

Lamb, Sarah; Schulz, Ann-Katrin; Kuhn, Jörg-Tobias
**Einflussfaktoren der Lehrkräteeinschätzung mathematikbezogener
Schwierigkeiten in der Grundschule**

Empirische Sonderpädagogik 17 (2025) 1, S. 35-49



Quellenangabe/ Reference:

Lamb, Sarah; Schulz, Ann-Katrin; Kuhn, Jörg-Tobias: Einflussfaktoren der Lehrkräteeinschätzung mathematikbezogener Schwierigkeiten in der Grundschule - In: Empirische Sonderpädagogik 17 (2025) 1, S. 35-49 - URN: urn:nbn:de:0111-pedocs-343647 - DOI: 10.25656/01:34364; 10.2440/003-0039

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-343647>

<https://doi.org/10.25656/01:34364>

in Kooperation mit / in cooperation with:



<https://www.psychologie-aktuell.com/journale/empirische-sonderpaedagogik.html>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nc/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und das Werk bzw. den Inhalt nicht für kommerzielle Zwecke verwenden.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by-nc/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work, provided that the work or its contents are not used for commercial purposes.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Einflussfaktoren der Lehrkräfteeinschätzung mathematikbezogener Schwierigkeiten in der Grundschule

Sarah Lamb, Ann-Katrin Schulz & Jörg-Tobias Kuhn

Technische Universität Dortmund

Zusammenfassung

Lehrkräfte stehen vor der Hausforderung, die Lernvoraussetzungen ihrer Schüler*innen (SuS) zutreffend einzuschätzen. Der Zusammenhang zwischen Lehrkräfteeinschätzungen und SuS-Leistungen in standardisierten Tests ist substanziell, doch die Urteilsgenauigkeit zwischen den Lehrkräften variiert. Zudem können urteilsrelevante sowie -irrelevante SuS-Merkmale das Lehrkräfteurteil beeinflussen. Bei der Einschätzung mathematikbezogener Schwierigkeiten können mathematikspezifische und -unspezifische SuS-Merkmale wie die Lesefertigkeit in das Lehrkräfteurteil einfließen. Zudem gelingt es Lehrkräften weniger gut, SuS im unterdurchschnittlichen Leistungsbereich präzise zu beurteilen, wodurch Förderbedarfe übersehen werden können. Basierend auf den Daten von $N = 377$ Grundschüler*innen und $N = 33$ Lehrkräften, wurde mittels generalisierter linearer gemischter Modelle (GLMM) untersucht, ob die Lehrkräfteeinschätzung über mathematische Schwierigkeiten von den tatsächlichen mathematikspezifischen Leistungen der SuS abhängt, oder ob dabei auch mathematikunspezifische SuS-Merkmale eine Rolle spielen. Zudem wurde geprüft, ob die Urteilsstrenge bzw. -milde zwischen den Lehrkräften variiert und ob das Ausmaß, in dem die SuS-Merkmale herangezogen werden, lehrkräfteabhängig ist. Die Ergebnisse zeigten, dass die tatsächliche mathematische Leistung der SuS prädiktiv für die Lehrkräfteeinschätzung war. Während die Einschätzung basisnumerischer Schwierigkeiten ausschließlich auf mathematikspezifischen Informationen beruhte, bezogen die Lehrkräfte bei der Einschätzung von Schwierigkeiten in den Rechenfertigkeiten auch die Intelligenz und Lesefertigkeit der SuS ein. Das Geschlecht der SuS hatte keinen Einfluss. Urteilsvariationen wiesen auf interindividuelle Unterschiede in der Urteilsstrenge hin, die Gewichtung verschiedener SuS-Merkmale variierte jedoch nicht. Insgesamt zeigen die Ergebnisse, dass die Lehrkräfteeinschätzung über die mathematischen Schwierigkeiten ihrer SuS vorrangig von deren tatsächlichen mathematischen Leistung bestimmt wird. Die Ergebnisse bieten wichtige Erkenntnisse für die Früherkennung von Rechenschwierigkeiten, die für inklusive Schul- und Unterrichtsprozesse relevant sind.

Schlagwörter: Diagnostische Kompetenz, Lehrkräfte, Einflussfaktoren, Rechenschwierigkeiten

Factors influencing teachers' assessments of math-related difficulties in primary school

Summary

Teachers face the challenge of accurately assessing the learning requirements of students. Even if the correlation between teachers' assessments and students' performance in standardized tests is substantial, diagnostic accuracy varies between teachers. Additionally, both relevant and irrelevant student characteristics, such as reading skills can influence the teachers' assessment of mathematical difficulties. Further, teachers are less successful at accurately assessing students in the below-average performance range, therefore learning difficulties can be overlooked. Based on data from $N = 377$ elementary school students and $N = 33$ teachers, generalized linear mixed models (GLMM) were used to investigate whether teachers' assessments of students' mathematical difficulties depend primarily on their actual mathematics-specific performance or whether -unspecific characteristics also play a role. We also examined whether the strictness or leniency varied between teachers and if the extent to which student characteristics are taken into account is teacher-dependent. The findings indicated that students' actual mathematics performance was predictive of the teachers' assessments. While the teachers' assessments of basic numerical difficulties were based exclusively on mathematics-specific information, teachers incorporated intelligence and reading skills when assessing difficulties in arithmetic performance. Students' gender had no influence. Variations in the teachers' assessments pointed to inter-individual differences in strictness. The weighting of different student characteristics did not vary. Overall, the findings indicate that teachers' assessments of students' mathematical difficulties are primarily determined by their actual mathematical performance. These results provide important insights for the early detection of mathematical difficulties, which are relevant for inclusive school and teaching practices.

Keywords: diagnostic competence, teachers, influencing factors, mathematical difficulties

Die Leistungsbewertung von Schüler*innen (SuS) gehört zum Lehrkräftealltag (Karst et al., 2014). Obwohl Lehrkräfteeinschätzungen generell recht präzise sind (Kaufmann, 2020; Südkamp et al., 2012), gelingt die Beurteilung leistungsschwacher SuS weniger gut (z.B. Lorenz, 2011; Wagner, 2024). Dadurch besteht die Gefahr, dass Minderleistungen übersehen werden und keine angemessenen Interventionen folgen. Gerade in aufeinander aufbauenden Unterrichtsfächern wie der Mathematik, in denen die Kompetenzentwicklung kumulativ verläuft (Salaschek et al., 2014), haben unentdeckte Lernschwierigkeiten massive Folgen für den weiteren Lern- und Entwicklungsverlauf der SuS (Schrader, 2013). Ein Vier-

tel der SuS weist am Ende der Grundschulzeit unterdurchschnittliche mathematische Kompetenzen auf (Selter et al., 2020). Etwa 5 % der SuS entwickeln eine persistierende Lernstörung (Fischbach et al., 2013). Auch wenn Rechenschwierigkeiten (RES) nicht einheitlich definiert sind, lassen sie sich durch charakteristische Schwierigkeiten präzisieren. Dazu zählen Beeinträchtigungen in den basalen (z.B. Mengenvergleiche und Aufgaben im einstelligen Bereich) und komplexen (z.B. Zahlenraumvorstellung im mehrstelligen Bereich) basisnumerischen Fertigkeiten (z.B. Lamb et al., 2024 a), die den Ausgangspunkt für die mathematische Kompetenzentwicklung bilden (Fischer et al., 2017). Weiter zeigen sich Defizite in der

Merkfähigkeit mathematischer Fakten (z.B. Einmaleins) sowie in den Rechenfertigkeiten und -strategien (z.B. Grundrechenarten; APA, 2013). Es ist die Aufgabe der Lehrkräfte, Lernschwierigkeiten frühzeitig zu erkennen (Hesse & Latzko, 2017). Dazu müssen die Leistungen der SuS zutreffend beurteilt werden (Karst, 2012). Bei der Beurteilung mathematischer Schwierigkeiten sollten mathematikspezifische SuS-Merkmale, darunter Rechen- und basisnumerische Fertigkeiten, das Lehrkräfteurteil bestimmen (Fischer et al., 2015). Domänengenerelle SuS-Merkmale, die nicht immanenter Bestandteil der mathematischen Fertigkeiten sind (z.B. Kuhn et al., 2019), sollten nicht substanziell in die Beurteilung einfließen. Doch Lehrkräfte tendieren dazu, die Ausprägung eines SuS-Merkmals (z.B. Intelligenz als zentrales kognitives Merkmal für gute Lernleistungen) als Indiz für ein anderes Merkmal (z.B. Mathematikleistung) heranzuziehen (z.B. Fischer et al., 2015; Kaiser et al., 2015; Mack et al., 2023). Zudem existieren Unterschiede in der Urteilsgenauigkeit und -strenge zwischen den Lehrkräften (z.B. Hesse & Latzko, 2017; Karing & Artelt, 2014; Südkamp et al., 2012). Daher wird in der vorliegenden Studie untersucht, welche mathematikspezifischen und -unspezifischen SuS-Merkmale prädiktiv für die Lehrkräfteeinschätzung von Schwierigkeiten in den mathematischen Fertigkeiten sind und ob sich die Lehrkräfte in ihren Einschätzungen unterscheiden.

Diagnosekompetenz von Lehrkräften

Die Diagnosekompetenz ist elementar für das Handeln von Lehrkräften (Baumert & Kunter, 2006) und eine Schlüsselkompetenz für inklusive Schul- und Unterrichtsprozesse (z.B. Schäfer & Rittmeyer, 2015). Das mehrdimensionale Konstrukt setzt sich aus verschiedenen Kompetenzbereichen und -facetten zusammen (z.B. Brunner et al., 2011). Im Kern handelt es sich dabei um die fachspezifische unterrichtliche Fähigkeit SuS zutreffend zu beurteilen und die

Lernanforderungen korrekt einzuschätzen (Karst, 2012). Dieses Verständnis ist gleichbedeutend mit dem Konstrukt der Urteilsgenauigkeit. Für den Bereich der Mathematik setzt dies u.a. Wissen über das mathematische Denken von SuS und Wissen über Leistungsbeurteilungen voraus (z.B. Brunner et al., 2011). Schrader (1989) unterteilt die Diagnosekompetenz von Lehrkräften in drei Komponenten: 1) die Niveauebene gibt Auskunft darüber, ob Lehrkräfte dazu neigen, die zu beurteilenden SuS-Merkmale zu über- oder zu unterschätzen. Dazu wird die Differenz zwischen dem Lehrkräfteurteil und der SuS-Leistung berechnet. 2) Die Differenzierungskomponente gibt an, ob Lehrkräfte die Leistungsunterschiedlichkeit von SuS in ihrer gesamten Bandbreite wahrnehmen oder unter- bzw. überdifferenzieren. Dazu wird das Verhältnis der Streuungen zwischen den Lehrkräfteeinschätzungen und den SuS-Merkmalen bestimmt. Die 3) Rangordnungskomponente als originärer Kennwert diagnostischer Kompetenz, gibt Aufschluss darüber, wie akkurat Lehrkräfte die Rangordnung der Leistung verschiedener SuS einschätzen können (Schrader & Helmke, 1987). Dazu wird der statistische (klassenspezifische) Zusammenhang zwischen den Lehrkräfteeinschätzungen (z.B. mathematische Fertigkeiten) und den objektiv gemessenen SuS-Merkmalen (z.B. Leistungstest) berechnet. Meta-Analysen, die die Urteilsgenauigkeit anhand von Korrelationen untersuchten, zeigten, dass die Einschätzungen der Lehrkräfte recht präzise sind (Kaufmann, 2020; Südkamp et al., 2012). Dennoch variiert die Urteilsgenauigkeit zwischen Lehrkräften (Lorenz 2011; Südkamp et al. 2012). Grund dafür sind verschiedene Faktoren, die Einfluss auf das Urteil haben können (z.B. Wagner, 2024).

Einflussfaktoren auf die Diagnosekompetenz

In dem heuristischen Modell der Akkuratheit diagnostischer Urteile von Lehrkräften (Südkamp et al., 2012) werden die Einfluss-

faktoren auf vier Bereiche zurückgeführt, die sich gegenseitig beeinflussen aber auch direkt auf die Urteilsgenauigkeit wirken können: Merkmale der Lehrkraft (z.B. Berufserfahrung), der SuS (z.B. Motivation), des Tests (z.B. Länge des Tests) und des Urteils (z.B. Bewertungsskala).

Kaiser et al. (2015) zeigten, dass höhere allgemeine kognitive Fähigkeiten und bessere Deutschleistungen mit positiveren Einschätzungen der Mathematikleistung durch Lehrkräfte einhergingen. Mack et al. (2023) stellten fest, dass die Lehrkräfteeinschätzungen in Mathematik und Deutsch signifikant durch die Leistungen in der jeweils anderen Domäne vorhergesagt werden. Diese Effekte lassen sich durch kognitive Verzerrungen, wie dem logischen Fehlschluss oder Halo-Effekt, erklären (z.B. Helmke, 2017). Dadurch besteht die Gefahr, dass SuS mit RES, die gute Deutschleistungen oder hohe kognitive Fähigkeiten zeigen, übersehen werden. Hinzu kommt, dass Lehrkräfte leistungsstarke SuS tendenziell unter- und leistungsschwache SuS häufiger überschätzen bzw. weniger akkurat beurteilen (z.B. Lorenz, 2011; Wagner, 2024), wodurch sich in der Konsequenz Unterschiede zwischen leistungsstarken und -schwachen SuS verstärken könnten.

Negative mathematikbezogene Stereotype zu Ungunsten von Schülerinnen (z.B. Holder & Kessels, 2017) können zudem dazu führen, dass Schüler von Lehrkräften als mathematisch begabter eingeschätzt werden (z.B. Lorenz, 2011; Mack et al., 2023). Damit beruht das Lehrkräfteurteil auf urteilsrelevanten sowie -irrelevanten Informationen.

Ein möglicher Grund dafür, dass Lehrkräfte leistungsgleiche SuS unterschiedlich einschätzen, ist der Strenge- oder Mildeffekt. Beim Strengeeffekt bewerten Lehrkräfte Leistungen systematisch strenger, während beim Mildeffekt Leistungen tendenziell weniger streng bewertet werden (z.B. Helmke, 2017; Hesse & Latzko, 2017).

Daher fokussiert sich diese Studie auf Einflussfaktoren und Lehrkräfteeffekte, die

prädiktiv für die Lehrkräfteeinschätzung mathematischer Schwierigkeiten von Grundschulkindern sein können.

Fragestellungen

Die Entwicklung und Ausbildung basaler basisnumerischer Fertigkeiten ist vorrangig mit dem Vorschulalter assoziiert (Fischer et al., 2017). Im Grundschulalter rückt die Ausbildung von Rechenfertigkeiten und -strategien in den Fokus. Daher wäre zu erwarten, dass Rechenfertigkeiten (z.B. Addition) sowie komplexere basisnumerische Fertigkeiten (z.B. Transkodieraufgaben) prädiktiv bedeutsamer für das Lehrkräfteurteil sind, während basale basisnumerische Fertigkeiten (z.B. Mengenvergleiche) aufgrund ihrer geringeren Relevanz im Unterricht auch in geringerem Ausmaß in das Lehrkräfteurteil einfließen. Bei der Einschätzung mathematischer Schwierigkeiten sollten die Rechen- und basisnumerischen Fertigkeiten der SuS das Lehrkräfteurteil bestimmen. Gemäß der Studienlage (z.B. Fischer et al., 2015; Kaiser et al., 2015; Mack et al., 2023) wird jedoch erwartet, dass die Lehrkräfte auch die Lesefertigkeit und Intelligenz ihrer SuS mit in das Urteil einfließen lassen. Möglicherweise neigen Lehrkräfte auch dazu das Geschlecht der SuS einzubeziehen und Schüler als mathematisch begabter einzuschätzen als Schülerinnen (z.B. Holder & Kessels, 2017; Mack et al., 2023). Daher untersucht die erste Forschungsfrage (FF1), *inwieweit Lehrkräfte mathematikspezifische und -unspezifische SuS-Merkmale bei der Einschätzung mathematischer Schwierigkeiten berücksichtigen*. Dazu schätzten die Lehrkräfte die Schwierigkeiten ihrer SuS in der Basisnumerik und den Rechenfertigkeiten mittels des Fragebogens zur Erfassung mathematischer Fertigkeiten (FERMAT; Lamb et al., 2024 b) ein. Der Screeningfragebogen adressiert die Diagnostikkompetenz. Die mathematikspezifischen und -unspezifischen SuS-Merkmale wurden über standardisierte Tests erhoben.

Erwartet wird, dass die Urteilsstrenge bzw. -milde zwischen den Lehrkräften variiert (Hesse & Latzko, 2017). Karing und Artelt (2014) zeigten, dass Lehrkräfte die Mathematikleistung ihrer SuS unterschiedlich präzise einschätzten. Sogar innerhalb einer Domäne (z.B. Mathematik) kann die individuelle Urteilsgenauigkeit variieren (Mack et al., 2023; Kolovou et al., 2021). Das Ausmaß, in dem mathematikspezifische und -unspezifische SuS-Merkmale herangezogen werden, könnte somit lehrkräfteabhängig sein. Daher untersucht die zweite Forschungsfrage (FF2), *a) ob die Urteilsstrenge bzw. -milde zwischen den Lehrkräften variiert und b) ob das Ausmaß, in dem mathematikspezifische und -unspezifische SuS-Merkmale herangezogen werden, lehrkräfteabhängig ist.*

Methode

Stichprobe

Erhoben wurden die Daten von $N = 377$ SuS und $N = 33$ Grundschullehrkräften aus 33 regulären Schulklassen in NRW. Die vorliegende Studie knüpft an die Arbeit von Lamb et al. (2024 b) an und bezieht sich auf dieselbe Stichprobe. Die SuS verteilen sich auf drei Jahrgangsstufen (Tab. 1). Die Testung der SuS erfolgte im Klassen- und Einzelsetting. Ein positives Ethikvotum sowie die Einverständniserklärungen liegen vor.

Messinstrumente

Mithilfe des computeradministrierten CODY-Mathetests 2 – 4 (CODY-M 2 – 4; Kuhn et al., 2017) wurden die basalen und komplexen basisnumerischen Fertigkeiten der SuS erhoben. Die Subskala „Basale Zahlenverarbeitung“ ($r_{tt} = .72$) besteht aus drei Untertests: Punkte zählen, symbolischer

Tabelle 1
Demografische Angaben und deskriptive Kennwerte

	Jahrgangsstufe		
	2	3	4
<i>n</i> (Jungen) ¹	172 (81)	154 (73)	50 (25)
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Alter in Monaten	98.01 _a (6.64)	110.66 _b (6.88)	121.31 _c (6.62)
<i>Basale und komplexe basisnumerische Fertigkeiten</i>			
CODY-M 2 – 4 (BZV)	99.41 _a (14.88)	100.15 _a (15.53)	101.58 _a (13.86)
CODY-M 2 – 4 (KZV)	100.28 _a (15.77)	99.35 _a (14.12)	101.02 _a (15.11)
<i>Rechenfertigkeiten</i>			
HRT 1 – 4	102.63 _a (14.37)	98.24 _b (14.71)	96.34 _b (16.86)
<i>Lesefertigkeit</i>			
SLS 1 – 4	99.56 _a (15.05)	100.24 _a (14.35)	100.78 _a (16.95)
<i>Intelligenz</i>			
CFT 1-R / 20-R	101.65 _a (14.86)	101.09 _a (14.06)	90.92 _b (15.39)

Anmerkungen. Alle Werte sind Normwerte. Skalierung der Erhebungsinstrumente ($M: 100, SD: 15$); ¹eine fehlende Geschlechtsangabe, $\chi^2(2) = 0.136, p = .935$; basale Zahlenverarbeitung (BZV); komplexe Zahlenverarbeitung (KZV); alle Gruppenvergleiche wurden nach Tukey korrigiert (ausgehend von einem Signifikanzniveau von $\alpha = .05$); Kleinbuchstaben, die von zwei Gruppen nicht geteilt werden, weisen auf statistisch signifikante Mittelwertunterschiede zwischen diesen Gruppen im Post-hoc-Vergleich (Jahrgangsstufe) hin.

und gemischter Mengenvergleich. Die vier Untertests Zahlendiktat, -steine und -strahl sowie Fehlende Zahl bilden die Skala „Komplexe Zahlenverarbeitung“ ($r_{tt} = .76$).

Mittels des Heidelberger Rechentests (HRT 1 – 4; Haffner et al., 2005) wurden die Rechenfertigkeiten erfasst. Eingesetzt wurden vier Subtests der Skala „Rechenoperationen“ (RO, $r_{tt} = .93$): Addition (z.B. $6 + 2$), Subtraktion (z.B. $8 - 2$), Ergänzungsaufgaben (z.B. $__ + 7 = 12$) und Größer-Kleiner-Vergleiche (z.B. $5 __ 61$).

Die Leseflüssigkeit wurde mittels des Salzburger Lese-Screenings (SLS 1 – 4; Mayringer & Wimmer, 2003) erhoben. Dazu wurde den SuS eine Liste mit 84 einfachen Sätzen vorgelegt. Diese mussten so schnell wie möglich gelesen und auf ihre Richtigkeit hin beurteilt werden ($r_{tt} = .87\text{--}90$).

Die Intelligenz der Zweit- und Drittklässler*innen wurde durch die Grundintelligenzskala 1 (CFT 1-R; Weiß & Osterland, 2012) erhoben. Verwendet wurden drei Untertests: Reihen fortsetzen, Klassifikationen und Matrizen ($\alpha = .95$). Die SuS der vierten Klasse bearbeiteten vier Untertests der Grundintelligenzskala 20-Revision (CFT 20-R; Weiß, 2006): Reihen fortsetzen, Klassifikationen, Matrizen und topologische Schlussfolgerungen ($r_{tt} = 0.80$).

Die Lehrkräfte schätzten die mathematischen Schwierigkeiten ihrer SuS mittels des FERMAT ein (Lamb et al., 2024 b). Die Subskala „Basisnumerische Fertigkeiten“ (FERMAT-BF, $\omega = .85$) umfasst acht zu bewertende Fertigkeiten: Zählen, das Lesen und Schreiben von Zahlen, Mengenverständnis, Zahlen ordnen und vergleichen, Stellenwertverständnis sowie die Zahlenraumvorstellung bis 10, bis 100 und bis 1000. In der Subskala „Rechenfertigkeiten“ (FERMAT-RF, $\omega = .91$, acht Items) werden Schwierigkeiten in den vier Grundrechenarten, Ergänzungsaufgaben, Textaufgaben, Rechenstrategien und Zählen als Rechenstrategie eingeschätzt. Da nicht sichergestellt werden konnte, dass die Themen Multiplikation, Division und der Zahlenraum bis 1000 bereits in allen Klassen Bestandteil

des Unterrichts waren, wurden diese Items von den Analysen ausgeschlossen. Die Subskala FERMAT-BF wurde daher aus sieben und die Skala FERMAT-RF aus sechs Items gebildet. Die Gesamtskala des FERMAT, die sich aus den beiden Teilbereichen zusammensetzt, wurde nicht untersucht, da diese mit der Subskala FERMAT-RF hoch korrelierte, $r = .97$, $p < .001$. Die Lehrkraft kreuzte ein Item an (Itemscore = 1), wenn SuS regelmäßig in den letzten drei Unterrichtsmonaten Schwierigkeiten in Bezug auf die beschriebenen Fertigkeiten zeigten. Die Einschätzung durch die Lehrkräfte erfolgte analytisch, nicht holistisch, und wurde durch klare Instruktionen und praktische Beispiele unterstützt (Abb. 1). Höhere Werte im FERMAT (Summe der Itemscores) deuten somit auf größere Schwierigkeiten der SuS in der Basisnumerik (FERMAT-BF: $n = 367$, $M = 0.31$, $SD = 0.94$, $Min = 0$, $Max = 7$) und/oder den Rechenfertigkeiten hin (FERMAT-RF: $n = 356$, $M = 0.94$, $SD = 1.63$, $Min = 0$, $Max = 6$).

Statistische Analysen

Die statistischen Analysen erfolgten mit R (Version 4.3.1; R Core Team, 2023). Um den Einfluss mathematikspezifischer und -unspezifischer SuS-Merkmale auf die Lehrkräfteeinschätzung von Schwierigkeiten in den mathematischen Fertigkeiten zu untersuchen (FF1) und dabei Lehrkräfteeffekte zu berücksichtigen (FF2), wurden Mehrebenenmodelle spezifiziert. Auf der ersten Ebene (Level 1) wurden die mathematikspezifischen (basale und komplexe basisnumerische Fertigkeiten sowie Rechenfertigkeiten) und -unspezifischen (Lesefertigkeit, Intelligenz und Geschlecht) SuS-Merkmale betrachtet. Auf der zweiten Ebene (Level 2) wurden die Lehrkräfte berücksichtigt. Ergänzend wurde geprüft, ob die Jahrgangsstufe prädiktiv für die Einschätzung der Lehrkräfte ist. Dazu wurde diese als fester Level-2-Prädiktor betrachtet. Als Kriterien wurden die beiden Teilbereiche Basisnumerik und Rechenfertigkeiten, in die sich

Abbildung 1.
Adaptierter Auszug aus dem FERMAT (Lamb et al., 2024 b)

Fragebogen zur Erfassung der mathematischen Fertigkeiten (FERMAT)	
Dieser Fragebogen bezieht sich auf Ihre Beobachtungen zu den mathematischen Fertigkeiten Ihrer Schüler*innen. Bitte machen Sie für jede*n Schüler*in im Fragebogen ein Kreuz, wenn Sie bei diesem*r Schüler*in die im Erklärungsbogen genannten typischen Schwierigkeiten regelmäßig beobachten. Bitte beziehen Sie sich dabei auf den Zeitraum der letzten drei Unterrichtsmonate. Falls etwas noch nicht im Unterricht behandelt wurde (z.B. Zahlenraum bis 1000, Division), streichen Sie bitte die entsprechende Spalte durch.	
Wann muss im Fragebogen ein Kreuz gemacht werden? Der*die Schüler*in zeigt bei der in der Spalte beschriebenen Fertigkeit mindestens eine der typischen Schwierigkeiten (s.u.) und/oder andere Schwierigkeiten regelmäßig.	
Fertigkeiten	Auszug aus dem Erklärungsbogen
Basisnumerische Fertigkeiten (BF)	
AK = Anzahlkonzept ☐	Vorstellung von Zahlen als Mengen (kardinaler Zahlenbegriff) Typische Schwierigkeiten: Unpräzises Mengenschätzen, Zahlen werden „mengenlos“ (nur als „Zählzahl“) gedacht
ZOV = Zahlen ordnen und vergleichen ☐	Orientierung im Zahlenraum (Zahlenvergleiche/Zahlenordnen, Vorgänger/Nachfolger) Typische Schwierigkeiten: Falsche Zahlenvergleiche/falsches Zahlenordnen (z.B. $975 > 1215$, da $9 > 1$), Vorgänger/Nachfolger werden falsch benannt (z.B. „welche Zahl kommt vor der 34?“ → „23“)
Rechenfertigkeiten (RF)	
TA = Textaufgaben ☐	„Vier Hunde liegen im Korb. Zwei kommen noch dazu und drei gehen weg. Wie viele Hunde sind noch im Korb?“ Typische Schwierigkeiten: Modellieren/Mathematisieren gelingt nicht (Text kann nicht in Rechenaufgabe „übersetzt“ werden), wesentliche Informationen zum Lösen der Aufgabe können nicht erkannt werden
R-ST = Rechenstrategien ☐	Verwendung von effizienten Rechenstrategien/operationales Verständnis <i>Kommutativgesetz:</i> Der*die Schüler*in weiß, dass $4 + 9 = 13$ ist, also muss $9 + 4$ auch $= 13$ sein <i>$n + 1 / n - 1$ Prinzip:</i> Der*die Schüler*in weiß, dass $23 + 44 = 67$ ist, also muss $23 + 45 = 68$ sein (analog bei Subtraktion) <i>$n \cdot 10 / n + 10$-Prinzip:</i> Der*die Schüler*in weiß, dass $26 + 72 = 98$ ist, dann ist auch $260 + 720 = 980$ (oder: $5 + 3 = 8 \rightarrow 15 + 3 = 18$) <i>Inversions-/Umkehrprinzip:</i> Wenn $46 + 27 = 73$, dann ist $73 - 27 = 46$ (auch: $11 - 3 = 8 \rightarrow 11 - 8 = 3$) Typische Schwierigkeiten: Der*die Schüler*in verwendet keine/extrem selten Rechenstrategien

der FERMAT gliedert, untersucht. Die kontinuierlichen Prädiktorvariablen auf Level-1 wurden basierend auf den klassenspezifischen Normwerten für die Gesamtstichprobe z-standardisiert. Bei den Kriterien handelt es sich um Rohwerte, die nicht z-standardisiert wurden.

Voranalysen. Erwartungsgemäß wählten die Lehrkräfte mit Abstand am häufigsten die Antwortkategorie 0 (= keine Schwierigkeiten) aus. Daraus resultierte eine extrem schiefe Verteilung der FERMAT-Testscores: FERMAT-BF: $g_1 = 4.05$, $z = 15.27$, $p < .001$; FERMAT-RF: $g_1 = 1.89$, $z = 10.25$, $p < .001$ (D'Agostino, 1970). Daher wurde auf eine lineare Modellierung verzichtet und stattdessen geprüft, ob die Datenstruktur besser durch eine Poisson- oder negative Binomialverteilung repräsentiert wird und ob die ‚zero inflation‘ statistisch zu berücksichtigen ist. Verwendet wurde das R-Paket „glmTMB“ (Brooks et al., 2017). Die generalisiert linearen gemischten Modelle (GLMM; z.B. Fahrmeir et al., 2007) mit den vier verschiedenen Verteilungsannahmen (Poisson-Modell und negatives Binomialmodell jeweils mit und ohne zero inflation) wurden mit einem Likelihood-Ratio-Test und anhand von Fit-Indizes verglichen (Dobson, 2002). Die Ergebnisse zeigten zusammengefasst, dass die Modellspezifikation mit dem negativen Binomialmodell ohne Berücksichtigung der ‚zero inflation‘ am besten zu den Daten passte.

Ergebnisse

Modellvergleiche. FF1 untersucht, inwieweit Lehrkräfte mathematikspezifische und -unspezifische SuS-Merkmale bei der Einschätzung von Schwierigkeiten in den mathematischen Fertigkeiten berücksichtigen. Dazu wurden zunächst Modelle mit konstantem Achsenabschnitt und konstanter Steigung mit mathematikspezifischen und -unspezifischen Level-1-Prädiktoren verglichen. Für das Kriterium FERMAT-RF verbesserte sich der Modellfit durch die Be-

rücksichtigung mathematikspezifischer SuS-Merkmale signifikant (Tab. 2). Für das Kriterium FERMAT-BF galt dies nicht.

In FF2a wurde getestet, ob die Urteilsstrenge bzw. -milde zwischen den Lehrkräften variiert. Dazu wurde ein klassen- bzw. lehrkraftbezogener Zufallseffekt bzgl. des Achsenabschnitts (random intercept) in die Zufallsstruktur des Modells aufgenommen. Der Modellvergleich für das Kriterium FERMAT-RF zeigte, dass sich die Modellanpassung erneut verbesserte (Tab. 2). Für das Kriterium FERMAT-BF verbesserte sich der Modellfit nicht. Ergänzend wurde die Jahrgangsstufe als fester Level-2-Prädiktor berücksichtigt, wodurch sich der Modellfit für beide Kriterien signifikant verbesserte (Tab. 2).

In FF2b wurde geprüft, ob das Ausmaß, in dem mathematikspezifische und -unspezifische SuS-Merkmale herangezogen werden, lehrkräfteabhängig ist. Aufgrund von Konvergenzproblemen mussten dazu alle kontinuierlichen Prädiktorvariablen nacheinander als zufällige Level-2-Prädiktoren modelliert werden. Infolgedessen wurden für das Kriterium FERMAT-RF fünf Modelle verglichen. Keiner der Modellvergleiche zeigte eine Verbesserung des Modellfits durch die Berücksichtigung eines zufälligen Level-2-Prädiktors. Daher wurde das random-intercept-constant-slope Modell mit mathematikspezifischen und -unspezifischen Prädiktoren als finales Analysemodell für das Kriterium FERMAT-RF gewählt (Tab. 3). Für das Kriterium FERMAT-BF wurde das constant-intercept-constant-slope Modell mit mathematikspezifischen Prädiktoren herangezogen (Tab. 3), da dieses am besten zu den Daten passte, d.h. Lesefertigkeiten und Intelligenz zeigten keinen systematischen Zusammenhang mit dem FERMAT-BF (Tab. 2).

FF1: Für die Lehrkräfteeinschätzung im Teilbereich FERMAT-BF erwiesen sich die basalen und komplexen basisnumerischen Fertigkeiten sowie die Rechenfertigkeiten als prädiktiv. Die Intelligenz und Lesefertigkeit sowie das Geschlecht flossen nicht in das Urteil ein (Tab. 3).

Tabelle 2
Modellvergleiche mit festen und zufälligen Effekten

Modellspezifikation	Modellfit			Likelihood-Ratio-Test	
	AIC	BIC	LL	df	χ^2
<i>FERMAT-BF</i>					
cics MS	410.10	429.59	-200.05		
cics MS + MU ^a	410.32	441.49	-197.16	3	5.79
rics MS ^a	410.19	433.58	-199.10	1	1.91
cics MS + Jgst ^a	398.19	425.47	-192.10	2	15.91***
<i>FERMAT-RF</i>					
cics MS	819.60	838.93	-404.80		
cics MS + MU	812.39	843.32	-398.19	3	13.21**
rics MS + MU	802.92	837.72	-392.46	1	11.47***
rics MS + MU + Jgst	800.29	842.82	-389.15	2	6.63*

Anmerkungen. *** $p < .001$; ** $p < .01$; * $p < .05$; constant-intercept-constant-slope (cics); random-intercept-constant-slope (rics); mathematikspezifische Prädiktoren (MS) auf Level-1 (SuS): basale und komplexe basisnumerische Fertigkeiten, Rechenfertigkeiten; mathematikunspezifische Prädiktoren (MU) auf Level-1 (SuS): Lesefertigkeit, Intelligenz, Geschlecht und Jahrgangsstufe (Jgst) als fester Level-2-Prädiktor; klassen- bzw. lehrkraftbezogener Zufallseffekt bzgl. des Achsenabschnitts (random intercept); ^a mit cics MS Modell verglichen.

Für die Lehrkräfteeinschätzung im Teilbereich FERMAT-RF waren die komplexen basisnumerischen Fertigkeiten, die Rechenfertigkeiten sowie die Intelligenz und Lesefertigkeiten der SuS prädiktiv. Das Geschlecht der SuS hatte keinen Effekt (Tab. 3).

Die negativen Koeffizienten der festen Effekte (*B*) zeigen, dass niedrigere objektive Testleistungen der SuS mit höheren FERMAT-Scores einhergingen (Tab. 3). Kinder die nach Einschätzung der Lehrkräfte mathematische Schwierigkeiten hatten, erreichten auch objektiv tendenziell niedrigere Testwerte. Dies galt für die mathematikspezifischen- und unspezifischen SuS-Merkmale.

Ergänzend wurde geprüft, ob die Jahrgangsstufe einen Effekt hat. Nach Einschätzung der Lehrkräfte hatten Kinder der dritten und vierten Jahrgangsstufe weniger Schwierigkeiten als Zweitklässler*innen. Ein tendenziell ähnlicher Effekt war auch zwischen den SuS der dritten und vierten Jahrgangsstufe zu beobachten (Tab. 3).

FF2a: Die geschätzte random intercept Varianz ($\tau_{00}^{Lehrkräfte}$) zeigte, dass sich die

Lehrkräfte in der Urteilsstrenge bzw. -milde unterschieden. Für identisch leistungsfähige SuS wurden unterschiedliche FERMAT-Scores (FERMAT-RF) vergeben (Tab. 3). Für den Teilbereich FERMAT-BF gab es keinen Hinweis darauf, dass sich die Lehrkräfte in ihrer Urteilsstrenge bzw. -milde unterscheiden (Tab. 2).

FF2b: Keiner der Modellvergleiche zeigte eine Verbesserung des Modellfits durch die Berücksichtigung variierender Steigungen auf der Lehrkräfteebene. Es gab keinen Hinweis darauf, dass die kontinuierlichen mathematikspezifischen- und unspezifischen Prädiktoren in ihrer Bedeutung für den FERMAT-Score zufällig zwischen den Lehrkräften variierten.

Tabelle 3
Ergebnisse der Analysemodelle

	FERMAT-BF		FERMAT-RF	
	B	SE(B)	B	SE(B)
Feste Effekte				
Intercept	-1.52	0.20	-0.54	0.20
Basale basisnumerische Fertigkeiten	-0.26	0.13	0.03	0.09
Komplexe basisnumerische Fertigkeiten	-0.52	0.16	-0.31	0.11
Rechenfertigkeiten	-0.91	0.18	-0.57	0.12
Lesefertigkeit	-	-	-0.23	0.11
Intelligenz	-	-	-0.35	0.10
Geschlecht ^a	-	-	0.27	0.18
Jgst 3 ^b	-1.21	0.30	-0.46	0.25
Jgst 4 ^b	-0.76	0.39	-0.92	0.37
Zufällige Effekte				
τ_{00} Lehrkräfte	-		0.18	
$N_{\text{Lehrkräfte}}$	32		31	
N_{SuS}	364		353	

Anmerkungen. *** $p < .001$, ** $p < .01$, * $p < .05$. Mathematikspezifische Prädiktoren auf Level-1 (SuS): basale basisnumerische und komplexe basisnumerische Fertigkeiten, Rechenfertigkeiten; mathematikunspezifische Prädiktoren auf Level-1 (SuS): Lesefertigkeit, Intelligenz, Geschlecht, ^a das männliche Geschlecht als Referenzkategorie; Jahrgangsstufe (Jgst) als fester Level-2-Prädiktor, ^b Jahrgangsstufe zwei als Referenzkategorie; zufällige Effekte: Lehrkräfte als zufällige Achsenabschnitte, $\tau_{00 \text{ Lehrkräfte}}$ (geschätzte Varianz des random intercept). Alle kontinuierlichen Prädiktorvariablen wurden z-standardisiert. Für das Kriterium FERMAT-BF erwies sich das Modell mit den mathematikspezifischen Prädiktoren ohne random intercept Varianz als am besten passend.

Diskussion

Untersucht wurde, ob die Einschätzung von mathematischen Schwierigkeiten durch Lehrkräfte von der tatsächlichen mathematikspezifischen Leistung der SuS abhängt oder ob auch mathematikunspezifische SuS-Merkmale eine Rolle spielen. Außerdem wurde geprüft, ob die Urteilsstrenge bzw. -milde zwischen den Lehrkräften variiert und ob das Ausmaß, in dem SuS-Merkmale herangezogen werden, lehrkraftabhängig ist. Die Lehrkräfteeinschätzungen wurden vorrangig von der tatsächlichen mathematischen Leistung, aber auch durch mathematikunspezifische SuS-Merkmale beeinflusst (FF1). Unterschiede zwischen den Lehrkräften deuten auf Variationen in der Urteilsstrenge hin, da objektiv gleiche SuS-Leistungen unterschiedlich bewertet wurden (FF2a), die Gewichtung verschiedener SuS-Merkmale variierte jedoch nicht (FF2b).

Der Einfluss mathematikspezifischer und -unspezifischer SuS-Merkmale auf die Lehrkräfteeinschätzung (FF1)

Die Lehrkräfteeinschätzungen über die Schwierigkeiten ihrer SuS in den basisnumerischen Fertigkeiten wurden ausschließlich durch mathematikspezifische SuS-Merkmale bestimmt, wobei die Testleistung in den Rechenfertigkeiten, wie erwartet, stärker in das Urteil einfluss als die basale und komplexe basisnumerische Leistung. Unspezifische SuS-Merkmale beeinflussten das Urteil nicht, was auf eine unverzerrte Einschätzung hindeutet. Dieses Ergebnis ist für die Früherkennung von RES praktisch bedeutsam, da Defizite in der Basisnumerik in engem Zusammenhang mit Beeinträchtigungen in höheren Rechenfertigkeiten stehen (Braeuning et al., 2021; Lyons et al., 2014) und charakteristisch für RES sind (Landerl et al., 2022). Ob das standardisierte Bewertungssystem das der FERMAT bietet, die Einschätzung präzisierte, sollte in Folgestudien überprüft werden, etwa durch

einen Vergleich der Lehrkräfteeinschätzung mit und ohne FERMAT.

Erwartungsgemäß flossen die Testergebnisse der SuS in den Rechenfertigkeiten und die komplexen basisnumerischen Fertigkeiten in die Lehrkräfteeinschätzung über die Schwierigkeiten ihrer SuS in den Rechenfertigkeiten ein. Auch die Lesefertigkeit und Intelligenz der SuS beeinflussten das Lehrkräfteurteil. Die Leistung in der basalen Basisnumerik wurde nicht berücksichtigt. Auf den ersten Blick deutet dieses Ergebnis auf einen logischen Fehlschluss hin (z.B. Mack et al., 2023). Da die Lesefertigkeit und Intelligenz der SuS keinen Einfluss auf die Lehrkräfteeinschätzung von Schwierigkeiten in der Basisnumerik hatte, widerspricht dieses Ergebnis einer Urteilsverzerrung. Plausibel wäre, dass die Lehrkräfte, die im FERMAT zu beurteilenden basisnumerischen Fertigkeiten (z.B. Zahlen vergleichen) im Vergleich zu den Rechenfertigkeiten (z.B. Textaufgaben) als sprachunabhängiger eingeschätzt haben. So könnten sie, wie teils empirisch gezeigt (z.B. Cowan & Powell, 2014), davon ausgegangen sein, dass die Intelligenz für die Beherrschung von Rechenfertigkeiten bedeutsamer ist, als für das Lösen basalerer Aufgaben. Im Umkehrschluss könnte der Einfluss der Lesefertigkeit und Intelligenz auf das Lehrkräfteurteil über die Schwierigkeiten in den Rechenfertigkeiten durch einen Zusammenhang zwischen diesen Fertigkeiten begründet sein (z.B. Akin, 2022; Peng et al., 2019). Auch, wenn domänengenerelle Fertigkeiten förderlich oder hemmend auf die mathematische Kompetenzentwicklung wirken (z.B. von Aster & Shalev, 2007) und mit der mathematischen Leistung korreliert sind, sind mathematikspezifische Fertigkeiten für die Einschätzung von Schwierigkeiten in den Rechenfertigkeiten insgesamt bedeutsamer (z. B. Chen & Li, 2014; Duncan et al., 2007; Kuhn et al., 2019). Daher sollten domänenspezifische Merkmale bzw. mathematikspezifische Fertigkeiten die Lehrkräfteeinschätzung von Schwierigkeiten in den Rechenfertigkeiten bestimmen.

Im Gegensatz zu einigen früheren Studien (z.B. Mack et al., 2023), hatte das Geschlecht der SuS in dieser Studie keinen Einfluss auf die Lehrkräfteeinschätzung, konform zur Übersichtsarbeit von Urhahne und Wijnia (2021). Sollte dieses Ergebnis dauerhaft repliziert werden, könnten Lehrkräfte durch ihren Einfluss auf geschlechtsspezifische Erwartungen langfristig zur Reduktion von Geschlechterunterschieden (z.B. Schütty, 2022) beitragen.

Nach Einschätzung der Lehrkräfte hatten die Zweitklässler*innen mehr basisnumerische Schwierigkeiten, obwohl ihre objektive Testleistung nicht substanziell schlechter war (Tab. 1), was durch eine Depriorisierung basisnumerischer Inhalte in der dritten und vierten Jahrgangsstufe erklärt werden könnte. Diese Interpretation muss in zukünftigen Studien überprüft werden. Da RES aus Defiziten in den Kernsystemen der Zahlenverarbeitung, die für die Verarbeitung der basisnumerischen Inhalte zuständig sind, resultieren können (z.B. Lamb et al., 2024 a), sollten basisnumerische Schwierigkeiten bei der Identifikation von RES auch in höheren Jahrgangsstufen berücksichtigt werden.

Unterschiede zwischen Lehrkräften (FF2)

Konform zum Strenge- bzw. Mildeeffekt (Hesse & Latzko, 2017) erhielten identisch leistungsfähige SuS unterschiedliche FERMAT-Scores. Bei gleicher objektiver Testleistung in den mathematikspezifischen und -unspezifischen Fertigkeiten gaben einige Lehrkräfte Schwierigkeiten in den Rechenfertigkeiten an, während andere Lehrkräfte keine Schwierigkeiten sahen. Dies deutet darauf hin, dass die gleiche Leistung von verschiedenen Lehrkräften unterschiedlich bewertet wird. Es gab jedoch keinen Hinweis darauf, dass das Ausmaß in dem mathematikspezifische und -unspezifische SuS-Merkmale herangezogen werden, lehrkräfteabhängig ist. Für den einzuschätzenden Bereich der Basisnumerik wurden keine Lehrkräfteeffekte festgestellt.

Limitationen

Die Ergebnisse sind von den eingesetzten Testverfahren abhängig. Daher sollten die Analysen mit anderen standardisierten Tests, die u.a. Textaufgaben beinhalten, repliziert werden, um alle im FERMAT zu beurteilenden Fertigkeiten abzudecken. Auch fehlende Informationen zu den Lehrkräften limitieren die Ergebnisse, da eine Vielzahl weiterer Einflussfaktoren nicht kontrolliert wurde. Zukünftige Forschungen sollten weitere Merkmale der SuS (z. B. Verhalten im Unterricht) und Lehrkräfte (z. B. stereotype Vorstellungen in Bezug auf den Migrationshintergrund) untersuchen, die die Lehrkräfteeinschätzung beeinflussen können (z.B. Urhahne & Wijnia, 2021; Wagner, 2024). Das gleiche gilt für Effekte auf SuS- und Klassenebene.

Fazit

Lehrkräfte entscheiden darüber, ob SuS gezielte Förderungen erhalten oder nicht (Fischer et al., 2015). Daher sollten Lehrkräfte die mathematischen Schwierigkeiten ihrer SuS basierend auf deren tatsächlichen mathematischen Leistung einschätzen. Urteilsirrelevante Informationen sollten die Einschätzung nicht beeinflussen. Die Studie zeigte, dass die Testleistung in den basalen und komplexen basisnumerischen Fertigkeiten ebenso wie die Leistung in den Rechenfertigkeiten, die bei Kindern mit RES beeinträchtigt sind (z.B. Lamb et al., 2024 a; Landerl, 2022), die Lehrkräfteeinschätzung über *Schwierigkeiten in der Basisnumerik* bestimmten. Das Urteil wurde nicht durch mathematikunspezifische SuS-Merkmale verzerrt. Bei der Einschätzung von *Schwierigkeiten in den Rechenfertigkeiten* identifizierten einige Lehrkräfte bei gleichen objektiven Testleistungen Schwierigkeiten, andere jedoch nicht. Neben den tatsächlichen komplexen basisnumerischen Fertigkeiten und der Leistung in den Rechenfertigkeiten beeinflussten auch die Intelligenz und Lesefertigkeit der SuS die Lehrkräfteein-

schätzung über die *Schwierigkeiten in den Rechenfertigkeiten*. Geschlechterbezogene Urteilsfehler wurden nicht gefunden. Insgesamt sprechen die Ergebnisse dafür, dass die Lehrkräfteeinschätzung über die mathematischen Schwierigkeiten ihrer SuS vorrangig von deren tatsächlicher mathematischer Leistung bestimmt wird. Die Ergebnisse bieten positive Erkenntnisse zur Früherkennung von RES, die für inklusive Schul- und Unterrichtsprozesse relevant sind.

Literatur

- Akin, A. (2022). Is reading comprehension associated with mathematics skills: A meta-analysis research. *International Online Journal of Primary Education*, 11(1), 47–61. <https://doi.org/10.55020/ijope.1052559>
- American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5. Aufl.). American Psychiatric Association.
- Von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental medicine & child neurology*, 49(11), 868–873. <https://doi.org/10.1111/j.1469-8749.2007.00868.x>
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Braeuning, D., Hornung, C., Hoffmann, D., Lambert, K., Ugen, S., Fischbach, A., ... & Moeller, K. (2021). Long-term relevance and interrelation of symbolic and non-symbolic abilities in mathematical-numerical development: Evidence from large-scale assessment data. *Cognitive Development*, 58, 101008. <https://doi.org/10.1016/j.cogdev.2021.101008>
- Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, 9(2), 378–400. <https://doi.org/10.3929/ethz-b-000240890>
- Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 215–234). Waxmann.
- Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta psychologica*, 148, 163–172. <https://doi.org/10.1016/j.actpsy.2014.01.016>
- 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(1), 214–229. <https://doi.org/10.1037/a0034097>
- D'Agostino, R. B. (1970). Transformation to Normality of the Null Distribution of g_1 . *Biometrika*, 57(3), 679–681. <https://doi.org/10.2307/2334794>
- Dobson, A. J. (2002). *An introduction to generalized linear models*. Chapman & Hall/CRC.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Fahrmeir, L., Kneib, T., & Lang, S. (2007). *Regression: Modelle, Methoden und Anwendungen*. Springer.
- Fischer, U., Rösch, S., Nuerk, H.C., & Moeller, K. (2015). Erkennen von Rechenschwäche durch LehrerInnen und Testungen im Klassenverband. *Lernen und Lernstörungen*, 4(4), 269–285. <https://doi.org/10.1024/2235-0977/a000116>
- Fischer, U., Rösch, S., & Moeller, K. (2017). Diagnostik und Förderung bei Rechenschwäche. *Lernen und Lernstörungen*, 6(1), 25–38. <https://doi.org/10.1024/2235-0977/a000160>
- Fischbach, A., Schuchardt, K., Brandenburg, J., Kleczewski, J., Balke-Melcher, C., Schmidt, C., ... & Hasselhorn, M. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien. *Lernen und Lernstörungen*, 2(2), 65–76. <https://doi.org/10.1024/2235-0977/a000035>
- Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest (HRT 1 – 4)*. Hogrefe.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrprofessionalität. Diagnose, Evaluation und Verbes-*

- serung des Unterrichts (7. Aufl.). Klett-Kallmeyer.
- Hesse, I., & Latzko, B. (2017). *Diagnostik für Lehrkräfte* (3., vollst. Überarb. und erw. Auflage). Buchdrich.
- Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social psychology of education*, 20, 471–490. <https://doi.org/10.1007/s11218-017-9384-z>
- Kaiser, J., Möller, J., Helm, F., & Kunter, M. (2015). Das Schülerinventar: Welche Schülermerkmale die Leistungsurteile von Lehrkräften beeinflussen. *Zeitschrift für Erziehungswissenschaft*, 18(2), 279–302. <https://doi.org/10.1007/s11618-015-0619-5>
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern*. Waxmann.
- Karst, K., Schoreit, E., & Lipowsky, F. (2014). Diagnostische Kompetenzen von Mathematiklehrern und ihr Vorhersagewert für die Lernentwicklung von Grundschulkindern. *Zeitschrift für pädagogische Psychologie*, 28, 237–248. <https://doi.org/10.1024/1010-0652/a000133>
- Kaufmann, E. (2020). How accurately do teachers' judge students? Re-analysis of Hoge and Coladarsi (1989) meta-analysis. *Contemporary Educational Psychology*, 63, 101902. <https://doi.org/10.1016/j.cedpsych.2020.101902>
- Kolovou, D., Naumann, A., Hochweber, J., & Praetorius, A.-K. (2021). Content-specificity of teachers' judgment accuracy regarding students' academic achievement. *Teaching and Teacher Education*, 100(4), 103298. <https://doi.org/10.1016/j.tate.2021.103298>
- Kuhn, J.-T., Schwenk, C., Raddatz, J., Dobel, C., & Holling, H. (2017). *CODY-Mathetest für die 2. – 4. Klasse (CODY-M 2 – 4)*. Kaasa health.
- Kuhn, J.-T., Schwenk, C., Souvignier, E., & Holling, H. (2019). Arithmetische Kompetenz und Rechenschwäche am Ende der Grundschulzeit. Die Rolle statusdiagnostischer und lernverlaufsbezogener Prädiktoren. *Empirische Sonderpädagogik*, 11(2), 95–117. <https://doi.org/10.25656/01:17773>
- Lamb, S., Krieger, F., & Kuhn, J.-T. (2024 a). Delayed development of basic numerical skills in children with developmental dyscalculia. *Frontiers in Psychology*, 14, 1187785. <https://doi.org/10.3389/fpsyg.2023.1187785>
- Lamb, S., Schulz, A.-K., & Kuhn, J.-T. (2024 b). Entwicklung eines Lehrkräftefragebogens zur Früherkennung von Rechenstörungen in der Grundschule. *Lernen und Lernstörungen*, 13(4), 165–177. <https://doi.org/10.1024/2235-0977/a000456>
- Landerl, K., Vogel, S., & Kaufmann, L. (2022). *Dyskalkulie: Modelle, Diagnostik, Intervention*. Ernst Reinhardt.
- Lorenz, C. (2011). *Diagnostische Kompetenz von Grundschullehrkräften: Strukturelle Aspekte und Bedingungen*. University of Bamberg Press.
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science*, 17(5), 714–726. <https://doi.org/10.1111/desc.12152>
- Mack, E., Gnäs, J., Vock, M., & Preckel, F. (2023). The Domain-Specificity of Elementary School Teachers' Judgment Accuracy. *Contemporary Educational Psychology*, 72, 102142. <https://doi.org/10.1016/j.cedpsych.2022.102142>
- Mayringer, H. & Wimmer, H. (2003). *Salzburger Les-Screening für die Klassenstufen 1 – 4 (SLS 1 – 4)*. Hans Huber.
- Karing, C., & Artelt, C. (2014). Urteilsgenauigkeit von Lehrer(inne)n im emotionalmotivationalen Bereich und im Leistungsbereich. In M. Mudiappa & C. Artelt (Hrsg.) *BiKS – Ergebnisse aus den Längsschnittstudien. Praxisrelevante Befunde aus dem Primar- und Sekundarschulbereich* (S. 111–118). University of Bamberg Press.
- Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, 145(2), 189–236. <https://doi.org/10.1037/bul0000182>
- R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Salaschek, M., Zeuch, N., & Souvignier, E. (2014). Mathematics growth trajectories in first grade: Cumulative vs. compensatory patterns and the role of number sense. *Learning and Individual Differences*, 35, 103–112. <https://doi.org/10.1016/j.lindif.2014.06.009>
- Schäfer, H., & Rittmeyer, C. (2015). Inklusive Diagnostik. In H. Schäfer & C. Rittmeyer (Hrsg.), *Handbuch Inklusive Diagnostik* (S. 103–134). Beltz.

Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Peter Lang.

Schrader, F.W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung*, 31(2), 154–165. <https://doi.org/10.25656/01:13843>

Schrader, F.W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.

Schütky, R. (2022). Der Einfluss nicht-kognitiver Merkmale (Selbstkonzept, Stereotypen und Stereotype Threat) auf Mathematikleistungen im Bereich der Größen im Verlauf der Grundschulzeit. *Mathematik im Unterricht*, 13, 31–46. <https://doi.org/10.25598/miu/2022-13-3>

Selter, C., Walter, D., Heinze, A., Brandt, J., & Jentsch, A. (2020). Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky, & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 25–56). Waxmann. <https://doi.org/10.31244/97838300993193>

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers’ judgments of students’ academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>

Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, 32, 100374. <https://doi.org/10.1016/j.edurev.2020.100374>

Wagner, L. (2024). Einflussfaktoren auf die Diag-

nosekompetenz (angehender) Lehrkräfte – ein systematisches Literaturreview. *Unterrichtswissenschaft*, 1–30. <https://doi.org/10.1007/s42010-024-00215-3>

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 – Revision (CFT 20-R)*. Hogrefe.

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

Autorinnen- und Autorenhinweis

-  Sarah Lamb
<https://orcid.org/0000-0003-4619-3269>
-  Jörg-Tobias Kuhn
<https://orcid.org/0000-0002-4399-9569>
-  Ann-Katrin Schulz

Korrespondenzadresse

Sarah Lamb
Fakultät für Rehabilitationswissenschaften,
Technische Universität Dortmund
Emil-Figge-Straße 50, D-44227 Dortmund,
sarah.lamb@tu-dortmund.de

Erstmals eingereicht: 22.12.2023
Überarbeitung eingereicht: 29.07.2024
Angenommen: 25.02.2025

Offene Daten	Bisher wurden die Daten nicht zur Verfügung gestellt, da weitere Publikationen geplant sind.
Offener Code	keine Angabe
Offene Materialien	Der in der Studie verwendete Fragebogen (FERMAT) ist öffentlich zugänglich.
Präregistrierung	Bisher nicht.
Votum Ethikkommission	Es liegt ein positives Ethikvotum der Ethikkommission des Fachbereichs 7 der Universität Münster vor.
Finanzielle und weitere sachliche Unterstützung	Diese Arbeit wurde finanziell durch das Bundesministerium für Bildung und Forschung unterstützt (Projektnummer: 01GJ1302).
Autorenschaft	SL hat den ersten Entwurf des Manuskripts geschrieben. A-KS und J-TK haben den Entwurf überarbeitet, J-TK hat die Studie geplant und die Daten erhoben, SL und J-TK haben die Daten analysiert.