careful when interpreting this, and we cannot draw well-founded conclusions about their relative meanings. Nevertheless, based on the Siegler model and on earlier studies on students' strategy use (e.g., Blöte et al., 2001; Klein, et al., 1998; Rittle-Johnson & Star, 2007, 2009; Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009a, 2009b) we consider both, experience or routine as well as making different strategies and their use a subject of discussion, as favorable for students' strategy use.

A further limitation relates to the measure of BAP-based strategy use, in three ways: First, the six items used do not allow a comprehensive analysis of students' use of addition and subtraction strategies. Second, a BAP pretest was not administered. In Germany, students newly learn BAP strategies in their first school year and there is no obligatory time slot when to teach BAP for teachers. We administered the HaRet 1 test to control for children's precursory arithmetic knowledge for BAP (see 2.2.3). Third, measuring students' strategy use is a difficult and multifaceted undertaking in general. Even if students' are asked immediately after solving a problem, it is not clear whether the strategy which they stated is actually the strategy they used (Verschaffel et. al, 2007). Although the additional interview study indicates a relation between the use of BAP-based strategies and the number of correct solutions, it is restrained in terms of

validity. Hence, it is possible, albeit unlikely, that some first graders solve three or even four of the equation of the corresponding items by unusually fast calculation.

Last but not least, we examined the textbook editions that were used by the students in our dataset. Since the data were collected in 2013/2014, the results might not be meaningful for revised editions of the respective textbook series.

5. Conclusion

In the presented study we have shown a significant relation of mathematics textbooks' quality concerning BAP to students' ability to use BAP-based strategies at the end of Grade 1. The findings contribute to textbook research by offering an approach for developing theory-driven quality criteria for textbooks for a specific topic which allows comparing textbooks. Based on a large-scale one-year dataset of students in the regular mathematics classroom taught by the same curriculum, our results further provide evidence for the relevance of textbook quality as a factor for mathematics education in primary schools.

Appendix A. Textbook quality scales for BAP-based strategies (Grade 1) and adaptive strategy use for addition and subtraction (Grade 2, Sievert et al., 2019)

	BAP-based strategies	Adaptive strategy use for addition and subtraction
	Grade 1	Grade 2
Denken und Rechnen	-1.71	2.50
Einstern	0.81	1.25
Flex und Flo	0.38	1.75
Welt der Zahl	0.53	3.00

References

- Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). Achievement effects of four early elementary school math curricula: Findings for first and second graders. NCEE 2011-4001. National Center for Education Evaluation and Regional Assistance.
- Bauer, R., & Maurach, J. (2011). *Einstern 1*. Berlin: Cornelsen.
- Baroody, A. J. (2016). Curricular approaches to introducing subtraction and fostering fluency with basic differences in grade 1. In R. Bracho (Ed.), *The development of number sense: From theory to practice*. Monograph of the Journal of Pensamiento Numérico y Algebraico (Numerical and Algebraic Thought), 10(3) 135–165. University of Granada.
- Bhatt, R. & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis*, 34(4), 391– 412.

- Bhatt, R., Koedel, C., & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review*, 34, 107–121.
- Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S., & Kurlaender, M. (2019). Learning by the book: Comparing math achievement growth by textbook in six Common Core states. Research report. Cambridge, MA: Center for Education Policy Research, Harvard University.
- Blöte, A. W., van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, 93, 627–638.
- Braithwaite, D. W., Pyke, A. A., & Siegler, R. S. (2017). A computational model of fraction arithmetic. *Psychological Review*, 124(5), 603–625. <http://doi.org/10.1037/rev0000072>
- Brall, C. (2010). *Flex und Flo 1*. Braunschweig: Diesterweg.
- Buschmeier, G. (2011). *Denken und Rechnen 1*. Braunschweig: Westermann.
- Chen, Z., & Siegler, R. S. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development*, 65 (2, serial no. 261).
- Clements, D. H., & Sarama, J. (2007). Early childhood mathematics learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 461–555). Charlotte, NC: Information Age.
- Cowen, R. (2003). Does It All Add Up? Changes in children's knowledge of addition combinations, strategies, and principles. In A.J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 35–74). Mahwah, NJ: Lawrence Erlbaum Associates.
- De Smedt, B., Torbeyns, J., Strassens, N., Ghesquière, P., & Verschaffel, L. (2010). Frequency, efficiency and flexibility of indirect addition in two learning environments. *Learning and Instruction*, 20, 205–215.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Publications.
- Fagginger Auer, M. F., Hickendorff, M., van Putten, C. M., Béguin, A. A., & Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data. Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, 29(2), 144–159. <http://doi.org/10.1080/08957347.2016.1138959>
- Fan, L. (2013). Textbook research as scientific research. Towards a common ground on issues and methods of research on mathematics textbooks. *ZDM – The International Journal on Mathematics Education*, 45(5), 765–777. <http://doi.org/10.1007/s11858-013-0530-6>
- Fan, L., Zhu, Y., & Miao, Z. (2013). Textbook research in mathematics education: Development status and directions. *ZDM – The International Journal on Mathematics Education*, 45(5), 633–646. <https://doi.org/10.1007/s11858-013-0539-x>
- Grønmo, L., Lindquist, M., Arora, A., & Mullis, I. (2015). TIMSS 2015 mathematics framework. Retrieved from https://timssandpirls.bc.edu/timss2015/downloads/T15_FW_Chap1.pdf
- Hadar, L. L. (2017). Opportunities to learn: Mathematics textbooks and students' achievements. *Studies in Educational Evaluation*, 55, 153–166.
- Hartung, T. (2014). Schulbuchauswahl und Lernmittelfreiheit in den deutschen Bundesländern im Kontext von Schülerpartizipation [Textbook choice and the free supply of educational aids in the German federal states in the context of student participation]. Eckert. Working Papers, 11, 1–15. Retrieved from <http://www.edumeres.net/urn/urn:nbn:de:0220-2014-00346>

- Heinze, A., Arend, J., Gruessing, M., & Lipowsky, F. (2018). Instructional approaches to foster third graders' adaptive use of strategies: An experimental study on the effects of two learning environments on multi-digit addition and subtraction. *Instructional Science*, 46(6), 869–891. <https://doi.org/10.1007/s11251-018-9457-1>
- Heinze, A., Marschick, F., & Lipowsky, F. (2009). Addition and subtraction of three-digit numbers. Adaptive strategy use and the influence of instruction in German third grade. *ZDM – The International Journal on Mathematics Education*, 41(5), 591–604. <http://doi.org/10.1007/s11858-009-0205-5>
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Quantitative methodology series. New York: Routledge.
- Kilpatrick, J., Swafford, J., Findell, B. (Eds.) (2001). *Adding it up*. Washington, D.C.: National Academies Press. <https://doi.org/10.17226/9822>
- Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades. Realistic versus gradual program design. *Journal for Research in Mathematics Education*, 29(4), 443–464.
- Knievel, I., Lindmeier, A. M. & Heinze, A. (2015). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education*, 13(2), 309–329. <https://doi.org/10.1007/s10763-014-9608-z>
- Koedel, C., Li, D., Polikoff, M. S., Hardaway, T., & Wrabel, S. L. (2017). Mathematics curriculum effects on student achievement in California. *AERA Open*, 3(1). Krammer, H. P. M. (1985). The textbook as classroom context variable. *Teaching and Teacher Education*, 1(4), 273–278. [http://doi.org/10.1016/0742-051X\(85\)90015-0](http://doi.org/10.1016/0742-051X(85)90015-0)
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change. *Contributions to children's learning of multiplication*. *Journal of Experimental Psychology: General*, 124(1), 83–96.
- Lorenz, J. H. (2007). *HaRet – Hamburger Rechentest für Klasse 1* [Hamburg calculation test for Grade 1]. Hamburg, Germany: Freie und Hansestadt Hamburg.
- Mannhaupt, G. (2013). *MÜSC – Münsteraner Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten* [Münster screening for an early detection of difficulties in reading and writing]. Berlin, Germany: Cornelsen.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Mullis, I. V. S., Martin, M. O., Foy, P., Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, Mass.: TIMSS & PIRLS International Study Center.
- Muthén L. K., & Muthén B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- National Research Council. (2002). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education*. In I. R. Weiss, M. S. Knapp, K. S. Hollweg, & G. Burrill (Eds.), *Committee on understanding the influence of standards in K-12 science, mathematics, and technology education*, center for education, division of behavioral and social sciences and education. Washington, DC: National Academy Press.
- Prather, R. W., Alibali, M. W. (2009). The development of arithmetic principle knowledge: How do we know what learners know? *Developmental Review*, 29(4), 221–248.
- Rinkens, H.-D., Hönisch, K., & Träger, G. (Eds.). (2011). *Welt der Zahl 1*. Braunschweig: Schroedel.
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561–573.

- 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), 529–543.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.
- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics (Vol. 1)*. Dordrecht: Kluwer.
- Schneider, W., Krajewski, K., & Küspert, P. (2013). *Die Entwicklung mathematischer Kompetenzen [The development of mathematical competence]*. Paderborn, Stuttgart: Schöningh; UTB GmbH. Retrieved from <http://www.utb-studie-book.de/9783838538990>
- Siegler, R. S., DeLoache, J. S., & Eisenberg, N. (2006). *How children develop*. New York, NY: Worth Publ.
- Sievert, H., van den Ham, A.-K., Niedermeyer, I. & Heinze, A. (2019). Effects of mathematics textbooks on the development of primary school children's adaptive expertise in arithmetic. *Learning and Individual Differences* 74, 101716, 1-13. doi:10.1016/j.lindif.2019.02.006.
- Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319–370). Charlotte, CT: Information Age.
- Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodriguez, M., & Reckase, M. (2012). Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M). Amsterdam: International Association for the Evaluation of Educational Achievement. Retrieved from <http://www.iea.nl/teds-m.html>
- Törnroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, 31(4), 315–327.
- Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009a). Jump or compensate? Strategy flexibility in the number domain up to 100. *ZDM – The International Journal on Mathematics Education*, 41(5), 581–590.
- Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009b). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics*, 71(1), 1–18.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen, & K. Elliot (Eds.). *But what does it mean? The use of effect sizes in educational research* (pp. 55 – 66). London: National Foundation of Educational Research.
- Valverde, G., Bianchi, L. J., Wolfe, R., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht: Kluwer Academic Publishers.
- van den Ham, A.-K. & Heinze, A. (2018). Does the textbook matter? Longitudinal effects of textbook choice on primary school students' achievement in mathematics. *Studies in Educational Evaluation*, 59, 133-140. doi:10.1016/j.stueduc.2018.07.005.
- van Steenbrugge, H., Valcke, M., & Desoete, A. (2013). Teachers views of mathematics textbook series in Flanders. Does it (not) matter which mathematics textbook series schools choose? *Journal of Curriculum Studies*, 45(3), 322–353.
- Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Charlotte, CT: Information Age.

Weiß, R. H., & Osterland, J. (2013). CFT 1-R – Grundintelligenztest Skala 1 Revision [basic cognitive abilities test scale 1 revision].
Göttingen, Germany: Hogrefe.

Wright, V. (2009). Mathematics standards for years 1-8. Wellington, N.Z.: Ministry of Education.

C Artikel zu Studie 3

Dieser Artikel wurde in der vorliegenden Fassung am 29.07.2020 erstmals bei einer Fachzeitschrift eingereicht und befindet sich im Begutachtungsverfahren.

What Role Does the Textbook Play When First Graders Solve Quantitative Comparisons? A Longitudinal Study

Henning Sievert, Ann-Katrin van den Ham, Aiso Heinze

Abstract: Students' need to create mental models for different types of addition and subtraction situations to develop a broad and viable understanding of these operations. While most students succeed when changing or combining sets, situations demanding a quantitative comparison of sets seem considerably more difficult in the first school years. Textbooks represent the most important learning resource for primary school mathematics teachers. However, research on their impact on students' knowledge is limited. Hence, we examined the textbooks' role for the students' ability to model quantitative comparisons by analyzing the learning opportunities presented by four German textbooks for Grade 1, and conducting a secondary analysis on a longitudinal dataset of 1513 students from 84 classes, who used one of these textbooks. The results reveal differences in the textbooks' topic-specific instructional quality as well as a significant relation of this quality to students' achievement in quantitative comparisons.

Keywords: Quantitative comparisons; Textbook quality; Primary school; Textbook effects

1. Introduction

A main focus of mathematics instruction during the first years of school entails students' obtaining a comprehensive understanding of addition and subtraction (CCSSI 2010; KMK 2005; MoE 2011). Hence, they must learn how to translate different situations representing an addition or subtraction (semantics) into a mathematical problem (syntax), and also assign these situations to a given problem (Kintsch and Greeno 1985). These addition and subtraction situations are referred to as word problems and typically classified as: change (add to/take from), combine, equalize, and compare (e.g., Van de Walle et al. 2016). However, empirical research suggests that usually only the first three types of problems occur in class or textbooks. In contrast, comparison problems are often neglected or even ignored (Despina and Harikleia 2014; Selter et al. 2012; Tarim 2017). Comparison problems are also considerably more difficult for students in the beginning grades (Stern 1993). This might stem from a lack of suitable learning opportunities.

Textbook constitutes the most important learning resource in primary school mathematics class (Mullis et al. 2012). It frames the teaching activities and learning opportunities provided accordingly and is considered enacted curriculum (Stein et al. 2007). Previously, studies provided first evidence for relations between mathematics textbooks used and the student's topic-

specific achievement (e.g., Schmidt et al. 2001; Sievert et al. 2019a, 2019b; Törnroos 2005). Regarding comparison problems, first hints indicate the textbook might affect students' achievement (Despina and Harikleia 2014; Tarim 2017; Xin 2007), but current evidence is based on studies with cross-sectional designs and small samples sizes. In view of the relevance of the ability to solve comparison problems for further mathematics learning (Selter et al. 2012), a possible influence of the textbook on the learning success of the students would underline the importance of the role of the textbook as an instrument for change in arithmetic lessons.

Reviewing the state of textbook research, Fan (2013) concludes a lack of relational and causal studies. Accordingly, in this study we examined the relation of textbooks and students' ability to solve comparison problems in Grade 1. Our study is based on a secondary analysis of a data set having 1513 students from 84 school classes during the first seven months of Grade 1. We conducted a theory-based analysis of the four textbooks used in this sample to determine their quality regarding comparison problems. Subsequently, we carried out a multilevel analysis to investigate the relation of the resulting textbook quality to the students' ability to solve comparison problems, while controlling for relevant covariates on individual and class level.

1.1. Quantitative Comparisons

To develop a comprehensive understanding of addition and subtraction of whole numbers, students must learn to make sense of different addition and subtraction situations. They should be able to comprehend those situations and translate them into a mathematical problem (the mathematical model), and conversely assign them to a given problem (Kintsch and Greeno 1985). As translator between these two – (real world) word problem and mathematical problem – students must hence create mental models as internalized ideas of an action for different kinds of problems or situations (Riley et al. 1983). Several classifications of such word problems exist with different subcategories (e.g., Carpenter and Moser 1984; Verschaffel et al. 2007; Van de Walle et al. 2016), but all of them distinguish between the situations *change* (add to, take away from), *combine* (part-part-whole), *equalize* (take from/add to difference), and *compare* (difference). While change and equalize situations are characterized as dynamic (i.e., an action is inherent in the situation), the situations combine and compare are described as static, without proposal for an action (Riley et al. 1983). Further, it can be differentiated which quantity of the situation or the mathematical model is unknown (Nunes et al. 2016).

From the four types of situations, comparisons pose the greatest difficulties for students across countries (Nunes et al. 2016; Riley and Greeno 1988; Stern 1993). A typical example is: *“Paul has 3 marbles, Mila has 5 marbles. How many more marbles does Mila have than Paul?”* We refer to these problems as quantitative comparisons (QC¹) to distinguish them from qualitative comparison (*“Paul has 3 marbles, Mila has 5 marbles. Who has more?”*), since the required abilities are distinguishable (Obersteiner et al. 2013). QC are more difficult than other types of problems although they may be modeled by an equivalent mathematical model. This may be due to the static character of a comparison: If the problem itself does not imply an action, to

¹ For ease of reading we use the abbreviation QC for both singular and plural.

solve it the students must either generate one themselves to create a mental model for the problem, or activate a suitable model (like equalizing). However, the static character alone does not explain the difficulty since the second static situation type *combine* is considerably easier for students. Another reason given is the relational nature of the difference in these problems: In contrast to the two above compared sets the solution of the example task given (*2 marbles*) does not describe a concrete, conceivable subset of the given sets but a relation between two other sets. Stern (1998) describes the ability to understand numbers as ratio between two sets as a concept students must develop over time. A third reason that particularly applies to Grade 1 could be problems in text comprehension if the problem is presented in written form (Gabler and Ufer 2020; Stern 1993). An example is signal word strategies where the operation is derived directly from the problem wording (“less” – subtraction, “more” – addition, Stern 1993). Stern (1993) concludes that a lack of understanding the linguistic symmetry of QC (A has two more than B \Leftrightarrow B has two less than A) and thus a missing flexibility in dealing with QC are possible influencing factors concerning text comprehension in addition to signal word strategies. One more reason, inevitably connected to the previous, is a lack of learning opportunities for comparison problems in class. Several textbook analyses revealed an under-representation of QC – or even no representation – in Grade 1 textbooks, and thus an overrepresentation of the other problem types (e.g. Despina and Harikleia 2014; Tarim 2017; Wessel 2015).

In former studies researchers repeatedly reported the long-term value of early numerical competence (e.g. Reeve et al. 2012; Selter et al. 2012) and thus strengthened the need for a broader dealing with operations in all aspects, including QC, from the beginning. However, although the difficulties are known and have been studied for more than three decades, the question for effective learning environments still remains unsatisfactorily answered (Huang et al. 2019). Nonetheless, some exploratory studies give insights how to address this issue. Stern (1993) attaches importance to linguistic flexibility, particularly the insight into the equivalence of “more than/less than” statements. Accordingly, Schumacher and Fuchs (2012) could show that students from an intervention focusing on word-problems outperformed students from two other groups, focusing on calculation and taught conventionally in solving QC. This intervention effect was partially mediated by the students’ understanding of relational terminology. Furthermore, Mwagi and Sweller (1998) report that studying worked examples conducive for students solving QC, particularly if these examples integrate different representations. Correspondingly, Múñez and colleagues (2013) also report an advantageousness of additional representations for the construction of mental models for QC. Similarly, the study of Huang et al. (2019) indicates value of lessons based on a combination of an assumed learning trajectory and variation pedagogy for solving QC, which inter alia focuses on using multiple representations. While Múñez et al. (2013) studied students from secondary school, the other studies considered students from grades 1–3.

Solving a QC requires several steps which we outlined in an idealized form for the example “Paul has 3 marbles, Mila has 5 Marbles, how many more marbles does Mila have than Paul?” (cf. Kintsch and Greeno 1985): First, students must build a situation model, that is, they need to gather the problem and its context from the text: *Two people named Paul and Mila both*

possess a certain number of marbles, Paul 3 and Mila 5. The difference is unknown and sought. Second, on this basis they need to create a mathematical model by abstracting from the concrete persons and objects (marbles) and considering the quantities 3 and 5. The part-whole relation between these quantities, i.e., the subset relationship between the two sets represented, has to be identified and established as does the meaning of the difference for solving the problem. The cardinality of the difference must be modelled by a corresponding addition or subtraction problem and its solution interpreted in the situation model. Although this description is small-step and some parts of it probably happen implicitly with most students, it gives first insights about abilities students need to acquire to solve QC.

1.2 Textbooks as learning resources

The textbook is one of the most relevant learning resource for teachers and students in mathematics class. It is conceptualized as a mediator between the official curriculum and the curriculum implemented by teachers, since it translates the abstract curriculum into concrete operations, which teachers and students can carry out (Valverde et al. 2002). Specific data on mathematics textbook use in schools were collected for TIMSS 2011 which revealed that most primary school teachers use textbooks as a basis for their instruction - even 86 % of the teachers in Germany, where the data of our study were collected (Mullis et al. 2012)². Few quantitative empirical studies exist examining the relation between mathematics textbooks and teachers' instruction (Krammer 1985; Schmidt et al. 1997, 2001). Most of them were conducted in secondary school. They report a relation between textbooks and teaching practice and lesson content. For instance, topics not covered in a textbook are unlikely to be covered in class (Schmidt et al. 1997). Also, indications exist for an effect of textbooks on the students' performance for both primary and secondary school (Schmidt et al. 2001; Sievert et al. 2019a, b; Törnroos 2005; van den Ham and Heinze 2018). Conversely, van Steenbrugge et al. (2013) did not find a relation between the mathematics textbook used and the students' achievement. Several more studies on textbook effectiveness exist, in particular from North America, not reported here. Due to differences in educational systems, these studies consider textbooks and curricula as synonyms (e.g., Blazar et al. 2019) and thus can be understood as curriculum studies (or at least the textbook is confounded with the curriculum). In the study presented here, however, we examine differences in the effects of textbooks following the same curriculum. Thus, our goal is to compare different realizations of one curriculum in the form of a textbook.

Looking at QC, first studies indicate a relation between the mathematics textbook used and students' performance. Despina and Harikleia (2014) analyzed the distribution of different types of word problems in Greek primary school textbooks for Grade 1 and 2 as well as the solutions of 80 students from first and second grade for these problem types. Their results show QC are underrepresented in the textbooks analyzed, and the students performed better on the problem types overrepresented in the textbooks than on those underrepresented or

² However, in a more recent study, Blazar and colleagues (2019) showed that teachers' general textbook use can still mean quite a variance in frequencies.

missing. Similarly, Tarim (2017) studied the distribution of word problems in six Turkish textbooks (grades 1–3) as well as the performance of 158 Turkish third-graders with such problems. Accordingly, her results show an underrepresentation of QC in the textbooks and greater student difficulties with these problems. Likewise, Xin (2007) examined one US and one Chinese textbook series as well as the performance of 111 students from the USA and China with multiplicative comparisons. She identified an unbalanced distribution of word problem types in the US textbook, in contrast to the Chinese one, and a corresponding unbalanced achievement in the different item types for US students. These cross-sectional studies indicate the textbook might influence students' performance with QC problems. We want to extend this research and to examine the relation of the textbooks' quality for the topic QC for students' performance in a longitudinal design with a large sample while also controlling for relevant covariates.

1.3 Present Study

We examine the leverage of the mathematics textbook as a learning resource regarding first graders' ability to solve QC. The added value of this analysis to current research is its large-scale longitudinal database of 1513 students from 84 classes over the first seven months of schooling. All textbooks used in this dataset follow the same statewide curriculum, thus relations between textbooks and student performance are not confounded by varying curricula. Building on the approaches of Törnroos (2005) and Sievert et al. (2019a, 2019b), we analyzed the learning opportunities in the four different textbook series of our sample. Based on the research results presented in Section 1.1 and on discussions with experts from mathematics education and school practice, we conceptualized textbook quality regarding QC. The obtained scale for textbook quality develops and refines the research of Despina and Harikleia (2014), Tarim (2017), and Xin (2007) on the relation between textbooks and students' ability to solve QC. Hence, we addressed the following research hypotheses:

Hypothesis on textbook quality

- (1) First grade textbooks that follow the same curriculum differ in the quality of their learning opportunities with respect to the different aspects of the subject quantitative comparisons.

Hypothesis on the relation between textbook quality and student performance

- (2) The quality of learning opportunities concerning quantitative comparisons presented in Grade 1 textbooks is related to students' ability to solve quantitative comparisons.

2. Method

2.1. Participants and design

The study is a secondary analysis of a dataset originally collected for a longitudinal evaluation of a mathematics support program for weak students. It was implemented within the regular

mathematics class³. The evaluation includes comparing a control group to two intervention groups in which teachers received additional material for weak students. This teaching material addressed aspects of basic arithmetic (e.g., concept of cardinal numbers, place-value system, concepts of the basic arithmetic operations) to identify and support low-achieving students. It also included material on the part-whole relations of whole numbers in general and the inverse relation of addition and subtraction. It did not address QC and only weak students were supported with this material. The classes that participated in the support program and which textbook they used are independent of each other. We controlled for possible effects of participation using dummy coded variables in the analysis (see Section 2.4). It did not show any effect on the students' performance in solving QC (see Section 3.2).

The original dataset involves longitudinal data of 2330 students of 40 schools and 127 classes in Schleswig-Holstein (state in Germany), throughout primary school. This is about 10 % of the cohort in this federal state. The schools were selected from rural and urban areas all over the federal state. A state-wide obligatory curriculum stipulates content (e.g., number concept, concept and strategies of addition and subtraction), skills (e.g., switch between representation, split numbers), and forms of representations (concrete, iconic, symbolic) to be addressed in the mathematics classroom. Within this dataset, 93 classes used one of the four textbook series "Denken und Rechnen", "Einstern", "Flex und Flo" and "Welt der Zahl", distributed quite equally. Other classes used either no textbook or different ones but only in a few classes. For this study, 9 classes were excluded due to an insufficient number of students ($n < 5$) with data for Grade 1. Further, we excluded 119 students because they did not participate in the QC test. Thus, the sample of this study consists of 1513 students (720 female, 48 %; 11 with no data) with a sufficient number of students for each of the four textbooks (Denken und Rechnen: 351; Einstern: 344, Flex und Flo: 485; Welt der Zahl: 333).

The four textbook series are similar in their content structure and follow the same curriculum. The series "Denken und Rechnen" and "Welt der Zahl" include all topics of one school year within one book, while "Flex und Flo" and "Einstern" are divided into three and six books, respectively. All of them cover the same topics for Grade 1 (e.g., beginning with numbers, operations, geometry, patterns). For each textbook the vast majority of teachers in the sample (90.8–100 %) stated using the textbook at least once a week for arithmetic instruction. Furthermore, 81.1–94.1 % reported a general orientation towards the content, methods and chronology of the textbook. Officially, the textbook choice for primary school in Schleswig-Holstein is made on the school level and not by individual teachers. Little is known about the criteria used (Hartung 2014). We analyzed whether textbook choice depended on teacher qualification by testing the distributions of textbooks between teachers who studied mathematics and out-of-field teachers (which was not the case, see Section 3.2).

³ The studies of van den Ham and Heinze (2018) as well as Sievert et al. (2019a, 2019b) used data from this data set as well. The differences in the subsamples of these studies go back to different test instrument administered at different points of time. The studies focused on different aspects of arithmetic and partly children in different grades.

All data were collected by an authority subordinated to the Ministry of Education of Schleswig-Holstein, assuring the official human subject related approval. We received all raw data fully pseudonymized and had no contact to students or teachers. The compliance of the human subject related guidelines were monitored by the Data-Protection Supervisor of Schleswig-Holstein.

2.2. Measures

2.2.1. Textbook quality with respect to QC

We developed a measure for the quality of learning opportunities concerning QC presented in the textbooks to analyze textbook quality and to test Research Hypothesis 1. In contrast to previous studies (Sievert et al. 2019a, 2019b), we were unable to build upon an established model, since the question for effective QC learning opportunities remains insufficiently answered (Huang et al. 2019). Hence, we evolved criteria from the present state of research and discussed this category system with expert teachers from school practice and researchers from mathematics education. Both groups gave us valuable suggestions and feedback which helped us revise and extend the criteria. Five categories represent the resulting system for the assessment of the learning opportunities: decomposition of numbers with missing part, complementarity of addition and subtraction, subtraction as difference, modeling of QC, and impulses for QC in the teacher manual. We will elaborate the precise criteria and their scoring in the following.

Category “decomposition of numbers with missing part”: As shown in the modelled solution process (Section 1.1), insight in the subset relation and the associated idea that whole numbers can be decomposed into smaller numbers, is crucial for solving QC. In addition, just as in QC, *decomposed numbers with a missing part* usually have a static character (see Fig. 1). We excluded number decompositions with missing whole as they stimulate the mental model for the “combine situation”.

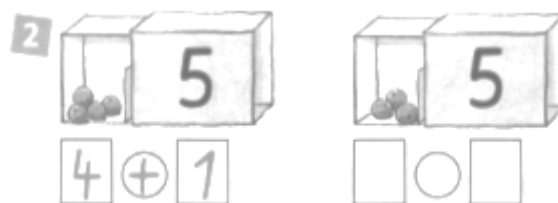


Figure 1: Example for the category decomposed numbers with a missing part (Buschmeier 2011)

To assess textbook quality in this category, we first identified textbook pages presenting learning opportunities for *decomposed numbers with a missing part* and divided these into the three subcategories introduction, practice and repeat. For each of these subcategories we then analyzed the learning opportunities and scored one point for each of the following criteria:

- a request for students to use manipulatives
- an iconic representation
- a symbolic representation
- a linguistic explanation or impulse (e.g., explain, exchange with others)
- a meaningful linking between these levels of representation.

These criteria take into account former research on the importance of linguistic flexibility (Schumacher and Fuchs 2012; Stern 1993) as well as on the meaning of multiple representation and their integration (Huang et al. 2019; Múñez et al. 2013; Mwagi and Sweller 1998). For each of the subcategories introduction, practice and repeat we assigned 0–5 points. Each textbook could reach 0–15 points in the category *decomposed numbers with a missing part*.

Category “complementarity of addition and subtraction”: The insight into the *complementarity of addition and subtraction* presents two ways to determine a difference (see Section 1.1). It also represents the equivalence of “more than/less than” in a mathematical model whose understanding was shown important for solving QC (Stern 1993). Examples for corresponding learning opportunities are representations of the invertibility of the operations (Fig. 2) or word problems with two possible interpretations. The scoring is analogous to the previous category (three subcategories, each scored 0–5, possible range of 0–15 points), since the understanding of such relations should be fostered by providing multiple representations as well as their linkage.



Figure 2: Example for the category complementarity of addition and subtraction (Buschmeier 2011)

Category “subtraction as a difference”: To understand that the result of a subtraction can not only represent the rest of a take-away action but also the difference between the two quantities dealt with is fundamental for a subtractive approach to QC (Selter et al. 2012). Moreover, the concept of the difference set as relation between two sets is key in the solution process of a QC (see Section 1.1). Thus, we rated whether the idea of *subtraction as a difference* is addressed in the textbooks. Again, the scoring scheme is the same as for the previous two categories. An example is given in Figure 3.

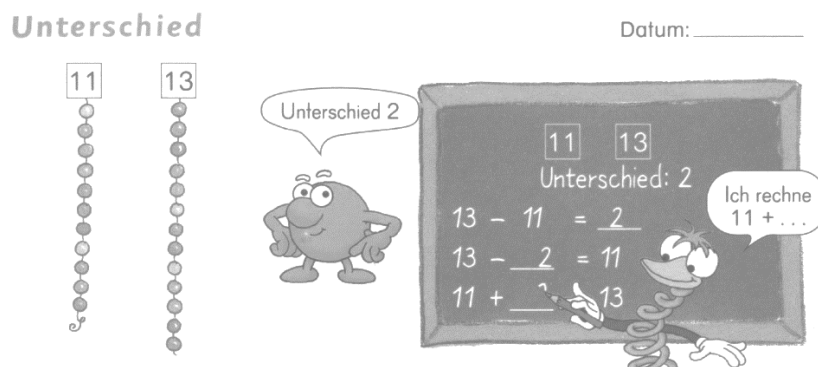


Figure 3: Example for the category subtraction as difference ("Difference" / "Difference 2" / "I calculate 11 + ...", Brall 2010)

Category "modeling of QC" Dealing with QC word problems itself is of course an important learning opportunity for this field (Schumacher and Fuchs 2012). In textbooks they are typically presented in written form or as pictures to interpret. We rated whether the textbooks contain such learning opportunities in the category *modeling of QC*. Since they are presented in one representation which must be translated and put into symbolic form, we did not account for different representations and their linkage in this category. Instead, we used the same three subcategories (introduction, practice, repeat) again, and scored whether QC problems are provided in each of them, leading to a possible range of 0–3 points in this category.

Category "impulses for QC in the teacher manual": Most learning opportunities in German Grade 1 textbooks are given nonverbally since the children are unable to read at this stage. While studying the Grade 1 textbooks, we recognized a series of pictures which could potentially be interpreted in a QC way but usually the context or task stimulates other interpretations, like take-away. Moreover, Figure 4 shows a multiple choice example task, in which the right task should be assigned to an image, while the others are deemed to be incorrect. The (potential & correct) comparative interpretation here is misrepresented as false. Hence, we decided to examine the teacher manual for possible indications in favor of QC. We did a partial credit scoring in which we decided whether such impulses are not given (0 points), QC impulses are given as a differentiation task for high achievers (1 point) or QC impulses are given for discussion with all students (2 points).

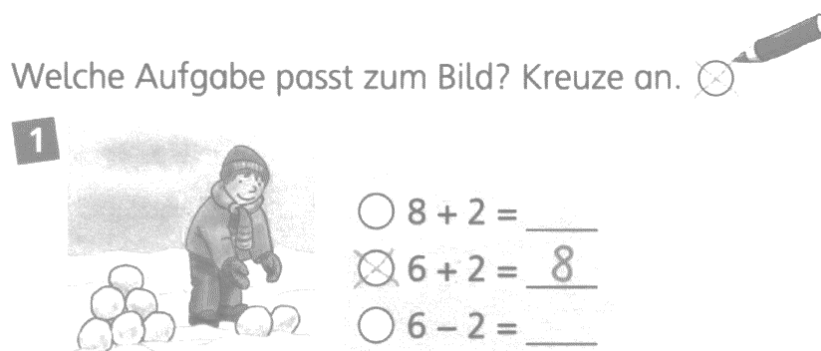


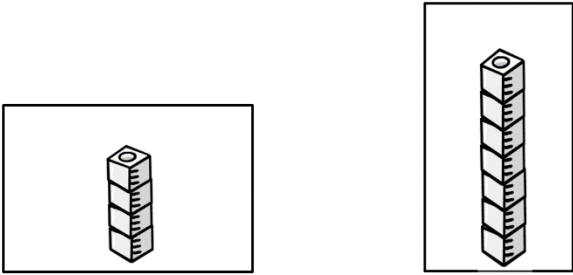
Figure 4: Comparative interpretation as wrong answer ("Which task fits the picture? Mark with an x.", Buschmeier 2011)

The five category scores were compiled to one textbook quality scale by weighting all categories equally (see Section 2.3.1).

2.2.2. Students' ability to solve QC

We assessed the students' ability to solve QC using students' solutions for QC items from an arithmetic test administered after seven months of schooling. The two item types both included a QC word problem and were identically structured, except that the second type also contained an iconic representation. See Figure 5.

Tom hat 4 Steckwürfel. Anna hat 7 Steckwürfel.



Tom

Anna

Wie viele Steckwürfel hat Anna mehr als Tom?

Schreibe die Aufgabe dazu auf.

Figure 5: Example for the second item type ("Tom has 4 cubes. Anna has 7 cubes" / "How many more cubes does Anna have than Tom?" / "Write down the corresponding problem.")

All items contained static QC problems with unknown difference. Students were presented a problem, asked for the difference, and then asked to write down a corresponding problem. The formulation direction was always how many "more". The test contained three items with and three without iconic representation. For each item we scored 1 point for a correct difference and 2 points for a correct difference in combination with an appropriate modelling. In the German mathematics class the term "Aufgabe" (=problem) in this item context is connoted with a symbolic expression. As appropriate modelling we accepted each problem that can be used to determine the correct difference, including solutions like " $7 - 4$ ", " $4 + _ = 7$ ", " $7 - ? = 4$ ", or " $4 + 3 = 7$ ".

The scores for the six items were totaled, leading to a scale from 0–12 points with a good reliability (Cronbach's $\alpha = .90$). It was standardized and used as a measure for students' ability to solve QC.

2.2.3. Individual and classroom covariates

We included a series of available individual and classroom-related covariates in our analysis to control for possible influencing factors. Various studies provide evidence regarding an influence of students' abilities at school entry on the development of their mathematics skills in primary school (e.g., Schneider et al. 2013). Hence, we controlled for students' basic cognitive abilities, basic numerical skills and German language skills at the beginning of Grade 1. The data were collected by approved standardized tests with sufficient reliabilities at school entry (basic cognitive abilities: CFT 1-R, Weiß and Osterland 2013, Cronbach's $\alpha = .91$; basic numerical skills: HaReT 1, Lorenz 2007, Cronbach's $\alpha = .74$; German language skills: Münsteraner Screening, MÜSC, Mannhaupt 2013, Cronbach's $\alpha = .72$). CFT 1-R included 45 items for non-verbal problem solving, HaRet 1 included 43 items for arithmetic-related precursory skills at school entrance, and MÜSC included 40 items testing students' phonological awareness. The standardized school entry tests were scored as defined in the test manuals. Section 3.1 shows the results for the different textbook groups.

On class level participation in one of the three groups of the arithmetic support program (see Section 2.1) were included as dummy variables. We further controlled for class composition effects by including the aggregated variable of basic numerical skills (HaReT 1), since differences between classes could also be caused by different levels of students' average ability. Aggregated basic cognitive abilities and German language skills were excluded due to multicollinearity which causes distorted estimates. We used mathematics-specific qualification as an indicator to control for teacher qualification for this covariate (as dummy variable indicating whether they studied mathematics or not). Primary school teachers in Germany are generalists, that is, all teachers teach mathematics but only some have studied mathematics. Different studies with primary school teachers revealed that the formal teacher qualification is strongly related to mathematics-specific professional knowledge (e.g., Blömeke et al. 2010; Knievel et al. 2015).

2.3. Procedure

2.3.1. Textbook analysis

Three trained persons independently assessed the four textbooks for Grade 1 based on the first four categories and following the scoring presented in Section 2.2.1. Their results indicated a substantial agreement (Fleiss' $\kappa: .74$, ICC: $.90$). Subsequently, we applied the consensus method for a uniform scoring of the four textbook categories. The teacher manuals were reviewed for the fifth category by one trained person. The three QC impulses for the teachers (none, only for high achievers, for all students) were given explicitly in writing so an interpretation was

unnecessary (low-inference rating). To weigh all five categories equally, the resulting scores of each were standardized and aggregated to a single standardized quality scale for the textbooks' learning opportunities with respect to QC.

2.3.2. Longitudinal study

Data collection began when the students entered primary school. Abilities at school entry were assessed in the beginning of Grade 1. The QC items were administered in March. Trained test administrators collected the data in accordance with the respective test manuals. All instructions were read out aloud. The administrator worked an example with the students for the QC items. At the end of Grade 1 teachers were given two questionnaires to collect data on their qualification and textbook use.

2.4. Data analysis

To address Research Hypothesis 1, we analyzed differences in textbook quality with respect to QC based on the category system by comparing and interpreting the scores of the four textbook series.

We applied multilevel random intercept models with the software Mplus 7.0 (Muthén and Muthén 1998–2012) to analyze the relation of textbook quality to students' performance (Research Hypothesis 2), while accounting for the nested data structure (Hox 2010). Model 0 without predictors (null model) was run to estimate the partition of variance between and within classes. Model 1 included individual and class composition covariates: students' basic cognitive abilities, basic numerical skills and German language skills at individual level to account for individual characteristics at school entry and students' basic numerical skills aggregated to a class mean at group level to control for class composition. Model 2 additionally included the classroom covariates participation in the support program (see Section 2.1) and teacher qualification. Model 3 included the textbook quality scale concerning QC. We could not include the five subscales of textbook quality separately due to multicollinearity.

The scores for the abilities at school entry, the aggregated class mean of basic numerical skills, and the dependent variable for students' ability to solve QC were standardized so the corresponding β -coefficients can be interpreted as effect sizes similar to Cohen's d (Tymms 2004). A full-information-maximum-likelihood (FIML) approach was applied for missing data on the independent variables. FIML combines missing data and parameter estimation in a single step and uses all the available information (Enders 2010). Due to sample selection there was no missing data for textbooks quality, support program and students' performance in QC.

3. Results

3.1. Quality of learning opportunities in textbooks regarding QC

Table 1 presents the results for the textbook quality concerning QC.

	Decomposition of numbers	Complementa- rity of add/sub	Subtraction as difference	Modeling of QC	Teacher manual	Overall score
Denken und Rechnen	0.82	-0.12	-0.93	0	-0.30	-0.18
Einstern	-1.63	0.35	-0.93	0	-1.51	-1.28
Flex und Flo	0.00	-1.50	0.42	0	0.90	-0.06
Welt der Zahl	0.82	1.27	1.44	0	0.90	1.52

Table 1: Results of the textbook analysis: z-scores for the five categories and the overall score (mean)

None of the four textbooks contains learning opportunities for modeling QC. Despite being surprising, it is consistent with earlier research (e.g., Despina and Harikleia 2014; Wessel 2015). Apart from this category, there is one series with the highest scores in all other categories, which is “Welt der Zahl”, and one series with the lowest score in all but one (“Einstern”). The overall score reflects this accordingly. In contrast, the other two series show specific strengths and weaknesses which mainly take place between the results of the aforementioned. While “Denken und Rechnen” shows the same score as “Welt der Zahl” in the *decomposition* category, “Flex und Flo” does so in the *teacher manual* category. Conversely, “Flex und Flo” has the lowest score in the category *complementarity* and “Denken und Rechnen” in the category *subtraction as difference*. Thus, both show a mid-range overall score. Summarizing, there is one textbook series with high and one with low quality in terms of QC, and there are two with medium quality concerning QC due to fluctuating scores in the categories.

Compared to the textbook quality determined for arithmetic principles for the same textbooks (Sievert et al. 2019a), the rankings vary widely. For instance, in representing arithmetic principles, the series “Einstern” showed the highest score. Hence, these results indeed reflect a topic-specific and not a general textbook quality.

3.2. The relation between textbook quality and students’ ability to solve QC

Table 2 shows the means and standard deviations of all covariates for the four textbook groups. We ran analyses of variance and subsequent post-hoc tests which showed significant differences for some covariates between the textbook groups with small effect sizes of $\eta^2 \leq .024$. Hence, it makes sense to include these covariates in the multilevel analysis.

	Denken und Rechnen	Einstern	Flex und Flo	Welt der Zahl
Basic cognitive abilities (0–45)	26.32 (8.62)	25.43 (8.68)	27.54 (9.23)	28.56 (9.04)
Basic numerical skills (0–43)	31.32 (7.01)	31.88 (6.74)	31.90 (7.19)	32.50 (6.57)
German language skills (0–40)	32.77 (6.09)	33.27 (5.92)	33.34 (6.20)	34.46 (5.41)
Basic numerical skills (aggregated) (0–43)	31.29 (1.76)	31.90 (2.41)	31.92 (2.91)	32.52 (3.27)
Support program participation (0–1)	0.82	0.77	0.95	0.63
Teacher qualification (0–1/ 0–1)	0.44 (0.50)	0.43 (0.50)	0.34 (0.47)	0.55 (0.50)

Table 2: Means and standard deviations of the covariates for the different textbook groups (theoretical range below the variable name)

The between-class variance in students' performance in solving QC as indicated by the null model is 13.7 %. Thus, students' ability to solve QC differ depending on the class students attended. Table 3 shows the results of the Models 1–3 (as described in Section 2.4).

Significant main effects of all three variables of the students' abilities exist on individual level. They explain about 20% of variance within classes. These effects are in line with current research on the meaning of students' abilities at the beginning of schooling for the development of mathematics competence (e.g., Schneider et al. 2013). The results also confirm this for QC.

On class level, the composition according to the aggregated numerical skills has no effect. Neither are there effects of participation in the support program, so that the analysis of textbook correlation is not influenced by the intervention. Also the teacher qualification shows no significant effects. The relation of textbook quality to the students' ability to solve QC is significant. For an interpretation of the coefficient ($\beta = .09$), we refer to the results of Table 1 showing a range of the textbook quality scale from -1.28 to 1.52 (z-scores). Students in classes with the lowest textbook quality and students in classes with the highest textbook quality on this scale differ in their ability to solve QC on average by about 0.25 standard deviations. Including the textbook quality leads to a considerable increase of explained variance between classes ($\Delta R^2 = 8.1$ percentage points). Regarding Research Hypothesis 2 we found a significant relation of textbook quality to the students' ability to solve QC.

	Model 1	Model 2	Model 3
Level 1 (students)			
Basic cognitive abilities	.25** (.03)	.25** (.03)	.25** (.03)
German language skills	.15** (.02)	.14** (.02)	.14** (.02)
Basic numerical skills	.11** (.03)	.11** (.03)	.11** (.03)
Level 2 (class)			
Basic numerical skills (aggregated)	.08 (.05)	.08 (.05)	.08 (.05)
Support program Group 1		.05 (.11)	.09 (.10)
Support program Group 2		.14 (.11)	.18 (.10)
Teacher qualification		-.11 (.08)	-.13 (.07)
Textbook quality			.09** (.03)
<i>Intercept</i>	<i>0.01 (0.04)</i>	<i>-0.01 (0.09)</i>	<i>-0.35 (0.09)</i>
Explained within class variance	20.3%	20.3%	20.2%
Explained between class variance	8.3%	13.2%	21.3%

Table 3: Results of the multilevel analysis of students' ability to solve QC (standard errors in brackets, all variables are standardized except Support program Group 1 & 2, and Teacher qualification (0,1), * $p < .05$, ** $p < .01$)

4. Discussion

In this analysis we studied the textbook's role as a learning resource on student achievement in solving QC in Grade 1. For this, we developed a set of indicators of textbook quality for this topic by combining current state of research and expert talks. Based on the resulting category system we analyzed the respective learning opportunities presented in four commonly used mathematics textbooks following the same arithmetic curriculum in a German state. We found differences in textbook quality concerning QC. In a subsequent multilevel analysis of students' performance in QC on a seven-month longitudinal dataset of 84 classes we could show a significant and relevant relation of the textbook quality to student performance.

4.1. Impact for educational research

Our findings suggest textbook quality as a relevant predictor for mathematics teaching and learning. Earlier textbook research (see Section 1.2) and our own teacher questionnaire data (see Section 2.1) suggest most arithmetic problems in mathematics class stem from or are caused by the chosen textbook. This study suggests the quality of textbooks might affect the learning environments teachers implement in their lessons. This holds true particularly when textbooks and teachers follow the same curriculum. Hence, one should consider the textbooks when investigating teaching effectiveness.

Regarding mathematics teachers' education, results propose teachers should recognize quality differences between textbooks and compensate for poor quality within a textbook. Thus, they should be able to identify a textbook's weaknesses and provide students with alternative learning opportunities as part of the teachers' pedagogical content knowledge. Consequently,

further research on the role of teachers' professional knowledge for textbook choice and use is necessary.

Concerning the understanding of operations, this study underlines the meaning of a broad instruction of addition and subtraction using the whole spectrum of different situation models. Although required by curricula and research (CCSSI 2010; KMK 2005; MoE 2011; Reeve et al. 2012), QC problems still appear underrepresented in classrooms when learning addition and subtraction (Despina and Harikleia 2014; Tarim 2017; Xin 2007). Based on data from the regular mathematics classroom, our findings indicate a lack of appropriate learning opportunities in several textbooks for Grade 1, and simultaneously a relation between textbooks and student achievement in solving QC. The results point out that the lower solution rates of QC problems are not only due to a possible inherent difficulty of the problem type but also due to missing learning opportunities in textbooks, which limit those provided in class.

Furthermore, this study contributes to research on the relation of mathematics textbooks' to student achievement. The results provide evidence regarding a relation of textbook choice to students' performance in primary school, in line with findings of Schmidt et al. (2001) and Törnroos (2005) for secondary school students. We developed specific criteria for textbook quality with respect to QC. The development of these criteria was based on empirical research on QC solving and on discussions with experts. The criteria are evident in the sense that they can explain an observed relation of textbooks to students' learning outcomes. The results show that a fine-grained, theory-based content analysis of mathematics textbooks is a rewarding approach to examine possible effects of these learning resources. Hence, this study meets the need for more relational textbook research as identified by Fan (2013) and extends former research on the meaning of topic-specific textbook quality by Sievert et al. (2019a, 2019b). A global measure of textbook quality is certainly a complex endeavor for future research, however, as these studies reveal, it is possible to identify relations between learning opportunities presented by textbooks and students' achievements for the subdomain QC. This could be a useful contribution to research on the quality and impact of mathematics textbooks. It is conceivable that this approach can be transferred to more mathematics topics to learn more about the strengths and weaknesses of textbooks.

4.2 Impact for educational practice

Concerning QC, the analysis revealed substantial differences in the quality of learning opportunities presented in four textbooks following the same curriculum and a subsequent relation of these differences to the students' performance. Hence, this study provides evidence students' achievement not only depends on the curricula given, but also on the realization of a curriculum as a textbook. Since textbooks are an easily and cost-effectively changed classroom determinant the potential improvement in student achievement makes the textbook a valuable instrument for improvement. It can further be regarded as instrument of quality assurance in education.

As previously stated, teachers should be aware of quality differences between textbooks, and be trained in the assessment of textbook quality. Finally, topic-specific criteria for mathematics textbook quality provide tools for textbook authors and publishing houses to analyze the strengths and weaknesses of their textbooks. This is associated with opportunities to improve the textbook quality.

4.3 Limitations

Some limitations of this study might influence the results and their interpretation. The secondary analysis of an existing dataset not based on an experimental setting with respect to textbook distribution provides correlational relations but does not allow causal interpretations. It also affects the sample used in our study as well as the number of textbooks. Further, we were unable to include certain covariates of interest possibly interacting or confounded with textbook quality. This includes data on the textbook choice in schools, observational data about teachers' actual textbook use (instead of teacher reports), or data on teachers' professional knowledge beyond their formal qualification. All these variables might influence the textbooks' role and should be considered in future studies to investigate their relevance. The choice of textbooks is particularly crucial for this research field since little is known about the criteria schools or teachers apply. For the United States, this process is described as highly idiosyncratic and market-driven (National Research Council 2002). For Germany, Hartung (2014) presumes rash decisions influenced by advertisement of publishers or a close relationship between sales representatives of publishers and the schools. Hence, possible influence of specific school characteristics on textbook choice requires additional research. A second limitation is multicollinearity of the single scales of our category system. The data does not provide details about the weighting of the five categories for QC quality. Although the rankings partly vary among categories, testing all five scales in a single model is not possible, that is, we cannot draw conclusions about the relative meanings. A third limitation relates to the measure of solving QC which is also due to the secondary analysis. Of course, six items are not a comprehensive basis to assess the students' ability. The items are restricted to QC problems with an unknown difference set. A more precise analysis of the students' performance must also include problems with unknown compare or reference set (cf. Gabler and Ufer 2020), including their specific difficulties and requirements. However, our measure shows solid and reliable psychometric characteristics, although it is restricted to unknown difference problems, and the corresponding relation to the textbooks is explainable by the quality of learning opportunities. Finally, we want to mention that we analyzed the textbook editions used by the students of our dataset. The results in Table 1 thus might not apply to revised or new editions of the respective textbook series.

5. Conclusion

In this study we have revealed a significant relation of mathematics textbooks quality to student performance in solving quantitative comparisons in Grade 1. The results contribute to textbook research by offering and further developing an approach for theory-driven and topic-specific

quality criteria for textbooks. Our findings provide further evidence regarding the importance of textbook quality for mathematics education in primary schools. They are based on a large-scale and longitudinal dataset of students from the regular mathematics class that were taught by the same curriculum.

References

- Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S., & Kurlaender, M. (2019). *Learning by the book: Comparing math achievement growth by textbook in six Common Core states. Research report*. Cambridge, MA: Center for Education Policy Research, Harvard University.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2010). *Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich [TEDS-M 2008- international comparison of future primary school teachers' professional competence and opportunities to learn]*. Münster: Waxmann.
- Brall, C. (2010). *Flex und Flo 1*. Braunschweig: Diesterweg.
- Buschmeier, G. (2011). *Denken und Rechnen 1*. Braunschweig: Westermann.
- Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15, 179–202.
- Common Core State Standards Initiative (CCSS). (2010). *Common core state standards for mathematics*. National Governors Association and the Council of Chief State School Officers, Washington, DC. <http://www.corestandards.org/Math/Practice>. Accessed 3 June 2020.
- Despina, D. & Harikleia, L. (2014). Addition and subtraction word problems in greek grade a and grade b mathematics textbooks: distribution and children's understanding. *International Journal for Mathematics Teaching and Learning*, 8, 340–354.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Publications.
- Fan, L. (2013). Textbook research as scientific research. Towards a common ground on issues and methods of research on mathematics textbooks. *ZDM Mathematics Education*, 45(5), 765–777. doi:10.1007/s11858-013-0530-6.
- Gabler, L., & Ufer, S. (2020). Flexibilität im Umgang mit mathematischen Situationsstrukturen – Eine Vorstudie für die Entwicklung eines Förderkonzepts zum Lösen additiver Textaufgaben [Flexibility in dealing with mathematical situation structures - A preliminary study for the development of a support concept for solving addition word problems]. *Journal für Mathematik-Didaktik*. doi:10.1007/s13138-020-00170-3.
- Hartung, T. (2014). Schulbuchauswahl und Lernmittelfreiheit in den deutschen Bundesländern im Kontext von Schülerpartizipation [Textbook choice and the free supply of educational aids in the German federal states in the context of student participation]. *Eckert. Working Papers*, 11, 1–15. http://www.pedocs.de/volltexte/2015/11061/pdf/EWP_2014_11_Hartung_Schulbuchauswahl.pdf. Accessed 3 June 2020.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications. Quantitative methodology series*. New York: Routledge.
- Huang, R., Zhang, Q., Chang, Y., & Kimmins, D. (2019). Developing students' ability to solve word problems through learning trajectory-based and variation task-informed instruction. *ZDM Mathematics Education*, 51, 169–181.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92(1), 109–129.

- Knievel, I., Lindmeier, A. M. & Heinze, A. (2015). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education, 13*(2), 309–329. doi:10.1007/s10763-014-9608-z.
- Kultusministerkonferenz (KMK). (2005). *Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Mathematik für den Primarbereich [Resolutions of the standing conference of the ministers of education and cultural affairs of the federal states of Germany: Educational standards in mathematics for primary education]*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Mathe-Primar.pdf. Accessed 3 June 2020.
- Lorenz, J. H. (2007). *HaRet – Hamburger Rechentest für Klasse 1 [Hamburg calculation test for Grade 1]*. Hamburg, Germany: Freie und Hansestadt Hamburg.
- Mannhaupt, G. (2013). *MÜSC – Münsteraner Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten [Münster screening for an early detection of difficulties in reading and writing]*. Berlin, Germany: Cornelsen.
- Ministry of Education, P. R. China (MoE). (2011). *Mathematics curriculum standards for compulsory education (Grades 1–9) (in Chinese)*. Beijing: Beijing Normal University Press.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, Mass.: TIMSS & PIRLS International Study Center.
- Múñez, D., Orrantia, J., & Rosales, J. (2013): The effect of external representations on compare word problems. Supporting mental model construction. *The Journal of Experimental Education 81*(3), 337–355. doi:10.1080/00220973.2012.715095.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide (7th ed.)*. Los Angeles, CA: Muthén & Muthén.
- Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16*(2), 173–199.
- National Research Council. (2002). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education*. Washington, DC: The National Academies Press. doi:10.17226/10023.
- Nunes, T., Dorneles, B.V., Lin, P.-J., Rathgeb-Schnierer, E. (2016). *Teaching and learning about whole numbers in primary school*. Cham: Springer.
- Obersteiner, A., Reiss, K., & Ufer, S. (2013): How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction 23*, 125–135. doi:10.1016/j.learninstruc.2012.08.004.
- Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology, 141*, 649–666.
- Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5*(1), 49–101.
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics (Vol. 1)*. Dordrecht: Kluwer.
- Schneider, W., Krajewski, K., & Küspert, P. (2013). *Die Entwicklung mathematischer Kompetenzen [The development of mathematical competence]*. Paderborn: Schöningh; UTB GmbH.
- Schumacher, R. F., & Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problem. *Journal of Experimental Child Psychology*, *111*, 607–628. doi:10.1016/j.jecp.2011.12.001.
- Selter, C., Prediger, S., Nührenböcker, M., & Hussmann, S. (2012). Taking away and determining the difference – a longitudinal perspective on two models of subtraction and the inverse relation to addition. *Educational Studies in Mathematics*, *79*, 389–408.
- Sievert, H., van den Ham, A.-K., & Heinze, A. (2019a). Why the textbook matters — A two-domain impact analysis. In S. Rezat, L. Fan, M. Hattermann, J. Schumacher, & H. Wuschke (Eds.), *Proceedings of the Third International Conference on Mathematics Textbook Research and Development: 16–19 September 2019 Paderborn, Germany* (pp. 311–316). Universitätsbibliothek Paderborn. doi: 10.17619/UNIPB/1-768.
- Sievert, H., van den Ham, A.-K., Niedermeyer, I., & Heinze, A. (2019b). Effects of mathematics textbooks on the development of primary school children's adaptive expertise in arithmetic. *Learning and Individual Differences* *74*, 101716, 1–13. doi:10.1016/j.lindif.2019.02.006.
- Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319–370). Charlotte, CT: Information Age.
- Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology*, *85*, 7–23.
- Stern, E. (1998). *Die Entwicklung des mathematischen Verständnisses im Kindesalter [The development of mathematical understanding in childhood]*. Lengerich: Pabst.
- Tarim, K. (2017). Problem solving levels of elementary school students on mathematical word problems and the distribution of these problems in textbooks. *Cukurova University Faculty of Education Journal*, *46*(2), 639–648. doi:10.14812/cuefd.306025.
- Törnroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, *31*(4), 315–327.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen, & K. Elliot (Eds.). *But what does it mean? The use of effect sizes in educational research* (pp. 55 – 66). London: National Foundation of Educational Research.
- Valverde, G., Bianchi, L. J., Wolfe, R., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht: Kluwer Academic Publishers.
- van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2016). *Elementary and middle school mathematics: Teaching developmentally*. Boston: Person Education Inc.
- van den Ham, A.-K., & Heinze, A. (2018). Does the textbook matter? Longitudinal effects of textbook choice on primary school students' achievement in mathematics. *Studies in Educational Evaluation*, *59*, 133–140. doi:10.1016/j.stueduc.2018.07.005.
- van Steenbrugge, H., Valcke, M., & Desoete, A. (2013). Teachers views of mathematics textbook series in Flanders. Does it (not) matter which mathematics textbook series schools choose? *Journal of Curriculum Studies*, *45*(3), 322–353.

- Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Charlotte, CT: Information Age.
- Weiß, R. H., & Osterland, J. (2013). *CFT 1-R – Grundintelligenztest Skala 1 Revision [basic cognitive abilities test Scale 1 revision]*. Göttingen, Germany: Hogrefe.
- Wessel, J. (2015): *Grundvorstellungen und Vorgehensweisen bei der Subtraktion [mental representations and procedures for subtraction]*. Dissertation. Fakultät für Mathematik, Technische Universität Dortmund. doi:10.1007/978-3-658-11386-5.
- Xin, Y. P. (2007). Word problem solving tasks in textbooks and their relation to student performance. *The Journal of Educational Research*, 100(6), 347–360. doi:10.3200/JOER.100.6.347-360.

Abbildungsverzeichnis

Abbildung 2.1: Operative Beziehungen am Beispiel der Aufgabe $3 + 5 = 8$	14
Abbildung 4.1: Beispiele für 1 Punkt- (a, Rinkens, 2010), 2 Punkte- (b, Eidt, 2010) und 3 Punktebewertungen (c, Bauer & Maurach, 2011b)	33
Abbildung 4.2: Beispiel für eine Lerngelegenheit für den Vergleich von Strategien (Eidt, 2010).....	35
Abbildung 4.3: Beispiele für ikonische Darstellungen (a & c), einen realitätsnahen Kontext (a), Einführung der Namen (a), sprachliche Beschreibung (b) und Darstellung anhand typischerweise vorhandener Arbeitsmaterialien (Wendeplättchen, c); (a): Tauschaufgabe, Buschmeier, 2011, (b): Nachbaraufgabe, Bauer & Maurach, 2011a, (c): Tauschaufgabe, Brall, 2010	38
Abbildung 4.4: Übungsaufgaben für die Strategien Tauschaufgabe (a, 2 Aufgaben: Buschmeier, 2011), Umkehraufgabe (b, 6 Aufgaben: Brall, 2010), und Nachbaraufgaben (c, 2 Aufgaben: Bauer & Maurach, 2011a).....	39
Abbildung 4.5: Beispiel für einen Strategievergleich (Rinkens et al., 2011)	39
Abbildung 4.6: Beispiel einer Lerngelegenheit mit einer 1–1-Tabelle (Rinkens et al., 2011)	40
Abbildung 4.7: Beispiel für eine Begründung für die Anwendung einer Strategie (Nutzen der Tauschaufgabe: Bauer & Maurach, 2011a).....	41
Abbildung 4.8: Beispiele für die beiden Item-Typen zur Nutzung Operativer Grundaufgaben unter Speedbedingungen.....	42
Abbildung 4.9: Beispiel für eine Lerngelegenheit in der Kategorie Zahlzerlegungen (Buschmeier, 2011)	44
Abbildung 4.10: Beispiel für eine Lerngelegenheit der Kategorie Komplementarität (Buschmeier, 2011)	45
Abbildung 4.11: Beispiel für eine Lerngelegenheit zur Subtraktion als Unterschied (Brall, 2010)	46
Abbildung 4.12: Vergleichende Interpretation als falsche Antwort (Buschmeier, 2011)	47
Abbildung 4.13: Beispiel-Item mit ikonischer Darstellung	48

Tabellenverzeichnis

Tabelle 2.1: Gängige im Unterricht eingeführte Additions- und Subtraktionsstrategien	11
Tabelle 2.2: Übersicht über die drei Strategien, welche die Operativen Grundaufgaben nutzen.....	16
Tabelle 4.1: Bewertungskategorien für die Schulbuchqualität für den Bereich adaptives Rechnen....	32
Tabelle 4.2: Zuordnung der Strategien zu den Effizienz kategorien am Beispiel der Aufgabe 702 – 698	36
Tabelle 4.3: Bewertungskategorien für die Schulbuchqualität für den Bereich Operative Grundaufgaben.....	37
Tabelle 4.4: Bewertungskategorien für die Schulbuchqualität für den Bereich quantitative Vergleiche	44
Tabelle 5.1: Ergebnisse der Schulbuchanalyse in Studie 1 (absolute Punktzahlen und Rangfolge)	49
Tabelle 5.2: Ergebnisse der Mehrebenenanalyse in Studie 1 (Standardfehler in Klammern, alle Variablen sind standardisiert mit Ausnahme der Teilnahme am Förderprogramm (0,1), der Lehrkraftqualifikation in Klasse 2 und 3 (0,1,2) und der Schulbuchqualität in Klasse 2 und 3 (mittlerer Rang))	50
Tabelle 5.3: Ergebnisse der Schulbuchanalyse in Studie 2 (z-Scores der vier Kategorien und z-standardisierter Summenscore).....	52
Tabelle 5.4: Ergebnisse der Mehrebenenanalyse in Studie 2 (Standardfehler in Klammern, alle Variablen sind standardisiert mit Ausnahme der Teilnahme am Förderprogramm (0,1) und der Lehrkraftqualifikation (0,1))	53
Tabelle 5.5: Ergebnisse der Schulbuchanalyse in Studie 3 (z-Scores der fünf Kategorien und z-standardisierter Summenscore).....	54
Tabelle 5.6: Ergebnisse der Mehrebenenanalyse in Studie 3 (Standardfehler in Klammern, alle Variablen sind standardisiert mit Ausnahme der Teilnahme am Förderprogramm (0,1) und der Lehrkraftqualifikation (0,1))	55

Literaturverzeichnis

- Agodini, R., Harris, B., Thomas, M., Murphy, R. & Gallagher, L. (2010). Achievement effects of four early elementary school math curricula: Findings for first and second graders. NCEE 2011-4001. *National Center for Education Evaluation and Regional Assistance*.
- Bauer, R. & Maurach, J. (2011a). *Einstern 1*. Berlin: Cornelsen.
- Bauer, R. & Maurach, J. (2011b). *Einstern 2*. Berlin: Cornelsen.
- Baroody, A. J. (2016). Curricular approaches to introducing subtraction and fostering fluency with basic differences in grade 1. In R. Bracho (Hrsg.), *The development of number sense: From theory to practice. Monograph of the Journal of Pensamiento Numérico y Algebraico (Numerical and Algebraic Thought)*, 10(3) (S. 135-165). University of Granada.
- Baroody, A. J. & Dowker, A. (Hrsg.) (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Benz, C. (2005). *Erfolgsquoten, Rechenmethoden, Lösungswege und Fehler von Schülerinnen und Schülern bei Aufgaben zur Addition und Subtraktion im Zahlenraum bis 100*. Hildesheim: Franzbecker.
- Bhatt, R. & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis*, 34(4), 391-412.
- Bhatt, R., Koedel, C. & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review*, 34, 107-121.
- Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S. & Kurlaender, M. (2019). *Learning by the book: Comparing math achievement growth by textbook in six Common Core states*. Research report. Cambridge, MA: Center for Education Policy Research, Harvard University.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.) (2010), *TEDS-M 2008 - Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich*. Münster: Waxmann.
- Blöte, A. W., Klein, A. S. & Beishuizen, M. (2000). Mental computation and conceptual understanding. *Learning and Instruction*, 10(3), 221-248.
- Blöte, A. W., van der Burg, E. & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, 93, 627-638.
- Blum, W. & vom Hofe, R. (2003). Welche Grundvorstellungen stecken in der Aufgabe? *Mathematik lehren*, 118, 14-18.
- Bodovski, K. & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal*, 108(2), 115-130.
- Braithwaite, D. W., Pyke, A. A., & Siegler, R. S. (2017). A computational model of fraction arithmetic. *Psychological Review*, 124(5), 603-625. doi:10.1037/rev0000072.
- Brall, C. (2010). *Flex und Flo 1*. Braunschweig: Diesterweg.
- Bruner, J. S. (1974). *Entwurf einer Unterrichtstheorie*. Berlin: Berlin-Verlag.

- Buschmeier, G. (2011). *Denken und Rechnen 1*. Braunschweig: Westermann.
- Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E. & Empson, S. B. (1997). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 29(1), 3-20.
- Carpenter, T. P. & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15, S. 179-202.
- Cattell, R., Weiß, R. & Osterland, J. (1997). *Grundintelligenztest Skala 1. CFT 1*. Göttingen: Hogrefe.
- Chen, Z. & Siegler, R. S. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development*, 65 (2, serial no. 261).
- Clements, D. H. & Sarama, J. (2007). Early childhood mathematics learning. In F. K. Lester (Hrsg.), *Second handbook of research on mathematics teaching and learning* (S. 461-555). Charlotte, NC: Information Age.
- Common Core State Standards Initiative (CCSSI). (2010). *Common core state standards for mathematics*. National Governors Association and the Council of Chief State School Officers, Washington, DC. Abgerufen am 24.06.2020, von <http://www.corestandards.org/Math/Practice>
- Cooper, T. J., Heirdsfield, A. M. & Irons, C. J. (1996). *Years 2 and 3 children's mental addition and subtraction strategies for 2 and 3 digit word problems and algorithmic exercises*, Paper presented at Topic Group 1 at ICME-8, Sevilla.
- Cowen, R. (2003). Does it all add up? Changes in children's knowledge of addition combinations, strategies, and principles. In A.J. Baroody & A. Dowker (Hrsg.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (S. 35-74). Mahwah, NJ: Lawrence Erlbaum Associates.
- Csikós, C. (2016). Strategies and performance in elementary students' three-digit mental addition. *Educational Studies in Mathematics*, 91(1), 123-139.
- De Smedt, B., Torbeyns, J., Strassens, N., Ghesquière, P. & Verschaffel, L. (2010). Frequency, efficiency and flexibility of indirect addition in two learning environments. *Learning and Instruction*, 20, 205-215.
- Despina, D. & Harikleia, L. (2014). Addition and subtraction word problems in Greek Grade a and Grade b mathematics textbooks: Distribution and children's understanding. *International Journal for Mathematics Teaching and Learning*, 8, 340.
- Dumont, H., Neumann, M., Maaz, K. & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen: Internationale und nationale Befunde. *Psychologie in Erziehung und Unterricht*, 60, 163-183.
- Eddy, R. M., Hankel, N., Hunt, A., Goldman, A. & Murphy, K. (2014). *Houghton Mifflin Harcourt GO Math! Efficacy study year one final report*. La Verne, CA: Cobblestone Applied Research & Evaluation, Inc.
- Education Institute. (2005). *Curriculum standards for the state of Qatar*. Doha, Qatar: Supreme Education Council.
- Eidt, H. (2010): *Denken und Rechnen 2*. Braunschweig: Westermann.

- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Publications.
- Fagginger Auer, M. F., Hickendorff, M., van Putten, C. M., Béguin, A. A. & Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data. Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, 29(2), 144-159. doi:10.1080/08957347.2016.1138959.
- Fan, L. (2013). Textbook research as scientific research. Towards a common ground on issues and methods of research on mathematics textbooks. *ZDM Mathematics Education*, 45(5), 765-777. doi:10.1007/s11858-013-0530-6.
- Fan, L., Zhu, Y. & Miao, Z. (2013). Textbook research in mathematics education: Development status and directions. *ZDM Mathematics Education*, 45(5), 633-646. doi:10.1007/s11858-013-0539-x.
- Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 243-275). New York: Macmillan.
- Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., Carpenter, T.P. & Fennema, E. (1997). Children's conceptual structures for multidigit numbers and methods of multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 28(2), 130-162.
- Gabler, L. & Ufer, S. (2020). Flexibilität im Umgang mit mathematischen Situationsstrukturen – Eine Vorstudie für die Entwicklung eines Förderkonzepts zum Lösen additiver Textaufgaben. *Journal für Mathematik-Didaktik*. doi:10.1007/s13138-020-00170-3.
- Gasteiger, H. (2019): Strategieverwendung bei Additionsaufgaben mit Zehnerübergang Ende Jahrgangsstufe 2. (Vortrag, 05.03.2019). Regensburg: 53. Jahrestagung der Gesellschaft für Didaktik der Mathematik, Universität Regensburg.
- Grønmo, L., Lindquist, M., Arora, A. & Mullis, I. (2015). TIMSS 2015 mathematics framework. Abgerufen am 24.06.2020, von https://timssandpirls.bc.edu/timss2015/downloads/T15_FW_Chap1.pdf
- Hadar, L. L. (2017). Opportunities to learn: Mathematics textbooks and students' achievements. *Studies in Educational Evaluation*, 55, 153-166.
- Hart, K.M. (Hrsg.). (1981). *Children's Understanding of Mathematics*. Murray, London, S. 11-16.
- Hartung, T. (2014). Schulbuchauswahl und Lernmittelfreiheit in den deutschen Bundesländern im Kontext von Schülerpartizipation. *Eckert. Working Papers*, 11, 1-15. Abgerufen am 03.06.2020, von http://www.pedocs.de/volltexte/2015/11061/pdf/EWP_2014_11_Hartung_Schulbuchauswahl.pdf
- Hatano, G. (1982). Cognitive consequences of practice in culture specific procedural skills. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 4, 15-18.
- Hatano, G. & Inagaki, K. (1986). Two courses of expertise. In H. W. Stevenson, H. Azuma & K. Hakuta (Hrsg.), *A series of books in psychology. Child development and education in Japan* (S. 262-272). New York: W H Freeman/Times Books/ Henry Holt & Co.
- Heinze, A., Arend, J., Gruessing, M. & Lipowsky, F. (2018). Instructional approaches to foster third graders' adaptive use of strategies: An experimental study on the effects of two

- learning environments on multi-digit addition and subtraction. *Instructional Science*, 6(6), 869-891. doi:10.1007/s11251-018-9457-1.
- Heinze, A., Arend, J., Gruessing, M. & Lipowsky, F. (2020). Systematisch einführen oder selbst entdecken lassen? Eine experimentelle Studie zur Förderung der adaptiven Nutzung von Rechenstrategien bei Grundschulkindern. *Unterrichtswissenschaft*, 48, 11-34. doi:10.1007/s42010-019-00063-6.
- Heinze, A., Marschick, F. & Lipowsky, F. (2009). Addition and subtraction of three-digit numbers. Adaptive strategy use and the influence of instruction in German third grade. *ZDM Mathematics Education*, 41(5), 591-604. doi:10.1007/s11858-009-0205-5
- Hickendorff, M. (2020). Fourth graders' adaptive strategy use in solving multidigit subtraction problems. *Learning and Instruction*, 67. doi:10.1016/j.learninstruc.2020.101311.
- Hiebert, J. & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction*, 14(3), 251-284.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications. Quantitative methodology series*. New York: Routledge.
- Huang, R., Zhang, Q., Chang, Y. & Kimmins, D. (2019). Developing students' ability to solve word problems through learning trajectory-based and variation task-informed instruction. *ZDM Mathematics Education* 51, 169-181.
- Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein (IQSH). (2020). Mathe macht stark – Grundschule. Abgerufen am 24.06.2020, von <https://nzi.lernnetz.de/index.php/mathe-macht-stark-grundschule.html>
- Jones, D. L., Brown, M., Dunkle, A., Hixon, L., Yorder, N. & Silbernack, Z. (2015). The statistical content of elementary school mathematics textbooks. *Journal of Statistics Education*, 23(3), 1-22.
- Kilpatrick, J., Swafford, J. & Findell B. (Hrsg.) (2001). *Adding it up*. Washington, D.C.: National Academies Press. doi:10.17226/9822.
- Kintsch, W. & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92(1), 109-129.
- Klein, A. S., Beishuizen, M. & Treffers, A. (1998). The empty number line in Dutch second grades. Realistic versus gradual program design. *Journal for Research in Mathematics Education*, 29(4), 443-464.
- Knievel, I., Lindmeier, A. M. & Heinze, A. (2015). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education*, 13(2), 309-329. doi:10.1007/s10763-014-9608-z.
- Koedel, C., Morgan, D. L., Polikoff, S., Hardaway, T. & Wrabel, S. L. (2017). Mathematics curriculum effects on student achievement in California. *AERA Open*, 3(1), 1-22.
- Kolovou, A., van den Heuvel-Panhuizen, M. & Bakker, A. (2009). Non-routine problem solving tasks in primary school mathematics textbooks: A needle in a haystack. *Mediterranean Journal for Research in Mathematics Education*, 8(2), 31-68.
- Krammer, H. P. M. (1985). The textbook as classroom context variable. *Teaching and Teacher Education*, 1(4), 273-278. doi:10.1016/0742-051X(85)90015-0.

- Krauthausen, G. (1993): Kopfrechnen, halbschriftliches Rechnen, schriftliche Normalverfahren, Taschenrechner. Für eine Neubestimmung des Stellenwertes der vier Rechenmethoden. *Journal für Mathematik-Didaktik* 14 (3-4), S. 189-219.
- Kuhnke, K. (2013). *Vorgehensweisen von Grundschulkindern beim Darstellungswechsel. Eine Untersuchung am Beispiel der Multiplikation*. Wiesbaden: Springer Spektrum.
- Kultusministerkonferenz (KMK). (2005). *Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Mathematik für den Primarbereich. Beschluss vom 15.10.2004*. Abgerufen am 03.06.2020, von https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Mathe-Primar.pdf
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2013). *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers - Results from the COACTIV Project*. New York: Springer.
- Lemaire, P. & Siegler, R. S. (1995). Four aspects of strategic change. Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124(1), 83-96.
- Lepik, M., Grevholm, B. & Viholainen, A. (2015). Using textbooks in the mathematics classroom – the teachers' view. *Nordic Studies in Mathematics Education*, 20(3-4), 129-156.
- Lipowsky, F., Faust, G. & Karst, K. (Hrsg.) (2011). *Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE) – Teil 2. Materialien zur Bildungsforschung*. Frankfurt am Main: GPF.
- Lorenz, J. H. (2007). *HaRet – Hamburger Rechentest für Klasse 1*. Hamburg: Freie und Hansestadt Hamburg.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, 58 (2), 103-117.
- Mannhaupt, G. (2013). *MÜSC – Münsteraner Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Berlin: Cornelsen.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein (Hrsg.) (1997). *Lehrplan Grundschule*. Kiel: Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein.
- Ministry of Education, P. R. China (MoE). (2011). *Mathematics curriculum standards for compulsory education (grades 1-9)* (in Chinese). Beijing: Beijing Normal University Press.
- Mullis, I. V., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's Third International mathematics and science study (TIMSS)*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, Mass.: TIMSS & PIRLS International Study Center.
- Mullis, I. V. S., Martin, M. O., Goh, S. & Cotter, K. (2016). *TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Abgerufen am 24.06.2020, von <http://timssandpirls.bc.edu/timss2015/encyclopedia/>

- Múñez, D., Orrantia, J. & Rosales, J. (2013): The Effect of External Representations on Compare Word Problems. Supporting Mental Model Construction. *The Journal of Experimental Education* 81(3), 337-355. doi:10.1080/00220973.2012.715095.
- Muthén L. K. & Muthén B. O. (1998–2012). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Mwangi, W. & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction*, 16(2), 173-199.
- National Council of Teachers of Mathematics (2003). *Standards for school mathematics: number and operations*. Abgerufen am 24.06.2020, von <http://standards.nctm.org/document/chapter3/numb.htm>
- National Research Council. (2002). Investigating the influence of standards: A framework for research in mathematics, science, and technology education. In I. R. Weiss, M. S. Knapp, K. S. Hollweg & G. Burrill (Hrsg.), *Committee on understanding the influence of standards in K-12 science, mathematics, and technology education, center for education, division of behavioral and social sciences and education*. Washington, DC: National Academy Press.
- Nunes, T., Carraher, D.W. & Dias Schlieman, A. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education* 18(2), 83-97.
- Nunes T., Dorneles B.V., Lin P.J. & Rathgeb-Schnierer E. (2016). *Teaching and Learning About Whole Numbers in Primary School. ICME-13 Topical Surveys*. Cham: Springer.
- Nunes, T., Schlieman, A.D. & Carraher, D.W. (1993). *Street Mathematics and School Mathematics*. Cambridge University Press: Cambridge.
- Obersteiner, A., Reiss, K. & Ufer, S. (2013): How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction* 23, 125-135. doi:10.1016/j.learninstruc.2012.08.004.
- Padberg, F. (2005). *Didaktik der Arithmetik*. Heidelberg: Spektrum.
- Padberg, F. & Benz, C. (2011): *Didaktik der Arithmetik*. Heidelberg: Spektrum.
- Pepin, B. & Haggarty, L. (2001). Mathematics textbooks and their use in English, French and German classrooms: A way to understand teaching and learning cultures. *ZDM Mathematics Education*, 33(5), 158-175.
- Prather, R. W. & Alibali, M. W. (2009). The development of arithmetic principle knowledge: How do we know what learners know? *Developmental Review*, 29(4), 221-248.
- Rabardel, P. (2002). *People and technology: A cognitive approach to contemporary instruments*. Abgerufen am 15.05.2020, von Paris: Université Paris. <https://hal.archives-ouvertes.fr/hal-01020705>
- Radatz, H. (1989). Schülervorstellungen von Zahlen und elementaren Rechenoperationen. In *Beiträge zum Mathematikunterricht 1989* (S. 306-309). Bad Salzdetfurth: Verlag Franzbecker.
- Radatz, H., Schipper, W., Ebeling, A. & Dröge, R. (1996). *Handbuch für den Mathematikunterricht*. Hannover: Schroedel.

- 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*. Dissertation, Pädagogische Hochschule Freiburg.
- Rathgeb-Schnierer, E. (2011). Warum noch rechnen, wenn ich die Lösung sehen kann? Hintergründe zur Förderung flexibler Rechenkompetenzen bei Grundschulkindern. In R. Haug & L. Holzäpfel (Hrsg.), *Beiträge zum Mathematikunterricht 2011* (S. 15–22). Münster: WTM.
- Rechtsteiner-Merz, C. (2013). *Flexibles Rechnen und Zahlenblickschulung: Entwicklung und Förderung von Rechenkompetenzen bei Erstklässlern, die Schwierigkeiten beim Rechnenlernen zeigen*. Münster: Waxmann.
- Reeve, R., Reynolds, F., Humberstone, J. & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology*, 141, 649-666.
- Rezat, S. & Sträßer, R. (2012): From the didactical triangle to the socio-didactical tetrahedron. Artifacts as fundamental constituents of the didactical situation. *ZDM Mathematics Education* 44(5), 641-651. doi:10.1007/s11858-012-0448-4.
- Riley, M. S. & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and instruction*, 5(1), 49-101.
- Riley, M. S., Greeno, J. G. & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Hrsg.), *The development of mathematical thinking* (S. 153-196). New York: Academic Press.
- Rinkens, H.-D. (Hrsg.) (2010): *Welt der Zahl 2*. Braunschweig: Schroedel.
- Rinkens, H.-D., Hönisch, K. & Träger, G. (Hrsg.). (2011). *Welt der Zahl 1*. Braunschweig: Schroedel.
- Rittle-Johnson, B. & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561-573.
- 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), 529-543.
- Rittle-Johnson, B., Star, J. R. & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836-852. doi:10.1037/a0016026
- Rittle-Johnson, B., Star, J. R. & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology*, 82, 436-455. doi:10.1111/j.2044-8279.2011.02037.x.
- Royar, T. (2013). *Handlung – Vorstellung – Formalisierung. Entwicklung und Evaluation einer Aufgabenreihe zur Überprüfung des Operationsverständnisses für Regel- und Förderklassen*. Hamburg: Verlag Dr. Kovac.
- Ruthven, K. (1998). The use of mental, written and calculator strategies of numerical computation by upper primary pupils within a 'calculator-aware' number curriculum. *British Educational Research Journal* 24(1), 21-42.

- Schäfer, J. (2005). *Rechenschwäche in der Eingangsstufe der Hauptschule. Lernstand, Einstellungen und Wahrnehmungsleistungen. Eine empirische Studie.* Hamburg: Verlag Dr. Kovac.
- Schipper, W. (2009). *Handbuch für den Mathematikunterricht an Grundschulen.* Braunschweig: Schroedel.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S. & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning.* San Francisco: Jossey-Bass.
- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T. & Wiley, D. E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics (Vol. 1).* Dordrecht: Kluwer.
- Schmidt, W. H., Raizen, S. A., Britton, E. D., Bianchi, L. J. & Wolfe, R. G. (1997). *A cross-national investigation of curricular intentions in school mathematics.* Dordrecht: Kluwer.
- Schneider, W., Krajewski, K. & Küspert, P. (2013). *Die Entwicklung mathematischer Kompetenzen.* Paderborn, Stuttgart: Schöningh; UTB GmbH.
- Schulz, A. (2018). Relational reasoning about numbers and operations – Foundation for calculation strategy use in multi-digit multiplication and division. *Mathematical Thinking and Learning*, 20(2), 108-141. doi:10.1080/10986065.2018.1442641.
- Schumacher, R. F. & Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problem. *Journal of Experimental Child Psychology*, 111, 607- 628. doi:10.1016/j.jecp.2011.12.001.
- Scribner, S. (1984). Studying working intelligence. In B. Rogoff and J. Lave (Hrsg.), *Everyday cognition: Its development in social context* (S. 9-40). Cambridge, MA: Harvard University Press.
- Selter, C. (2001). Addition and subtraction of three-digit numbers. German elementary children's success, methods and strategies. *Educational Studies in Mathematics*, 47(2), 145-174.
- Selter, C. (2009). Creativity, flexibility, adaptivity, and strategy use in mathematics. *ZDM Mathematics Education*, 41(5), 619-625. doi:10.1007/s11858-009-0203-7.
- Selter, C., Prediger, S., Nührenbörger, M. & Hussmann, S. (2012). Taking away and determining the difference – a longitudinal perspective on two models of subtraction and the inverse relation to addition. *Educational Studies in Mathematics*, 79, 389-408.
- Siegler, R. S., DeLoache, J. S. & Eisenberg, N. (2006). *How children develop.* New York, NY: Worth Publ.
- Siegler, R. S. & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication. Testing predictions of ASCM using the choice-no-choice method. *Journal of Experimental Psychology: General*, 126(1), 71-92.
- Sievert, H., van den Ham, A.-K. & Heinze, A. (2019). Why the textbook matters — A two-domain impact analysis. In S. Rezat, L. Fan, M. Hattermann, J. Schumacher & H. Wuschke (Hrsg.), *Proceedings of the Third International Conference on Mathematics Textbook Research and Development: 16-19 September 2019 Paderborn, Germany* (S. 311-316). Universitätsbibliothek Paderborn. doi:10.17619/UNIPB/1-768.

- Sievert, H., van den Ham, A.-K. & Heinze, A. (2020). Mit guten Schulbüchern zum flexiblen Rechnen? Lerngelegenheiten zum flexiblen Rechnen in Schulbüchern und ihr Zusammenhang zum Lernerfolg. *Die Grundschulzeitschrift*, Nr. 324.
- Sievert, H., van den Ham, A.-K., Niedermeyer, I. & Heinze, A. (2019). Effects of mathematics textbooks on the development of primary school children's adaptive expertise in arithmetic. *Learning and Individual Differences* 74, 101716, 1-13. doi:10.1016/j.lindif.2019.02.006.
- Star, J. R. & Newton, K.J. (2009). The nature and development of experts' strategy flexibility for solving equations. *ZDM Mathematics Education*, 41, 557-567.
- Star, J. R. & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, 31, 280-300.
- Staub, F. C. & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains. Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94(2), 344-355.
- Stein, M. K., Remillard, J. & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Hrsg.), *Second handbook of research on mathematics teaching and learning* (S. 319-370). Charlotte, CT: Information Age.
- Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology*, 85, 7-23.
- Stern, E. (1998). *Die Entwicklung des mathematischen Verständnisses im Kindesalter*. Lengerich: Pabst.
- Stöber, G. (2010). Schulbuchzulassung in Deutschland: Grundlagen, Verfahrensweisen und Diskussionen. Eckert.Beiträge 2010/3. Abgerufen am 24.06.2020 von <http://repository.gei.de/handle/11428/92>
- Tarim, K. (2017). Problem solving levels of elementary school students on mathematical word problems and the distribution of these problems in textbooks. *Cukurova University Faculty of Education Journal*, 46 (2), 639-648. doi:10.14812/cuefd.306025.
- Tarr, J. E., Reys, R. E., Reys, B. J., Chávez, Ó., Shih, J. & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education*, 39(3), 247-280.
- Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodriguez, M. & Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Amsterdam: International Association for the Evaluation of Educational Achievement. Abgerufen am 24.06.2020 von <http://www.iea.nl/teds-m.html>
- Thompson, I. (1994). Young children's idiosyncratic written algorithms for addition. *Educational Studies in Mathematics*, 26, 323-345.
- Threlfall, J. (2002). Flexible mental calculation. *Educational Studies in Mathematics*, 50, 29-47.
- Threlfall, J. (2009). Strategies and flexibility in mental calculation. *ZDM Mathematics Education*, 41(5), 541-556.

- Törnroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, 31(4), 315-327.
- Torbeyns, J., De Smedt, B., Ghesquière, P. & Verschaffel, L. (2009a). Jump or compensate? Strategy flexibility in the number domain up to 100. *ZDM Mathematics Education*, 41(5), 581-590.
- Torbeyns, J., De Smedt, B., Ghesquière, P. & Verschaffel, L. (2009b). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics*, 71(1), 1-18.
- Torbeyns, J., Hickendorff, M. & Verschaffel, L. (2017). The use of number-based versus digit-based strategies on multi-digit subtraction: 9–12-year-olds' strategy use profiles and task performance. *Learning and Individual Differences*, 58, 64-74.
- Torbeyns, J. & Verschaffel, L. (2016). Mental computation or standard algorithm? Children's strategy choices on multi-digit subtractions. *European Journal of Psychology of Education*, 31(2), 99-116. doi:10.1007/s10212-015-0255-8.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Hrsg.). *But what does it mean? The use of effect sizes in educational research* (S. 55-66). London: National Foundation of Educational Research.
- Valverde, G., Bianchi, L. J., Wolfe, R., Schmidt, W. H. & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht: Kluwer Academic Publishers.
- van de Walle, J. A., Karp, K. S. & Bay-Williams, J. M. (2016). *Elementary and middle school mathematics: Teaching developmentally*. Boston: Person Education Inc.
- van den Ham, A.-K. & Heinze, A. (2018). Does the textbook matter? Longitudinal effects of textbook choice on primary school students' achievement in mathematics. *Studies in Educational Evaluation*, 59, 133-140. doi:10.1016/j.stueduc.2018.07.005.
- van den Heuvel-Panhuizen, M. (2001). Realistic mathematics education in the Netherlands. In J. Anghileri (Hrsg.), *Principles and practices in arithmetic teaching* (S. 49-64). Buckingham: Open University Press.
- van Steenbrugge, H., Valcke, M. & Desoete, A. (2013). Teachers views of mathematics textbook series in Flanders. Does it (not) matter which mathematics textbook series schools choose? *Journal of Curriculum Studies*, 45(3), 322-353.
- Verboom, L. (2010). „Ich habe drei Plättchen mehr als du“. *Grundschule Mathematik*, 25, S. 6-7.
- Verschaffel, L., Greer, B. & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester (Hrsg.), *Second handbook of research on mathematics teaching and learning* (S. 557-628). Charlotte, CT: Information Age.
- Verschaffel, L., Luwel, K., Torbeyns, J. & van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24(3), 335-359.
- Weiß, R. H. & Osterland, J. (2013). *CFT 1-R – Grundintelligenztest Skala 1 Revision*. Göttingen: Hogrefe.

- Wessel, J. (2015): *Grundvorstellungen und Vorgehensweisen bei der Subtraktion*. Dissertation. Fakultät für Mathematik, Technische Universität Dortmund. doi:10.1007/978-3-658-11386-5.
- Winter, H. (1996). *Praxishilfe Mathematik. Didaktik im Überblick, kreatives Üben, Beispiele für die Klasse 1 bis 4*. Berlin: Cornelsen Scriptor.
- Wright, V. (2009). *Mathematics standards for years 1-8*. Wellington, N.Z.: Ministry of Education.
- Xin, Y. P. (2007). Word problem solving tasks in textbooks and their relation to student performance. *The Journal of Educational Research*, 100:6, 347-360. doi:10.3200/JOER.100.6.347-360.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit, abgesehen von der Beratung durch die Betreuerinnen und Betreuer meines Projekts, in Inhalt und Form selbstständig verfasst habe. Die Arbeit ist unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Die vorliegende Arbeit wurde weder im Ganzen noch in Teilen an einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt. Teile dieser Arbeit wurden bereits in Form von Zeitschriftenartikeln publiziert oder wurden zur Publikation eingereicht. Mir wurde kein akademischer Grad entzogen.

Kiel, den _____

Henning Christoph Sievert

