

Ressourcenorientiert Testen? Effekte des Instruktionsdesigns auf die Testleistung

Sarah Schulze  und Jan Kuhl 

Fachgebiet Unterrichtsentwicklungsforschung mit dem Schwerpunkt Inklusion, Fakultät Rehabilitationswissenschaften,
Technische Universität Dortmund, Deutschland

Zusammenfassung: Testaufgaben stellen – ebenso wie Lernmaterialien und Lernsettings – hohe Anforderungen an die Informationsverarbeitung von Schülerinnen und Schülern. Mit dieser kognitiven Belastung beim Lernen und Problemlösen befasst sich die Cognitive Load Theory bereits seit einiger Zeit, sodass mittlerweile zahlreiche Befunde vorliegen. Auf dieser Basis wurden Gestaltungsprinzipien formuliert, an denen Unterricht und Lernmaterialien ausgerichtet werden sollten. Dabei geht es stets darum, die kognitiven Ressourcen möglichst effizient für die Auseinandersetzung mit dem Lerninhalt zu nutzen. Im Sinne der Konstruktvalidität sollte dies auch bei diagnostischen Verfahren angestrebt werden. So sollten die Testergebnisse ausschließlich von dem Zielkonstrukt, z. B. von den mathematischen Kompetenzen, bestimmt sein und nicht von entfernten Merkmalen wie dem Arbeitsgedächtnis (AG). Welchen Einfluss die ressourcenorientierte Gestaltung auf die Testergebnisse in diagnostischen Tests haben können, ist bisweilen kaum untersucht. Daher wurde in der vorliegenden Arbeit ein standardisierter Test zur Erfassung mathematischer Basiskompetenzen nach ressourcenschonenden Prinzipien adaptiert. Zur Prüfung der Effekte auf die Testleistung wurde eine Gruppenstudie mit 261 Schülerinnen und Schülern ($M = 6.95$ Jahre) durchgeführt. Zusätzlich wurden AG-Funktionen als Kontrollvariablen erfasst. Die regressionsanalytischen Ergebnisse bilden einen signifikanten Effekt der Bedingung zu Gunsten der ressourcenschonenden Adaption ab. Eine Interaktion zwischen Testversion und AG wurde nicht gefunden. Somit schneiden die Schülerinnen und Schüler im adaptierten Test unabhängig von ihren AG-Funktionen besser ab. Lehrkräfte und andere Praktikerinnen und Praktiker sollten für derartige Effekte sensibilisiert sein und Testergebnisse auch unter dem Gesichtspunkt der erforderlichen Zugangsfertigkeiten einordnen.

Schlüsselwörter: Arbeitsgedächtnis, Cognitive Load, extrinsische Belastung, Instruktionsdesign, Testfairness

Resource-oriented testing? Effects of instructional design on test performance

Abstract: Test items – as well as learning materials and learning settings – place high demands on students' information processing. Cognitive load theory has been dealing with this cognitive load during learning and problem solving for some time, so there are many results by now. On this basis, design principles have been formulated to guide teaching and learning materials. The aim is to always use cognitive resources as efficiently as possible for the learning content. In line with construct validity, this should be the goal in diagnostic procedures as well. Thus, test results should be determined exclusively by the target construct, e.g. mathematical competencies, and not by distant characteristics such as working memory. The influence of a resource-oriented design on the results of diagnostic tests is hardly investigated so far. In the present study, a standardized test for assessing basic mathematical competencies is adapted according to resource-oriented principles. To test the effects, a group study was conducted with 261 students ($M = 6.95$ years). In addition, working memory functions were recorded as control variables. The results of the regression analyses show a significant effect of the condition in favour of the resource-saving adaptation. There is no interaction between test type and working memory, so students perform better on the adapted test regardless of their working memory functions. Teachers and other practitioners should be aware of such effects and should also consider test results in terms of the access skills required.

Keywords: cognitive load, extrinsic load, instructional design, testfairness, working memory

Einleitung

In der schulischen Förderung kommt eine Vielfalt von Lernmaterialien und Aufgabenformaten zum Einsatz. Diese können mit hohen Anforderungen an die Informationsverarbeitung einhergehen (Sweller & Chandler, 1991). Aus

jeder Aufgabe resultiert eine Anforderung für die Informationsverarbeitung, wobei sich die Gesamtanforderung aus Anforderungen des Inhalts und Anforderungen des Formats bzw. Aufgabendesigns zusammensetzt (Sweller, 1989). An der Verarbeitung der Informationen sind hier als grundlegende Komponenten Arbeitsgedächtnis (AG),

Langzeitgedächtnis und die Aufmerksamkeit beteiligt. AG und Aufmerksamkeit sind in ihrer Kapazität begrenzt, sodass eine effiziente Nutzung verfolgt werden sollte, damit Wissen aufgebaut und Problemstellungen bearbeitet werden können.

Das Verhältnis von kognitiver Belastung und kognitiven Ressourcen beim Lernen wird in Theorien wie der Cognitive Load Theory (CLT, Paas, Renkl & Sweller, 2004; Sweller, 1989; Sweller & Chandler, 1991) beschrieben. Da jede Aufgabe mit Anforderungen an die Informationsverarbeitung einhergeht, ist zu erwarten, dass sich die kognitive Belastung auch bei Testleistungen im Rahmen schulischer Diagnostik zeigt. Ein unklares Instruktionsdesign könnte sich etwa ungünstig auf Bearbeitungszeit und Aufgabebearbeitung auswirken. Anknüpfend an diese Annahme wird mit der vorliegenden Studie die Frage bearbeitet, inwiefern sich die Gestaltung des Testmaterials, auf die Testleistung auswirkt.

Grundannahmen der kognitiven Belastung beim Lernen

Die Effekte des Instruktions- und Aufgabendesigns sind bereits vielfach untersucht worden (Hecht, 2014; Mayer & Moreno, 2003). Die CLT (Sweller, 1989; Sweller & Chandler, 1991) befasst sich vor dem Hintergrund der Modelle menschlicher Informationsverarbeitung mit der kognitiven Belastung beim Lernen. Vor der Informationsverarbeitung müssen die relevanten Informationen zunächst selektiert werden, wozu Aufmerksamkeitsprozesse erforderlich sind. Der Informationsfluss zwischen den Hauptkomponenten des Gedächtnissystems wird bereits von Atkinson und Shiffrin (1968) beschrieben. Die beteiligten kognitiven Funktionen sind in ihrer Kapazität begrenzt und werden von einem Übermaß an Informationen überlastet – was mit negativen Effekten auf die Lernleistung einhergeht. Für Lernen und die Bewältigung von Problemstellungen ist es relevant, dass die Anforderungen an die Speichervorgänge im Arbeits- und Langzeitgedächtnis die vorhandenen Ressourcen nicht übersteigen.

Das Arbeitsgedächtnis (AG, Baddeley, 1992) ist als kognitives System zur Aufrechterhaltung und Verarbeitung von Informationen an allen bewussten mentalen Prozessen beteiligt (Hasselhorn & Schumann-Hengsteler, 2001) und somit auch bei schulischen Aktivitäten von Relevanz. Zur Beschreibung der Funktionsweise des AG wird im europäischen Raum häufig das Modell von Baddeley (Baddeley, 1992, 1996, 2000) herangezogen. Es handelt sich um ein Mehrkomponentenmodell in dem eine Leitzentrale (zentrale Exekutive) zwei modalitätsspezifische Sub-

systeme, für akustisch-verbale Informationen (phonologische Schleife) und für visuell-räumliche Informationen (visuell-räumlicher Notizblock), steuert (Baddeley, 2012).

Die zentrale Exekutive umfasst verschiedene Funktionen, die die Kontrolle, Steuerung und Regulierung kognitiver Prozesse betreffen (Miyake et al., 2000). Sie kontrolliert auch eine weitere Komponente des Modells: den episodischen Puffer (Baddeley, 2000). Die zentrale Exekutive fügt Informationen aus verschiedenen Quellen zu einer einheitlichen episodischen Repräsentation zusammen, die laut Theorie im episodischen Puffer in einem multimodalen Code repräsentiert sind. Viele Aufgaben sind so komplex, dass sie es erfordern, dass Inhalte im AG aktualisiert werden müssen, auch hier ist die zentrale Exekutive gefordert. Baddeley beschreibt mit dem Begriff „zentrale exekutive“ Prozesse, die auch unter dem Oberbegriff der exekutiven Funktionen geführt werden könnten, der aus der neuropsychologischen Forschung stammt (Rauch, 2022). Hier handelt es sich ebenfalls um einen Sammelbegriff für kognitive Prozesse, die der Kontrolle von Aufmerksamkeitsprozessen sowie der Kontrolle von Emotionen und Verhalten dienen (Rauch, 2022). Drei exekutive Funktionen, die auch faktoranalytisch bereits untersucht wurden, werden in der Literatur besonders hervorgehoben: Inhibition, Shifting und Updating (Miyake et al., 2000). Mit Inhibition ist das Hemmen einer automatischen oder dominanten Reaktion gemeint. Updating bezeichnet das gleichzeitige Aufrechterhalten und Manipulieren von AG-Inhalten. Und die Fähigkeit zum flexiblen Aufgabenwechsel – z.B. wenn sich die Anforderungen einer Aufgabe ändern – wird als Shifting bezeichnet (Miyake et al., 2000).

Das AG bildet zudem die Brücke zum Langzeitgedächtnis, welches eine unbegrenzte Kapazität hat und in dem das gesamte Wissen abgespeichert ist (Alloway & Alloway, 2015; Gruber, 2011). Auch mathematisches Faktenwissen, das bereits automatisiert vorliegt, ist hier abgelegt (z.B. $4 + 4 = 8$). Dieses Wissen kann nur mittels AG abgerufen werden. Allein der Wissensabruf erfordert daher kognitive Ressourcen, jedoch tendenziell weniger als die Konstruktion neuen Wissens (Büttner, 2003).

Für eine erfolgreiche Aufgabenbewältigung sollten möglichst viele kognitive Ressourcen zur Verfügung stehen. Werden jedoch bereits viele Ressourcen für die Auseinandersetzung mit der Aufgabenstellung aufgebracht, bleiben weniger Ressourcen für weitere Prozesse übrig. Das kann schließlich dazu führen, dass Aufgaben abgebrochen oder falsch bearbeitet werden, oder, dass sich die Bearbeitungszeit erhöht (Gathercole & Alloway, 2008). In der CLT wird dies als kognitive Überlastung (Sweller, van Merriënboer & Paas, 1998) bezeichnet. Die gesamte kognitive Belastung ergibt sich der Theorie zufolge aus zwei Belastungsarten: Der *intrinsischen* und der *extrinsischen kognitiven Belastung*.

Die intrinsische kognitive Belastung ist das Resultat aus der Komplexität des Lerninhalts (Sweller, van Merriënboer & Paas, 2019). Ihr Ausmaß bestimmt sich durch die Menge an Informationen, die im AG aufrechterhalten und in Beziehung gesetzt werden müssen. Die extrinsische Belastung ergibt sich durch die Art und Weise, wie die Informationen präsentiert werden, z.B. durch Anforderungen, die durch die Instruktion entstehen (Sweller et al., 2019). Anzustreben ist, dass möglichst viele Ressourcen auf die Verarbeitung von inhaltlichen Anforderungen verwertet werden. Wie die Befunde zur CLT zeigen, ist der überwiegende Teil der Effekte auf die Variation der extrinsischen Belastung zurückzuführen (Sweller et al., 2019).

Effekte des Instruktionsdesigns

In der Vergangenheit wurde eine Reihe von Effekten (z. B. Split-attention effect, Redundancy effect) beschrieben, die aus der Manipulation des Instruktionsdesigns hervorgehen (Überblick bei Sweller et al., 2019). Allerdings wird die kognitive Belastung während der Aufgabenbearbeitung selten direkt gemessen. Vielmehr werden Theorien wie die CLT herangezogen, um Lehrmethoden und Instruktionsdesigns zu entwickeln und zu prüfen. So wurden bereits verschiedene Designprinzipien für Lernmaterial entwickelt. Wenn diese die erwarteten Auswirkungen auf die Lernergebnisse haben, wird davon ausgegangen, dass sie die Theorie untermauern (Sweller et al., 2019, S. 265). Da es zu jedem Effekt eine Reihe von Experimenten gibt, soll an dieser Stelle nur beispielhaft skizziert werden, welche Anforderungen sich aus dem Instruktionsdesign ergeben können. Eine Reihe von Befunden existiert mittlerweile zum sogenannten split-attention effect (Ayres & Sweller, 2014; Tarmizi & Sweller, 1988). Häufig liegen bei Aufgaben Informationen im Material getrennt vor, z.B. wenn wie in Abbildung 1 ein Bild präsentiert wird, auf dem

eine Anzahl von Fingern zu sehen ist und aus 4 Lösungsalternativen das Bild ausgewählt werden soll auf dem genau eins weniger abgebildet ist. In dem Beispiel ist es erforderlich, beide Informationen aufeinander zu beziehen. Dabei muss das Ausgangsbild immer wieder fokussiert und mit den Lösungsbildern abgeglichen werden.

Es lässt sich nicht immer vermeiden, dass verschiedene Informationen integriert werden müssen, um eine Aufgabe zu lösen. Jedoch können weitere Anforderungen hinzukommen. Der Split-Attention-Effekt bezieht sich auch auf die zeitliche Organisation von Informationen (Sweller et al., 2019), z.B. wenn Instruktionen mündlich dargeboten werden und im Anschluss die Aufgabenbearbeitung folgt – so wie in viele standardisierte Tests. Schmidt, Stiller und Wilde (2019) haben die Effekte verschiedener Anleitungen während des Experimentierens in Biologie geprüft und untersucht, wie sich die Manipulation der Anleitung auf die wahrgenommene extrinsische kognitive Belastung auswirkt. Die geringste Belastung empfanden die Lernenden bei einer schrittweisen Anleitung.

Hecht (2014) hat für die Bereiche Rechtschreiben und Mathematik untersucht, wie sich das Instruktionsdesign auf die Übungsleistung auswirkt. Dabei hat die Autorin das konventionelle Lernmaterial umfassend adaptiert und Effekte zugunsten ihrer Adaption nachwiesen.

Ein zentrales Element vieler Aufgaben ist die schriftsprachlich dargebotene Instruktion. Durch sprachliche Anforderungen können für einige Personen Barrieren entstehen – auch dort, wo nicht direkt die Sprachkompetenz geprüft werden soll. Noll, Roth und Scholz (2020) haben untersucht, durch welche Maßnahmen der Zugang zu Arbeitsaufträgen erleichtert werden kann und welchen Effekt dies auf die Performanz bei mathematischen Aufgaben hat. Ihre Ergebnisse zeigen, dass die Kombination aus *Leichter Sprache* und zusätzlichen Fotos, auf denen die Instruktion dargeboten ist, einen Effekt auf das Testergebnis hat. Zudem zeigte sich, unabhängig von der Richtigkeit der Aufgaben, ein Effekt auf die Lösungsversuche. Der Vor-

Was ist eins weniger?

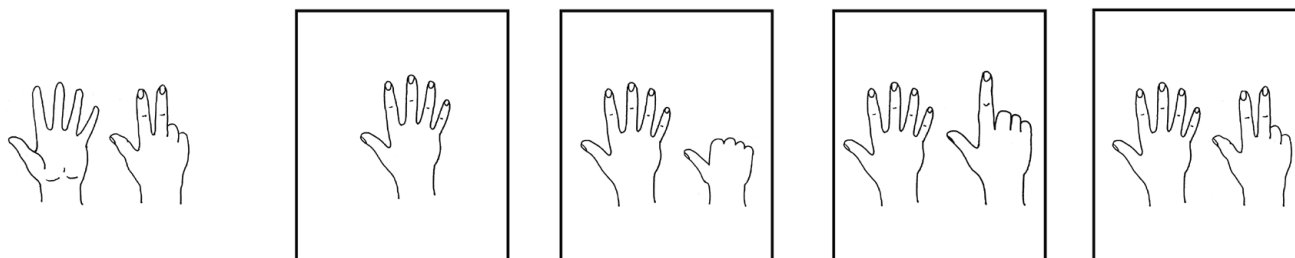


Abbildung 1. Aufgabe *Eins weniger* aus MBK 1 + Testheft Form A (Ennemoser, Krajewski & Sinner, 2017, S. 7).

teil von Fotos besteht vor allem darin, dass ein komplexer Inhalt in einem einzigen Bild dargeboten werden kann (Noll et al., 2020). In der Studie ist der Effekt der Unterstützungsmaßnahme bei starken und schwachen Leserinnen und Lesern vergleichbar.

Die Anpassung der Instruktion und des Aufgabendesigns scheint vielversprechend, um das Lernen und die Problemlösung zu optimieren. In Hinblick auf die Diagnostik in Mathematik sind diese Befunde von Noll et al. (2020) besonders interessant, wenn es um den Bereich der frühen mathematischen Kompetenzen geht, da die Lernenden hier tendenziell über weniger fundierte schriftsprachliche Kompetenzen verfügen. Zudem geht es in Tests aus diesem Bereich nicht darum, prozessbezogene Kompetenzen wie das Argumentieren oder Erklären von Beziehungen zu erfassen, bei denen die Sprachkompetenz unabdingbar ist.

Auf Basis der Ergebnisse aus dem Bereich der CLT wurden Prinzipien für die Gestaltung von Lernmaterialien herausgearbeitet. Im deutschsprachigen Raum haben Krajewski und Ennemoser (2010) und Hecht (2014) entsprechende Gestaltungsprinzipien formuliert. Hier werden u. a. die Abstimmung auf das Vorwissen und die kognitiven Ressourcen, den Aufbau und die Automatisierung inhaltspezifischen Basiswissens, die Ausrichtung der Darstellungen an die Struktur des Lerngegenstands und den Verzicht auf unnötige Formatwechsel genannt.

Bedeutung des Designs von diagnostischen Testaufgaben

Die bisher beschriebenen Effekte gelten ebenso für diagnostische Tests. Auch hier ist es für die erfolgreiche Aufgabenbewältigung relevant, Informationen zu fokussieren, aufrechtzuerhalten und zueinander in Beziehung zu setzen. Müssen hierfür bereits viele Ressourcen aufgebracht werden, kann es dazu kommen, dass für die eigentliche Aufgabe zu wenig Ressourcen übrigbleiben. In diesem Szenario ist die Leistung im Test nicht ausschließlich von der Beherrschung des Ausgabensinhalts abhängig. Damit misst der Test nicht nur das Merkmal, das er messen soll, sondern auch die kognitive Belastung durch das Aufgabendesign. Damit ist die Validität des Tests eingeschränkt (Grabowski, Castello & Brodersen, 2022). Wird eine bestimmte Gruppe, nämlich Personen mit geringen kognitiven Ressourcen, durch einen hohen extrinsischen Load von Testaufgaben in besonderem Maße benachteiligt, handelt es sich zudem um eine eingeschränkte Testfairness (Renner & Scholz, 2022).

In Hinblick auf Lernaufgaben vermuten Kuhl et al. (2021), dass es vor allem bei geringen kognitiven Fähigkeiten einen positiven Effekt von ressourcenschonend gestal-

teten Aufgaben gibt, da die Lernzeit effektiver genutzt würde. Der zeitliche Aspekt spielt auch in der Diagnostik eine Rolle (z. B. Speed-Tests), sodass ein ungünstiges Instruktionsdesign zu schlechteren Ergebnissen führen kann. Elliott, Kratochwill und McKeivitt (2001) untersuchten den Einsatz von Anpassungsmaßnahmen bei standardisierten Tests für Personen mit Behinderungen. Ihre Ergebnisse zeigen, dass zusätzliche Zeit für die Bearbeitung, Hilfe bei den Instruktionen und Unterstützung beim Lesen am häufigsten empfohlen wird. Diese Anpassungen haben eine moderate bis starke Auswirkung auf die Testergebnisse. In der Studie geht es allerdings um zusätzliche Testanpassungen für eine spezifische Zielgruppe, bei der bestimmte Barrieren direkt identifiziert werden können. Untersuchungen, die Effekte für eine breite Zielgruppe prüfen, lagen uns zum Zeitpunkt der vorliegenden Studie nicht vor.

Da bereits vielfach gezeigt wurde, dass Lernschwierigkeiten mit einem tendenziell schwächeren AG einhergehen (z. B. Kleszczewski et al., 2018), bedeutet das für die Erfassung des Lernstandes im unteren Leistungsbereich, dass Kompetenzen ggf. nicht valide erfasst werden. Zudem stellen sich Fragen in Hinblick auf die Messinvarianz (Schwab & Helm, 2015) von Verfahren, wobei es darum geht, inwieweit ein Verfahren bei verschiedenen Personengruppen ein und dasselbe Konstrukt erfasst.

Ziele und Forschungsfragen

Der skizzierte theoretische und empirische Hintergrund führt zu der Annahme, dass Instruktionsdesign und Aufgabengestaltung sich auch bei diagnostischen Tests auf die Bearbeitung der Lernenden auswirken könnte. Nun stellt sich die Frage, wie Testmaterialien gestaltet sein müssen, wenn sie die kognitiven Ressourcen berücksichtigen sollen. Mit dem Ziel, die Effekte des Aufgaben- und Instruktionsdesigns zu untersuchen, wurde für die vorliegende Studie ein Test zur Erfassung mathematischer Basiskompetenzen – MBK 1 + (Ennemoser et al., 2017) – im Sinne der Ressourcenorientierung adaptiert, wobei die folgenden Forschungsfragen im Fokus standen:

Inwiefern wirkt sich die Adaption auf die Performanz aus?

Auf Basis der CLT sollten sich im Durchschnitt bessere Testleistungen zeigen, wenn der ressourcenschonend adaptierten Test bearbeitet wird.

Welchen Einfluss hat der aktuelle Lernstand in Mathematik auf die Testleistung?

Es wird vermutet, dass ein besserer Lernstand in Mathematik – unabhängig von der Testversion – mit besseren Testleistungen einhergeht, da das Wissen in einem Inhaltsbereich der stärkste Prädiktor für die Leistung im gleichen Inhaltsbereich ist (Neubauer & Stern, 2008).

Welchen Einfluss hat die Funktionstüchtigkeit des AG auf die Testleistung?

Vermutlich geht eine bessere Funktionstüchtigkeit des AG – unabhängig von der Testversion – mit besseren Testleistungen einher, da es sich um eine wichtige individuelle Lernressource handelt, die sowohl während der Aufgabenbearbeitung als auch beim Aufbau von Vorwissen von Bedeutung ist. Schwierigkeiten im Lernen hängen tendenziell mit einem schwächeren AG zusammen.

Gibt es Wechselwirkungen zwischen der Testversion und dem aktuellen Lernstand in Mathematik sowie zwischen der Testversion und der Funktionstüchtigkeit des AG?

Bei schwächeren individuellen Lernvoraussetzungen (mathematisches Vorwissen und AG) sollte sich ein stärkerer Effekt einer ressourcenschonenden Adaption zeigen, da die kognitive Belastung für Lernende mit einer schwächeren Ausgangslage der Theorie nach höher ist als für Lernende mit einer soliden Wissensbasis und gutem AG. Es wird vermutet, dass die Gesamtbelastung für Letztere geringer ist, wodurch sie mehr Ressourcen für die Verarbeitung – etwa von unklaren Instruktionen – haben.

Methode

Stichprobe und Design

Die Forschungsfragen sollten durch ein Zwei-Gruppen-Design mit zwei experimentellen Bedingungen beantwortet werden. Als unabhängige Variable diente die Testversion (konventionell oder ressourcenschonend), wobei Aufgaben- und Instruktionsdesign der Langform des *Test mathematischer Basiskompetenzen ab Schuleintritt* (MBK 1 +, Ennemoser et al., 2017) manipuliert wurden. Die Rekrutierung erfolgte an insgesamt 15 Schulen in Nordrhein-Westfalen. Für die Untersuchung wurden eine Gesamtstichprobe von $N = 261$ Schülerinnen und Schülern (51% weiblich) der ersten und zweiten Klassenstufe rekrutiert.

Das Durchschnittsalter betrug zum Zeitpunkt der Testung 6 Jahre; 11 Monate ($SD = 1$ Monat). Es wurde stets nur ein Teil der Klasse in die Untersuchung einbezogen, wobei die Teilnehmenden zufällig ausgewählt wurden. Die Zuweisung zu den experimentellen Bedingungen erfolgte randomisiert.

Eine Gruppe bearbeitete die Langversion des MBK 1 + (Ennemoser et al., 2017) in der Standardform ($n = 125$), wohingegen die zweite Gruppe die ressourcenschonend adaptierte Version bearbeitete ($n = 133$). Um den Unterschied in der Performanz zu prüfen, wird als abhängige Variable die Leistung in Rohwertpunkten herangezogen. Da sich die mathematischen Anforderungen nicht zwischen den Testversionen unterscheiden, wird die aV also gleichzeitig mit der jeweiligen Form des MBK 1 + gemessen. Zusätzlich wurden die AG-Leistung und der aktuelle Lernstand in Mathematik erfasst.

Instrumente

MBK 1 + und MBK 1 + Adaption

Als Grundlage der Untersuchung dient der MBK 1 + (Ennemoser et al., 2017). Vor dem Hintergrund des Entwicklungsmodells der Zahl-Größen-Verknüpfung (Krajewski, 2003, 2013) dient der Test der Diagnostik früher mathematischer Kompetenzen. Dies umfasst Aufgaben zum einfachen und tiefen Zahlverständnis, wie etwa die Kenntnis von Zahlen, das Vornehmen von Zahlvergleichen (welche von zwei Zahlen ist größer?) oder das Verständnis für die Größe von Zahlen anhand von Zahlenstrahlaufgaben. Zudem gibt es Aufgaben, die Operationsvorstellungen ansprechen (z.B. Textaufgaben). Der MBK 1 + ist ab 6 Wochen nach Schuleintritt und über den Verlauf der ersten Klasse einsetzbar. Der Test ist an einer großen Stichprobe ($N = 6.086$) normiert und es liegen separate Normtabellen für die Gesamttestleistung und die Leistung auf den drei Kompetenzebenen vor. Die Retest-Reliabilität liegt zwischen $r = .67$ und $r = .77$.

Setze die fehlenden Zahlen ein!

3 ...	4	... 5
-------	---	-------

Schreibe die fehlende Zahl auf.



9		11
---	--	----

Abbildung 2. Gegenüberstellung der Aufgabe *Zahlenlücken* aus MBK 1 + Testheft Form A (Ennemoser et al., 2017, S. 3) (links) und adaptierte Aufgabe (rechts).

Stellenweise sind die sprachlichen Anweisungen lang wodurch – z. B. bei Messung zu Beginn der ersten Klasse – Barrieren entstehen könnten. Zudem gibt es verschiedene Aufgabendesigns und Darstellungswechsel, worin ein Ansatzpunkt für die vorliegende Studie besteht.

Auf Grundlage der CLT sowie den Erkenntnissen von Noll et al. (2020) wurde eine Adaption des MBK 1 + im Sinne der Ressourcenorientierung vorgenommen. Dazu wurden die Seiten sparsamer gestaltet, sodass pro Seite im Arbeitsheft weniger Reize dargeboten werden. Auf Piktogramme wurde, sofern sie für die eigentliche Aufgabebearbeitung irrelevant sind, verzichtet. Instruktionen wurden sprachlich vereinfacht und durch ein zusätzliches Foto unterstützt. Auf den Fotos ist beispielhaft die Aufgabebearbeitung zu erkennen.

Bei der Aufgabe *Zahlenlücken* in Abbildung 2 ist zu sehen, wie die Adaption der Aufgabe aussehen kann. Die drei Punkte vor und nach der Zahlenlücke (konventioneller Test) wurden ausgelassen, da diese mit unterschiedlichen Bedeutungen gefüllt werden können. Die Beispielaufgabe wurde in ein Foto übertragen, das gleichzeitig das Eintragen der Lösung zeigt.

Im gesamten Test wurden die Darstellungswechsel reduziert. Die verwendeten Darstellungen sind sparsam und strukturiert. Die Strukturierung ermöglicht ein einfaches Erfassen der Anzahlen, was gleichzeitig dazu dient, das Abzählen zu verhindern. Das Testheft ist unter <https://tudortmund.sciebo.de/s/DRanEFQ5IyimoEm> einsehbar.

Die Abfolge der Aufgaben ist in beiden Testversionen gleich. Ebenso wurde auf die Beibehaltung des mathematischen Inhalts geachtet.

Arbeitsgedächtnis (AG)

Als Kontrollvariablen wurden drei AG-Funktionen erfasst. Als Indikator für die phonologische Schleife wurden *Zahlen nachsprechen Aufgaben* (vorwärts und rückwärts, ähnlich z. B. im WISC-V, deutschsprachige Adaption von Petermann, 2017) genutzt. Dabei werden Ziffernfolgen verbal dargeboten, die anschließend von den Lernenden reproduziert werden sollen. Die Testung beginnt auf der ersten Ebene mit einer Ziffernfolge, die aus zwei Ziffern besteht (1–8). Wird die Ziffernfolge korrekt wiedergegeben, wird die nächste Ziffernfolge präsentiert, die aus drei Ziffern besteht. Der visuell-räumliche Notizblock wurde durch zwei Aufgaben erfasst. Die statische Komponente wurde durch eine selbst entwickelte *Matrixspannenaufgabe* (ähnlich bei Schuchardt, 2008) gemessen. Hierbei wird den Teilnehmenden eine 5-mal-5-Matrix, in der jeweils eine bestimmte Anzahl von Feldern schwarz ist, präsentiert. Nach einer standardisierten Präsentationszeit soll angezeigt werden, wo die schwarzen Felder waren. Als Indikator für den dynamischen Teil des visuell-räumlichen Notizblocks wurde eine *Corsiblockaufgabe* (ähnlich

bei Schuchardt, 2008), eingesetzt. Hierbei wurden auf einem Bildschirm neun unregelmäßig angeordneten Kästchen dargeboten. Diese färben sich nacheinander blau, wodurch ein Weg produziert wird, den die Lernenden sich merken müssen. Der Weg steigt von Aufgabe zu Aufgabe um einen Block (von zwei auf acht Blöcke). Die Abbruchkriterien waren bei allen drei Aufgaben gleich: bei einer falschen Antwort wurde ein zweiter Versuch unternommen, bei zwei falschen Reproduktionen wurde der Test abgebrochen und die Gedächtnisspanne auf die Anzahl an maximal korrekt reproduzierten Items festgelegt.

Lernstand Mathematik

Der aktuelle Lernstand in Mathematik wurde durch die Einschätzung der Fachlehrkraft erfasst. Dabei wurde auf die Skala Schulnoten (1 bis 6) zurückgegriffen, da es sich um eine vertraute Skala handelt.

Datenanalyse

Mittels *t*-Test wurde überprüft, ob sich die beiden Untersuchungsgruppen in ihren individuellen Lernvoraussetzungen, Lernstand Mathematik und AG-Funktionen, unterscheiden. Anschließend wurde für alle drei Ebenen des MBK 1 + sowie für den Gesamtwert geprüft, inwiefern sich die Gruppen voneinander unterscheiden. Da nicht alle Daten die Anwendungsvoraussetzungen erfüllen, wurde die Zahl der Freiheitsgrade z. T. korrigiert.

Zur Prüfung des Haupteffekts der Testbedingung und der Interaktionseffekten zwischen Testbedingung und den Kontrollvariablen wurden schrittweise lineare Regressionsanalysen gerechnet, bei denen der Testwert im (ressourcenschonenden) MBK 1 + als *aV* in die Analyse einging. Da es sich um eine Regressionsanalyse mit Interaktionsterm handelt, erfolgte eine Zentrierung der metrischen und der kategorialen Prädiktorvariablen (Richter, 2007). Die metrischen Variablen wurden z-standardisiert. Bei der kategorialen Variablen erfolgte eine einfache Kodierung (MBK 1 + Adaption = 1; MBK 1 + = 0). Es ist davon auszugehen, dass der Lernstand in Mathematik den stärksten Einfluss auf das Testergebnis hat, daher wurden im ersten Schritt die durch die Lehrkraft eingeschätzte Mathematikleistung und die Gruppenvariable als Prädiktoren in die Regression aufgenommen. Da eine starke empirische Evidenzlage für Zusammenhänge zwischen AG und mathematischen Kompetenzen vorliegt (Friso-van den Bos, van der Ven, Kroesbergen & van Luit, 2013), wurde im nächsten Schritt die Leistung des AG als Prädiktor in die Regression aufgenommen. Die AG-Funktionen wurden für die Analyse zu einem Gesamtwert zusammengefasst. Um zu klären, ob eine Interaktion des Lernstands Mathematik mit dem der Gruppenvariable besteht, wurde im nächsten

Schritt die Variable „Gruppe*Lernstand“ in das Modell aufgenommen. Schließlich wurde mit der Interaktion der AG-Funktionen und der Gruppenzugehörigkeit ebenso verfahren (Gruppe*AG).

Ergebnisse

Der *t*-Test zeigt, dass sich die Gruppen in den erfassten individuellen Lernvoraussetzungen nicht signifikant unterscheiden (Tabelle 1).

Die mittels MBK 1+ und MBK 1+ Adaption erfassten mathematischen Basiskompetenzen unterscheiden sich auf der rein deskriptiven Ebene zwischen den Gruppen (Tabelle 2). Durchschnittlich hat die Gruppe, die die adaptierte Testform bearbeitet hat (MBK 1+ Adaption) einen um etwa 5 Punkte besseren Gesamt-Rohwert. Betrachten wir die Ergebnisse auf Untertestebene, so zeigt die Gruppe MBK 1+ Adaption auf jeder Ebene einen höheren Testwert. Auf Ebene 1 ist dieser Unterschied mit einer Differenz von 0.41 Punkten marginal. Mit Ausnahme von Ebene 1 ($t(259) = -1.77, p = .08$) unterstützt die Signifikanzprüfung diese Gruppenunterschiede ($p < .00$).

Die Ergebnisse der schrittweisen linearen Regressionsanalyse (Tabelle 3) zeigen, dass bereits die ersten beiden Prädiktoren (Lernstand Mathematik und Gruppe) 29% der Varianz erklären. Nach den Konventionen von Cohen (1988) entspricht dies einer starken Varianzaufklärung.

Im zweiten Modell bildet sich der Einfluss der AG-Funktionen ab, so erklären die Prädiktoren im dritten Modell 35% der Varianz, wobei die Änderung in R^2 auf dem Niveau $< .01$ signifikant ist. Die zusätzliche Aufnahme der beiden Interaktionsterme in Modell 3 und 4 führt zu keiner signifikanten Änderung der Varianzaufklärung. Somit ist das zweite Modell das sparsamste Modell mit der höchsten Varianzaufklärung. Anhand der Betagewichte in Modell 2 zeigt sich, dass der Lernstand in Mathematik den größten Einfluss auf das Testergebnis hat ($\beta = 0.42$). Auch das AG hat einen Einfluss auf die Testleistung ($\beta = 0.25$). Schließlich bildet sich hier ebenso ein signifikanter – wenn auch geringerer – Einfluss der Gruppe ab ($\beta = 0.17$). Dabei handelt es sich um einen Einfluss zugunsten der mit 1 kodierten Gruppe MBK 1+ Adaption. Somit ist die Gruppenzuweisung ein signifikanter Prädiktor für die Testleistung.

Die in den Modellen 3 und 4 geprüften Interaktionen zwischen der Gruppe und dem Lernstand sowie zwischen der Gruppe und dem AG zeigen keine signifikanten Betagewichte.

Diskussion

In der vorliegenden Arbeit wurde untersucht, inwieweit sich Aufgaben- und Instruktionsdesign auf die Leistung in einem diagnostischen Test auswirken. Dazu wurde ein Test zur Erfassung mathematischer Basiskompetenzen

Tabelle 1. Prüfung der Voraussetzungen der beiden Untersuchungsgruppen (MBK 1+ $n = 125$; MBK 1+ Adaption $n = 133$)

Geprüfte Variablen	MBK 1+	MBK 1+ Adaption	t-Test	Signifikanz ^b	Effekt
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)			
Alter	7.02 (0.86)	6.90 (0.77)	$t(242) = 1.14$.62	0.15
Lernstand Mathematik	2.92 (1.31)	2.66 (1.30)	$t(180) = 1.35$.62	0.20
Corsi-Block	1.71 (1.55)	1.88 (1.38)	$t(256) = -0.92^a$.62	-0.14
Matrixspanne	1.02 (1.51)	0.73 (1.27)	$t(256) = 1.70^a$.45	0.21
Phonologische Schleife	4.37 (2.05)	4.71 (1.75)	$t(256) = -1.43$.62	-0.18

Anmerkung: ^a Wegen eines signifikanten Levene-Tests auf Varianzgleichheit wurde die Zahl der Freiheitsgrade korrigiert. ^b Aufgrund multipler Vergleiche wurde mittels Bonferroni-Holm korrigiert.

Tabelle 2. Deskriptive Statistiken und Mittelwertsunterschiede zwischen den Untersuchungsgruppen

Abhängige Variablen	MBK 1+	MBK 1+ Adaption	t-Test	Signifikanz ^b	Effekt
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)			
Gesamttestwert (MBK)	32.50 (10.32)	37.80 (8.58)	$t(259) = -4.52^a$	<.001	-0.56
Ebene 1	7.59 (2.09)	8.00 (1.60)	$t(259) = -1.77^a$.08	-0.22
Ebene 2	11.15 (3.52)	13.07 (2.99)	$t(259) = -4.76^a$	<.001	-0.59
Ebene 3	13.76 (6.36)	16.73 (5.75)	$t(259) = -3.96^a$	<.001	-0.49

Anmerkung: ^a Wegen eines signifikanten Levene-Tests auf Varianzgleichheit wurde die Zahl der Freiheitsgrade korrigiert. ^b Aufgrund multipler Vergleiche wurde mittels Bonferroni-Holm korrigiert.

Tabelle 3. Schrittweise lineare Regression mit den Variablen Lernstand Mathematik, Gruppe (MBK 1+; MBK 1 + Adaption) als Prädiktoren sowie den Testergebnissen als Kriterium

Modellzusammenfassung					
Modell	R^2	korrigiertes R^2	ΔR^2	F	df
1	0.30	0.29		38.4**	2, 179
2	0.35	0.35	0.05**	32.6**	3, 178
3	0.36	0.36	0.00**	24.7**	4, 177
4	0.36	0.36	0.00**	19.8**	5, 176
Koeffizienten					
Modell	Prädiktor	B	$SE(B)$	β	
1	Lernstand Mathematik	-0.59	0.07	-0.51**	
	Gruppe	0.39	0.15	0.17**	
2	Lernstand Mathematik	-0.49	0.08	-0.42**	
	Gruppe	0.40	0.14	0.17**	
	AG	0.28	0.07	0.25**	
3	Lernstand Mathematik	-0.57	0.10	-0.49**	
	Gruppe	0.40	0.14	0.17**	
	AG	0.30	0.07	0.26**	
	Gruppe*Lernstand	0.16	0.14	0.10	
4	Lernstand Mathematik	-0.58	0.10	-0.49**	
	Gruppe	0.39	0.14	0.13**	
	AG	0.25	0.09	0.22**	
	Gruppe*Lernstand	0.19	0.15	0.12	
	Gruppe*AG	0.09	0.15	0.06	

Anmerkung: ** $p < .01$.

(MBK 1+) auf Basis von Erkenntnissen aus dem Bereich der Cognitive Load Theory ressourcenschonend adaptiert. In der ersten Forschungsfrage stand der Effekt der Adaption im Fokus. Die Hypothese, dass der ressourcenschonend adaptierte Test zu besseren Testleistungen, hat sich hier bestätigt. Da der MBK 1 + die mathematischen Basiskompetenzen über drei Kompetenzebenen erfasst, wurden die Unterschiede zudem auf Untertestebene analysiert. Hier zeigen sich die Unterschiede nur auf zwei von drei Ebenen. So gibt es keinen Unterschied in der Testleistung auf Ebene 1. Für dieses Ergebnismuster gibt es mindestens zwei Erklärungen. Zum einen weisen die Aufgabenformate der Tests der Ebene 1 eine geringere extrinsische Belastung auf. Dadurch gibt es auf Ebene 1 möglicherweise keinen bedeutenden Effekt. Zum anderen ist es möglich, dass die Items der Ebene 1 hier generell sehr einfach waren. Dann werden tendenziell weniger AG-Ressourcen benötigt (Büttner, 2003) und auch ein weniger gutes Aufgabendesign hat keinen Einfluss auf die Testleistung. Auf Ebene 2 und 3 gibt es einen Gruppenunterschied zugunsten der Gruppe MBK 1 + Adaption, womit in dieser Arbeit gezeigt wurde, dass sich die Adaption auf die Performanz auswirkt.

Die Forschungsfragen zwei und drei beziehen sich auf den Einfluss der individuellen Lernvoraussetzungen – Vorwissen (bzw. Lernstand Mathematik) und AG – auf die Testleistungen. Die Hypothese, dass ein besserer Lernstand in Mathematik – unabhängig von der Testversion – mit besseren Testleistungen einhergeht, ist durch die Ergebnisse vorläufig bestätigt. Erwartungsgemäß zeigte sich in der Regression, dass der Lernstand in Mathematik den größten Einfluss auf das Testergebnis hat. Aber auch das AG eignet sich zur Vorhersage der Testleistung, womit die dritte Hypothese vorläufig bestätigt ist. Das Ergebnis passt in den aktuellen Forschungsstand, da Zusammenhänge zwischen verschiedenen mathematischen Kompetenzen und AG-Funktionen vielfach gezeigt wurden (Bull, Espy & Wiebe, 2008; Friso-van den Bos et al., 2013) und AG als ein inhaltsunspezifischer Vorläufer für mathematisches Lernen gehandelt wird (Passolunghi & Costa, 2019).

Die Zusammenhänge zwischen AG(-Belastung) und mathematischen Kompetenzen bzw. mathematischer Performanz führen zu der Frage nach den Wechselwirkungen zwischen dem Lernstand in Mathematik und der Testversion sowie zwischen der Testversion und dem AG. Angenommen wurde, dass Lernende mit schwächeren indivi-

duellen Lernvoraussetzungen (mathematisches Vorwissen und AG) stärker von der einer ressourcenschonenden Adaption profitieren, da die kognitive Belastung für sie höher sein sollten als für Lernende mit einer soliden Wissensbasis und gutem AG. Letzteren stehen mehr Ressourcen für die Verarbeitung – etwa von unklaren Instruktionen – zur Verfügung. Die Ergebnisse zeigen jedoch keine signifikanten Wechselwirkungen, sowohl für den Lernstand in Mathematik als auch für die AG-Funktionen. Somit bestätigte sich die vierte Hypothese nicht und die gefundenen Effekte der Adaption bestehen unabhängig von den kognitiven Lernvoraussetzungen. Dieses Ergebnis ist überraschend, unterstützt aber die Ergebnisse vorangegangener Untersuchungen (z. B. Hecht, 2014, Noll et al., 2020). Bei Noll et al. (2020) zeigen sich die Effekte der Unterstützungsmaßnahme unabhängig von der basalen Lesefertigkeit sowie des IQ. Die fehlende Interaktion zwischen Testversion und AG könnte zudem durch die eingesetzten AG-Tests erklärt werden. Die deskriptiven Statistiken deuten an, dass die visuell-räumlichen AG-Tests nicht bei allen Teilnehmenden gut funktioniert haben, so konnten einige von ihnen keine Reizfolge korrekt wiedergeben.

Möglicherweise hat die Adaption des MBK 1 + in der vorliegenden Untersuchung dazu geführt, dass alle Schülerinnen und Schüler weniger kognitive Ressourcen zur Aufgabenbewältigung benötigten, wodurch alle Teilnehmenden im Sinne eines Fahrstuhleffekts bessere Testergebnisse zeigten. Positiv ist, dass es sich in der vorliegenden Studie nicht um eine eingeschränkte Testfairness für bestimmte Gruppen von Lernenden handelt.

Einschränkend muss darauf hingewiesen werden, dass wir keine Erkenntnisse über das tatsächliche Ausmaß der kognitiven Belastung während der Aufgabenbearbeitung haben. Der Effekt der Adaption könnte auch dadurch zustande kommen, dass die Instruktionen besser und eindeutiger verstanden wurden – etwa durch die Fotos oder die Aufgabenformate – und dass auf diese Weise weniger Fehler gemacht wurden. Demzufolge können wir keine Aussagen zu den kognitiven Mechanismen treffen.

Die Ergebnisse der vorliegenden Studie sind aufschlussreich, müssen jedoch vor dem Hintergrund weiterer Limitationen betrachtet werden. In zukünftige Untersuchungen sollte den Lernstand in Mathematik durch einen weiteren standardisierten Test erfasst werden. Auch das AG sollte noch umfangreicher getestet werden. Der Einfluss der zentralen Exekutive auf die mathematische Leistung ist vielfach gezeigt (z. B. Imbo & Vandierendonck, 2007; Röhm, Starke & Ritterfeld, 2017) und es besteht die Vermutung, dass Wechselwirkungen zwischen zentraler Exekutive und Testversion bestehen.

Trotz der Limitationen liefern unsere Ergebnisse einen ersten Hinweis auf die Relevanz der ressourcenorientierten Gestaltung von diagnostischen Verfahren. Positiv ist, dass

die vorliegenden Ergebnisse nicht für eine eingeschränkte Testfairness für Schülerinnen und Schüler mit schwächeren AG-Funktionen sprechen. Allerdings muss die „Testfairness immer in Bezug auf bestimmte Personengruppen betrachtet werden“ (Renner & Scholz, 2022, S. 260) und es ist denkbar, dass die Instruktionen und Aufgaben des Tests besondere Anforderungen an Lernende stellen, die eingeschränkte sprachliche Kompetenzen haben.

Allerdings deuten unsere Ergebnisse auf eine eingeschränkte Konstruktrelevanz und damit auf eine eingeschränkte Validität (Renner & Scholz, 2022) hin. Denn: Die Testergebnisse scheinen bei allen Schülerinnen und Schülern der Stichprobe nicht ausschließlich von dem Zielkonstrukt mathematische Basiskompetenzen bestimmt zu sein. Dies kann Auswirkungen auf diagnostische Entscheidungen haben. Daher betreffen die praktischen Implikationen sowohl die schulische Diagnostik als auch die Testentwicklung. Die gestalterischen Merkmale eines Tests sollten einen möglichst geringen Einfluss auf die Testleistung haben, was bei der Testentwicklung berücksichtigt werden muss. Zudem sind Gütekriterien an einer breiten Stichprobe zu prüfen. Bestimmte Personengruppen (sonderpädagogische Unterstützungsbedarfe) sind hier unterrepräsentiert. Lehrkräfte und andere Praktikerinnen und Praktiker sollten hierfür sensibilisiert sein und Testergebnisse unter dem Gesichtspunkt der erforderlichen Zugangsfertigkeiten (z. B. Gedächtnis, Selbststeuerung, Überblick bei Renner & Scholz, 2022) einordnen.

Stellungnahmen und Erklärungen

Die Studie wurde in Übereinstimmung mit der lokalen Gesetzgebung und den institutionellen Anforderungen durchgeführt. Eine Teilnahme an der Studie erfolgte nur mit schriftlicher Zustimmung (informierte Einwilligung) der Erziehungsberechtigten. Die Testungen entsprachen der Deklaration von Helsinki.

Literatur

- Alloway, T. P. & Alloway, R. (2015). *Understanding working memory* (2nd edition). Los Angeles, London: SAGE.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Hrsg.), *The psychology of learning and motivation. Advances in research and theory* (Psychology of Learning and Motivation, Bd. 2, S. 89 – 195). New York: Academic P. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Ayres, P. & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (Cambridge handbooks in psychology, S. 135 – 146). New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.009>

- Baddeley, A. D. (1992). Working memory. *Science*, 255(5044), 556 – 559. <https://doi.org/10.1126/science.1736359>
- Baddeley, A. D. (1996). Exploring the Central Executive. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 5 – 28. <https://doi.org/10.1080/713755608>
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 11(4), 417 – 423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1 – 29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Bull, R., Espy, K. & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, 33(3), 205 – 228. <https://doi.org/10.1080/87565640801982312>
- Büttner, G. (2003). Gedächtnisentwicklung im Kindes- und Jugendalter. *Sprache Stimme Gehör*, 27(1), 24 – 30. <https://doi.org/10.1055/s-2003-37876>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Auflage). Hillsdale, NJ: Erlbaum.
- Elliott, S. N., Kratochwill, T. R., McKeivitt, B. C. (2001). Experimental Analysis of the Effects of Testing Accommodations on the Scores of Students with and without Disabilities, *Journal of School Psychology*, 39(1), 3 – 24. [https://doi.org/10.1016/S0022-4405\(00\)00056-X](https://doi.org/10.1016/S0022-4405(00)00056-X)
- Ennemoser, M., Krajewski, K. & Sinner, D. (2017). *MBK 1+. Test mathematischer Basiskompetenzen ab Schuleintritt* (1. Aufl.). Göttingen: Hogrefe.
- Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H. & van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, 10, 29 – 44. <https://doi.org/10.1016/j.edurev.2013.05.003>
- Gathercole, S. E. & Alloway, T. P. (2008). *Working memory and learning. A practical guide for teachers*. Los Angeles, London: SAGE.
- Grabowski, F., Castello, A. & Brodersen, G. (2022). Diagnostische Gütekriterien bei Statustests. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Version 1.0* (S. 247 – 258). Universität Regensburg.
- Gruber, T. (2011). *Gedächtnis* (1. Aufl.). Wiesbaden: Springer. <https://doi.org/10.1007/978-3-531-92096-2>
- Hasselhorn, M. & Schumann-Hengsteler, R. (2001). Arbeitsgedächtnis. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 17 – 22). Weinheim: Beltz.
- Hecht, A. T. (2014). *Ressourcenorientierte Lernförderung in der Grundschule: der Einfluss des Aufgabendesigns auf die Übungsleistungen von Zweitklässlern in Rechtschreiben und Mathematik*. Dissertation, Justus-Liebig-Universität Gießen. Verfügbar unter: <https://jlu.pub.uni-giessen.de/server/api/core/bitstreams/f7d33add-41ad-42f6-bc19-bdea8e805c45/content>
- Imbo, I. & Vandierendonck, A. (2007). The role of phonological and executive working memory resources in simple arithmetic strategies. *European Journal of Cognitive Psychology*, 19, 910 – 933. <https://doi.org/10.1080/09541440601051571>
- Kluszczewski, J., Brandenburg, J., Fischbach, A., Schuchardt, K., Grube, D., Hasselhorn, M. et al. (2018). Development of working memory from grade 3 to 5: Differences between children with and without mathematical learning difficulties. *International Journal of Disability, Development and Education*, 20, 1 – 17. <https://doi.org/10.1080/1034912X.2017.1419555>
- Krajewski, K. (2003). *Vorhersage von Rechenschwäche in der Grundschule*. Hamburg: Verlag Dr. Kovač.
- Krajewski, K. (2013). Wie bekommen die Zahlen einen Sinn? Ein entwicklungspsychologisches Modell der zunehmenden Verknüpfung zwischen Zahlen und Größen. In M. von Aster & J. H. Lorenz (Hrsg.), *Rechenstörungen bei Kindern. Neurowissenschaft, Psychologie, Pädagogik; mit 8 Tabellen* (2., überarb. und erw. Aufl., S. 155 – 179). Göttingen: Vandenhoeck & Ruprecht. <https://doi.org/10.13109/9783666462580.155>
- Krajewski, K. & Ennemoser, M. (2010). Die Berücksichtigung begrenzter Arbeitsgedächtnisressourcen in Unterricht und Lernförderung. In H.-P. Trollenier, W. Lenhard & P. Marx (Hrsg.), *Brennpunkte der Gedächtnisforschung. Entwicklungs- und pädagogisch-psychologische Perspektiven* (S. 337 – 365). Göttingen: Hogrefe.
- Kuhl, J., Hecht, T. & Vossen, A. (2021). Evidenzbasierte Förderung bei Lernschwierigkeiten. In J. Kuhl, A. Vossen, N. Hartung & C. Wittich (Hrsg.), *Evidenzbasierte Förderung bei Lernschwierigkeiten in der Grundschule* (S. 40 – 49). München: Ernst Reinhardt Verlag.
- Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43 – 52. https://doi.org/10.1207/S15326985EP3801_6
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49 – 100. <https://doi.org/10.1006/cogp.1999.0734>
- Neubauer, A. & Stern, E. (2008). *Lernen macht intelligent. Warum Begabung gefördert werden muss* (2. Aufl.). München: Dt. Verlags-Anstalt.
- Noll, A., Roth, J. & Scholz, M. (2020). Lesebarrieren im inklusiven Mathematikunterricht überwinden – visuelle und sprachliche Unterstützungsmaßnahmen im empirischen Vergleich. *Journal für Mathematik-Didaktik*, 41(1), 157 – 190. <https://doi.org/10.1007/s13138-020-00158-z>
- Paas, F., Renkl, A. & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, 32(1/2), 1 – 8. <https://doi.org/10.1023/B:TRUC.0000021806.17516.d0>
- Passolunghi, M. C. & Costa, H. M. (2019). Working Memory and Mathematical Learning. In A. Fritz, V. G. Haase & P. Räsänen (Hrsg.), *International Handbook of Mathematical Learning Difficulties. From the Laboratory to the Classroom* (vol. 29, S. 407 – 421). Cham: Springer. https://doi.org/10.1007/978-3-319-97148-3_25
- Petermann, F. (2017). *WISC-V. Wechsler Intelligence Scale for Children – Fifth Edition. Deutschsprachige Adaptation der WISC-V von David Wechsler*. Pearson Assessment.
- Rauch, W. A. (2022). Exekutive Funktionen. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 163 – 174). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149>
- Renner, G. & Scholz, M. (2022). Fair oder nicht fair, das ist hier die Frage! Die Sicherung der Testfairness als Aufgabe der sonderpädagogischen Diagnostik. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Version 1.0* (S. 259 – 274). Universität Regensburg.
- Richter, T. (2007). Wie analysiert man Interaktionen von metrischen und kategorialen Prädiktoren? Nicht mit Median-Splits! *Zeitschrift für Medienpsychologie*, 19, 116 – 125. <https://doi.org/10.1026/1617-6383.19.3.116>
- Röhm, A., Starke, A. & Ritterfeld, U. (2017). Die Rolle von Arbeitsgedächtnis und Sprachkompetenz für den Erwerb mathematischer Basiskompetenzen im Vorschulalter. *Psychologie in Erziehung und Unterricht*, 64(2), 1. <https://doi.org/10.2378/peu2016.art26d>
- Schmidt, S., Stiller, C. & Wilde, M. (2019). Hilfen beim Experimentieren. Auswirkungen unterschiedlicher Arten der Unterstützung auf den extraneous Cognitive Load. *Erkenntnisweg Biologie*, 18, 9 – 23.
- Schuchardt, K. (2008). *Arbeitsgedächtnis und Lernstörungen. Differenzielle Analysen der Funktionstüchtigkeit des Arbeitsgedächtnisses bei Kindern mit Lernstörungen*. Göttingen.

- Schwab, S. & Helm, C. (2015). Überprüfung von Messinvarianz mittels CFA und DIF-Analysen. *Empirische Sonderpädagogik*, 3, 175 – 193.
- Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, 81, 457 – 466. <https://doi.org/10.1037/0022-0663.81.4.457>
- Sweller, J. & Chandler, P. (1991). Evidence for Cognitive Load Theory. *Cognition and Instruction*, 8, 351 – 362. https://doi.org/10.1207/s1532690xci0804_5
- Sweller, J., van Merriënboer, J. J. G. & Paas, F. G. W. C. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, 31(2), 261 – 292. <https://doi.org/10.1007/s10648-019-09465-5>
- Sweller, J., van Merriënboer, J. J. G. & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251 – 296. <https://doi.org/10.1023/A:1022193728205>
- Tarmizi, R. & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80, 424 – 436. <https://doi.org/10.1037/0022-0663.80.4.424>

Historie

Manuskript eingereicht: 04.10.2023

Manuskript nach Revision angenommen: 29.04.2024

Onlineveröffentlichung: 23.05.2024

Förderung

Open-Access-Veröffentlichung ermöglicht durch die Technische Universität Dortmund.

ORCID

Sarah Schulze

 <https://orcid.org/0000-0001-9036-5254>

Jan Kuhl

 <https://orcid.org/0000-0002-5500-0281>

Dr. Sarah Schulze

Technische Universität Dortmund

Fakultät Rehabilitationswissenschaften

Fachgebiet Unterrichtsentwicklungsforschung

mit dem Schwerpunkt Inklusion

Otto-Hahn-Str. 6

44227 Dortmund

Deutschland

sarah.schulze@tu-dortmund.de