Empirische Sonderpädagogik, 2019, Nr. 2, S. 95-117
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Dieser Artikel wurde unter der Gastherausgeberschaft von Prof. Dr. Antje Ehlert und Prof. Dr. Annemarie Fritz-Stratmann begutachtet und angenommen.

Arithmetische Kompetenz und Rechenschwäche am Ende der Grundschulzeit: die Rolle statusdiagnostischer und lernverlaufsbezogener Prädiktoren

Jörg-Tobias Kuhn[1], Christin Schwenk[1], Elmar Souvignier[2] & Heinz Holling[2]

[1] Technische Universität Dortmund, [2] Westfälische Wilhelms-Universität Münster

Zusammenfassung

Die präzise Vorhersage der mathematischen Kompetenz am Ende der Grundschulzeit ist für die Unterrichts- und Förderplanung entscheidend. Unspezifische Prädiktoren der mathematischen Kompetenz (z. B. Intelligenz) können dabei von solchen unterschieden werden, die spezifische mathematische Fertigkeiten messen (z. B. arithmetischer Faktenabruf). Individuelle Lernverläufe bieten eine weitere, bislang selten genutzte Informationsquelle, um Leistungsstände am Schuljahresende vorherzusagen. In diesem Beitrag wurde die Rolle spezifischer, unspezifischer und lernverlaufsbezogener Prädiktoren für die arithmetische Kompetenz sowie die Vorhersage einer Rechenschwäche untersucht. Eine Stichprobe von $N = 196$ Grundschulkindern der dritten und vierten Klasse (Alter: $M = 106.12$ Monate, $SD = 7.53$, 107 Mädchen) bearbeitete zu Beginn des Schuljahres Testverfahren zur Erfassung der fluiden Intelligenz (Kurzform CFT 1-R/CFT 20-R), der Leseflüssigkeit (SLS 2-9), der arithmetischen Kompetenz (DEMAT 2+/DEMAT 3+, Arithmetik) und des arithmetischen Faktenabrufs (DIRG Multiplikation). Im zweiten Schulhalbjahr wurde in etwa zweiwöchigen Abständen zehnmal ein Lernverlaufstest Mathematik (LVD-M 2-4) durchgeführt. Zur Vorhersage der arithmetischen Kompetenz am Schuljahresende (DEMAT 3+/4, Arithmetik) bzw. einer Rechenschwäche (DEMAT Arithmetik $T \leq 40$) wurden gemischte lineare und logistische Modelle berechnet. Zudem wurden indirekte Effekte über die Lernverlaufsparameter mittels Mediatoranalysen untersucht. Eine statistische Dominanzanalyse ergab, dass die Lernverlaufsparameter gegenüber den spezifischen und unspezifischen Prädiktoren signifikant bedeutsamer waren, die spezifischen Prädiktoren dominierten zudem die unspezifischen. Die Vorhersage von Rechenschwäche am Ende des Schuljahres wurde anhand des arithmetischen Faktenwissens (DIRG) durch die Lernverlaufsparameter partiell mediiert. Die Befunde werden im Hinblick auf praktische Konsequenzen diskutiert.

Schlüsselwörter: Mathematische Kompetenz, Rechenschwäche, Lernverlaufsdiagnostik

Arithmetic skills and mathematical learning difficulties at the end of elementary school: The role of summative and formative predictors

Abstract

A precise prediction of mathematical skills at the end of elementary school is a decisive factor in planning teaching activities and interventions. In this context, unspecific predictors of mathematical skills (e.g. intelligence) can be distinguished from predictors that are specific for mathematical skills (e.g. arithmetic fact retrieval). Individual patterns of growth in achievement over time offer an additional source of information for predicting learning outcomes. In this study, we assessed the role of specific, unspecific, and growth-related predictors for arithmetic skills or mathematical learning difficulties at the end of the school year. At the beginning of the school year, N = 196 elementary school students in third and fourth grades (age: M = 106.12 months, SD = 7.53, 107 girls) were administered tests of fluid intelligence, reading fluency, curricular arithmetic skills, and arithmetic fact retrieval. During the second term, growth in arithmetic skills was assessed by a learning progress assessment. Prediction of arithmetic skills or mathematical learning disabilities at the end of the school year was based on linear or logistic mixed models, respectively. Indirect effects transmitted by growth parameters were assessed using mediation analysis. A statistical dominance analysis showed that growth-related parameters were substantially more important than specific or unspecific predictors, and specific predictors dominated unspecific predictors. Mathematical learning disabilities were indirectly affected by arithmetic fact retrieval, mediated by growth parameters. Results are discussed with respect to practical consequences in educational settings.

Keywords: Mathematical skills, mathematical learning disabilities, learning progress assessment

Einleitung

In der Literatur herrscht Konsens darüber, dass schulisches Lernen sowohl auf domänenspezifischem Wissen und Fertigkeiten als auch auf domänengenerellen, allgemeineren Fähigkeiten beruht (Deary, Strand, Smith & Fernandes, 2007; Gustafsson & Undheim, 1992). Der relative Beitrag domänenspezifischer und domänengenereller Fertigkeiten zur Vorhersage erfolgreichen Lernens, aber auch von Lernschwierigkeiten, wird in der Literatur jedoch kontrovers diskutiert, da zum Teil widersprüchliche Ergebnisse vorliegen. Die Befunde unterscheiden sich je nach untersuchter Altersgruppe, nach Expertise der Versuchspersonen und variieren häufig längsschnittlich (Ferrer & McArdle, 2004). Die vorliegende längsschnittliche Studie verfolgt das Ziel, einen Beitrag zur Klärung dieser und weiterer Fragen (z. B. die Untersuchung direkter und indirekter Effekte) zur Vorhersage arithmetischer Fertigkeiten und Schwierigkeiten am Ende der Grundschulzeit zu leisten.

Theoretischer Hintergrund

Unspezifische Prädiktoren mathematischer Kompetenz

Fertigkeiten, die nicht nur die Entwicklung der mathematischen Kompetenz, sondern auch anderer Bereiche (z. B. schriftsprachliche Leistungen) signifikant beeinflussen, werden im Folgenden als unspezifische Prädiktoren mathematischer Kompetenz bezeichnet. Ihre Bedeutung für die Erforschung schulischer, insbesondere auch mathematischer, Entwicklung spiegelt sich in zahlreichen Primärstudien und Übersichtsartikeln wider (s. z.B. Geary, Nicholas, Yi & Sun, 2017; Vanbinst & De Smedt, 2016). Auch in theoretischer Hinsicht spielen unspezifische kognitive Fertigkeiten eine Rolle: Beispielweise berücksichtigen von Aster und Shalev (2007) die Arbeitsgedächtniskapazität in ihrem vierstufigen Entwicklungsmodell.

Neben der Arbeitsgedächtnisleistung ist die allgemeine Intelligenz einer der am besten erforschten unspezifischen Prädiktoren der Mathematikleistung (z. B. Fuchs et al., 2016; Geary, 2011a). In etlichen Studien zeigte sich, dass während der Grundschulzeit die allgemeine Intelligenz die spätere Mathematikleistung auch bei Berücksichtigung einer Reihe kognitiver und nicht-kognitiver Kovariaten (z. B. Arbeitsgedächtnis, frühe mathematische Fertigkeiten, Lesefertigkeiten) vorhersagt (Cowan & Powell, 2014; Fuchs et al., 2016; Geary et al., 2017). In der über acht Jahre angelegten Längsschnittstudie von Geary et al. (2017) war zum Ende der Grundschulzeit der Beitrag unspezifischer Prädiktoren zur Vorhersage mathematischer Kompetenz sogar größer als der Beitrag mathematikspezifischer Prädiktoren. Allerdings scheint die Bedeutung der Intelligenz für die Vorhersage komplexerer mathematischer Fertigkeiten (z. B. Lösen von Textaufgaben) größer zu sein als für basalere Aufgaben (z. B. Kopfrechnen im Zahlenraum bis 20; Cowan & Powell, 2014).

Andere Arbeiten weisen allerdings nach, dass die Intelligenz die spätere Mathematikleistung nicht oder nur indirekt beeinflusst. So konnte in einer Längsschnittstudie von Krajewski und Schneider (2009) gezeigt werden, dass die Intelligenz zu Beginn der Grundschulzeit die mathematische Kompetenz am Ende der Grundschulzeit nicht direkt, sondern nur indirekt über mathematische Basiskompetenzen (Zahl-Größen-Verknüpfung) vorhersagt (s. aber die vergleichbare Arbeit von Östergren & Träff, 2013, in der keine vollständige, sondern nur partielle Mediation des Intelligenzeffekts gefunden wurde). In der Longitudinalstudie zur Genese individueller Kompetenzen (LOGIK) zeigte sich zudem, dass das Lösen von Textaufgaben in der 11. Klasse höher mit dem Lösen von Textaufgaben in der 2. Klasse als mit der Intelligenz in der 11. Klasse korrelierte (Stern, 2013), was die Bedeutung spezifischer Fertigkeiten unterstreicht.

Das Lösen von Textaufgaben, aber auch grundlegende Rechenprozesse sind mit Lesefertigkeiten assoziiert (r = .55 in der Meta-Analyse von Singer & Strasser, 2017). Bereits frühe sprachliche Fertigkeiten wie Wortschatz und phonologische Bewusstheit, die eng mit den Lesefertigkeiten verbunden sind, sind zu Beginn der Schulzeit bedeutsame Prädiktoren für spätere mathematische Fertigkeiten (Lefevre et al., 2010). Fuchs et al. (2016) konnten nachweisen, dass frühe Lesefertigkeiten zu Beginn der Schulzeit die spätere Mathematikleistung auch dann vorhersagten, wenn zahlreiche weitere kognitive und nicht-kognitive Fertigkeiten berücksichtigt wurden. Lese- und Rechenschwierigkeiten treten zudem häufig gleichzeitig auf (Moll, Kunze, Neuhoff, Bruder & Schulte-Körne, 2014). Einige Arbeiten zeigen allerdings, dass die Vorhersage späterer Lesekompetenz durch frühe mathematische Fertigkeiten besser gelingt als umgekehrt (Duncan et al., 2007).

Zusammenfassend kann festgestellt werden, dass unspezifische Prädiktoren in Form von allgemeiner Intelligenz und Lesefertigkeiten die Entwicklung mathematischer Fertigkeiten beeinflussen. Unklar bleibt jedoch ihre Relevanz im Verhältnis zu spezifischen Prädiktoren, auf die als nächstes eingegangen wird.

Spezifische Prädiktoren mathematischer Kompetenz

Unter spezifischen Prädiktoren verstehen wir solche, die primär die Vorhersage späterer mathematischer Kompetenz erlauben, aber für andere (schulische) Kompetenzen von eher nachrangiger Bedeutung sind. Mittlerweile existiert eine große Zahl von Studien, die längsschnittlich die mathematische Kompetenzentwicklung, beginnend vom späten Kindergartenalter bis in die weiterführende Schule, untersuchen (s. Schneider, Küspert & Krajewski, 2016). Die theoretische Grundlage für Studien im späten Kindergarten- und frühen Grundschulalter bieten Entwicklungsmodelle der Zahlverarbeitung (vgl. Fischer, Roesch & Moeller, 2017). Diese fokussieren auf für die arithmetische Kompetenzentwicklung grundlegende Fertigkeiten wie das (kardinale) Zahlenverständnis, das Aspekte wie die Zahlen-Größen-Verknüpfung (kardinales Zahlenverständnis) sowie das Verständnis von Zahlensymbolen umfasst (Göbel, Watson, Lervåg & Hulme, 2014; Schwenk et al., 2017).

Neben diesen basalen mathematischen Fertigkeiten ist im schulischen Kontext mit zunehmendem Schulalter vor allem die Rechenflüssigkeit von Bedeutung. Flüssiges Rechnen ist zum einen abhängig von effektiven Rechenprozeduren (z.B. Ableitung von Ergebnissen; Baroody, Bajwa & Eiland, 2009). Zum anderen ist für eine effiziente Lösung arithmetischer Aufgaben die Fertigkeit, einfache mathematische Fakten zu speichern und bei Bedarf abzurufen, zentral (De Smedt, 2016). Geary (2011b) schlussfolgerte, dass es sich dabei um das konsistenteste Defizit bei Rechenschwäche handelt.

Bei der arithmetischen Kompetenzentwicklung vollzieht sich allgemein ein Wandel vom „zählenden Rechnen“, d. h. dem Lösen einfacher Rechenaufgaben durch Abzählen, zu effizienteren Lösungsstrategien wie Ableiten (Herleitung des Ergebnisses aus bereits bekanntem Wissen) oder Ergebnisabruf aus dem Gedächtnis. Diese Entwicklungsstadien können sich teilweise überlappen, sodass eine Zeitlang sowohl zählende als auch Abrufstrategien eingesetzt werden („overlapping waves theory“; Siegler, 2006). Es wird davon ausgegangen, dass arithmetische Fakten, insbesondere Multiplikationsfakten (das „kleine Einmaleins“), im Langzeitgedächtnis in Form phonologischer Codes abgespeichert sind (Dehaene, Piazza, Pinel & Cohen, 2003). Dadurch ließe sich erklären, warum Kinder mit Legasthenie oft eine isolierte Schwäche beim mathematischen Faktenabruf zeigen (Simmons & Singleton, 2008). Kinder mit Rechenschwäche verharren in der Regel beim zählenden Rechnen als dominierende Lösungsstrategie (Ostad, 1997), wodurch sie bei einfachen Rechenaufgaben deutlich mehr Zeit benötigen. Aus diesem Grunde stellt die Geschwindigkeit des arithmetischen Faktenabrufs einen guten Prädiktor für spätere, komplexere mathematische Fertigkeiten dar (Hecht, Torgesen, Wagner & Rashotte, 2001), und zwar auch am Ende der Grundschulzeit, wenn die curricularen Inhalte bereits komplexer sind (Nelson, Parker & Zaslofsky, 2016).

Zusammenfassend kann festgehalten werden, dass im späteren schulischen Kontext (am Ende der Grundschulzeit und darüber hinaus) die Geschwindigkeit des arithmetischen Faktenabrufs einen wichtigen Indikator für Rechenschwäche sowie einen guten Prädiktor für spätere Mathematikleistungen darstellt. Effekte basisnumerischer Kompetenzen werden über den Faktenabruf mediiert (z.B. Cirino, Tolar, Fuchs & Huston-Warren, 2016).

Die vorliegende Studie fokussiert auf den Faktenabruf, der empirisch betrachtet eine zunehmende Relevanz in der Vorhersage mathematischer Leistung während der Grundschulzeit gewinnt (vgl. Geary et al., 2017). Theoretisch ist diese Bedeutung mit der Entwicklung effizienter Strategien für die Lösung arithmetischer Aufgaben zu begründen. Verzögerungen in dieser Entwicklung sind vor allem, aber nicht nur, für Kinder mit einer Rechenschwäche indikativ.

Lernverlaufsbezogene Prädiktoren mathematischer Kompetenz

Die bisher diskutierten spezifischen und unspezifischen Prädiktoren mathematischer Kompetenz wurden in den dargestellten Studien zu diagnostischen Zwecken vornehmlich statusdiagnostisch verwendet, d. h. zu einzelnen Zeitpunkten erfasst. Einen komplementären Ansatz stellt die Lernverlaufsdiagnostik dar, bei der in regelmäßigen Abständen durch kurze Tests die interessierende Kompetenz gemessen wird und somit der individuelle Lernfortschritt, das Stagnieren oder das Abfallen der individuellen Leistung dokumentiert werden kann. Ziel der Lernverlaufsdiagnostik ist weniger die Klassifikation, sondern eher das Abbilden der Lernentwicklung und das daran orientierte formative Anpassen pädagogischer Aktivitäten (Souvignier, Förster & Zeuch, 2016). Deshalb beziehen sich Lernverlaufsmaße in der Regel auf domänen*spezifische* Kriterien, die einer Förderung potentiell zugänglich sind. Dazu zählen sowohl (a) empirisch valide, meist basale (vgl. „robust indicator"-Ansatz, Fuchs, 2004) als auch (b) curriculare (vgl. „curriculum-sampling"-Ansatz, ebd.) Aspekte des jeweiligen Kompetenzbereichs. Auch eine (c) entwicklungstheoretisch begründete Konstruktion von Lernverlaufsdiagnostika ist möglich (vgl. Balt, Ehlert & Fritz, 2017 für ein aktuelles Beispiel). Im deutschsprachigen Raum ist im Bereich Mathematik derzeit vor allem die zweite, curriculare, Klasse von Lernverlaufsdiagnostika gebräuchlich. Dies trifft auch auf diese Studie zu. Der Vorteil dieses Ansatzes ist, dass mathematische bzw. arithmetische Kompetenzen sehr curriculumsnah erfasst werden können, wodurch zugleich eine kriteriumsorientierte Messung ermöglicht wird (Klauer & Strathmann, 2013).

Um die individuellen Lernverläufe einzelner Schüler abzubilden, werden meist lineare Trends auf Schülerebene geschätzt. Dazu wird auf lineare Regressionsmodelle oder gemischte lineare Modelle zurückgegriffen (Snijders & Bosker, 2012). Bei der Schätzung individueller linearer Trends sind die beiden Parameter Regressionsinterzept (Intercept) sowie Regressionssteigung (Slope) von Interesse. Der Slope stellt dabei den mittleren Lernzuwachs einer Person dar, in den Intercept gehen sowohl Informationen zum individuellen Lernausgangsniveau („Startpunkt") als auch zum mittleren Leistungsniveau („Durchschnitt") ein.

Einige Studien untersuchten die prädiktive Validität individueller Slopes und berichten heterogene Ergebnisse. Die meisten Studien wurden zudem im Zusammenhang mit Lesefertigkeiten durchgeführt. So fanden Stage und Jacobsen (2001) eine Korrelation von $r = .26$ zwischen dem Slope bei *Oral Reading Fluency* (Lernverlauf Lesen) und dem Textverständnis. Schatschneider, Wagner und Crawford (2008) berichten hingegen, dass aus individuellen Lernverläufen im Lesen geschätzte Slopes keine inkrementelle Validität für die Vorhersage des Leseverständnisses am Ende des Schuljahres beitragen (s. auch Yeo, Fearrington & Christ, 2012). Im Bereich Mathematik fanden Keller-Margulis, Shapiro und Hintze (2008) zwar substanzielle Korrelationen zwischen den individuellen Slopes, die auf einem Lernverlaufstest zum Kopfrechnen basierten, und einem späteren mathematischen Kompetenztest ($r = .35$ - $.45$), die Korrelationen der individuellen Slopes mit dem Kompetenztest waren allerdings geringer als die Korrelationen der einzelnen Lernverlaufstests mit dem Kompetenztest. In einigen Studien zum Lesen zeigte sich zudem, dass die individuellen Intercepts stärkere Zusammenhänge mit einem statusdiagnostischen Kriterium aufwiesen als die – häufig statistisch nicht bedeutsamen – individuellen Slopes (Codding, Petscher & Truckenmiller, 2015).

Zusammenfassend kann festgestellt werden, dass mathematikspezifische Prädiktoren insgesamt einen Erklärungswert für die mathematische (hier: arithmetische) Leistungsentwicklung haben, der über unspezifische, d. h. allgemeine kognitive, Fähigkeiten hinausgeht. Von diesem inkrementellen Beitrag spezifischer Prädiktoren gehen wir auch in unserer Studie aus. Lernverlaufsbezogene Parameter weisen zwar in etlichen Arbeiten Zusammenhänge mit späterer schulischer Kompetenz auf, allerdings sind die Befunde inkonsistent.

Zudem gibt es bislang im Bereich Mathematik kaum Arbeiten, die gleichzeitig die prädiktive Validität von unspezifischen, spezifischen und lernverlaufsbezogenen Prädiktoren für die spätere mathematische Kompetenz untersuchen. Dies ist Gegenstand der vorliegenden Studie, die sich auf Kinder am Ende der Grundschulzeit konzentriert. Die Lernverlaufsdiagnostik erfolgte dabei zeitlich zwischen der Statusdiagnostik zu Schuljahresbeginn und der Messung des Kriteriums am Schuljahresende. Mit diesem Design kann mittels Mediatoranalysen geprüft werden, ob die spezifischen und unspezifischen Kompetenzen der Grundschulkinder direkt oder (auch) indirekt – über die spezifischen Lernverläufe vermittelt – zur Prognose arithmetischer Leistung am Schuljahresende dienen.

Mathematische Kompetenz und Rechenschwäche

Mathematische Kompetenz ist ein Konstrukt mit vielen Dimensionen, wovon Arithmetik nur eine ist (Dowker, 2015). Arithmetik, im Spezifischen die Minderleistung im Bereich der Grundrechenarten, bildet jedoch den definitorischen Kern des klinischen Konzepts „Rechenstörung“ bzw. „Dyskalkulie“. Damit eine klinische Diagnose vergeben werden kann, muss die Minderleistung substantiell sein, d. h. per Konvention mindestens 1,5 SD – oder 1 SD, wenn zusätzliche schulische und biographische Informationen vorliegen (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, 2018) – von der Norm abweichen. Einem solchen kategorialen diagnostischen Vorgehen, wie es in der klinischen Praxis nötig ist, steht die Perspektive auf mathematische Leistung als kontinuierlich verteiltes Merkmal gegenüber. In der vorliegenden Studie werden beide Perspektiven eingenommen: Arithmetikleistung wird sowohl dimensional als auch kategorial ausgewertet. Für die kategoriale Definition wird ein Cut-Off von -1 SD ($T \leq 40$) angelegt. Da keine umfangreiche klinische Diagnostik erfolgte, wird der Begriff „Rechenschwäche“ (statt „Rechenstörung“) verwendet.

Methode

Design und Stichprobe

Zu Beginn des Schuljahres 2015/16 wurde im Rahmen einer längsschnittlichen *Response-to-Intervention* (RTI)-Studie ein breit angelegtes Screening in 30 dritten und 28 vierten Grundschulklassen im Bundesland Nordrhein-Westfalen durchgeführt. Insgesamt wurde eine Gesamtstichprobe von N = 723 Grundschulkindern mit einer Screeningbatterie von kognitiven und Schulleistungstests untersucht. In der vorliegenden Studie wurden nur die Daten der Teilstichprobe analysiert, die während der RTI-Studienphase keine Förderung erhielt (N = 196 Kinder, n = 18 Schulklassen). Diese Teilstichprobe umfasst 99 Drittklässler und 97 Viertklässler, davon 107 Mädchen und 89 Jungen.

Tabelle 1: Deskriptive Stichprobenmerkmale

Variable	***M***	***SD***	**Range (*Min – Max*)**
Alter (in Monaten)	106.12	7.53	91.70 – 128.00
DEMAT Arithmetik, Prätest (T-Wert)	49.45	9.26	24 – 73
DEMAT Arithmetik, Posttest (T-Wert)	50.16	10.16	20 – 76
CFT (IQ-Wert)	105.21	14.68	61 – 148
SLS (Lesequotient)	98.08	13.95	67 – 131
DIRG M100 (T-Wert)	43.73	11.02	24 – 66
LVD-M[1], Zeitpunkt 1	13.32	5.65	0 – 24
LVD-M[1], Zeitpunkt 2	13.66	5.66	0 – 24
LVD-M[1], Zeitpunkt 3	14.02	5.84	0 – 24
LVD-M[1], Zeitpunkt 4	14.74	6.00	0 – 24
LVD-M[1], Zeitpunkt 5	15.18	6.08	0 – 24
LVD-M[1], Zeitpunkt 6	15.72	5.90	0 – 24
LVD-M[1], Zeitpunkt 7	16.14	5.81	0 – 24
LVD-M[1], Zeitpunkt 8	16.14	5.86	0 – 24
LVD-M[1], Zeitpunkt 9	16.65	5.55	0 – 24
LVD-M[1], Zeitpunkt 10	16.07	6.06	1 – 24
LVD-M[2] Intercept-Parameter	0.09	0.08	-2 – 2.91
LVD-M[2] Slope-Parameter	-0.08	0.92	-2.9 – 1.96

Anmerkungen: [1]Anzahl korrekt gelöster Aufgaben. [2]Empirische Bayes Schätzer (Normwerte, s. u.). DEMAT = Deutscher Mathematiktest (DEMAT 2+, DEMAT 3+, DEMAT 4), CFT = Grundintelligenztest (Skala 1 bzw. Skala 2), SLS = Salzburger Lesescreening, DIRG M100 = Modul M100 aus dem Diagnostischen Inventar zu Rechenfertigkeiten im Grundschulalter, LVD-M = Lernverlaufsdiagnostik Mathematik für zweite bis vierte Klassen.

Neben der lehrplanbezogenen mathematischen Kompetenz im Bereich Arithmetik wurden zu Beginn des Schuljahres grundlegende Rechenfertigkeiten (arithmetisches Faktenwissen im Bereich des Einmaleins), die sprachfreie Intelligenz und die Leseflüssigkeit erfasst. Im zweiten Schulhalbjahr wurde die individuelle Entwicklung curricularer Rechenleistung anhand regelmäßig durchgeführter Lernverlaufstests erfasst. Am Ende des Schuljahres wurde erneut die curriculare mathematische Kompetenz im Bereich Arithmetik untersucht. Deskriptive Merkmale der Stichprobe sind in Tabelle 1 zusammengefasst. Die eingesetzten Testverfahren werden in den folgenden Abschnitten erläutert.

Spezifische Prädiktoren

Curriculare Mathematikleistung (Arithmetik). Die curriculare Mathematikleistung wurde zu Schuljahresbeginn mit Subtests bzw. Aufgaben aus der Reihe Deutscher Mathematiktests (DEMAT), die an den Lehrplänen aller deutschen Bundesländer orientiert ist, erfasst (DEMAT Prätest). Dabei absolvierten die Viertklässler die Skala *Arithmetik* aus dem DEMAT 3+ (Roick, Gölitz & Hasselhorn, 2004), bestehend aus den Aufgaben Zahlenstrahlen, Additionen, Subtraktionen und Multiplikationen. Die Grundrechenarten müssen schriftlich bearbeitet werden, mit zwei- bis dreistelligen Zahlen. Die Zahlenstrahlaufgabe prüft die Konzepte Nachbarzahl, Zehner- und Hunderterübergänge. Die Drittklässler bearbeiteten zu Beginn des Schuljahres ausgewählte arithmetische Aufgaben aus dem DEMAT 2+ (Krajewski, Liehm & Schneider, 2004): Addition und Subtraktion, Verdoppeln, Division, Halbieren und Rechnen mit Geld. Die Aufgaben bewegen sich im Zahlenraum bis 100 bzw. 1€ (Rechnen mit Geld).

Zum Ende des Schuljahres bearbeiteten die Drittklässler die oben beschriebene Skala *Arithmetik* aus dem DEMAT 3+, während die Viertklässler mit der Skala *Arithmetik* aus dem DEMAT 4 (Gölitz, Roick & Hasselhorn, 2006) getestet wurden (DEMAT Posttest). Letzterer umfasst eine Zahlenstrahlaufgabe sowie Aufgaben zur schriftlichen Addition, Subtraktion, Multiplikation und Division. Alle Aufgaben des DEMAT 4 bewegen sich im Zahlenraum > 1000. Auswertungsgrundlage war die Anzahl korrekt bearbeiteter Aufgaben.

Es wurden entweder die *T*-Wert-Normen zum jeweiligen Schuljahresbeginn bzw. -ende für die Arithmetik-Skala (DEMAT 3+, DEMAT 4) oder der mittlere *T*-Wert für die einzelnen arithmetischen Aufgaben (DEMAT 2+) verwendet.

Arithmetischer Faktenabruf (Multiplikation). Der arithmetische Faktenabruf wurde mit dem Modul M100 aus dem Diagnostischen Inventar zu Rechenfertigkeiten im Grundschulalter (DIRG; Grube, Weberschock, Blum & Hasselhorn, 2010) erfasst, welches Aufgaben zum kleinen Einmaleins beinhaltet. Die Bearbeitungszeit beträgt 4 Minuten. In dieser Zeit sollen so viele Aufgaben wie möglich gelöst werden. Insgesamt umfasst der Test 110 zufällig angeordnete Aufgaben, wobei *tie problems* (z. B. 4 · 4) enthalten sind, jedoch keine Multiplikationsaufgaben mit den Faktoren 0 oder 1.

Auswertungsgrundlage war die Anzahl korrekt bearbeiteter Aufgaben. Da erst ab dem Ende des dritten Schuljahres Testnormen für das Modul M100 vorliegen, wurden für die zu Beginn des Schuljahres getesteten Drittklässler basierend auf der Gesamtstichprobe der RTI-Studie eigene Normalrangwerte nach dem Vorgehen von Blom (1958) berechnet.

Unspezifische Prädiktoren

Intelligenz. Die fluide Intelligenz wurde über die für die jeweilige Klassenstufe geeignete Variante des sprachfreien, kulturfairen CFT erhoben. Die Drittklässler bearbeiteten den zweiten Teil des CFT 1-R (Weiß & Osterland, 2012) bestehend aus den Subtests Reihenfortsetzen, Klassifikationen und Matrizen. Die Viertklässler wurden mit der Kurzform (Teil 1) des CFT 20-R (Weiß, 2006) getestet, welcher topologisches Schlussfolgern als vierte Aufgabe umfasst. Die mit den klassenstufenspezifischen Normen ausgewerteten Ergebnisse werden auf der IQ-Skala berichtet.

Lesefertigkeiten. Die basale Leseflüssigkeit wurde mit dem Salzburger Lese-Screening für die Klassenstufen 2-9 (SLS 2-9; Mayringer & Wimmer, 2014) ermittelt. Bei diesem Test sollten die Kinder einfache Sätze lesen und angeben, ob die Aussagen korrekt oder falsch sind. Auswertungsgrundlage war die Anzahl korrekt bearbeiteter Sätze (Bearbeitungszeit 3 Minuten). Für das SLS 2-9 liegen klassenstufen- und geschlechtsspezifische Normwerte in Form eines Lesequotienten ($M = 100$, $SD = 15$) vor.

Lernverlaufsbezogene Prädiktoren

Um die arithmetische Kompetenzentwicklung im Laufe des zweiten Schulhalbjahrs zu erfassen, wurde zehnmal jeweils im Abstand von zwei Wochen die Lernverlaufsdiagnostik Mathematik für zweite bis vierte Klassen (LVD-M 2-4; Strathmann & Klauer, 2012) eingesetzt. Dieses Testverfahren umfasst Aufgaben zu den Grundrechenarten, die sich nach Komplexität und Schwierigkeit am jeweiligen klassenstufenbezogenen Lehrplan orientieren. Die Aufgabenmengen für die Klassenstufen 3 und 4 bestehen sowohl aus Kopfrechenaufgaben als auch Aufgaben zum schriftlichen Rechnen. Dem

Lehrplan entsprechend bearbeiten Kinder der dritten Klasse nur Aufgaben zum schriftlichen Addieren und Subtrahieren, während die Aufgabenmenge der vierten Klasse auch schriftliche Multiplikationen und Divisionen beinhaltet. Individuelle Testhefte des LVD-M 2-4 werden mittels stratifiziertem Itemsampling generiert, d. h. die theoretische Grundstruktur aller erzeugten Testhefte war identisch, allerdings unterschieden sich jeweils die konkret „gezogenen“ Testaufgaben. In dieser Studie wurden zehn verschiedene Testhefte mit zufälligen Aufgabesamples eingesetzt. Mit Hilfe eines ausbalancierten Designs wurde sichergestellt, dass a) alle zehn Versionen zu jedem Messzeitpunkt berücksichtigt wurden und b) jedes Kind über den Zeitraum hinweg jedes Heft genau einmal bearbeitete. Als Testscore wurde die Anzahl korrekt gelöster Aufgaben je Testzeitpunkt verwendet. Die lernverlaufsbezogenen Parameter (Intercept, Slope) wurden basierend auf den individuellen Lernverläufen geschätzt (s. nächster Abschnitt). Den Lehrkräften sowie den Schülern wurde nach fünf Lernverlaufsmessungen eine Rückmeldung über die Lernverläufe gegeben (Lehrkräfte: Schüler- und Klassenebene, Schüler: Rückmeldung individueller Ergebnisse).

Statistische Analysen

Für alle hier beschriebenen Analysen wurde die Statistiksoftware R (Version 3.4.3; R Core Team, 2017) verwendet. In einem ersten Analyseschritt wurden mit der Funktion lmer aus dem R-Paket lme4 (Bates, Maechler, Bolker & Walker, 2015) zunächst zwei verschiedene gemischte lineare Modelle an die Lernverlaufsdaten (LVD-M 2-4) angepasst und verglichen: Random-Intercept-Modell (RI) und Random-Intercept-Random-Slope-Modell (RIRS; z. B. Snijders & Bosker, 2012). Auf der Basis des RIRS-Modells wurde dann für jedes Kind der sich individuell aus dem Modell ergebende Intercept- und Slope-Parameter vorhergesagt (empirische Bayes-Schätzer; Snijders & Bosker, 2012). Im Anschluss wurden, basierend auf der Gesamtstichprobe, eigene klassenstufenspezifische Normwerte für die Intercept- und Slope-Parameter berechnet, um eine vergleichbare Metrik für die beiden Klassenstufen zu erhalten. Die Berechnung der Normalrangwerte erfolgte nach dem Vorgehen von Blom (1958). Zur Parameterschätzung wurde ein Restricted Maximum Likelihood (REML)-Ansatz verwendet.

In einem zweiten Analyseschritt wurden alle Variablen mittels der adjustierten Mahalanobis-Distanz zunächst auf multivariate Ausreißer überprüft (R-Paket MVN; Korkmaz, Goksuluk & Zararsiz, 2016). Dann wurde die curriculare Mathematikleistung am Ende des Schuljahres anhand der beschriebenen spezifischen, unspezifischen und lernverlaufsbezogenen Prädiktoren vorhergesagt. Alle Variablen, die durch die Verwendung klassenstufenspezifischer Testhefte und/ oder Normwerte für Unterschiede zwischen den Jahrgangsstufen adjustiert wurden, wurden zuvor in z-Werte überführt. Es wurden beide Klassenstufen (3 und 4) gemeinsam analysiert, da keine signifikanten Interaktionseffekte zwischen Prädiktorkoeffizienten und Klassenstufe nachweisbar waren. Zur Berücksichtigung der hierarchischen Datenstruktur (Schüler in Schulklassen) wurden ebenfalls gemischte lineare Modelle mittels lme4 geschätzt. Um die unterschiedlichen, genesteten Modelle mittels Likelihood-Ratio-Tests vergleichen zu können, wurden Maximum Likelihood (ML)-basierte Schätzer verwendet. Aufgrund der eher kleinen Anzahl von Klassen (n = 18) wurde bei der Schätzung der festen Effekte im vollständigen Modell der REML-Schätzer in Kombination mit der Kenward-Roger Korrektur (Kenward & Roger, 1997) verwendet. Dazu wurde das R-Paket pbkrtest genutzt (Halekoh & Højsgaard, 2014). Dieses Vorgehen führt in der Regel auch bei einer geringeren Zahl von Clustern zu unverzerrten Parameterschätzungen (McNeish, 2017a). Um die relative Bedeutsamkeit von spezifischen, unspezifischen und lernverlaufsbezogenen Prädiktoren zu vergleichen, wurde zudem eine für gemischte lineare Modelle geeignete Variante der Dominanzanalyse (Luo & Azen, 2013) durchgeführt. Dafür wurde auf eine Generalisierung des R^2-Index für die erklärte Varianz in gemischten und verallgemeinerten linearen Modellen zurückgegriffen (marginales R^2; Nakagawa & Schielzeth, 2013), der mittels des R-Pakets piecewiseSEM berechnet wurde (Lefcheck, 2016).

Die im zweiten Analyseschritt beschriebene Vorgehensweise wurde mittels eines gemischten logistischen Modells für die abhängige, dichotome Variable „Rechenschwäche ja/nein" wiederholt, wobei Rechenschwäche hier definiert wurde als ein *T*-Wert ≤ 40 im DEMAT am Ende des Schuljahres (Fischbach et al., 2013), während die Intelligenz nicht als Ausschlusskriterium diente (nur ein Schüler wies einen IQ < 70 auf). Eine Rechenschwäche wiesen am Ende des Schuljahres N=35 Kinder auf (N=7 Drittklässler/innen, N=28 Viertklässler/innen; N=12 Jungen, N=23 Mädchen). Unterschiede zur vorherigen Analyse bestanden darin, dass zur Parameterschätzung die Laplace-Approximation verwendet wurde, so dass die Kenward-Roger Korrektur nicht anwendbar war.

In einem dritten Schritt wurde eine Mediatoranalyse durchgeführt (MacKinnon, Fairchild & Fritz, 2007), um zu überprüfen, ob die spezifischen und unspezifischen Prädiktoren im Modell neben direkten Effekten auch indirekte, über die Lernverlaufsparameter vermittelte Effekte auf die Mathematikleistung (Arithmetik) am Ende des Schuljahres ausübten. Spezifische und unspezifische Prädiktoren stellten entsprechend die unabhängigen Variablen dar, die Mathematikleistung am Ende des Schuljahres die abhängige Variable. Als Mediatorvariablen dienten die zeitlich zwischen der abhängigen und den unabhängigen Variablen erhobenen Lernverlaufsparameter (Random-Intercept, Random-Slope). Aufgrund der hierarchischen Datenstruktur wurde eine Erweiterung des klassischen Vorgehens nach Baron und Kenny (1986) auf gemischte lineare Modelle verwendet (Krull & MacKinnon, 2001). Zur Schätzung der Modellgleichungen wurde erneut auf den im lme4-Paket implementierten REML-Schätzer mit Kenward-Roger-Korrektur zurückgegriffen. Anstatt die Standardfehler der indirekten Effekte zu berechnen, was im vorliegenden Fall komplex ist, wurden für letztere mittels der distribution-of-product-Methode (MacKinnon, Lockwood, Hoffman, West & Sheets, 2002) Konfidenzintervalle approximiert (R-Paket RMediation; Tofighi & MacKinnon, 2011). Bei geringerer Clusterzahl und kleineren Stichproben in hierarchischen Datensätzen hat sich dieses Vorgehen bei Mediatoranalysen als optimal erwiesen (z. B. McNeish, 2017b). Auf die Verwendung strukturgleichungsbasierter bzw. pfadanalytischer Mediatoranalysen (Preacher, Zyphur & Zhang, 2010) wurde angesichts der vergleichsweise geringen Zahl von Level2-Clustern (Schulklassen) in dieser Studie verzichtet. Aufgrund der großen Variabilität von Effektstärkemaßen für den indirekten Effekt bei Stichproben < 500 (Hayes, 2013) werden keine derartigen Maße berechnet. Die im dritten Analyseschritt beschriebene Vorgehensweise wurde ebenfalls mittels eines gemischten logistischen Modells für die abhängige, dichotome Variable „Rechenschwäche: ja/nein" wiederholt.

Ergebnisse

Basierend auf der adjustierten Mahalanobis-Distanz ergaben sich keine multivariaten Ausreißer, daher wurde der gesamte Datensatz verwendet. Zunächst wurden gemischte lineare Modelle zur Bestimmung der individuellen Lernverlaufsparameter berechnet (Vorhersage Testscore LVD-M durch Testzeitpunkt). Ein Vergleich zwischen RI-Modell und RIRS-Modell ergab einen substanziell besseren Modellfit für das RIRS Modell, $\Delta\chi^2(2) = 32.33$, $p < .001$, mit BIC(RI) = 9184.4, BIC(RIRS) = 9167.1. Anhand des RIRS Modells wurden entsprechend die Lernverlaufsparameter Intercept und Slope für jedes Kind vorhergesagt. Intercept und Slope korrelierten leicht negativ, $r = -.16$, $p = .02$. Der durchschnittliche Lernzuwachs (fester Effekt Slope) lag bei .31 ($SE = .03$, $p < .01$), während der durchschnittliche Intercept (fester Effekt) einen Wert von 13.70 aufwies ($SE = .37$, $p < .01$). Der Mittelwert von R^2 über alle Kinder lag bei $R^2 = .29$ ($SD = .23$), das R^2 für den festen Effekt Messzeitpunkt in der Gesamtstichprobe (Edwards et al., 2008) lag bei .09.

Im Anschluss wurden die Produkt-Moment-Korrelationen sowie die Semipartialkorrelationen zwischen den interessierenden Variablen berechnet (s. Tabelle 2) und die Multikollinearität mittels Variance Inflation-Faktoren (VIF) untersucht. Die Semipartialkorrelationen stellen die Korrelationen der um alle anderen Prädiktoren bereinigten Einzelprädiktoren mit dem unbereinigten Kriterium (DEMAT Arithmetik Posttest) dar und spiegeln (in quadrierter Form) wider, welcher Varianzanteil im Kriterium sich allein auf den jeweiligen einzelnen Prädiktor zurückführen lässt. Die VIF wurden anhand eines Multilevel-Regressionsmodells (random intercept) berechnet, in das alle Prädiktoren eingingen und welches den DEMAT Arithmetik Posttestwert mittels der dargestellten Prädiktoren vorhersagte.

Es zeigte sich, dass die vorhandene Multikollinearität der Prädiktoren für weitere statistische Analysen unproblematisch war (alle VIF < 4). Neben dem DEMAT Arithmetik Prätest wiesen nur die beiden Lernverlaufsparameter Intercept und Slope eine substanzielle Semipartialkorrelation mit dem DEMAT Arithmetik Posttest auf, was die „isolierte“ Bedeutsamkeit dieser Prädiktoren (bereinigt um alle anderen Prädiktoren) bei der Vorhersage der zukünftigen arithmetischen Leistung unterstreicht. Beim Lernverlaufsparameter Slope war im Gegensatz zu allen anderen Prädiktoren die Semipartialkorrelation statistisch bedeutsam größer als die Produkt-Moment-Korrelation ($z = -4.54$, $p < .001$), d. h. die Bedeutsamkeit des Slopes war nach statistischer Kontrolle der anderen Prädiktoren größer.

Tabelle 2: Produkt-Moment-Korrelationen (Prädiktoren, Kriterium) und Semipartialkorrelationen (um alle anderen Prädiktoren bereinigter Einzelprädiktor mit Kriterium)

	DEMAT Prätest	**DIRG M100**	**CFT**	**SLS**	**LVD-M Int[1]**	**LVD-M Sl[1]**	**DEMAT Posttest (*sr*)**
DEMAT Prätest	1						.16*
DIRG M100	.37**	1					.12
CFT	.39**	-.07	1				.08
SLS	.44**	.33**	.28**	1			-.01
LVD-M Int	.70**	.37**	.48**	.45**	1		.31**
LVD-M Sl	-.09	.13	-.12	-.05	-.16*	1	.23**
DEMAT Posttest	.63**	.43**	.37**	.37**	.70**	.15*	1

Anmerkungen: Produkt-Moment-Korrelationen unter der Diagonalen, Semipartialkorrelationen (*sr*) in der rechten Spalte. [1]Empirische Bayes Schätzer (s. Text). DEMAT = Deutscher Mathematiktest (DEMAT 2+, DEMAT 3+, DEMAT 4), CFT = Grundintelligenztest (Skala 1 bzw. Skala 2), SLS = Salzburger Lesescreening, DIRG M100 = Modul M100 aus dem Diagnostischen Inventar zu Rechenfertigkeiten im Grundschulalter, LVD-M Int = Intercept Lernverlaufsdiagnostik Mathematik für zweite bis vierte Klassen, LVD-M Slope = Slope Lernverlaufsdiagnostik Mathematik für zweite bis vierte Klassen. *$p < .05$, **$p < .01$.

Modellvergleiche

Im nächsten Schritt wurden gemischte lineare und logistische Modelle berechnet, um zu überprüfen, welche Prädiktoren (spezifisch: DEMAT Prätest und DIRG M100, unspezifisch: CFT und SLS 2-9, lernverlaufsbezogen: LVD-M 2-4 Intercept und Slope) für die Vorhersage der arithmetischen Kompetenz am Ende des Schuljahres bzw. für die Vorhersage einer Rechenschwäche von Bedeutung waren. Es ergab sich eine substanzielle Intraklassenkorrelation (ICC) von .26 (95% Konfidenzintervall für die RI-Varianz, basierend auf der Profile Likelihood: .33, .83) im gemischten linearen Modell, die ICC im gemischten logistischen Modell lag bei .43 (95% Konfidenzintervall für die RI-Varianz, basierend auf der Profile Likelihood: .83, 2.97). Damit ist in beiden Fällen mindestens ein RI-Modell zur adäquaten Beschreibung der hierarchischen Daten notwendig.

Für jeden Prädiktor wurde mittels Likelihood-Ratio-Test überprüft, ob ein RIRS-Modell besser auf die Daten passte als ein RI-Modell, was allerdings für keinen Prädiktor der Fall war. Im Folgenden wurden daher nur Analysen berechnet, die auf dem RI-Modell beruhen. Zunächst wurde ein gemischtes lineares bzw. logistisches RI-Modell berechnet (Tabelle 3). Dabei ergab sich das konsistenteste Bild für die lernverlaufsbezogenen Parameter, die sowohl für die Vorhersage arithmetischer Kompetenz als auch Rechenschwäche substanzielle Beiträge lieferten. Die Modelle unterschieden sich jedoch in anderer Hinsicht: Während die arithmetische Kompetenz zum Ende des Schuljahres erwartungsgemäß von der Leistung im DEMAT zu Schuljahresbeginn abhing, lieferte letztere keinen Beitrag zur Vorhersage einer Rechenschwäche. Im Unterschied dazu war die Leistung im DIRG prädiktiv für Rechenschwäche. Während die Leseflüssigkeit ohne Auswirkung blieb, trug die Intelligenzleistung zumindest zur Vorhersage der arithmetischen Kompetenz bei. Die aufgeklärte Varianz lag bei $R^2 = .57$ im linearen gemischten Modell und bei $R^2 = .69$ im logistischen gemischten Modell.

Im nächsten Schritt wurden für die drei Prädiktorklassen *spezifisch*, *unspezifisch* und *lernverlaufsbezogen* alle möglichen Sequenzen der schrittweisen Prädiktoraufnahme miteinander verglichen (z. B. unspezifisch → unspezifisch + spezifisch → unspezifisch + spezifisch + lernverlaufsbezogen). Die Modellvergleiche der genesteten Modelle wurden mittels Likelihood-Ratio-Test sowie der Informationskriterien AIC und BIC vorgenommen (Raftery, 1995). Die Prädiktoren wurden hier blockweise aufge-

Tabelle 3: Ergebnisse des gemischten linearen bzw. logistischen Modells zur Vorhersage der arithmetischen Kompetenz bzw. der Rechenschwäche (feste Effekte)

	Arithmetische Kompetenz				Rechenschwäche			
Parameter	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	-.03	.08	-.34	.74	-4.36	.96	-4.52**	< .01
DEMAT	.23	.08	3.03**	< .01	-.26	.49	-.53	.60
DIRG	.09	.06	1.58	.12	-1.07	.40	-2.65**	< .01
CFT	.12	.06	2.06*	.04	-.80	.41	-1.94	.05
SLS	-.01	.06	-.18	.86	.06	.46	.13	.89
LVD-M Int	.57	.08	6.82**	< .01	-2.72	.82	-3.31**	< .01
LVD-M Sl	.27	.05	5.01**	< .01	-.74	.37	-2.01*	.04

Anmerkungen: LVD-M Int = Interceptschätzung aus dem individuellen Lernverlauf, LVD-M Sl = Slopeschätzung aus dem individuellen Lernverlauf. $^*p < .05$, $^{**}p < .01$.

nommen, um die Analyse überschaubar zu halten. Die Ergebnisse sind in Tabelle 4 dargestellt.

Die Ergebnisse für die linearen und logistischen gemischten Modelle waren vergleichbar. Es zeigte sich, dass die lernverlaufsbezogene und spezifische Prädiktorklasse den Modellfit unabhängig davon substanziell verbesserten, welche anderen Prädiktoren bereits im Modell waren. Sowohl eine hohe arithmetische Kompetenz zu Beginn des Schuljahres als auch eine positive arithmetische Kompetenzentwicklung während des Schuljahres sind damit wichtige Voraussetzungen für gute arithmetische Fertigkeiten am Ende des Schuljahres bzw. dafür, dass eine Rechenschwäche vermieden werden kann. Die unspezifische Prädiktorklasse trug nur dann zur Varianzaufklärung bei, wenn die lernverlaufsbezogenen Prädiktoren noch nicht im Modell berücksichtigt waren, und wies auch bei alleiniger Betrachtung den schlechtesten Fit aller Prädiktorklassen auf.

Dominanzanalyse

Mittels einer Dominanzanalyse wurde der Frage nachgegangen, ob sich die relative Wichtigkeit der Prädiktorklassen in den vorliegenden Daten voneinander unterschied. Die Wichtigkeit von Prädiktoren im Regres-

Tabelle 4: Modellvergleiche zur Einschätzung der Bedeutsamkeit der Prädiktorklassen für die arithmetische Kompetenz und Rechenschwäche

	Arithmetische Kompetenz				Rechenschwäche			
Modell[1]	AIC	BIC	$\Delta\chi^2$	R^2	AIC	BIC	$\Delta\chi^2$	R^2
U	516.19	532.58		.196	149.92	163.03		.220
U + S	459.15	482.09	61.04**	.433	126.43	146.09		.523
U + S + L	407.06	436.56	56.09**	.568	107.88	134.10	22.55**	.694
U	516.19	532.58		.196	149.92	163.03		.220
U + L	416.22	439.16	103.97**	.520	111.85	131.52	42.07**	.610
U + L + S	407.06	436.56	13.16**	.568	107.88	134.10	7.97*	.694
S	471.45	487.84		.400	130.19	143.31		.477
S + U	459.15	482.09	16.30**	.433	126.43	146.09	7.77*	.523
S + U + L	407.06	436.56	56.09**	.568	107.88	134.10	22.55**	.694
S	471.45	487.84		.400	130.19	143.31		.477
S + L	407.26	430.21	68.19**	.567	108.11	127.78	26.08**	.668
S + L + U	407.06	436.56	4.20	.568	107.88	134.10	4.23	.694
L	417.25	433.64		.518	111.83	124.94		.573
L + U	416.22	439.16	5.03	.520	111.85	131.52	3.98	.610
L + U + S	407.06	436.56	13.16**	.568	107.88	134.10	7.97*	.694
L	417.25	433.64		.518	111.83	124.94		.573
L + S	407.26	430.21	13.99**	.567	108.11	127.78	7.72*	.668
L + S + U	407.06	436.56	4.20	.568	107.88	134.10	4.23	.694

Anmerkungen: Arithmetische Kompetenz = T-Wert DEMAT (gemischte lineare Modelle), Rechenschwäche = T-Wert DEMAT ≤ 40 vs. > 40 (gemischte logistische Modelle), $\Delta\chi^2$ = Teststatistik des Likelihood-Ratio-Tests, R^2 = marginales R^2 (Nakagawa & Schielzeth, 2013). [1]U = unspezifische Prädiktoren (CFT, SLS), S = spezifische Prädiktoren (DEMAT Prätest, DIRG M100), L = lernverlaufsbezogene Prädiktoren (LVD-M Intercept, Slope). *p < .05, **p < .01.

sionsmodell kann über verschiedene Maße bestimmt werden, z. B. bei der multiplen linearen Regression durch ΔR^2, also die Varianz, die ein ergänzend ins Modell aufgenommener Prädiktor zusätzlich aufklärt. Bei korrelierten Prädiktoren ist dieser Wert abhängig vom Zeitpunkt bzw. der Reihenfolge der Aufnahme eines Prädiktors ins Modell (Luo & Azen, 2013). Daher ist für die Bestimmung der Wichtigkeit von Prädiktoren bzw. Prädiktorklassen die Berücksichtigung aller möglichen Reihenfolgen der Prädiktoraufnahme notwendig. Genau das ist das Prinzip der Dominanzanalyse: Es werden alle möglichen Regressionsmodelle berechnet und für jeden einzelnen Prädiktor wird bestimmt, wie groß jeweils die inkrementelle Varianzaufklärung ΔR^2 gegenüber den anderen Prädiktoren im jeweiligen Modell ist. Durch den Vergleich der inkrementellen Varianzaufklärung zwischen zwei Prädiktoren über alle Modelle, in denen diese beiden betrachteten Prädiktoren nicht gemeinsam vorkommen, kann bestimmt werden, ob ein Prädiktor einen anderen „dominiert", d. h. für die Vorhersage des Kriteriums eine höhere Wichtigkeit aufweist. In einem Regressionsmodell mit drei Prädiktoren X_1, X_2, X_3 würde z. B. beim Vergleich der relativen Wichtigkeit von X_1 und X_2 geprüft, ob $R^2_{X_1} > R^2_{X_2}$ und $R^2_{X_1X_3} > R^2_{X_2X_3}$ ist. Trifft dies zu, dann wird X_2 von X_1 (vollständig) dominiert.

Mit der Dominanzanalyse (Tabelle 5) konnten die Ergebnisse der vorherigen Analysen repliziert werden: In Tabelle 4 ist erkennbar, dass die Aufnahme unspezifischer Prädiktoren die Modellgüte nicht verbesserte, wenn bereits spezifische und lernverlaufsbezogene Prädiktoren im Modell waren. Entsprechend zeigt Tabelle 5, dass in diesem Fall die inkrementelle Varianzaufklärung (U + S + L vs. S + L) bei .001 (lineares gemischtes Modell) bzw. bei .025 (logistisches gemischtes Modell) lag. Interessanter sind die Ergebnisse im Hinblick auf die relative Wichtigkeit der Prädiktorklassen. Hier ergab sich sowohl bei der Vorhersage arithmetischer Kompetenz als auch von Rechenschwäche eine Ordnung, bei der die lernverlaufsbezogene Prädiktorklasse die größte Wichtigkeit aufwies, gefolgt von der spezifischen Prädiktorklasse, welche wiederum die unspezifische Prädiktorklasse dominierte (L > S > U). Dies spiegelt sich insbesondere auch im Gesamtdurchschnitt von ΔR^2 für jede Prädiktorklasse über alle Modelle hinweg wider (letzte Zeile von Tabelle 5).

Mediatoranalyse

Um zu prüfen, in welchem Ausmaß spezifische und unspezifische Prädiktoren die Mathematikleistung zum Ende des Schuljahres direkt oder indirekt beeinflussen, wurde eine Mediatoranalyse durchgeführt. In dieser Analyse dienten die lernverlaufsbezogenen Parameter als Mediatorvariablen. Zunächst wurden sowohl für die arithmetische Kompetenz am Ende des Schuljahres (Y = DEMAT Posttest) bzw. das Vorliegen einer Rechenschwäche (Y = Rechenschwäche: ja/nein) die drei für Mediatoranalysen zentralen RI-Regressionsmodelle M1-M3 berechnet:

(1) M1: $Y = \text{DEMAT}_{\text{prä}} + \text{DIRG} + \text{CFT} + \text{SLS} + \text{LVDM}_{\text{int}} + \text{LVDM}_{\text{sl}}$

(2) M2: $\text{LVDM}_{\text{int}} = \text{DEMAT}_{\text{prä}} + \text{DIRG} + \text{CFT} + \text{SLS}$

(3) M3: $\text{LVDM}_{\text{sl}} = \text{DEMAT}_{\text{prä}} + \text{DIRG} + \text{CFT} + \text{SLS}$

Auf dieser Basis konnten dann die Parameter und Schätzfehler der indirekten Effekte für jeden Prädiktor berechnet werden (MacKinnon, 2008, S. 106ff.). Die direkten Effekte der Prädiktoren auf die abhängige Variable (M1) entsprechen den Werten in Tabelle 3, die Effekte der Prädiktoren auf die Mediatorvariablen (M2 und M3) sind in Tabelle 6 zusammengefasst (identische Gleichungen und Ergebnisse für lineare und logistische gemischte Modelle). Tabelle 7 zeigt die Ergebnisse für die indirekten Effekte.

Tabelle 6 zeigt, dass sich der Lernverlaufsparameter Intercept besser vorhersagen ließ als der Lernverlaufsparameter Slope

Tabelle 5: Dominanzanalyse zur Bestimmung der relativen Wichtigkeit der Prädiktorklassen im linearen (arithmetische Kompetenz) und logistischen (Rechenschwäche) gemischten Modell

Prädiktor-subset	**Arithmetische Kompetenz: zusätzlicher Beitrag (ΔR^2) durch…**				**Rechenschwäche: zusätzlicher Beitrag (ΔR^2) durch…**			
	R^2	**Unspezifisch (U)**	**Spezifisch (S)**	**Lernverlaufsbezogen (L)**	R^2	**Unspezifisch (U)**	**Spezifisch (S)**	**Lernverlaufsbezogen (L)**
k = 0 Durchschnitt		.196	.400	.518		.220	.477	.573
U	.196		.237	.324	.220		.303	.390
S	.400	.033		.167	.477	.046		.191
L	.518	.001	.049		.573	.037	.096	
k = 1 Durchschnitt		.017	.143	.246		.042	.200	.291
U + S	.433			.135	.523			.171
U + L	.520		.048		.610		.084	
S + L	.567	.001			.668	.025		
k = 2 Durchschnitt		.001	.048	.135		.025	.084	.171
U + S + L	.568				.694			
Gesamtdurchschnitt		.071	.197	.300		.096	.254	.345

Anmerkungen: R^2 = marginales R^2 (Nakagawa & Schielzeth, 2013). Durch k wird angegeben, wie viele Prädiktoren bereits im Modell sind, bevor der interessierende Prädiktor ins Modell aufgenommen wird (k = 0: noch kein Prädiktor, k = 1: ein Prädiktor etc.). So ist z. B. R^2 = .196 bei der Vorhersage arithmetischer Kompetenz allein durch die unspezifischen Prädiktoren, jedoch R^2 = .400 für die spezifischen Prädiktoren ($R^2_{X_1} > R^2_{X_2}$). Auch ist ΔR^2 für die spezifischen Prädiktoren größer, wenn sie ins Modell mit den lernverlaufsbezogenen Prädiktoren aufgenommen werden, als dies für die unspezifischen Prädiktoren der Fall ist (ΔR^2 = .049 vs. ΔR^2 = .001), so dass in diesem Fall $R^2_{X_1X_3} > R^2_{X_2X_3}$ gilt. Damit dominiert das spezifische Prädiktorset das unspezifische Prädiktorset.

(Intercept: R^2 = .576, Slope: R^2 = .044). So kann der eigentliche Lernzuwachs in Form des Slope-Parameters nur durch das arithmetische Faktenwissen zu Beginn der Studie vorhergesagt werden, während der Intercept zusätzlich durch die Intelligenz und curriculare arithmetische Fertigkeiten prädiziert wird. Anhand von Tabelle 7 wird klar, dass nur bei Prädiktoren mit substanziellen direkten Effekten (Tabelle 3) auch indirekte Effekte nachweisbar sind. In drei Fällen substanzieller indirekter Effekte (DEMAT und CFT als Prädiktoren für die arithmetische Kompetenz, DIRG als Prädiktor für Rechenschwäche) werden diese durch den Intercept-Lernverlaufsparameter vermittelt. Bei der Vorhersage der Rechenschwäche läuft der indirekte Effekt von DIRG zusätzlich über den Lernverlaufsparameter Slope.

Tabelle 6: Ergebnisse der Vorhersage der Mediatorvariablen Intercept und Slope (Modelle M2 und M3, feste Effekte)

Parameter	*b*	*SE*	*t*	*p*
Lernverlaufsparameter Intercept (R^2 = .576)				
Intercept	.13	.06	2.05	.05
DEMAT	.46	.06	8.05**	< .01
DIRG	.16	.05	3.35**	< .01
CFT	.24	.05	5.04**	< .01
SLS	.09	.05	1.81	.07
Lernverlaufsparameter Slope (R^2 = .044)				
Intercept	-.16	.11	-.15	.88
DEMAT	-.15	.09	-1.66	.10
DIRG	.16	.08	2.08*	.04
CFT	-.03	.08	-.35	.73
SLS	-.05	.08	-.64	.52

Anmerkungen: **p < .01, *p < .05.

Tabelle 7: Indirekte Effekte der Prädiktoren in der Mediatoranalyse

	Arithmetische Kompetenz			Rechenschwäche		
Parameter	*b*	*SE*	*95% CI*	*b*	*SE*	*95% CI*
DEMAT → Int → Y	**.10**	.04	**[.04, .18]**	-.12	.23	[-.57, .32]
DEMAT → Sl → Y	-.03	.03	[-.09, .001]	.04	.09	[-.12, .24]
DIRG → Int → Y	.02	.01	[.00, .04]	**-.17**	.08	**[-.36, -.03]**
DIRG → Sl → Y	.02	.01	[.00, .04]	**-.17**	.11	**[-.42, -.01]**
CFT → Int → Y	**.03**	.01	**[.01, .06]**	-.19	.11	[-.43, .00]
CFT → Sl → Y	.00	.01	[-.03, .02]	.02	.07	[-.11, .17]
SLS → Int → Y	.00	.01	[-.02, .01]	.01	.05	[-.10, .11]
SLS → Sl → Y	.00	.01	[-.01, .01]	.00	.05	[-.11, .09]

Anmerkungen: Int = Lernverlaufsparameter Intercept (LVD-M), Sl = Lernverlaufsparameter Slope (LVD-M), Y = abhängige Variable (DEMAT Posttest bzw. Rechenschwäche ja/nein), SE = Standardfehler (Delta-Methode), 95% CI = 95% Konfidenzintervallschätzung für indirekten Effekt (Tofighi & MacKinnon, 2011). Substanzielle indirekte Effekte sind fett markiert.

Diskussion

Zusammenfassung und Relevanz der Ergebnisse

Die vorliegende Arbeit hatte das Ziel, die arithmetische Kompetenz (curriculare Fertigkeiten) bzw. das Vorliegen einer Rechenschwäche anhand von unspezifischen, spezifischen und lernverlaufsbezogenen Prädiktoren vorherzusagen. Zusätzlich sollte die relative Wichtigkeit dieser drei Prädiktorklassen beurteilt werden. Abschließend sollte überprüft werden, inwieweit spezifische bzw. unspezifische Prädiktoren – neben ihren direkten Effekten – auch indirekt über die während des Schuljahres gemessenen individuellen Lernverläufe auf die arithmetische Kompetenz einwirken. Wir gehen im Folgenden auf die einzelnen Punkte ein.

Ein zentrales Ergebnis ist darin zu sehen, dass die lernverlaufsbezogenen Prädiktoren Intercept und Slope zusammengenommen sowohl spezifische als auch unspezifische Prädiktoren dominierten. Die beste Vorhersage der arithmetischen Kompetenz und der Rechenschwäche gelang also anhand einer einzelnen Prädiktorklasse: über Parameter, die die individuelle arithmetische Lernentwicklung während des Schuljahres repräsentieren. Dieses etwas überraschende Ergebnis unterscheidet sich von den Befunden von Schatschneider et al. (2008) und Yeo et al. (2012), bei denen Lernverlaufsparameter keine inkrementelle Validität gegenüber Statusdiagnostika aufwiesen. Eine Erklärung für diesen Befund könnte darin liegen, dass das hier verwendete Verfahren der Lernverlaufsdiagnostik kein einfaches Maß der Rechenflüssigkeit war, sondern die Curriculumsinhalte des jeweiligen Schuljahres abbildete und damit konzeptuell dem Kriterium (DEMAT) ähnelte. Auch in der Studie von Jitendra, Dupuis und Zaslofsky (2014), in der komplexere Lernverlaufsmaße benutzt wurden (Textaufgaben), war der Slope ein reliabler Prädiktor eines späteren mathematischen Kompetenztests.

Die beiden Lernverlaufsparameter waren in der vorliegenden Studie die einzigen Prädiktoren, die sowohl für die Vorhersage der arithmetischen Kompetenz als auch für die Vorhersage einer Rechenschwäche am Ende des Schuljahres bedeutsam waren. Anders gelagert waren die Ergebnisse für die spezifischen Prädiktoren: Während der DEMAT-Wert zu Beginn des Schuljahres den DEMAT-Wert am Ende des Schuljahres erwartungsgemäß vorhersagte, war die Vorhersage einer Rechenschwäche durch den DEMAT-Prätest im Gesamtgefüge aller Prädiktoren nicht möglich. Spiegelbildlich dazu erscheinen die Ergebnisse im Faktenabruf (DIRG): der Abruf von Multiplikationsfakten prädizierte zwar im vollständigen Modell nicht die arithmetische Kompetenz, wohl aber das Vorhandensein einer Rechenschwäche. Einfache Multiplikationsaufgaben scheinen im unteren Leistungsbereich damit einen höheren Informationsgehalt aufzuweisen als für das gesamte Leistungsspektrum.

Die sich anhand der Modellvergleiche sowie der Dominanzanalyse ergebende Prädiktorordnung „Lernverlaufsbezogene Prädiktoren > spezifische Prädiktoren > unspezifische Prädiktoren“ (L > S > U) ist aus Sicht der Praxis aus mehreren Gründen von Bedeutung: Erstens legt sie nahe, dass curriculumsorientierte Lernverlaufsdiagnostika im Hinblick auf die Kompetenzentwicklung einen hohen Stellenwert besitzen sollten. Trotz des größeren Aufwandes der Durchführung erhöht diese formative Ergänzung üblicher diagnostischer Bemühungen die Wahrscheinlichkeit für adäquates pädagogisches Handeln und das frühzeitige Erkennen von Defiziten. Dies ist beispielsweise bei gegenwärtig diskutierten und implementierten Förderansätzen wie „Response to Intervention“ relevant (Huber & Grosche, 2012). Zweitens spielen unspezifische Prädiktoren keine Rolle, sofern lernverlaufsbezogene und spezifische Prädiktoren berücksichtigt werden. Zumindest die Intelligenz scheint zwar mit der arithmetischen Kompetenz assoziiert, angesichts der eher wenig überzeugenden teststatistischen Ergebnisse ist die Priorität der Intelligenz bei der direkten Prognose der Arithmetikleistung allerdings als gering einzustufen. Dies gilt nicht zuletzt auch im Hinblick auf gegenwärtig diskutierte strengere Maßstäbe für statistische Signifikanz, um die Belastbarkeit empirischer Befunde im Hinblick auf die statistische Hypothesenprüfung zu verbessern (Benjamin et al., 2018), sowie die Ergebnisse der lernverlaufsbezogenen und spezifischen Prädiktoren. Die Leseflüssigkeit zeigte keinen substanziellen Zusammenhang mit der curricularen Arithmetikleistung am Ende der Grundschulzeit.

Die Mediatoranalyse ergab, dass manche Prädiktoren indirekt über die Lernverlaufsparameter die spätere arithmetische Kompetenz oder das Bestehen einer Rechenschwäche beeinflussten. Allgemein wirken drei der vier indirekten Effekte über den lernverlaufsbezogenen Intercept-Parameter, was die Bedeutung dieses Parameters unterstreicht und im Einklang mit den meisten Ergebnissen in der Literatur steht (Codding et al., 2015). Einzig der Faktenabruf (DIRG) wirkt sich indirekt über den Slope-Parameter auf das Vorliegen einer Rechenschwäche aus: ein effizienterer Faktenabruf ist mit höheren Lernzuwächsen assoziiert, was wiederum das Risiko für eine Rechenschwäche senkt. Dennoch scheinen die Intercept-Parameter bedeutsamer, wobei dieser Befund mit der vergleichsweise geringeren Reliabilität der Slopes erklärt werden könnte (Schatschneider et al., 2008). Angesichts der hohen Bedeutsamkeit stellt sich die Frage, was die individuellen Intercept-Parameter genau repräsentieren: Bei stark linearen Zusammenhängen stellen sie den Ausgangspunkt (Lernausgangsniveau) dar, bei geringer ausgeprägt linearen Verlaufsmustern eher das mittlere Leistungsniveau. Das Ausmaß der Linearität des Lernverlaufs variierte jedoch von Schüler zu Schüler, so dass auch die konzeptuelle Bedeutung dieses Parameters unterschiedlich ist. Post-Hoc-Analysen ergaben, dass der individuelle Intercept-Parameter mit dem Mittelwert aller LVD-M Messungen über das Schuljahr hoch korrelierte (r = .75), der Slope-Parameter korrelierte mit diesem Mittelwert deutlich geringer (r = .16). Das spricht dafür, dass der Intercept-Parameter die durchschnittliche Leistung approximiert. Trotz der hohen Bedeutsamkeit des Lernverlaufs-Intercepts in der vorliegenden Studie ist zu betonen, dass der Slope-Parameter ebenfalls einen substanziellen Beitrag zur Prädiktion arithmetischer Kompetenz aufwies (vgl. Tabellen 2 und 3). Vor allem in Kontexten, die über eine reine Klassifikationsdiagnostik hinausgehen, z. B. bei der Förderplanung, ist der Slope eine wichtige Informationsquelle (Kovalevski, VanDerHeyden & Shapiro, 2013).

Zusammenfassend empfiehlt sich für die Praxis mathematikspezifische Diagnostik mit wiederholten Messungen. Bei der Abbildung von Lernverläufen mittels Regressionsgleichungen sollte nicht nur auf den Slope-, sondern auch auf den Intercept-Parameter geachtet werden. Für die Vorhersage einer arithmetischen Minderleistung im Sinne einer Rechenschwäche kann die Kompetenz im arithmetischen Faktenwissen zusätzlich informativ sein.

Limitationen

In der vorliegenden Studie konnte nur eine eng umrissene Menge an spezifischen und unspezifischen Prädiktoren untersucht werden. Die Ergebnisse in Regressionsmodellen hängen allgemein stark davon ab, welche Prädiktoren in das Modell eingehen, vor allem aber auch davon, welche Prädiktoren *nicht* im Modell sind (Harrell, 2015). So konnten hier keine Maße zur Arbeitsgedächtniskapazität, einem der bedeutsamsten unspezifischen Prädiktoren der mathematischen Kompetenz, eingesetzt werden (Peng, Namkung, Barnes & Sun, 2016). Allerdings wurde die Intelligenz berücksichtigt, die allgemein hoch mit der Arbeitsgedächtniskapazität korreliert (Oberauer, Schulze, Wilhelm & Süß, 2005). Auch nicht-kognitive Parameter wie der sozio-ökonomische Status oder die Unterrichtsqualität konnten nicht berücksichtigt werden, wobei Effekte der Klassenzusammensetzung zumindest indirekt über gemischte Modelle abgebildet wurden. Insbesondere im Hinblick darauf, dass Lernverlaufsdiagnostik als formatives Instrument zur Verbesserung pädagogischer Entscheidungen dienen sollte, wäre es für zukünftige Studien wünschenswert, die Unterrichtsqualität sowie die Verwertung der Lernverlaufsergebnisse durch die Lehrkräfte zu dokumentieren und in die Analyse einzubeziehen.

Die längsschnittliche Messung erfolgte innerhalb eines relativ eng umrissenen Zeitraums (ein Schuljahr, Ende der Grundschulzeit). Die Interpretation der Ergebnisse muss auf diese Zeitspanne bezogen werden und unterscheidet sich damit von Studien, die die Effekte von frühen Vorläuferfähigkeiten untersuchen (z. B. Krajewski & Schneider, 2009). Das Design dieser Studie ist also nicht dazu geeignet, Entwicklungspfade aufeinander aufbauender mathematischer Kompetenzen abzubilden, wie es für die Validierung ätiologischer Modelle nötig wäre. Der Schwerpunkt der Arbeit lag stattdessen auf dem Vergleich unspezifischer und spezifischer Prädiktoren in der prognostischen (diagnostischen) Validität zur Vorhersage von Arithmetikleistung bzw. -schwäche. Dabei wurde die Rolle lernverlaufsbezogener Daten als partieller Mediator zwischen spezifischem und unspezifischem Leistungsstatus zu Beginn des Schuljahres und Arithmetikleistung am Schuljahresende untersucht. So konnte zwischen direkten und indirekten Effekten unterschieden werden.

Obwohl Rechenschwäche in dieser Studie in Anlehnung an die Literatur als unterdurchschnittliche Leistung ($T \leq 40$) definiert wurde und damit inhaltlich gerechtfertigt werden kann, basiert die Klassifikation auf der Diskretisierung eines stetigen Merkmals, was mit einigen potenziellen Nachteilen (z. B. Verlust von Power) einhergehen kann (Harrell, 2015). Aus dieser Perspektive sind die Analysen mit Vorsicht zu interpretieren. Folgt man jedoch der in der Praxis etablierten Vorgehensweise der Kategorisierung von Leistungsdaten, können die Ergebnisse der gemischten logistischen Regressionsmodelle als hilfreiche Ergänzung im Hinblick auf die Aufnahme von Förderentscheidungen betrachtet werden.

Danksagung
Diese Arbeit wurde finanziell durch das Bundesministerium für Bildung und Forschung unterstützt (FKZ: 01GJ1302).

Literatur

Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (2018). S3-Leitlinie: Diagnostik und Behandlung der Rechenstörung. URL: http://www.awmf.org/uploads/tx_szleitlinien/028-046l_S3_Rechenstörung-2018-03_1.pdf (Zugriff: 29.05.2018).

Balt, M., Ehlert, A. & Fritz, A. (2017). Theoriegeleitete Testkonstruktion dargestellt am Beispiel einer Lernverlaufsdiagnostik für den mathematischen Anfangsunterricht. *Empirische Sonderpädagogik*, 2, 165-183.

Baron, R. M. & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Baroody, A. J., Bajwa, N. P. & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews, 15*, 69-79.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Benjamin, D. J., Berger, J., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R. et al. (2018). Redefine statistical significance. *Nature Human Behavior, 2*, 6-10.

Blom, G. (1958). *Statistical estimates and transformed beta-variables*. NY: Wiley.

Cirino, P. T., Tolar, T. M., Fuchs, L. S. & Huston-Warren, E. (2016). Cognitive and numerosity predictors of mathematical skills in middle school. *Journal of Experimental Child Psychology, 145*, 95-119.

Codding, R. S., Petscher, Y. & Truckenmiller, A. (2015). CBM reading, mathematics, and written expression at the secondary level: Examining latent composite relations among indices and unique predictions with a state achievement test. *Journal of Educational Psychology, 107*, 437-450.

Cowan, R. & Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *Journal of Educational Psychology, 106*, 214-229.

Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13-21.

Dehaene, S., Piazza, M., Pinel, P. & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*, 487–506.

De Smedt, B. (2016). Individual differences in arithmetic fact retrieval. In D. B. Berch, D. C. Geary & K. Mann Koepke (Eds.), *Development of Mathematical Cognition* (pp. 219-243). London: Academic Press.

Duncan, G .J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P. et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446.

Edwards, L. L., Muller, K. E., Wolfinger, R. D., Qaqish, B. F. & Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine, 27*, 6137-6157.

Ferrer, E. & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology, 40*, 935–952.

Fischbach, A., Schuchardt, K., Brandenburg, J., Klesczewski, J., Balke-Melcher, C., Schmidt, C. et al. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien. *Lernen und Lernstörungen, 2*, 65–76.

Fischer, U., Roesch, S. & Moeller, K. (2017). Diagnostik und Förderung bei Rechenschwäche: Messen wir, was wir fördern wollen? *Lernen und Lernstörungen, 6*, 25–38.

Fuchs, L. S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research. *School Psychology Review*, *33,* 188-193.

Fuchs, L. S., Geary, D. C., Fuchs, D., Compton, D. L. & Hamlett, C. L. (2016). Pathways to third-grade calculation versus word-reading competence: Are they more alike or different? *Child Development, 87*, 558–567.

Geary, D. C. (2011a). Cognitive predictors of individual differences in achievement growth in mathematics: A five-year longitudinal study. *Developmental Psychology, 47*, 1539–1552.

Geary, D. C. (2011b). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental & Behavioral Pediatrics, 32*, 250-263.

Geary, D. C., Nicholas, A., Li, Y. & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology, 109*, 680-693.

Göbel, S., Watson, S. E., Lervåg, A. & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate sense, that counts. *Psychological Science, 25,* 789-798.

Gölitz, T., Roick, D. & Haselhorn, M. (2006). *Deutscher Mathematiktest für vierte Klassen (DEMAT 4)*. Göttingen: Hogrefe.

Grube, D., Weberschrock, U., Blum, M. & Hasselhorn, M. (2010). *Diagnostisches Inventar zu Rechenfertigkeiten im Grundschulalter (DIRG)*. Göttingen: Hogrefe.

Gustafsson, J. E. & Undheim, J. O. (1992). Stability and change in broad and narrow factors of intelligence from ages 12 to 15 years. *Journal of Educational Psychology, 84,* 141–149.

Halekoh, U. & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models: The R package pbkrtest. *Journal of Statistical Software, 59*, 1-32.

Harrell, F. E., Jr. (2015). *Regression modeling strategies* (2nd ed.). NY: Springer.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. NY: Guilford.

Hecht, S. A., Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grade. *Journal of Experimental Child Psychology, 79*, 192-227.

Huber, C. & Grosche, M. (2012). Das response-to-intervention Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik, 8*, 312-322.

Jitendra, A. K., Dupuis, D. N. & Zaslofsky, A. S. (2014). Curriculum-based measurement and standards-based mathematics: Monitoring the arithmetic word problem-solving performance of third-grade students at risk for mathematics difficulties. *Journal of Learning Disabilities, 37*, 241-251.

Keller-Margulis, M. A., Shapiro, E. S. & Hintze, J. M. (2008). Long-term diagnostic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review, 37*, 374-390.

Kenward, M. G. & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997.

Klauer, K. J. & Strathmann, A. M. (2013). Lernverlaufsdiagnostik Mathematik: Test auf Änderungssensibilität bei rechenschwachen Grundschülern. *Psychologie in Erziehung und Unterricht, 60*, 241-252.

Kovalevski, J. F., VanDerHeyden, A. M. & Shapiro, E. S. (2013). *The RTI approach to evaluating learning disabilities*. New York: Guilford.

Korkmaz, S., Goksuluk, D. & Zararsiz, G. (2016). *MVN: Multivariate normality tests*. R package version 4.0.2.

Krajewski, K., Liehm, S. & Schneider, W. (2004). *Deutscher Mathematiktest für zweite Klassen (DEMAT 2+)*. Göttingen: Hogrefe.

Krajewski, K. & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction, 19*, 513-526.

Krull, J. L. & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249-277.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution, 7*, 573-579.

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D. et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*, 1753–1767.

Luo, W. & Azen, W. (2013). Determining predictor importance in hierarchical linear models using dominance analysis. *Journal of Educational and Behavioral Statistics, 38*, 3-31.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. NY: Erlbaum.

MacKinnon, D. P., Fairchild, A. J. & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.

Mayringer, H. & Wimmer, H. (2014). *Salzburger Lesescreening für die Schulstufen 2-9 (SLS 2-9)*. Göttingen: Hogrefe.

McNeish, D. (2017a). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research, 52*, 661-670.

McNeish, D. (2017b). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling, 24*, 609-625.

Moll, K., Kunze, S., Neuhoff, N., Bruder, J. & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PLOS ONE, 9*, e103537.

Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effect models. *Methods in Ecology and Evolution, 4*, 133-142.

Nelson, P. M., Parker, D. C. & Zaslofsky, A. F. (2016). The relative value of growth in math fact skills across late elementary and middle school. *Assessment for Effective Intervention, 41*, 184-192.

Oberauer, K., Schulze, R., Wilhelm, O. & Süß, H.-M. (2005). Working memory and intelligence – their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61-65.

Ostad, S. A. (1997). Developmental differences in addition strategies: a comparison of mathematically disabled and mathematically normal children. *British Journal of Educational Psychology, 67*, 345-357.

Östergren, R. & Träff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology, 115*, 405-421.

Peng, P., Namkung, J., Barnes, M. & Sun, C. (2016). A Meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*, 455-473.

Preacher, K. J., Zyphur, M. J. & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209–233.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wien. https://www.r-project.org/

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-163.

Roick, D., Gölitz, T. & Haselhorn, M. (2004). *Deutscher Mathematiktest für dritte Klassen (DEMAT 3+)*. Göttingen: Hogrefe.

Schatschneider, C. Wagner, R. K. & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and individual differences, 18*, 308-315.

Schneider, W., Küspert, P. & Krajewski, K. (2016). *Die Entwicklung mathematischer Kompetenzen (2. Aufl.)*. Paderborn: Schöningh.

Schwenk, C., Sasanguie, D., Kuhn, J.-T., Kempe, S., Doebler, P. & Holling, H. (2017). (Non-) symbolic magnitude processing in children with mathematical difficulties: a meta-analysis. *Research in Developmental Disabilities*, *165*, 152-167.

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon & R. M. Lerner (Series Eds.) D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology, Volume 2: Cognition, perception, and language* (6th ed., pp. 464–510). Hoboken, NJ: Wiley.

Simmons, F. R. & Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia, 14*, 77–94.

Singer, V. & Strasser, K. (2017). The association between arithmetic and reading performance in school: A meta-analytic study. *School Psychology Quarterly, 32*, 435-448.

Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel analysis (2nd ed.)*. London: Sage.

Souvignier, E., Förster, N. & Zeuch, N. (2016). Lernverlaufsdiagnostik. In K. Seifried, S. Drewes & M. Hasselhorn (Hrsg.), *Handbuch Schulpsychologie* (2. Aufl., S. 140-149). Stuttgart: Kohlhammer.

Stage, S. A. & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30*, 407-419.

Stern, E. (2013). Kognitive Entwicklungspsychologie des mathematischen Denkens. In M. von Aster & H. Lorenz (Hrsg.), *Rechenstörungen bei Kindern – Neurowissenschaft, Psychologie, Pädagogik* (S. 141-154). Göttingen: Vandenhoeck & Ruprecht.

Strathmann, A. M. & Klauer, K. J. (2012). *Lernverlaufsdiagnostik – Mathematik für zwei-*

te bis vierte Klassen (LVD-M 2-4). Göttingen: Hogrefe.

Tofighi, D. & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692-700.

Vanbinst, K. & De Smedt, B. (2016). Individual differences in children's mathematics achievement: The roles of symbolic numerical magnitude processing and domain-general cognitive functions. *Progress in Brain Research, 227*, 105-130.

Von Aster, M. G. & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology*, *49*, 868-873.

Weiß, R. H. & Osterland, J. (2012). *Grundintelligenztest Skala 1 – Revision (CFT 1-R*). Göttingen: Hogrefe.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 - Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest - Revision (WS/ZF-R)*. Göttingen: Hogrefe.

Yeo, S., Fearrington, J. Y. & Christ, T. J. (2012). Relation between CBM-R and CBM-mR slopes: An application of latent growth curve modeling. *Assessment for Effective Intervention, 37*, 147-158.

Prof. Dr. Jörg-Tobias Kuhn
Fakultät für Rehabilitationswissenschaften
TU Dortmund
Emil-Figge-Straße 50
44227 Dortmund
E-Mail: tobias.kuhn@tu-dortmund.de