

Entwicklung eines Lehrkräftefragebogens zur Früherkennung von Rechenstörungen in der Grundschule

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

Methoden der empirischen Bildungsforschung, Fakultät Rehabilitationswissenschaften, Technische Universität Dortmund, Deutschland

Zusammenfassung: *Hintergrund:* Bislang fehlen standardisierte Screeninginstrumente zur Identifikation von Rechenstörungen (RS) im Grundschulalter durch Lehrkräfte, die sich effektiv in die Unterrichtspraxis integrieren lassen und zeitgleich allgemein anerkannte Gütekriterien erfüllen. Mit dem Fragebogen zur Erfassung mathematischer Fertigkeiten (FERMAT) wurde ein theoriebasiertes Screeningverfahren für Lehrkräfte entwickelt, das Kinder mit einem erhöhten Risiko für RS ökonomisch, reliabel und valide identifiziert. *Methoden:* Anhand der Daten von $N = 377$ Schüler_innen aus Nordrhein-Westfalen (Klassenstufe zwei bis vier), wurden mittels psychometrischer Analysen (u.a. Receiver-Operating-Characteristics-Analysen) verschiedene Gütekriterien bestimmt. Zur Prüfung der Treffsicherheit wurden verschiedene Cut-Offs verwendet. *Ergebnisse:* Der FERMAT verfügt über gute bis überwiegend zufriedenstellende Screeningeigenschaften. Die Sensitivität und Spezifität variierten je nach Kriterium zwischen 57.6% und 69.8% bzw. 82.5% und 91.4%. Die RATZ-Indizes lagen zwischen .511 und .613. Die Screeningeigenschaften der Teilbereiche Basisnumerik und Rechenfertigkeiten, in die sich der FERMAT gliedert, sind ebenfalls überwiegend zufriedenstellend. *Diskussion:* Der FERMAT identifiziert Kinder mit und ohne erhöhtes Risiko für RS mit angemessener Treffsicherheit. Die praktische Bedeutung des Instruments wird vor dem Hintergrund langfristiger Folgen unbehandelter RS und den Chancen frühzeitiger Risikoidentifikation diskutiert.

Schlüsselwörter: Rechenstörung, Screening, Früherkennung, Grundschule

Development of a Teacher Questionnaire for the Early Identification of Mathematics Disorders in Primary School

Abstract: *Background:* There is a lack of screening instruments for the early identification of mathematics disorders (MD) in primary school that can be effectively integrated into classroom practices while simultaneously meeting widely accepted quality criteria. The questionnaire to assess mathematical skills of children in grades two to four ("Fragebogen zur Erfassung mathematischer Fertigkeiten von Grundschulkindern, FERMAT") has been developed as a theory-based screening instrument to identify children at risk of developing MD in an effective, reliable, and valid manner. *Methods:* Based on data from $N = 377$ students from North Rhine Westphalia in grades two to four, psychometric quality criteria (e.g., Receiver-Operating-Characteristics-analyses) were investigated. Different cut-off criteria were employed to examine the predictive accuracy of the FERMAT. *Results:* The FERMAT exhibits good to mostly satisfactory screening characteristics. For the criteria used, sensitivity ranged from 57.6% to 69.8%, while specificity ranged from 82.5% to 91.4%. The RATZ index ranged from .511 to .613. FERMAT is divided into two subscales (basic numerical skills and numeracy) whose screening characteristics are also predominantly satisfactory. *Discussion:* Thus, the FERMAT can accurately identify children with and without an increased risk for MD. The practical importance of early identification of MD is discussed regarding long-term consequences of untreated MD and the potential benefits of early risk identification.

Keywords: mathematics disorder, screening, early identification, primary school

Einleitung

Basierend auf dem angeborenen Zahlensinn beginnt die Entwicklung mathematischer Vorläuferfertigkeiten bereits im Säuglingsalter (Dehaene, 1992). Im Vorschulalter verfügen die meisten Kinder bereits über ein Repertoire an Fertigkeiten, das von der Mengenunterscheidung bis hin zur Zahlwortreihe und basaler Arithmetik reicht (Ennemoser,

Sinner & Krajewski, 2015). Die Entwicklung setzt sich unter formeller Beschulung meistens erfolgreich fort. Allerdings weisen etwa 25% der Kinder am Ende der Grundschulzeit unterdurchschnittliche mathematische Kompetenzen auf (Selter, Walter, Heinze, Brandt & Jentsch, 2020), und etwa 3 bis 5% entwickeln eine klinisch relevante Rechenstörung (RS) (Fischbach et al., 2013). Die Prävalenz ist damit ähnlich hoch wie bei Beeinträchtigungen im schriftsprach-

lichen Bereich (z.B. Fischbach et al., 2013), das zugehörige Forschungsfeld sowie die Anerkennung in Bildungspraxis und -politik jedoch weniger ausgereift (Ennemoser et al., 2015; Lorenz, 2012). Ohne Intervention besteht die Gefahr, dass sich anfängliche Schwierigkeiten zu persistierenden Lernstörungen entwickeln (Lorenz, 2012), die bis in das Erwachsenenalter bestehen bleiben (Butterworth, Varma & Laurillard, 2011), sekundäre emotionale soziale Komorbiditäten verursachen (Kohn, Wyschkon, Ballaschk, Ihle & Esser, 2013) und in einer benachteiligten Bildungsbiografie münden (Parsons & Bynner, 2005). Diese nachteiligen Folgen unterstreichen die Wichtigkeit der Früherkennung von Kindern mit einem Risiko für RS.

Im Sinne des Response-to-Intervention-Ansatzes ermöglicht die frühzeitige Risikoidentifikation eine weiterführende Diagnostik und gezielte Förderung (z.B. Grosche & Huber, 2012), noch bevor sich anfängliche Schwierigkeiten in einer negativen Abwärtsspirale manifestieren (Vaughn & Fuchs, 2003). Um den Prozess von der Früherkennung über die Diagnostik bis hin zur Intervention (Tröster, 2009) besser zu verzahnen, ist es wichtig, Screeninginstrumente auf der theoretischen Basis mathematischer Kompetenzentwicklungsmodelle zu konstruieren (vgl. Kuhn & Schwenk, 2018). Screenings, die sich leicht in die Unterrichtspraxis integrieren lassen, sind besonders relevant, da Lehrkräfte als primäre Informationsquelle für den akademischen Entwicklungsverlauf von Kindern ein wichtiger Bestandteil des diagnostischen Prozesses sind (Fischer, Rösch, Nuerk & Moeller, 2015). Doch die derzeit existierenden Screeningverfahren sind vorrangig auf das Kindergarten- und frühe Schulalter ausgerichtet (Landerl, Vogel & Kaufmann, 2022), obwohl bei einigen Kindern Rechenschwierigkeiten erst im Laufe der Grundschulzeit auftreten (Kohn et al., 2013). Zudem basieren die meisten Verfahren auf standardisiert gemessenen Testleistungen (Lorenz, 2012), weshalb die Anwendung teils zeitaufwendig, komplex und damit schwierig in den Lehrkräftealltag integrierbar ist. Andere Instrumente wiederum wurden nicht auf ihre psychometrischen Eigenschaften hin untersucht, so dass keine Evidenz über die tatsächliche Vorhersagegenauigkeit vorliegt. Zudem werden basisnumerische Fertigkeiten teils vernachlässigt, obwohl diese den Ausgangspunkt der mathematischen Entwicklung bilden (z.B. Fischer, Rosch & Moeller, 2017) und auch am Ende der Grundschule gut zwischen Kindern mit und ohne RS differenzieren (Gaupp, Zoelch & Schumann-Hengstler 2004). Ziel dieser Arbeit ist es daher, einen theoretisch fundierten Screeningfragebogen für Lehrkräfte zur Erfassung mathematischer Fertigkeiten (FERMAT) vorzustellen. Der ökonomisch anwendbare und in dieser Studie psychometrisch geprüfte Fragebogen soll Lehrkräfte bei der Früherkennung von Grundschulkindern mit einem Risiko für RS unterstützen.

Rechenschwierigkeiten

Rechenschwierigkeiten sind kein einheitlich definiertes Konstrukt. Nicht nur die Begrifflichkeiten (z.B. Rechenschwäche/-störung, Dyskalkulie), sondern auch die diagnostischen Kriterien variieren (DSM-5: American Psychiatric Association [APA], 2013; ICD-11: World Health Organization [WHO], 2020). Klinisch-diagnostische Vertreter_innen sind sich weitgehend einig, dass allgemein von Rechenschwierigkeiten die Rede ist, wenn die mathematische Leistung ohne das Vorliegen einer Intelligenzminderung ($IQ > 70$) und trotz angemessener Bildung mehr als eine Standardabweichung unter dem Durchschnitt liegt (z.B. Kuhn, 2017). Der davon abzugrenzende Begriff der RS, wird gemäß der S3-Leitlinie zur Diagnostik und Intervention bei Rechenstörung (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften [AWMF], 2018) gebraucht, wenn die mathematische Leistung in einem standardisierten Test trotz ausreichender Bildung und ohne Intelligenzminderung einem Prozentrang ($PR \leq 7$) entspricht. Im internationalen Raum wird häufig ein PR von ≤ 10 verwendet (z.B. Mazzocco, Feigenson & Halberda, 2011). Obwohl das Profil der RS selbst heterogen ist (z.B. Kißler, Schwenk & Kuhn, 2021), sind auf behavioraler Ebene einige Leitsymptome identifizierbar (z.B. Haberstroh & Schulte-Körne, 2019; AWMF, 2018). Diese manifestieren sich in beeinträchtigten basisnumerischen Fertigkeiten und defizitären Rechenfertigkeiten (z.B. Kuhn, 2017). Basisnumerische Fertigkeiten bilden die Grundlage für die weitere Ausbildung des rechnerischen Denkens und Manipulierens von Zahlen im frühen Schulalter (Ennemoser et al., 2015; Fischer et al., 2017) und sind prädiktiv für die mathematische Kompetenzentwicklung (z.B. Krajewski & Schneider, 2009). Die Fertigkeiten werden i.d.R. durch leicht zu lösende Aufgaben erfasst. Ein Beispiel hierfür ist das Größenvergleichsparadigma (z.B. Halberda & Feigenson, 2008), bei dem die Aufgabe darin besteht, aus zwei Mengen (z.B. Punkten oder Zahlen) möglichst schnell die numerisch größere auszuwählen, ohne zu zählen. Ein weiteres Beispiel sind Zahlenstrahlaufgaben, die die Fähigkeit testen, einen numerischen Wert auf einer visuellen Linie zu positionieren. Diesen Aufgabenparadigmen liegt die Annahme zugrunde, dass die Testleistungen Informationen über angeborene Kernmechanismen der Zahlenverarbeitung generieren, die wiederum Rückschlüsse auf die domänenspezifischen Ursachen von RS ermöglichen (Andersson & Östergren, 2012). Daher sollte die theoretische Konstruktion eines Instruments zur Früherkennung von RS auf Entwicklungsmodellen zur Zahlenverarbeitung basieren und neben den Rechenfertigkeiten auch basisnumerische Fertigkeiten erfassen.

Die Entwicklung der Zahlenverarbeitung

Mehrere Modelle beschreiben die Entwicklung der Zahlenverarbeitung von Kindern (Fischer et al., 2017). Ausgangspunkt der Modelle ist die Annahme, dass die Zahlenverarbeitung auf präverbalen Kernsystemen beruht. Unterschieden werden das *Approximate Number System*, das für die ungefähre Erkennung und Schätzung von Mengen zuständig ist, sowie das *Object Tracking System*, das dem simultanen Erkennen von max. drei bis vier Objekten dient (Feigenson, Dehaene & Spelke, 2004).

Fischer et al. (2017) kontrastierten die bekanntesten deutschsprachigen Entwicklungsmodelle der Zahlbegriffsentwicklung und synthetisierten daraus acht *mathematische Kompetenzbereiche*: Kompetenzbereich 1) beschreibt den Ausgangspunkt der numerischen Entwicklung: die Fähigkeit *Mengen zu schätzen und zu vergleichen*. Darauf aufbauend entwickelt sich Kompetenzbereich 2) *Zahlwortreihe und Zählen*, wodurch ein ordinales Zahlverständnis erworben wird. Kompetenz 3) *Zahlensymbole lesen und schreiben* beschreibt die Fähigkeit arabische Zahlen zu schreiben und zu lesen. Kernkompetenz 4) *Zahlengröße verstehen* beschreibt das Verständnis darüber, dass jede Zahl auch für eine bestimmte Menge steht (kardinales Zahlverständnis). Basierend darauf entwickelt sich die Fähigkeit 5) *Zahlbeziehungen verstehen*, d. h. die Erkenntnis, dass Zahlen zueinander in Beziehung gesetzt werden können (relationale Zahlverständnis). Kompetenz 6) *Zahlenraumvorstellung*, bezieht sich auf die zunehmende Ausdifferenzierung des mentalen Zahlenstrahls, auf dem Zahlen entsprechend ihrer numerischen Größe lokalisiert werden. Die Kompetenzen 7) *Rechnen (Addition/Subtraktion)* und 8) *Stellenwertverständnis* entwickeln sich parallel zu den übrigen Kompetenzen und beschreiben keine separaten Entwicklungsstufen. Lösungsstrategien wie das Zählen beim Addieren und Subtrahieren ermöglichen Rückschlüsse auf die Entwicklungsstufen. Das Stellenwertverständnis beschreibt schließlich das Verständnis dafür, dass sich größere Zahlen aus der Zusammensetzung von Einern, Zehnern usw. ergeben.

Instrumente zur (Früh-)Erkennung von Rechenstörungen in der Grundschule

Es existieren bereits einige deutschsprachige Instrumente zur (Früh-)Erkennung von RS, die im Folgenden exemplarisch im Hinblick auf ihre Vor- und Nachteile dargestellt werden.

Das *Elementar Mathematische Basisinterview* (EMBI; Peter-Koop, Wollring, Spindeler & Grüßing, 2007) dient einer curricular orientierten qualitativen/förderdiagnostischen Einschätzung und ermöglicht eine gezielte Be-

trachtung der Entwicklung mathematischer Fertigkeiten. Allerdings ist die Durchführung des EMBI aufgrund seines standardisierten Interviewleitfadens mit 20 Seiten als zeitintensiv zu bewerten, was die Integrierbarkeit in den Alltag der Lehrkräfte erschwert.

Ökonomisch verwendbare Checklisten (z. B. Jacobs & Petermann, 2012) und Fragebögen (z. B. Dögnitz, 2022) für Lehrkräfte erscheinen auf den ersten Blick aufgrund ihrer inhaltlichen Konstruktion vielversprechend. Allerdings lässt sich aufgrund fehlender psychometrischer Angaben ihre Eignung hinsichtlich der Testgüte nicht abschließend beurteilen.

Screeninginstrumente wie das *Flensburger-Schulspiel* (FleSch; Clausen-Suhr & Walter, 2022) sind hingegen zwar prognostisch valide, basieren jedoch auf standardisierten Testleistungen und können daher nicht als Fremdbeurteilungsverfahren eingesetzt werden.

Daher war es das Ziel der vorliegenden Studie, ein standardisiertes, theoriebasiertes und ökonomisches Screeninginstrument mit angemessener Treffsicherheit für Lehrkräfte zur Früherkennung von RS zu konstruieren und psychometrisch zu validieren.

Fragebogenkonstruktion

Der FERMAT zielt nicht darauf ab, eine klinische RS (gemäß der ICD-11; WHO, 2020) festzustellen, sondern Kinder zu identifizieren, die einer detaillierteren diagnostischen Untersuchung zugewiesen werden sollten. Angelehnt an Opitz und Bern (2008) werden folgende Anforderungen an den FERMAT gestellt: 1) Der FERMAT soll zentrale mathematische Kompetenzen und häufig beobachtbare Leitsymptome von RS abbilden. 2) Lehrkräfte sollen den FERMAT ohne umfangreiche Schulung selbstständig ausfüllen, auswerten und interpretieren können. 3) Der Fragebogen soll möglichst wenige, aber inhaltlich und testtheoretisch informative Items beinhalten. 4) Das Testergebnis soll leicht interpretiert werden können und klare Handlungsempfehlungen ermöglichen. 5) Der FERMAT soll allgemein anerkannte psychometrische Gütekriterien (z. B. Reliabilität, Objektivität, Validität; American Educational Research Association, American Psychological Association & National Council for Measurement in Education, 2014) erfüllen und entsprechend Kinder mit einem erhöhten RS-Risiko identifizieren.

Der FERMAT (s. Fragebogen im elektronischen Supplement ESM1) setzt sich aus 19 Items zusammen. Je acht Items dienen der Einschätzung *basisnumerischer Fertigkeiten* und *Rechenfertigkeiten*. Drei weitere Items zielen auf eine globale Einschätzung ab. Die theoretische Konstruktion des FERMAT basiert auf den von Fischer et al. (2017)

synthetisierten numerischen Kompetenzen und wird durch häufig beobachtete Defizite bei RS und curricular relevante Themen ergänzt (s. Tab. 1 im ESM2). Nachfolgend werden die im FERMAT verwendeten Items und die damit erfassten Fertigkeitensbereiche dargestellt.

Basisnumerische Fertigkeiten

Das Verständnis der Zahlwortreihe und das damit vorrangig verbundene ordinale Zahlenverständnis wird über das Item *Zählen* (ZÄ) eingeschätzt. Beeinträchtigungen bei RS spiegeln sich häufig in Schwierigkeiten beim Vorwärts- und Rückwärtszählen (30, 31, 32 ...) und/oder im Zählen in Schritten (10, 12, 14 ...) wider (Gaupp et al., 2004). Das Item *Lesen und Schreiben von Zahlensymbolen* (ZLS) bezieht sich auf die Fähigkeit, Zahlensymbole korrekt zu lesen und zu schreiben. Diese Schlüsselkompetenz ist relevant für die Bewältigung visuell dargestellter Rechenaufgaben (Fischer et al., 2017). Bei RS sind häufig Transkodierfehler zu beobachten. Dabei werden verbal dargebotene Zahlen fehlerhaft aufgeschrieben (Kuhn, Raddatz, Holling & Dobel, 2013), bspw. in Form von Inversionsfehlern („vierunddreißig“ statt 43) oder Kompositionsfehlern („50015“ statt 515) (Zuber, Pixner, Moeller & Nuerk, 2009). Das Verständnis von Zahlengrößen und deren Repräsentation als Anzahl wird über das *Anzahlkonzept* (AK) erfasst. Kinder mit mangelndem kardinalen Zahlenverständnis interpretieren Zahlen vorrangig als Zählzahlen und erkennen nicht, dass diese auch Mengen darstellen können ($\dots =$ „fünf“) (Kuhn, 2017). Dies führt u.a. zu Schwierigkeiten bei der Anwendung von Rechenstrategien (Kaufmann & Wessolowski, 2021). Das Verständnis für Zahlensymbole und die Orientierung im Zahlenraum wird über das Item *Zahlen ordnen und vergleichen* (ZOV) abgebildet. Typische Fehler zeigen sich beim Zahlenvergleich ($835 > 1358$, da 8 größer 1) oder auch darin, dass Vorgänger und Nachfolger einer Zahl falsch benannt werden (welche Zahl kommt vor 43? \rightarrow 32). Um das häufig bei RS auftretende mangelnde Stellenwertverständnis zu berücksichtigen (Scherer & Moser Opitz, 2010), wurde das Item *Stellenwert* (STW) konstruiert, welches das Verständnis über das Dezimalsystem abdeckt. Um die *Zahlenraumvorstellung* (ZR) zu berücksichtigen, wird die Fähigkeit eingeschätzt, eine numerische Größe auf einem Zahlenstrahl zu positionieren. Kinder mit RS positionieren die Zahlen meist weniger präzise als Kinder ohne RS (Decarli et al., 2023). Die Zahlenraumvorstellung wird separat für den Zahlenraum bis zehn (ZR-10), hundert (ZR-100) und tausend (ZR-1000) eingeschätzt.

Rechenfertigkeiten

Eingeschätzt werden die Fertigkeiten in den Grundrechenarten: *Addieren* (AD), *Subtrahieren* (SU), *Multiplizieren*

(MU) und *Dividieren* (DI), da Schwierigkeiten in diesen Bereichen zu den Leitsymptomen der RS zählen (WHO, 2020). Schwierigkeiten manifestieren sich häufig in Problemen beim Zehnerübergang oder durch einen beeinträchtigten Faktenabruf (Busch, Oranu, Schmidt & Grube, 2013). Dies zeigt sich darin, dass auch bei einfachen Rechenaufgaben wie $8 + 2$ nicht auf Faktenwissen zurückgegriffen werden kann, sondern die Aufgabe immer wieder neu berechnet werden muss (Busch et al., 2013). Zur Beurteilung des relationalen Zahlverständnisses wurde das Item *Ergänzungsaufgabe* (ER, z.B. $? - 2 = 5$ statt der 7 wird eine 3 eingetragen) konstruiert. Schwierigkeiten beim Lösen dieser Aufgabe zeigen sich typischerweise in einem mangelnden Verständnis für die Verwendung mathematischer Operationen bzw. Rechenzeichen. Das Lösen von *Textaufgaben* (TA), das u.a. curricular relevant ist, erfordert i.d.R. ein relationales Zahlenverständnis, da Kinder verstehen müssen, dass Zahlen in Beziehung zueinander gesetzt werden können (Fischer et al., 2017). Um den Inhalt einer Sachaufgabe zu erfassen, wird ein Verständnis für mathematische Operationen benötigt, ohne dieses werden die Zahlen willkürlich verknüpft (Kaufmann & Wessolowski, 2021). *Rechenstrategien* (R-ST) zählen nicht explizit zu den von Fischer et al. (2017) formulierten Kompetenzen, doch Kinder mit RS weisen teils eine verzögerte Strategieentwicklung auf (z.B. Torbeyns, Verschaffel & Ghesquière, 2004) und zeigen wenig bis gar keine Variabilität in ihren Rechenstrategien (z.B. Kuhn, 2017). Vor allem weniger ausgereifte Zählstrategien, z.B. verbales Zählen oder das Zählen mit den Fingern, sind bei RS zu beobachten (z.B. Kuhn 2017), weshalb das *zählende Rechnen* (ZÄ-R) erfasst wird.

Zielsetzung und Fragestellung

Für das Grundschulalter existiert bislang kein ökonomisches, psychometrisch validiertes Instrument zur Früherkennung von RS durch Lehrkräfte. Daher wird ein Fragebogen zur Erfassung mathematischer Fertigkeiten von Grundschulkindern der Klassenstufen zwei bis vier vorgestellt und hinsichtlich psychometrischer Eigenschaften geprüft. Untersucht werden die faktorielle Struktur des Fragebogens (Fragestellung 1[F1]), das psychometrische Messmodell auf Itemebene bzw. die Rasch-Konformität (F2) sowie die Güte der Identifikation von Kindern mit RS-Risiko (F3). Hinsichtlich F1 wird erwartet, dass sich die theoriebasiert entwickelte zweifaktorielle Struktur (basisnumerische Fertigkeiten versus Rechenfertigkeiten) des FERMAT empirisch bestätigt. In Zusammenhang mit F2 wird exploriert, ob die globale Modellpassung des FERMAT für ein ein- oder zweiparametrisches logistisches Testmodell spricht. F3 untersucht vorrangig, inwie-

fern eine zuverlässige Identifikation von Kindern mit einem Risiko für RS anhand der Lehrkräfteeinschätzung im FERMAT gelingt. Der FERMAT gilt als valide, wenn möglichst viele Risikokinder korrekt identifiziert werden, sowohl solche mit tatsächlichem Risiko für RS (richtig-positiv) als auch solche ohne ein Risiko für RS (richtig-negativ). Gleichzeitig sollten die Fehlentscheidungen (falsch-positiv und falsch-negativ) möglichst gering sein. Die Ergebnisse der Treffsicherheit sind von der Operationalisierung des vorherzusagenden Kriteriums (RS-Risiko) abhängig. Da in der Praxis und Forschung unterschiedliche Kriterien angewandt werden, wird die Risikoidentifikation in Anlehnung an gängige internationale (z.B. Mazzocco et al., 2011) und nationale (AWMF, 2018) diagnostische Kriterien exploriert. Kinder zählten zur RS-Risikogruppe, deren Leistung im Heidelberger Rechentest 1–4 (HRT 1–4; Haffner, Baro, Parzer & Resch, 2005) oder im CODY-Mathetest für die 2. bis 4. Klasse (CODY-M 2–4; Kuhn, Schwenk, Raddatz, Dobel & Holling, 2017) einem PR von ≤ 10 oder ≤ 7 entsprach. Während der HRT 1–4 (über die Skala Rechenoperationen) insbesondere die Rechenfertigkeiten in den Blick nimmt, fokussiert der CODY-M 2–4 (Kuhn et al., 2017) die basisnumerischen Fertigkeiten.

Methode

Stichprobe

Die Erhebung erfolgte zu Beginn des Schuljahres an regulären Grundschulen in Nordrhein-Westfalen. Ein positives Ethikvotum sowie die Einverständniserklärungen liegen vor. Die $N = 377$ Schüler_innen aus $n = 33$ Schulklassen (weiblich: $n = 197$ [52.25%]) verteilen sich auf drei Jahrgangsstufen wie in Tabelle 1 dargestellt.

Design

Über die Dauer von drei Schulstunden bearbeiteten die Kinder Aufgaben aus vier verschiedenen Testverfahren. Die Verfahren wurden in einem ausbalancierten Design vorgegeben, sodass die Reihenfolge zwischen verschiedenen Klassen variierte und jedes Testverfahren an jeder Position vorkam. Während die Testung mittels papierbasierter Instrumente im Klassenverband stattfand, wurde der computerbasierte CODY-M 2–4 Test (Kuhn et al., 2017) in kleineren Gruppen von max. 13 Kindern durchgeführt. Um sicherzustellen, dass die Tests standardisiert durchgeführt werden, wurden die Aufgaben durch geschulte Testleitungen instruiert. Details zu den Instruktionen können den jeweiligen Testmanualen entnommen werden. Die Instruktionen zum unter Aufsicht selbstadministrierten CODY-M 2–4 Test erhielten die Kinder via Kopfhörer, während die Instruktion der übrigen Testverfahren von der Testleitung im Klassenverband vorgelesen wurde.

Die Lehrkräfte schätzten die mathematischen Fertigkeiten ihrer Schüler_innen nach sorgfältiger Durcharbeitung der Instruktion mittels des FERMAT ein. Die Testleitung stand den Lehrkräften im Vorfeld für Rückfragen zur Verfügung. Alle Instruktions- und Ausfüllhinweise (inkl. Beispiele) sind auf dem FERMAT-Fragebogen vermerkt.

Die Rechenfertigkeiten der Kinder wurden mittels des HRT 1–4 erfasst (Haffner, Baro, Parzer & Resch, 2005). Der HRT 1–4 umfasst 11 Untertests, die sich auf zwei Subskalen verteilen: 1) *Rechenoperationen* (6 Untertests, $r_{tt} = .93$): Addition (z.B. $2 + 4$), Subtraktion (z.B. $10 - 4$), Multiplikation (z.B. $2 \cdot 5$), Division (z.B. $10 : 5$), Ergänzungsaufgaben (z.B. $_ + 5 = 12$), Größer-Kleiner-Vergleiche (z.B. $4 _ 50$) und 2) *numerisch-logische und räumlich-visuelle Fähigkeiten* (5 Untertests, $r_{tt} = .87$): Zahlenreihen (z.B. $10 \ 8 \ 6 \ _$), Längenschätzen, Würfelzählen, Mengenzählen,

Tabelle 1. Demografische Angaben und Testergebnisse in Mathematik, Intelligenz und Leseflüssigkeit gruppiert nach Klassenstufen

Details	Klassenstufe 2	Klassenstufe 3	Klassenstufe 4
n (Jungen) ^{1,2}	172 (81)	154 (73)	50 (25)
	M (SD)	M (SD)	M (SD)
Alter in Monaten ²	98.01 _a (6.64)	110.66 _b (6.88)	121.31 _c (6.62)
HRT 1–4	101.81 _a (14.31)	99.05 _a (15.33)	96.68 _a (15.75)
CODY-M-2-4	100.11 _a (15.33)	99.74 _a (14.74)	100.43 _a (14.93)
CFT 1-R/20-R	101.65 _a (14.86)	101.09 _a (14.06)	90.92 _b (15.39)
SLS 1–4	99.56 _a (15.05)	100.24 _a (14.35)	100.78 _a (16.95)

Anmerkungen: ¹ $\chi^2(2) = 0.13559$, $p = .935$. ² Teils fehlende Alters- und Geschlechtsangaben. Alle Werte sind Normwerte. Skalierung für alle Instrumente ($M: 100$, $SD: 15$). HRT 1–4 (Rechenfertigkeiten); CODY-M-2-4 (vorrangig basisnumerische Fertigkeiten); CFT (Intelligenz), CFT 1-R für die Klassenstufen zwei und drei, CFT 20-R für Klassenstufe vier; SLS 1–4 (Leseflüssigkeit). Alle Gruppenvergleiche wurden nach Tukey korrigiert ($\alpha = .05$). Kleinbuchstaben, die von zwei Gruppen nicht geteilt werden, weisen auf statistisch signifikante Mittelwertunterschiede im Post-hoc-Vergleich hin.

Zahlenverbinden. Jeder Subtest besteht aus mind. 10 und höchstens 40 Aufgaben, die in der Reihenfolge mit zunehmender Schwierigkeit angeordnet sind. Pro Subtest sollten innerhalb von zwei Minuten (Min.) so viele Aufgaben wie möglich korrekt gelöst werden. Als Testergebnis wurde die Gesamtpunktzahl beider Subskalen kombiniert, wobei die Aufgaben Multiplikation, Division, Zahlenverbinden und Würfelzählen nicht berücksichtigt wurden (Haffner et al., 2005).

Die basisnumerischen Fertigkeiten der Kinder wurden mithilfe des CODY-M 2-4 erfasst (Kuhn et al., 2017). Die Subskala *basale Zahlenverarbeitung* ($r_{tt} = .72$) besteht aus drei Untertests: Punkte Zählen (18 Aufgaben), Mengenvergleich symbolisch und gemischt (je 24 Aufgaben). Die Testwerte basieren auf reaktionszeitbasierten Effizienzmaßen. Die vier Untertests Zahlendiktat (8 Aufgaben), Zahlensteine (max. 140 Aufgaben), Fehlende Zahl (16 Aufgaben) und Zahlenstrahl (23 Aufgaben) lassen sich der Subskala *komplexe Zahlenverarbeitung* ($r_{tt} = .76$) zuordnen. Die Testergebnisse der ersten drei Aufgaben basieren auf der Anzahl richtig und falsch gelöster Aufgaben, während beim Zahlenstrahl die mittlere Abweichung von der Zielzahl als Testergebnis dient. Die Subskala *Rechnen* ($r_{tt} = .85$) umfasst vier Untertests: Addition, Subtraktion (je sieben Aufgaben), Multiplikation und Platzhalteraufgaben (je vier Aufgaben). Als Testscore wird jeweils die Anzahl der Richtigantworten verwendet. Die Skala *visuell-räumliches Arbeitsgedächtnis* ($r_{tt} = .61$) besteht aus dem Test Matrixspanne, der 16 Aufgaben beinhaltet. Als Testscore wird die Anzahl vollständig korrekt reproduzierter Muster verwendet. Als Testergebnis (Gesamttest: $r_{tt} = .88$) wurden die Ergebnisse der vier Subskalen kombiniert (Kuhn et al., 2017). Die Bearbeitungszeit des CODY-Tests variierte je nach individueller Lösungsgeschwindigkeit zwischen 25 und 35 Min.

Zur Erfassung der allgemeinen Intelligenz der Zweit- und Drittklässler_innen wurde die Kurzform der *Grundintelligenztest Skala 1-Revision* (CFT 1-R; Weiß & Osterland, 2013) eingesetzt, die aus drei Untertests besteht: Reihen fortsetzen, Klassifikationen und Matrizen ($\alpha = .95$). Die Viertklässler_innen wurden anhand von vier Untertests der *Grundintelligenztest Skala 2-Revision* (CFT 20-R; Weiß, 2006) getestet: Reihen fortsetzen, Klassifikationen, Matrizen und topologische Schlussfolgerungen ($r_{tt} = .80$).

Zur Beurteilung der Leseflüssigkeit wurde das *Salzburger Lese-Screening* (SLS 1-4; Mayringer & Wimmer, 2003) verwendet. Den Kindern wurde eine Liste mit 84 einfachen Sätzen (z. B. „Bananen sind rot“) vorgelegt ($r_{tt} = .87-.90$). Innerhalb von drei Min. sollten möglichst viele Sätze auf ihre Korrektheit hin beurteilt werden. Das Testergebnis (Lesequotient [LQ]) ist wie der Intelligenzquotient skaliert ($M: 100, SD: 15$) und basiert auf der Gesamtzahl richtig bewerteter Sätze.

Zur Früherkennung von Kindern mit einem RS-Risiko wurde der FERMAT, basierend auf den Beobachtungen der Lehrkräfte der letzten drei Unterrichtsmonate, genutzt. Die 19 einzuschätzenden Items gliedern sich in drei Bereiche: 1) *Basisnumerische Fertigkeiten* (FERMAT-BF), 2) *Rechenfertigkeiten* (FERMAT-RF) (je acht Items, s. Abschnitt Fragebogenkonstruktion) und 3) *Ergänzungsfragen* (drei Items zur globalen Einschätzung). Ein beigefügter Erklärungsbogen illustriert typische Schwierigkeiten, die bei Kindern mit RS beobachtbar sind (s. Fragebogen im ESM1). Der FERMAT kann auf einem DIN-A4-Bogen ausgefüllt werden. Ein Item wird angekreuzt, wenn ein Kind regelmäßig Schwierigkeiten in einer beschriebenen Fertigkeit hat. Themen, die noch nicht im Unterricht behandelt wurden, können ausgelassen werden. Der Gesamtscore ergibt sich aus der Summe der gesetzten Kreuze. Die Ergänzungsfragen fließen nicht in den Testscore ein. Pro Klasse beträgt die Bearbeitungszeit für die Lehrkräfte ca. 10 bis 15 Min.

Ergebnisse

Häufigkeitsverteilungen für Rechenstörungen

Von 377 Kindern erfüllten 45 (11.94 %) im HRT 1-4 das Kriterium eines RS-Risikos ($PR \leq 10$). Davon erfüllten 32 (71.11 %) Kinder das strengere Kriterium und erreichten nur einen $PR \leq 7$. Ermittelt durch den CODY-M 2-4 erfüllten 40 der 377 Kinder (10.61 %) das Kriterium eines RS-Risikos ($PR \leq 10$), wobei davon 31 (77.5 %) nur einen $PR \leq 7$ erreichten. Von den Kindern, die im HRT 1-4 Minderleistungen ($PR \leq 10$) zeigten, wiesen 40 % zusätzlich Schwierigkeiten ($PR \leq 10$) im Lesen auf, wobei 13 % gravierende Minderleistungen ($PR \leq 2$) entspricht auf der IQ-Metrik etwa einem Wert von 70) zeigten. Etwa 11 % erreichten einen IQ von ≤ 70 . Von den Kindern ohne Minderleistungen (HRT 1-4, $PR > 10$), zeigten ca. 8 % erwartungswidrige Leseleistungen ($PR \leq 10$), wobei etwa 1 % nur einen $PR \leq 2$ erreichte. 1 % der Kinder ohne Beeinträchtigungen im Rechnen erreichten einen IQ-Wert von ≤ 70 . Die Kinder wurden in dieser Studie nicht ausgeschlossen, da diese zusätzliche Beeinträchtigung für die reine Identifikation von Kindern mit einem RS-Risiko zunächst irrelevant ist.

Psychometrische Eigenschaften

F1: Die postulierte Struktur des theoriebasiert entwickelten FERMAT wurde mittels einer konfirmatorischen Faktorenanalyse für kategoriale Daten untersucht (z. B. Wirth & Edwards, 2007). Die Items MU, DI und ZR-1000 wurden von

den Analysen ausgeschlossen, da nicht sichergestellt werden konnte, dass alle Kinder in diesen Kompetenzbereichen bereits formal unterrichtet wurden. Der Likelihood-Ratio-Test (LRT) des Vergleichs eines einfaktoriellen mit einem zweifaktoriellen Modell bevorzugt das zweifaktorielle Modell: $\chi^2(1) = 6.7114, p < .01$. Die Fitindizes sprechen, interpretiert nach den Richtwerten von Hu und Bentler (1999), für eine gute Passung: $\chi^2 = 76.900, df = 64, p < .129$; CFI = .997; TLI = .995; RMSEA = .024). Da die beiden Faktoren (FERMAT-BF und -RF) hoch korreliert sind $r = .93$ [.87; .98], wurde in den nachfolgenden Analysen neben den Subskalen auch die Gesamtskala berücksichtigt.

Als Maß für die interne Konsistenz wurde McDonald's Omega (McDonald, 1999) herangezogen. Der FERMAT ($\omega = .91$) und die Subskalen (FERMAT-BF, $\omega = .85$; FERMAT-RF, $\omega = .91$) erwiesen sich als hoch reliabel.

F2: Zur Exploration der Rasch-Konformität wurde die globale Modellpassung des Rasch-Modells (one parameter logistic model/1PL-Modell; Rasch, 1960) und eines zweiparametrischen IRT-Modells (2PL-Modell; Birnbaum, 1968), welches zusätzlich zur Itemschwierigkeit eine itemspezifische Trennschärfe schätzt, verglichen. Der LRT ergab eine bessere Passung für das 2PL-Modell, während das informationstheoretische Fit-Maß des BIC für den FERMAT-Gesamtscore und die Subskala FERMAT-BF eher einen besseren Fit für das Rasch-Modell nahelegte (Tab. 2). Da die Schätzungen der Personenparameter verschiedener unidimensionaler IRT-Modelle (1PL/2PL) sehr hoch korrelieren (> 0.9) (von Davier, 2016), wird der FERMAT im Folgenden Rasch-konform ausgewertet. Die Itemschwierigkeiten für den FERMAT-Gesamtscore reichten von 1.993 ($SE = 0.219$) bis 6.309 ($SE = 0.423$) (Tab. 2 im ESM3).

F3: Um die Treffsicherheit von Kindern mit einem RS-Risiko mittels des FERMAT zu bestimmen, wurde ein

klassifikatorischer Ansatz verwendet. Mit Receiver Operating Characteristic Curve Analysen wurde die Beziehung zwischen Sensitivität (SN) und Spezifität (SP) mit verschiedenen Cut-Off-Werten eruiert. Als Cut-Off-Werte wurden die Werte herangezogen, die basierend auf der max. Trennschärfe identifiziert wurden: die datenbasierten Cut-Off-Werte repräsentieren den Punkt, an dem die max. Differenzierung zwischen Kindern mit und ohne Anzeichen für RS erreicht wird (Tröster, 2009). Abhängig von den Kriterien zur Klassifikation des RS-Risikos (HRT 1-4, CODY-M 2-4, $PR \leq 7$ und $PR \leq 10$) erreichten der Gesamtscore bzw. die Subskalen des FERMAT ihre max. Trennschärfe bei Cut-Off-Werten von ein bis vier Testpunkten. Basierend darauf sollten Kinder, je nach Kriterium, bereits beim Vorliegen von Schwierigkeiten in einem der im FERMAT erfassten Fertigkeiten einer weiterführenden Diagnostik zugeteilt werden.

Anschließend wurden die Testergebnisse des FERMAT (positiv = Risiko für RS/negativ = kein Risiko für RS) und die Ergebnisse der standardisierten Tests (HRT 1-4 und CODY-M 2-4), d.h. das tatsächliche Vorliegen eines RS-Risikos (vorhanden/nicht vorhanden) in einer Vierfeldertafel zusammengefasst. Die Spalten der Tabelle geben die Häufigkeiten des Auftretens vier möglicher Kombinationen an: die richtigen Vorhersagen (richtig-positiv bzw. richtig-negativ) und die falschen Vorhersagen (falsch-positiv bzw. falsch-negativ).

Für den FERMAT-Gesamtscore sowie für die beiden Subskalen FERMAT-BF und FERMAT-RF wurden folgende Gütekriterien berechnet: SN und SP, die Positive und Negative Korrektheit (PK und NK), der AUC-Wert (Area Under the Curve), der prävalenzunabhängige Youden-Index sowie der RAZ-Index, der den Relativen Anstieg der Trefferquote gegenüber der Zufallstrefferquote beschreibt (Marx,

Tabelle 2. Globale Modellpassung des 1PL und 2PL-Modells

	AIC	SABIC	HQ	BIC	logLik	χ^2	df	p
<i>FERMAT-Gesamtscore</i>								
1PL	1983.531	1994.164	2005.383	2038.583	-977.766			
2PL	1942.876	1962.623	1983.457	2045.115	-945.438	64.655	12	.001
<i>FERMAT-BF</i>								
1PL	797.316	803.392	809.802	828.774	-390.658			
2PL	787.430	798.063	809.282	842.482	-379.715	21.886	6	.001
<i>FERMAT-RF</i>								
1PL	1347.403	1352.719	1358.328	1374.928	-666.701			
2PL	1309.992	1319.106	1328.722	1357.179	-642.996	47.41	5	.001

Anmerkungen: 1PL = one parameter logistic model/1PL-Modell; 2PL = two parameter logistic model/2PL-Modell; AIC = Akaike Information Criterion; SABIC = Sample-Size Adjusted Bayesian Information Criterion; HQ = Hannan-Quinn Criterion; BIC = Bayesian Information Criterion; logLik = Log-Likelihood-Wert; χ^2 = Wert des Likelihood-Ratio-Tests.

Tabelle 3. Ergebnisse der ROC-Analysen

	HRT 1–4 PR ≤ 10	HRT 1–4 PR ≤ 7	CODY-M 2–4 PR ≤ 10	CODY-M 2–4 PR ≤ 7
<i>FERMAT-Gesamtscore</i>				
AUC [KI]	.835 [.765, .905]	.802 [.714, .890]	.795 [.706, .884]	.828 [.733, .924]
Cut-Off (YI)	2	2	4	4
YI	.546	.503	.489	.547
SN	.698	.677	.576	.640
SP	.848	.825	.914	.907
PK	.395	.276	.413	.348
NK	.952	.963	.953	.970
FP	46	55	27	30
FN	13	10	14	9
RATZ-Index	.613	.587	.511	.585
<i>FERMAT-BF</i>				
AUC [KI]	.733 [.655, .812]	.722 [.629, .815]	.726 [.634, .818]	.774 [.670, .877]
Cut-Off (YI)	1	1	1	2
YI	.446	.432	.415	.500
SN	.545	.548	.529	.538
SP	.901	.884	.886	.962
PK	.429	.304	.321	.519
NK	.936	.955	.949	.965
FP	32	39	38	13
FN	20	14	16	12
RATZ-Index	.464	.467	.445	.502
<i>FERMAT-RF</i>				
AUC [KI]	.829 [.758, .900]	.793 [.704, .882]	.803 [.729, .878]	.825 [.747, .903]
Cut-Off (YI)	2	3	1	1
YI	.544	.529	.527	.563
SN	.682	.625	.821	.867
SP	.862	.904	.707	.696
PK	.411	.392	.256	.208
NK	.951	.961	.970	.983
FP	43	31	93	99
FN	14	12	7	4
RATZ-Index	.600	.562	.723	.795

Anmerkungen: AUC = Area Under the Curve; KI = 95%-Konfidenzintervall; YI = Youden-Index; SN = Sensitivität; SP = Spezifität; PK = Positive Korrektheit; NK = Negative Korrektheit; FP = Falsch-Positiv; FN = Falsch-Negativ; RATZ-Index = Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote.

1992). Alle Youden-Indizes waren positiv und sind damit als gut zu bewerten (Youden, 1950). Alle AUC-Werte lagen nach Backhaus, Erichson, Plinke und Weiber (2018) im guten bis sehr guten Bereich und bestätigten damit, dass der FERMAT in der Lage ist, zwischen Kindern mit und ohne Anzeichen eines Risikos für RS zu unterscheiden.

Die SN des *FERMAT-Gesamtscores* variierte je nach Kriterium und lag zwischen 57.6% und 69.8% (Tab. 3). D.h. durch den FERMAT werden max. ca. 70% der Kinder mit testdiagnostischen Anzeichen für RS korrekt klassifiziert. Der höchste positive prädiktive Wert wurde für das Kriterium CODY-M 2-4 $PR \leq 10$ ermittelt und betrug 0.413, was bedeutet, dass ca. 40% der im FERMAT als auffällig erkannten Kinder auch auffällige Leistungen im CODY-M 2-4 zeigten. Die ermittelten RAZ-Indizes bewegen sich in einem Bereich von 51.1% bis 61.3%. Werte über 60% zeigen eine deutliche Verbesserung der Vorhersagegenauigkeit im Vergleich zu einer rein zufälligen Zuordnung (Marx, 1992). Hingegen wurden mind. etwa 80% der Kinder ohne Anzeichen für RS korrekt durch den FERMAT erkannt. Der niedrigste negative prädiktive Wert wurde für das Kriterium HRT 1-4 $PR \leq 10$ ermittelt und lag bei 0.952, d.h. ca. 95% der Kinder, die im FERMAT als unauffällig identifiziert wurden, erzielten auch im standardisierten Leistungstest (HRT 1-4 $PR \leq 10$) unauffällige Ergebnisse (Tab. 3).

Für den Teilbereich *FERMAT-BF* lag die SN je nach Kriterium zwischen 52.9% und 54.8%. Somit wurden max. etwa 55% der Kinder korrekt klassifiziert. Die ermittelten RAZ-Indizes lagen im Bereich von 44.5% bis 50.2%. Der niedrigste positive prädiktive Wert zeigte, dass mind. ca. 30% der Kinder, die im Teilbereich FERMAT-BF als rechenwach identifiziert wurden, tatsächlich Minderleistungen im Rechnen zeigten (HRT 1-4 $PR \leq 7$). Der höchste positive prädiktive Wert wurde für das Kriterium CODY-M 2-4 $PR \leq 7$ ermittelt und betrug 51.9%. Der niedrigste negative prädiktive Wert zeigt, dass ca. 94% der Kinder, die als unauffällig identifiziert wurden, auch im standardisierten Leistungstest (HRT 1-4 $PR \leq 10$) unauffällige Ergebnisse erzielten (Tab. 3).

Für den Teilbereich *FERMAT-RF* lag die SN je nach Kriterium zwischen 62.5% und 86.7%. Damit wurden ca. 63% bis 87% der Kinder korrekt identifiziert. Die ermittelten RAZ-Indizes lagen in einem Bereich von 56.2% bis 79.5%. Der niedrigste positive prädiktive Wert zeigt, dass etwa ein Fünftel der Kinder, die im Teilbereich FERMAT-RF als auffällig identifiziert wurden, tatsächlich Minderleistungen im Rechnen aufwiesen (CODY-M 2-4 $PR \leq 7$). Der höchste positive prädiktive Wert wurde für das Kriterium HRT 1-4 $PR \leq 10$ ermittelt und betrug ca. 41%. Der niedrigste negative prädiktive Wert zeigte, dass ca. 95% der Kinder, die als unauffällig identifiziert wurden, auch im HRT 1-4 ($PR \leq 10$) unauffällige Ergebnisse erzielten (Tab. 3).

Diskussion

Anfängliche Rechenschwierigkeiten sollten früh erkannt werden (AWMF, 2018). Die meisten Screeninginstrumente konzentrieren sich daher auf den Vorschulbereich und das frühe Schulalter (Landerl et al., 2022), allerdings entwickeln einige Kinder erst im Verlauf der Grundschulzeit Schwierigkeiten (z. B. Kohn et al., 2013). Für diese Zielgruppe fehlen Screeninginstrumente mit angemessener Treffsicherheit, die Lehrkräfte zeiteffizient in ihren Unterricht integrieren können. Die vorliegende Arbeit schließt an diese Forschungslücke durch die Entwicklung des FERMAT an und konzentriert sich auf die Überprüfung der psychometrischen Eigenschaften des Instruments. Im Folgenden werden die Befunde entlang der Forschungsfragen diskutiert.

F1: Beeinträchtigungen in der Basisnumerik zählen neben Defiziten in den Rechenfertigkeiten zu den Kernsymptomen von RS (z. B. Busch et al., 2013). Daher berücksichtigt der FERMAT im Vergleich zu anderen Instrumenten zur (Früh-)Erkennung nicht nur Rechenfertigkeiten, sondern auch basisnumerische Fertigkeiten. Diese theoretische zweidimensionale Struktur wurde empirisch bestätigt, wobei eine Differenzierung hinsichtlich der Risikoidentifikation mittels des FERMAT nicht sinnvoll ist, denn unabhängig davon, in welchem Teilbereich die Schwierigkeiten von der Lehrkraft eingeschätzt werden, sollte die weitere diagnostische Abklärung beide Teilbereiche berücksichtigen, um geeignete Interventionen auszuwählen (AWMF, 2018; Kuhn & Schwenk, 2018). Dennoch unterstützt der FERMAT bei der Auswahl geeigneter Tests. Werden im FERMAT z. B. vorrangig Schwierigkeiten in der Basisnumerik festgestellt, empfiehlt es sich, in der weiteren Diagnostik den Schwerpunkt auf diesen Bereich zu legen.

F2: Basierend auf dem LRT wurde die globale Modellpassung des FERMAT besser durch das 2PL- als durch das 1PL-Modell abgebildet. Da die Schätzungen der Personenparameter verschiedener unidimensionaler IRT-Modelle hoch korrelieren (von Davier, 2016) und informationstheoretische Maße (BIC) das überwiegend 1PL-Modell bevorzugen, wurden in der vorliegenden Studie Rasch-konform variierende Schwierigkeitsparameter, nicht aber unterschiedliche Trennschärfen berücksichtigt. Die Rasch-Konformität sollte in Folgestudien klassenstufenspezifisch erneut überprüft werden.

F3: Durch den FERMAT werden Kinder mit einem erhöhten RS-Risiko mit angemessener Treffsicherheit identifiziert. Ca. 70% der Kinder, die unterdurchschnittliche Leistungen ($PR \leq 10$) im HRT 1-4 zeigten, wurden durch den FERMAT (Gesamtscore) identifiziert. Etwa 55% der Kinder, die Minderleistungen im HRT 1-4 ($PR \leq 7$) zeigten, wurden durch die Subskala FERMAT-BF identifiziert. Durch die Subskala FERMAT-RF konnten fast 90% der Kinder, die im CODY-M 2-4 einen $PR \leq 7$ erreichten, rich-

tig erkannt werden. Dass die Treffsicherheit der Subskala FERMAT-RF höher ausfiel, könnte mit der diagnostischen Kompetenz der Lehrkräfte in Zusammenhang stehen. Der Mathematikunterricht in der Grundschule schließt an die frühen mathematischen Alltagserfahrungen der Kinder an, doch der Fokus liegt auf der Vermittlung mathematischer Inhalte. Daher wäre es plausibel, dass die Lehrkräfte die Rechenfertigkeiten akkurater einschätzen als die basisnumerischen Fertigkeiten. Der FERMAT könnte dazu beitragen, das Bewusstsein der Lehrkräfte für Schwierigkeiten in der Basisnumerik zu schärfen, die selbst in der weiterführenden Schule noch substanziell mit der Mathematikleistung assoziiert sind (Ennemoser, Krajewski & Schmidt, 2011). Studien in anderen Bereichen zeigten, dass fragebegünstigte Beurteilungsverfahren zuverlässigere Einschätzungen ermöglichen (Begeny, Eckert, Montarello & Storie 2008) und explizite Informationen über den zu beurteilenden Bereich die Genauigkeit der Beurteilungen verbessern können (Demaray & Elliott, 1998).

Die datenbasierten Cut-Off-Werte des FERMAT reichen von einem bis vier Testpunkten, je nach Betrachtung des Gesamtwerts oder der Subskalen. Um schwerwiegendere falsch-negative Ergebnisse zu vermeiden, empfiehlt sich die Nutzung eines konservativeren Cut-Off-Wertes von zwei. Denn auch wenn ein falsch-positiver Befund negative Auswirkungen wie Stigmatisierung oder schulische Ängste haben kann (Fischbach et al., 2013; Tröster, 2009), sind falsch-negative Ergebnisse im Kontext von Lernschwierigkeiten weitaus gravierender. Setzen sich anfängliche Schwierigkeiten in einer negativen Abwärtsspirale fort und führen zu sekundären Auffälligkeiten, sind bspw. rechenspezifische Interventionen weniger wirksam (Herzog & Casale, 2022). Weist ein Kind in zwei der durch den FERMAT abgedeckten Fertigkeiten Schwierigkeiten auf, ergibt sich eine SN von 69,8%, d. h. ca. 70% aller Kinder, die im HRT 1–4 Minderleistungen aufwiesen ($PR \leq 10$), ließen sich durch den FERMAT korrekt identifizieren. Durch die frühe Risikoidentifikation können Lehrkräfte mit einzelnen Kindern weitere Tests durchführen (eine Übersicht über geeignete Test- und Förderverfahren bietet bspw. das LONDI-Hilfssystem; Schulte-Körne & Hasselhorn, 2022), um die Erfordernis weiterer Maßnahmen zur Abklärung einer klinisch relevanten RS zu prüfen und um auf dieser Basis adaptive Lehr-Lernsituationen (Fischer et al., 2017) bzw. Fördermaterialien (LONDI-Hilfssystem; Schulte-Körne & Hasselhorn, 2022) bereit zu stellen.

Einige Kinder wiesen unterdurchschnittliche Leistungen in der Leseflüssigkeit und Intelligenz auf. Obgleich diese zusätzliche Beeinträchtigung für die RS-Risikoidentifikation zunächst irrelevant ist, belegen die Befunde erneut, dass einige Kinder multiple schulische Schwierigkeiten aufweisen (z. B. Fischbach et al., 2013). Dies unterstreicht die Relevanz der Berücksichtigung weiterer

Ein- und Ausschlusskriterien in einer nachfolgenden Diagnostik. Des Weiteren konnte beobachtet werden, dass etwa 8% der Kinder zwar keine Schwierigkeiten im Rechnen, jedoch im Lesen aufwiesen, weshalb auch Screening-Instrumente für die Früherkennung von Lese- und Rechtschreibstörungen sinnvoll erscheinen.

Wiederum wurden ca. 85% der Kinder, die keine Schwierigkeiten (HRT 1–4, $PR > 10$) aufwiesen, im FERMAT als unauffällig identifiziert und damit ebenfalls richtig erkannt. Die Anzahl der fälschlicherweise als rechen-schwach identifizierten Kinder, beläuft sich auf ca. 15%. Auf Basis qualitativer Daten oder förderdiagnostischer Ansätze könnte zukünftig untersucht werden, ob sich ein systematisches Muster bezüglich der Fehlklassifikation identifizieren lässt. Da das Testergebnis im FERMAT keine Diagnose mit lerntherapeutischer Intervention, sondern lediglich eine weiterführende Diagnostik anregt, kann die Treffsicherheit des FERMAT aus beiden Perspektiven als ausreichend bewertet werden.

Limitationen

Die ermittelten Cut-Off-Werte basieren auf der max. Trennschärfe. Künftige Studien sollten den Einfluss der Cut-Off-Werte im Hinblick auf weitere Gütekriterien (z. B. Aufdeckungsrate) kontrastieren. Obwohl eine geringe positive Korrektheit bei hoher SN und ausreichenden RATZ-Indizes nicht unüblich ist, ist eine Verbesserung der positiven Korrektheit anzustreben. Stichprobenspezifische Effekte, insbesondere bei der positiven und negativen Korrektheit, sind möglich, da sie nicht die Testeigenschaften widerspiegeln, sondern von der Grundquote der Stichprobe abhängen (Tröster, 2009). Zukünftige Studien sollten sicherstellen, dass die Kinder bereits formal in allen vom FERMAT erfassten Kompetenzen unterrichtet wurden, um das Ausschließen von Items in den Analysen zu vermeiden. Die prognostische Validität sollte durch zwei unabhängige Messzeitpunkte ergänzend überprüft werden. Ob andere Kriterien zur Klassifikation von RS die Genauigkeit, mit der der FERMAT Kinder mit RS-Risiko identifiziert beeinflusst und ob die Ergebnisse in Abhängigkeit der Klassenstufen variieren, sollte künftig ebenfalls getestet werden. Da die Daten aus einem Bundesland stammen, könnte die Generalisierbarkeit der Ergebnisse eingeschränkt sein. Aufgrund der freiwilligen Teilnahme könnten selektiert Lehrkräfte an der Befragung teilgenommen haben, die sich in ihrer Einschätzung über die mathematischen Fertigkeiten ihrer Schüler_innen sicher fühlten. Künftig sollten Informationen erhoben werden, die einen Einfluss auf die Lehrkräfteeinschätzung und damit auf die Güte und Bearbeitungszeit des FERMAT haben könnten (z. B. Berufserfahrung, diagnostische Kompetenz etc.).

Relevanz für die Praxis

Der FERMAT kann Kinder mit einem erhöhten Risiko für RS ökonomisch und valide identifizieren. Die Anwendung erfordert keine umfangreiche Schulung. Alle Ausfüllhinweise inkl. Beispiele für die typisch zu beobachtenden Schwierigkeiten bei RS sind auf dem Fragebogen vermerkt. Die Auswertung erfolgt durch eine einfache Summenbildung, dessen Interpretation eindeutig ist: A) Weist ein Kind in weniger als zwei Bereichen Schwierigkeiten auf, liegt kein erhöhtes RS-Risiko vor. B) Weist ein Kind in mind. zwei Bereichen Schwierigkeiten auf, sollten die mathematischen Fertigkeiten des Kindes in einem standardisierten Test überprüft werden. Der ggf. ins Digitale übertragbare FERMAT kann somit dazu beitragen, dass personelle Ressourcen im schulischen Setting optimal genutzt und Kinder mit einem RS-Risiko frühzeitig erkannt und gefördert werden.

Elektronische Supplemente (ESM)

Die elektronischen Supplemente sind mit der Online-Version dieses Artikels verfügbar unter <https://doi.org/10.1024/2235-0977/a000456>.

ESM1. Fragebogen zur Erfassung mathematischer Fertigkeiten (FERMAT).

ESM2. Tabelle E1. Verortung der numerischen Kompetenzbereiche des FERMAT im Kompetenzentwicklungsmodell nach Fischer et al. (2017).

ESM3. Tabelle E2. Itemschwierigkeiten und Standardfehler des FERMAT-Gesamtscores basierend auf dem 1 PL-Modell.

Literatur

American Educational Research Association, American Psychological Association & National Council for Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.

American Psychiatric Association (APA). (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American psychiatric association. Verfügbar unter <https://doi.org/10.1176/appi.books.9780890425596>

Andersson, U. & Östergren, R. (2012). Number magnitude processing and basic cognitive functions in children with mathematical learning disabilities. *Learning and Individual Differences*, 22(6), 701–714. <https://doi.org/10.1016/j.lindif.2012.05.004>

Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) (2018). *S3-Leitlinie: Diagnostik und Behandlung der Rechenstörung*. Verfügbar unter: <https://www.awmf.org/leitlinien/detail/ll/028-046.html>

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2018). *Multivariate Analysemethoden*. Berlin: Springer.

Begeny, J. C., Eckert, T. L., Montarello, S. A. & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43–55. <https://doi.org/10.1037/1045-3830.23.1.43>

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Hrsg.), *Statistical Theories of Mental Test Scores* (S. 397–479). Frankfurt a.M.: Addison-Wesley.

Busch, J., Oranu, N., Schmidt, C. & Grube, D. (2013). Rechenschwäche im Grundschulalter: Reduzierte Verfügbarkeit basalen arithmetischen Faktenwissens und Belastung des Arbeitsgedächtnisses bei Drittklässlern. *Lernen und Lernstörungen*, 2, 217–227. <https://doi.org/10.1024/2235-0977/a000043>

Butterworth, B., Varma, S. & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science*, 332(6033), 1049–1053. <https://doi.org/10.1126/science.1201536>

Clausen-Suhr, K. & Walter, J. (2022). Entwicklung und Evaluation eines Screening-Verfahrens zur Prognose von Rechenschwierigkeiten in der Grundschule. *Das Flensburger Schulspiel (FleSch)*. *Lernen und Lernstörungen*, 11, 125–138. <https://doi.org/10.1024/2235-0977/a000373>

Von Davier, M. (2016). Rash model. In W. J. van der Linden (Hrsg.), *Handbook of Item Response Theory* (Vol 1, S. 31–48). Boca Raton, FL, USA: Chapman and Hall/CRC. <https://doi.org/10.1201/9781315374512>

Decarli, G., Sella, F., Lanfranchi, S., Gerotto, G., Gerola, S., Cossu, G. et al. (2023). Severe Developmental Dyscalculia Is Characterized by Core Deficits in Both Symbolic and Nonsymbolic Number Sense. *Psychological Science*, 34(1), 8–21. <https://doi.org/10.1177/09567976221097947>

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1–2), 1–42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N)

Demaray, M. K. & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13(1), 8–24. <https://doi.org/10.1037/h0088969>

Dögnitz, S. (2022). *Diagnostik von besonderen Rechenschwierigkeiten in der Sekundarstufe I*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-40071-2_9

Ennemoser, M., Krajewski, K. & Schmidt, S. (2011). Entwicklung und Bedeutung von Mengen-Zahlen-Kompetenzen und eines basalen Konventions- und Regelwissens in den Klassen 5 bis 9. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 228–242. <https://doi.org/10.1026/0049-8637/a000055>

Ennemoser, M., Sinner, D. & Krajewski, K. (2015). Kurz- und langfristige Effekte einer entwicklungsorientierten Mathematikförderung bei Erstklässlern mit drohender Rechenschwäche. *Lernen und Lernstörungen*, 4, 43–59. <https://doi.org/10.1024/2235-0977/a000091>

Feigenson, L., Dehaene, S. & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>

Fischbach, A., Schuchardt, K., Brandenburg, J., Kleszczewski, J., Balke-Melcher, C., Schmidt, C. et al. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien. *Lernen und Lernstörungen*, 2, 65–76. <https://doi.org/10.1024/2235-0977/a000035>

Fischer, U., Rösch, S. & Moeller, K. (2017). Diagnostik und Förderung bei Rechenschwäche. *Lernen und Lernstörungen*, 6, 25–38. <https://doi.org/10.1024/2235-0977/a000160>

Fischer, U., Rösch, S., Nuerk, H. C. & Moeller, K. (2015). Erkennen von Rechenschwäche durch LehrerInnen und Testungen im Klassenverband. *Lernen und Lernstörungen*, 4, 269–282. <https://doi.org/10.1024/2235-0977/a000116>

- Fusion, K. C. (2012). *Children's counting and concepts of number*. Wiesbaden: Springer.
- Gaupp, N., Zoelch, C. & Schumann-Hengsteler, R. (2004). Defizite numerischer Basiskompetenzen bei rechenschwachen Kindern der 3. und 4. Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, 18, 31 – 42. <https://doi.org/10.1024/1010-0652.18.1.31>
- Grosche, M. & Huber, C. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik*, 81(2), 312 – 322.
- Haberstroh, S. & Schulte-Körne, G. (2019). The Diagnosis and Treatment of Dyscalculia. *Deutsches Ärzteblatt international*, 116(7), 107 – 114. <https://doi.org/10.3238/arztebl.2019.0107>
- Haffner, J., Baro, K., Parzer, P. & Resch, F. (2005). *Heidelberger Rechentest (HRT 1 – 4)*. Göttingen: Hogrefe.
- Halberda, J. & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5), 1457 – 1465. <https://doi.org/10.1037/a0012682>
- Herzog, M. & Casale, G. (2022). The effects of a computer-based mathematics intervention in primary school students with and without emotional and behavioral difficulties. *International Electronic Journal of Elementary Education*, 14(3), 303 – 317. <https://iejee.com/index.php/IEJEE/article/view/1730>
- Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1 – 55. <https://doi.org/10.1080/10705519909540118>
- Jacobs, C. & Petermann, F. (2012). *Diagnostik von Rechenstörungen*. Göttingen: Hogrefe.
- Kaufmann, S. & Wessolowski, S. (2021). *Rechenstörungen: Diagnose und Förderbausteine* (5. Aufl.). Hannover: Klett Kallmeyer.
- Kißler, C., Schwenk, C. & Kuhn, J. T. (2021). Two dyscalculia subtypes with similar, low comorbidity profiles: A mixture model analysis. *Frontiers in Psychology*, 12, 589506. <https://doi.org/10.3389/fpsyg.2021.589506>
- Kohn, J., Wyschkon, A., Ballaschk, K., Ihle, W. & Esser, G. (2013). Verlauf von umschriebenen Entwicklungsstörungen: eine 30-monats-follow-up-studie. *Lernen und Lernstörungen*, 2, 77 – 89. <https://doi.org/10.1024/2235-0977/a000032>
- Krajewski, K. & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year-longitudinal study. *Journal of Experimental Child Psychology*, 103(4), 516 – 531. <https://doi.org/10.1016/j.jecp.2009.03.009>
- Kuhn, J.-T. (2017). Rechenschwäche – eine interdisziplinäre Einführung. In A. Fritz, S. Schmidt & G. Ricken (Hrsg.), *Handbuch Rechenschwäche: Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie* (3. Aufl., S. 14 – 29). Weinheim: Beltz.
- Kuhn, J.-T., Raddatz, J., Holling, H. & Dobel, C. (2013). Dyskalkulie vs. Rechenschwäche: Basisnumerische Verarbeitung in der Grundschule. *Lernen und Lernstörungen*, 2, 229 – 247. <https://doi.org/10.1024/2235-0977/a000044>
- Kuhn, J.-T., & Schwenk, C. (2018). Onlinebasierte Diagnostik mathematischer Kompetenzen: Möglichkeiten und Grenzen. *Lernen und Lernstörungen*, 7, 231 – 235. <https://doi.org/10.1024/2235-0977/a000232>
- Kuhn, J.-T., Schwenk, C., Raddatz, J., Dobel, C. & Holling, H. (2017). *CODY-M 2–4. CODY-Mathetest für die 2.–4. Klasse. Manual*. Düsseldorf: Kaasa health.
- Landerl, K., Vogel, S. & Kaufmann, L. (2022). *Dyskalkulie: Modelle, Diagnostik, Intervention* (4. überarb. und erw. Auflage). UTB: Bd. 3066. München: Ernst Reinhardt. <https://doi.org/10.36198/9783838557342>
- Lorenz, J. H. (2012). *Kinder begreifen Mathematik: Frühe mathematische Bildung und Förderung. Entwicklung und Bildung in der Frühen Kindheit*. Stuttgart: Kohlhammer.
- Marx, H. (1992). Methodische und inhaltliche Argumente für und wider einer frühen Identifikation und Prädiktion von Leserechtschreibschwierigkeiten. *Diagnostica*, 38, 249 – 268.
- Mayringer, H. & Wimmer, H. (2003). *Salzburger Lese-Screening für die Klassenstufen 1 – 4 (SLS 1 – 4)*. Bern: Huber.
- Mazzocco, M. M., Feigenson, L. & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child development*, 82(4), 1224 – 1237. <https://doi.org/10.1111/j.1467-8624.2011.01608.x>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence.
- Opitz, E. M. & Bern, P. H. (2008). *Rechenschwäche erfassen: Screening für die Schuljahre 4 – 8*. Universitätsbibliothek Dortmund. <https://eldorado.tu-dortmund.de/bitstream/2003/31691/1/141.pdf>
- Parsons, S. & Bynner, J. (2005). *Does numeracy matter more?* London: NRDC.
- Peter-Koop, A., Wollring, B., Spindeler, B. & Grüßing, M. (2007). *Elementarmathematisches Basisinterview*. Offenburg: Mildenerberger.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Scherer, P. & Moser Opitz, E. (2010). *Fördern im Mathematikunterricht der Primarstufe*. Heidelberg: Springer.
- Schulte-Körne, G. & Hasselhorn, M. (2022). *LONDI-Hilfssystem. Unterstützung bei Diagnostik und Förderung von Lernstörungen bei Kindern*. Verfügbar unter: <https://hilfssystem.londi.de/>
- 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, N., C. Selter, M. Steffensky & H. Wendt (Hrsg.), *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 25 – 56). Münster; New York: Waxmann. <https://doi.org/10.31244/9783830993193>
- Torbeyns, J., Verschaffel, L. & Ghesquière, P. (2004). Strategy development in children with mathematical disabilities: Insights from the choice/no-choice method and the chronological-age/ability-level-match design. *Journal of Learning Disabilities*, 37(2), 119 – 131. <https://doi.org/10.1177/00222194040370020301>
- Tröster, H. (2009). *Früherkennung im Kindes- und Jugendalter. Strategien bei Entwicklungs-, Lern- und Verhaltensstörungen*. Göttingen: Hogrefe.
- Vaughn, S. & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction. The promise and potential problems. *Learning Disabilities Research & Practice*, 18(3), 137 – 146. <https://doi.org/10.1111/1540-5826.00070>
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – Revision (WS/ZFR)*. Göttingen: Hogrefe.
- Weiß, R. H. & Osterland, J. (2013). *Grundintelligenztest Skala 1 – Revision (CFT 1-R)*. Göttingen: Hogrefe.
- Wirth, R. J. & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, 12(1), 58 – 79. <https://doi.org/10.1037/1082-989X.12.1.58>
- World Health Organization (WHO) (2020). International Classification of Diseases for Mortality and Morbidity Statistics (11th revision). Retrieved from <https://icd.who.int/browse11/l-m/en>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32 – 35. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)
- Zuber, J., Pixner, S., Moeller, K. & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory ca-

capacity. *Journal of Experimental Child Psychology*, 102(1), 60–77.
<https://doi.org/10.1016/j.jecp.2008.04.003>

Historie

Manuskript eingereicht: 05.12.2023
Manuskript angenommen: 04.07.2024
Onlineveröffentlichung: 25.09.2024

Danksagung


Wir danken allen Kindern und Lehrkräften, die an der Studie teilgenommen haben.

Förderung


Diese Arbeit wurde finanziell durch das Bundesministerium für Bildung und Forschung unterstützt (Projektnummer: 01GJ1302). Open-Access-Veröffentlichung ermöglicht durch die Technische Universität Dortmund.

ORCID

Sarah Lamb

 <https://orcid.org/0000-0003-4619-3269>

Jörg-Tobias Kuhn

 <https://orcid.org/0000-0002-4399-9569>



Sarah Lamb

Methoden der empirischen
Bildungsforschung
Fakultät Rehabilitationswissenschaften
Technische Universität Dortmund
Emil-Figge-Straße 50
44227 Dortmund
Deutschland
sarah.lamb@tu-dortmund.de