

SAMT

Validierung eines diagnostischen Instruments zur Erfassung der sprachlichen Ausdrucksfähigkeit in Mathematik für die dritte und vierte Klassenstufe

Alexandra Merkert^{1,2} und Gerlinde Lenke²

¹Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU), Deutschland

²Institut für Bildungswissenschaft, Fakultät Bildung, Leuphana Universität Lüneburg, Deutschland

Zusammenfassung: Sprachliche Fähigkeiten erweisen sich als relevant für das mathematische Lernen (siehe u. a. Bochnik & Ufer, 2016; Daroczy et al., 2015; Prediger et al., 2019). Die Diagnostik sprachlicher Fähigkeiten in Mathematik kann eine Grundlage für die individuelle und fachspezifische Förderung bilden (Häsel-Weide & Prediger, 2017). Das Diagnoseinstrument SAMT ermöglicht die Diagnostik der sprachlichen Ausdrucksfähigkeit von Primarstufenschülerinnen und -schülern im Fach Mathematik. Gegenstand des Beitrags ist die Validierung des Diagnoseinstruments und damit die Darstellung der Testgüte. Die Validierung war in das Forschungs- und Evaluationsprojekt Eva-Prim eingebettet, das durch das BMBF gefördert wurde. Die der Validierungsstudie zugrundeliegende Stichprobe umfasst insgesamt 384 Schülerinnen und Schüler der 3. und 4. Klassenstufe. Zusammenfassend liefern die Ergebnisse den Nachweis für eine zufriedenstellende Reliabilität, faktorielle Validität und kriteriale Validität sowie Änderungssensitivität. Zusätzlich konnte die Messinvarianz zwischen den Geschlechtern nachgewiesen werden (siehe dazu auch Merkert, 2022).

Schlüsselwörter: Diagnostik, diagnostische Instrumente, Sprachkompetenz, Sprachförderung, sprachsensibler Mathematikunterricht

SAMT. Validation of a Diagnostic Instrument for Assessing Language Proficiency in Mathematics in Third and Fourth Graders

Abstract: Language proficiency has proved to be relevant for mathematical learning (e.g., Bochnik & Ufer, 2016; Daroczy et al., 2015; Prediger et al., 2019). A diagnosis of language proficiency in mathematics can form the basis for promoting individual and subject-specific language measures (Häsel-Weide & Prediger, 2017). The diagnostic instrument SAMT enables the diagnosis of language proficiency in mathematics in primary-school children. This article concerns the validation of the diagnostic instrument and with that the representation of its psychometric quality. The validation was embedded in the research and evaluation project Eva-Prim, funded by the BMBF. The sample on which the validation study is based comprises 384 students in the 3rd and 4th grades. The results provide evidence for satisfactory reliability, factorial validity, and criterion validity as well as sensitivity to change. Further, we could substantiate measurement invariance between the sexes (see Merkert, 2022).

Keywords: diagnosis, diagnostic instruments, language proficiency, language promotion, language-sensitive mathematics education

Um Sprachförderung ein Fundament zu bieten, bedarf es einer durchgängigen systematischen Diagnostik der sprachlichen Fähigkeiten von Schülerinnen und Schülern. Was bereits vor Schuleintritt durch eine Diagnostik der sprachlichen Eingangsvoraussetzungen bundesweit stattfindet (siehe Lisker, 2013; Kultusministerkonferenz, 2009; Wildemann & Vach, 2018), kann im Primarbereich mit Blick auf die Bedeutung für das fachliche Lernen fortgesetzt werden. Die hohe Relevanz von Sprache für das Lernen und den Schulerfolg (OECD, 2007) konnten empirische Studien für den Mathematikunterricht mehrfach belegen (u. a. Abedi & Lord, 2010; Bochnik & Ufer, 2016; Dröse & Prediger, 2020; Gürsoy, Benholz, Renk, Prediger & Büchter, 2013; Paetsch, Radmann, Felbrich, Lehmann &

Stanat, 2016; Prediger, Wilhelm, Büchter, Gürsoy & Benholz, 2015). Geht es jedoch um die Diagnostik sprachlicher Fähigkeiten im Fach, die auf die Besonderheiten der Mathematik und ihre unterschiedlichen Darstellungsformen (Brown, 2005; Maier & Schweiger, 1999; Prediger & Wessel, 2011) zugeschnitten ist, besteht nach wie vor ein Bedarf an diagnostischen Verfahren (Bochnik & Ufer, 2017). Um ein diagnostisches Instrument für den Primarbereich bereitzustellen, das die Spanne sprachlicher Ausdrucksfähigkeit in Mathematik von Alltagssprachlichen bis hin zu differenzierten fachsprachlichen Äußerungen sowie auch fachspezifischen nonverbalen Darstellungsformen beachtet, wurde der SAMT-Test (SAMT = Sprachliche Ausdrucksfähigkeit in Mathematik) entwickelt (Details zur

Entwicklung des Instruments: siehe Merkert, 2022). Ziel dieser Studie ist die Überprüfung der psychometrischen Eigenschaften des Instruments im Hinblick auf Reliabilität, faktorielle Validität, Messinvarianz und kriteriale Validität.

Sprachliche Fähigkeiten im Mathematikunterricht

Sprachliche Fähigkeiten und fachliche Leistung

Schwierigkeiten beim Mathematiklernen stehen nicht selten mit sprachlichen Faktoren in Verbindung (Prediger, Erath & Moser Opitz, 2019). Insbesondere in einem sprachlich heterogenen Kontext können unterschiedliche Sprachstände der Schülerinnen und Schüler bzw. verschiedene Kompetenzniveaus in Bezug auf die Unterrichtssprache, z. B. hinsichtlich des Leseverstehens, zu Leistungsunterschieden in Mathematik führen (Greisen, Georges, Hornung, Sonnleitner & Schiltz, 2021). Sprachlich bedingte Hürden (siehe dazu u. a. Daroczy, Wolska, Meurers & Nuerk, 2015; Franke & Ruwisch, 2010; Weis, 2013) betreffen nicht nur Schülerinnen und Schüler mit Deutsch als Zweitsprache, sondern auch erstsprachig deutsch aufwachsende Lernende (Prediger, 2013; Prediger et al., 2019). Insgesamt sind Leistungsunterschiede in Mathematik zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund allerdings sowohl aufgrund sprachlicher Faktoren als auch des sozioökonomischen Status festzustellen, was für den Primar- (Bochnik & Ufer, 2016; Greisen et al., 2021; Heinze, Herwartz-Emden & Reiss, 2007; Kempert, Schalk & Saalbach, 2019; Paetsch et al., 2016, Stanat et al., 2021) und Sekundarbereich (Gebhardt, Rauch, Mang, Sälzer & Stanat, 2013; Gürsoy et al., 2013; Leiss, Plath & Schwippert, 2019; Paetsch et al., 2016; Prediger et al., 2015; Quasthoff, Heller, Prediger & Erath, 2022) gezeigt werden konnte. Beleuchtet man den Einfluss der sprachlichen Fähigkeiten von Schülerinnen und Schülern auf ihre Mathematikleistung genauer, so zeigt sich jedoch, dass dieser höher ausfällt als der des sozioökonomischen Status (Prediger et al., 2015; Wessel, 2015). Beachtlich erscheint außerdem der Befund, dass allgemeinsprachliche Fähigkeiten zur Erklärung von Leistungsunterschieden in Mathematik in Bezug auf Primarstufenschülerinnen und -schüler stärker beitragen als der Intelligenzquotient (Bochnik & Ufer, 2016; Heinze et al., 2007). Über die allgemeinsprachlichen Fähigkeiten hinaus kann auch der Fachwortschatz weitere Varianz hinsichtlich der Mathematikleistung aufklären, was dessen Relevanz für

schulisches Lernen unterstreicht (Bochnik & Ufer, 2016; Ufer & Bochnik, 2020). Die Grenzen zwischen Alltags- und Fachsprache sind dabei nicht immer eindeutig. Hinsichtlich der Lexik wird dies bspw. an dem Phänomen ersichtlich, dass manche scheinbar aus der Alltagssprache bekannte Wörter im fachlichen Kontext eine andere Bedeutung aufweisen (Steinbring, 2017), was potenziell Missverständnisse verursachen kann. Als besondere Herausforderung bezeichnet Brown (2005) außerdem die syntaktische Komplexität mathematischer Texte. Satzverkürzungen können im Mathematikunterricht bspw. im Kontext von Erklär- oder Aufgabentexten zu Irritationen führen (Weis, 2013). Für das Textverständnis erweisen sich darüber hinaus, auch in Bezug auf das Fach Mathematik, Präpositionen und Adverbien als zentral, da durch sie wesentliche Beziehungen ausgedrückt werden (Dröse & Prediger, 2020; Duarte, Gogolin & Kaiser, 2011; Gürsoy et al., 2013; Niederhaus, Pöhler & Prediger, 2016; Prediger, 2013; Weis, 2013). Zu beachten sind ebenfalls sogenannte domänenübergreifende bildungssprachliche Merkmale (siehe dazu u. a. Ahrenholz, 2013; Gogolin, Neumann & Roth, 2007; Meyer & Prediger, 2012; Ortner, 2009; Riebling, 2013). Bildungssprache vermittelt zwischen der Alltags- und der Fachsprache, womit ihr eine zentrale Bedeutung in Bildungs- bzw. Unterrichtsdiskursen zukommt (Ahrenholz, 2017; Ortner, 2009). In diesen entfaltet sie ihre kommunikative und epistemische Funktion (Feilke, 2012; Morrek & Heller, 2012; Schleppegrell, 2001). Bildungssprache stellt eine Art „Ausbauvarietät“ (Ortner, 2009, S. 2228) der Alltagssprache dar. Die Grenzen zwischen beiden können als fließend bezeichnet werden (Ahrenholz, 2010; Berendes, Dragon, Weinert, Heppt & Stanat, 2013). Bei der Zuordnung sprachlicher Äußerungen zu einem der Register ist daher auch deren Verwendungszusammenhang zu beachten (siehe dazu Fornol, 2020). Als typisch bildungssprachlich werden differenzierende und abstrakte Ausdrücke, Komposita, trennbare und reflexive Verben oder auch komplexe und unpersönliche Satzkonstruktionen benannt (Ahrenholz, 2013; Gogolin et al., 2007; Ortner, 2009). Der Wortschatz ist dabei ein wichtiger, jedoch nicht der alleinige Faktor. Tendenziell steht Bildungssprache konzeptioneller Schriftlichkeit nahe (Koch & Oesterreicher, 1985). Im englischsprachigen Raum wurden unter den Begriffen *academic language* (Bailey, 2007) oder *language of schooling* (Schleppegrell, 2004) ebenfalls entsprechende Merkmale herausgearbeitet und untersucht, die denen im Deutschen zum Teil entsprechen. Eine Gleichsetzung der Begrifflichkeiten sowie der dahinterliegenden Konzepte soll an dieser Stelle dennoch vermieden werden. Fachsprache ist, anders als Bildungssprache, domänenspezifisch (siehe u. a. Hopf & Eckhardt, 2013; Rincke, 2010). Fachsprachlichkeit betrifft nach Roelcke (2019) außerdem nicht nur die Lexik, sondern auch die gram-

matische, textliche, semiotische und pragmatische Ebene. In Bezug auf Möhn (1981) ist bei der Bestimmung ihrer Merkmale der mündliche und schriftliche Kontext stets mit zu denken. Fachsprache baut ebenfalls auf alltags- und bildungssprachlichen Ausdrücken, Strukturen und Praktiken auf und kann deshalb, insbesondere im schulischen Kontext, kaum in Reinform beobachtet werden. Bildungssprache, als eine Erweiterung der Alltagssprache, fungiert somit als Basis (Thürmann, 2013). Es ist in Anbetracht dessen nicht anzunehmen, dass es sich bei Alltags-, Bildungs- und Fachsprache um disjunkte Sprachbereiche handelt (siehe auch Niederhaus et al., 2016). Vielmehr kann von einem fließenden Verlauf ausgegangen werden, der die Spanne der sprachlichen Ausdrucksfähigkeit der Lernenden im fachlichen Kontext abbildet.

Zusammenfassend sind im Mathematikunterricht sowohl rezeptiv als auch produktiv umfassende sprachliche Fähigkeiten gefordert, die den alltagssprachlichen Bereich übersteigen und als bildungs- bzw. fachsprachlich bezeichnet werden können (Prediger, 2013; Prediger, Clarkson & Bose, 2016; Prediger & Wessel, 2011). Zum Tragen kommen diese insbesondere im Kontext diskursiver Praktiken bzw. bei der Realisierung verschiedener einfacher und komplexer Sprachhandlungen (Quasthoff et al., 2022; Vollmer & Thürmann, 2010; Vollmer, 2011). Dazu zählt bspw. das Erklären (Quasthoff et al., 2022) und Begründen eines Lösungsweges (Krummheuer, 2010) oder auch das Verallgemeinern von Mustern (Akinwunmi, 2012). Ein Spezifikum des Mathematikunterrichts besteht zudem darin, dass Lernende mit verschiedenen und zum Teil auch miteinander vernetzten sprachlichen Darstellungsebenen umgehen müssen (siehe Prediger, 2013; Prediger et al., 2016; Prediger & Wessel, 2011). Vor diesem Hintergrund unterscheiden Prediger und Wessel (2011) symbolische, verbale, graphische und gegenständliche Darstellungen, wobei sie nicht nur fachsprachliche, sondern auch alltags- und bildungssprachliche Realisierungen mit einbeziehen.

Ohne potenzielle sprachbedingte Hürden bzw. sprachliche Spezifika im Mathematikunterricht an dieser Stelle erschöpfend behandelt zu haben, wird deren Stellenwert für das Lernen und in der Folge auch die Leistung in Mathematik deutlich. Dabei geht es nicht nur um das Leseverstehen im Kontext von Textaufgaben (Prediger et al., 2015). In Bezug auf den Primarbereich erscheint auch der Einfluss der sprachlichen Komplexität bestimmter fachlicher Gegenstände bedeutsam, was bspw. am Einfluss des sprachlichen Aufbaus der Zahlworte auf numerische Kognitionen und arithmetische Leistungen erkennbar wird (siehe u. a. Daroczy et al., 2015; Mark & Dowker, 2015; Prior, Katz, Mahajna & Rubinsten, 2015). So konnten Zuber, Pixner, Moeller und Nuerk (2009) in einer Untersuchung an siebenjährigen Schülerinnen und Schülern aus

Österreich nachweisen, dass die Inversion von Zahlworten und Zifferschreibweise im Deutschen zu Problemen bei der Transkodierung führen kann, woraus wiederum Fehlerquellen beim Lösen von Mathematikaufgaben resultieren. Hervor tritt an diesem Beispiel noch einmal die untrennbare Beziehung zwischen den fachlichen Inhalten und ihrer sprachlichen Repräsentation, wobei sich letztere auf den Erwerb ersterer auswirkt.

Schließlich sind es jedoch die individuellen Voraussetzungen der Lernenden, die darüber entscheiden, ob und welche Besonderheiten der im Mathematikunterricht verwendeten Sprache (siehe dazu u. a. Brown, 2005) zu Lernschwierigkeiten führen (Dröse & Prediger, 2020; Leiss et al., 2019). Notwendig erscheint daher eine individuelle Förderung (fach-)sprachlicher Kompetenzen in Mathematik, um entsprechend der Zielsetzung eines sprachsensiblen Mathematikunterrichts den mathematischen Kompetenzerwerb sprachlich zu unterstützen (siehe Bochnik & Ufer, 2017). Dabei sollte individuelle Förderung möglichst diagnosegestützt erfolgen (Häsel-Weide & Prediger, 2017). Dies betrifft im Mathematikunterricht nicht nur die Diagnostik und Förderung bspw. numerischer bzw. rechnerischer Kompetenzen, sondern sollte in der Konsequenz auch auf die spezifisch fachbezogene sprachliche Ausdrucksfähigkeit als ebenfalls wesentlichem fachlichen Ziel (siehe Kultusministerkonferenz, 2004, 2022) übertragen werden, damit Schülerinnen und Schüler sprachliche Handlungsfähigkeit (siehe dazu Klieme et al., 2003) erlangen.

In Konzentration auf den produktiv schriftsprachlichen Bereich wird sprachliche Ausdrucksfähigkeit in speziellem Bezug auf das Fach Mathematik im Folgenden als die Fähigkeit definiert, „verschiedene einfache und komplexe Sprachhandlungen im fachlichen Kontext angemessen und funktional zu nutzen. Dazu zählen nicht nur verbalsprachliche Darstellungen, die alltags-, bildungs- und fachsprachlich realisiert werden können, sondern auch nonverbale Darstellungen auf symbolischer, graphischer und numerisch-tabellarischer Ebene“ (Merkert, 2022, S. 132). Das definierte Konstrukt orientiert sich somit am fach- und sprachintegrierten Modell der vernetzten Darstellungsregister nach Prediger und Wessel (2011) und lässt sich nach Jude und Klieme (2007) im Bereich Schreiben verorten. Auch wenn sprachliche Ausdrucksfähigkeit in Mathematik verschiedene Facetten verbaler (wie z. B. die Nutzung bestimmter Begriffe) und nonverbaler Art (bspw. Tabellen, Terme und Gleichungen) beinhaltet (diese werden im Folgenden als Teilbereiche bezeichnet), wird sie in der vorliegenden Studie dennoch als eindimensional definiert. Auf theoretischer Basis können sprachliche bzw. bildungssprachliche Merkmale z. B. in Anlehnung an Morek und Heller (2012) auch auf lexikalisch-semantischer, syntaktischer und diskursiver Ebe-

ne unterschieden werden. Diese Ebenen weisen jedoch teilweise Überlappungen auf. In Anbetracht ihrer sprachpraktischen Verschränkung wird von einer künstlichen Aufspaltung und damit Unterscheidung mehrerer Dimensionen abgesehen (siehe dazu Merkert, 2022).

Diagnostik fachbezogener sprachlicher Fähigkeiten

In der alltäglichen Praxis fokussiert sich die Diagnostik sprachlicher Voraussetzungen im Fach Mathematik meist auf einzelne Bereiche wie den (Fach-)Wortschatz oder das Verstehen von Textaufgaben (Prediger, 2017). Zudem unterliegen Fach- und Sprachunterricht im Schulalltag oftmals einer Trennung, wodurch Sprachdiagnostik und -förderung schnell hinter andere Anforderungen des Fachs zurücktreten (Riebling, 2013; Tajmel, 2017). Dennoch sollte die Bedeutung sprachlicher Fähigkeiten angesichts der fachlichen Ansprüche (Kultusministerkonferenz, 2004, 2022) besonders berücksichtigt werden. Zu einer systematischen Erfassung sprachlicher Fähigkeiten im Fach Mathematik benötigen Lehrkräfte allerdings ein entsprechendes diagnostisches Instrumentarium, das zur Planung und Evaluation von Fördermaßnahmen dienen kann. Für die Diagnostik sprachlicher und mathematischer Leistungen kann inzwischen auf ein breites Spektrum an Testinstrumenten geblickt werden (siehe dazu Merkert, 2022). Mit dem Verfahren Mathes 3 (Sikora & Hartke, 2020) liegt außerdem ein Instrument zur Lernverlaufdiagnostik im Fach Mathematik für die dritte Klasse vor, das sich an den Kompetenzbereichen der Bildungsstandards (Kultusministerkonferenz, 2004) orientiert. In Bezug auf die Erfassung sprachlicher Leistungen im Fach zeigt sich allerdings weiterhin Bedarf. Sprachdiagnostische Verfahren, wie bspw. das *Salzburger Lese-Screening 2-9* (Mayringer & Wimmer, 2014) oder der ökonomisch einsetzbare *C-Test* (Kniffka, Linnemann & Thesen, 2007), erfassen nur begrenzte Teilbereiche bzw. mehr allgemeinsprachliche Fähigkeiten, die besondere Spezifika des Fachs, wie die Realisierung bestimmter Sprachhandlungen mit fachlichem Bezug oder auch den Umgang mit nonverbalen Darstellungsformen, nicht abdecken. In Bezug auf den Sekundarbereich kann bspw. der *TeMaText* (Jordan & Stein, 2012) genannt werden, der zur Diagnostik des mathematischen Textverständnisses in der 9. und 10. Klasse dient. Weitere Instrumente zur Messung fachbezogener sprachlicher Fähigkeiten legen ihren Fokus oftmals nur auf Teilaspekte, wie bspw. die Lexik (Paetsch, Felbrich & Stanat, 2015; Riccomini, Smith, Hughes & Fries, 2015; Ufer & Bochnik, 2020). Insbesondere im Primarbereich mangelt es daher weiterhin an standardisierten Instrumenten, die eine Messung fachbezogener sprachlicher Fähigkeiten in Ma-

thematik unter Berücksichtigung verschiedener verbaler und nonverbaler Darstellungsformen ermöglichen. Bochnik und Ufer (2017) plädieren angesichts dieses Mangels für eine (Weiter-) Entwicklung der Instrumente zur Messung (fach-)sprachlicher Kompetenzen.

Zielsetzung

Um einen Beitrag zur Sprachdiagnostik im Mathematikunterricht der Primarstufe zu leisten, wurde der SAMT-Test entwickelt (siehe Merkert, 2022). Er richtet sich an Schülerinnen und Schüler der dritten und vierten Klassenstufe und soll dazu dienen, verschiedene Niveaus der sprachlichen Ausdrucksfähigkeit in Mathematik bei Lernenden abzubilden, um a) den Ist-Stand zu erfassen und b) Entwicklungsprozesse nachzuzeichnen.

Der vorliegende Beitrag beschäftigt sich mit der Frage nach den Gütekriterien des entwickelten Testinstruments. Im Zentrum stehen die psychometrischen Testeigenschaften im Hinblick auf Reliabilität, faktorielle Validität, Messinvarianz und kriteriale Validität.

Reliabilität

Um sowohl für gruppen- als auch individualdiagnostische Zwecke geeignet zu sein, sollte der SAMT-Test in Anlehnung an Lienert und Raatz (1998) ein Cronbachs α von .70 nicht unterschreiten.

Faktorielle Validität

Das zu messende Konstrukt, die sprachliche Ausdrucksfähigkeit in Mathematik, berücksichtigt zwar unterschiedliche Facetten, sollte sich jedoch aufgrund deren unmittelbaren Zusammenwirkens als eindimensional abbilden lassen.

Messinvarianz zwischen Gruppen und Messzeitpunkten

Geschlechtsspezifische Differenzen zeigen sich hinsichtlich sprachlicher Leistungen oftmals zugunsten der Mädchen (May, 2006; Mullis, Martin, Gonzalez & Kennedy, 2003) und hinsichtlich mathematischer Leistungen zugunsten der Jungen (Pietsch & Krauthausen, 2006; Walther, Geiser, Langeheine & Lobemeier, 2003). Dies kann bereits im Primarbereich festgestellt werden (Lühe, Becker, Neumann & Maaz, 2017). Um die SAMT-Ergebnisse

von Jungen und Mädchen adäquat vergleichen zu können, ist Messinvarianz zwischen den Geschlechtern nachzuweisen.

Darüber hinaus wird in Bezug auf die Zielsetzung, ein Instrument zu entwickeln, das sich zur Nachzeichnung von Entwicklungsprozessen eignet, zum einen der Frage nach longitudinaler Messinvarianz und zum anderen nach Änderungssensitivität nachgegangen.

Kriteriale Validität

In Bezug auf bisherige Befunde zum Zusammenhang sprachlicher und mathematischer Fähigkeiten kann angenommen werden, dass erstere den Erwerb letzterer positiv beeinflussen und zwar stärker als umgekehrt. Dies sollte insbesondere für die fachbezogene sprachliche Ausdrucksfähigkeit in Mathematik gelten.

Testentwicklung

Wie dargestellt, soll der SAMT-Test die sprachliche Ausdrucksfähigkeit von Schülerinnen und Schülern der dritten und vierten Klasse mit Fokus auf den Bereich Schreiben messen. Eine Besonderheit des Tests ist, dass er dabei sowohl verbale und diskursive Aspekte berücksichtigt als auch nonverbale Darstellungen, wie Zeichnungen, Tabellen, Terme und Gleichungen. Er erfasst somit die im Mathematikunterricht vernetzten Darstellungsregister (Prediger & Wessel, 2011) in umfassender Weise.

Die Entwicklung von SAMT gliederte sich in insgesamt drei aufeinander aufbauende Teilstudien (Übersicht siehe elektronisches Supplement ESM 1): 1. Konzeption und Pilotierung der Testaufgaben ($n = 82$), 2. Entwicklung der Skala ($n = 79$) und 3. Pilotierung der Skala ($n = 50$). Auf die Teilstudien folgte die Validierung, auf der der Fokus des vorliegenden Beitrags liegt. Die Entwicklung des Tests wird hier lediglich kurz zusammengefasst, eine ausführliche Darstellung findet sich bei Merkert (2022).

Konzeption und Pilotierung der Testaufgaben (erste Teilstudie)

Zur Konzeption des Tests wurden Mathematikaufgaben aus den Bereichen Arithmetik, Geometrie und Sachrechnen konzipiert und pilotiert. Verfolgt wurde das Ziel der Evokation spezifischer Sprachhandlungen unter Verwendung einer im mathematischen Kontext funktionalen und differenzierenden Lexik sowie Morphosyntax. Dabei geht es nicht nur um verbalsprachliche Äußerungen, sondern

auch um nonverbale Darstellungen. Die Pilotierung diente dazu, den sprachlichen Anregungsgehalt der Aufgaben, die Klarheit der Instruktionen und auch die Bearbeitungsdauer je Aufgabe zu überprüfen. Auf der Grundlage dieser Anforderungen wurden aus den ursprünglich sieben Aufgaben sechs ausgewählt, womit eine entfiel. Im Rahmen dieser sechs verbliebenen Testaufgaben (siehe ESM 2) müssen bspw. geometrische Formen benannt, Lagebeziehungen und Muster beschrieben, Rechenwege dargestellt, Lösungen begründet sowie Rechenoperationen erklärt bzw. Regeln verallgemeinert werden. Der gesamte Test einschließlich der dazugehörigen Materialien und Beispiele kann bei Merkert (2022) eingesehen werden.

Entwicklung der Skala (zweite Teilstudie)

Die auf der Basis der ersten Teilstudie ausgewählten Aufgaben wurden als Test konzipiert und eingesetzt. Die erhobenen Lösungen der Schülerinnen und Schüler wurden gemäß der strukturierenden skalierenden qualitativen Inhaltsanalyse nach Mayring (2015) ausgewertet, um verschiedene Niveaus der sprachlichen Ausdrucksfähigkeit in Mathematik theoriebasiert bestimmen und empirisch besetzen zu können. Dabei wurden die Teilbereiche in Anlehnung an klassische linguistische Kategorien (siehe u. a. Zifonun, Hoffmann & Stecker, 1997), sprach- und mathematikdidaktische Modelle und Unterscheidungen (u. a. Prediger & Wessel, 2011; Vollmer & Thürmann, 2010) sowie curriculare Standards (Kultusministerkonferenz, 2004) deduktiv abgeleitet und hinsichtlich ihrer Ausprägung abgestuft bzw. skaliert. Letzteres erfolgte in Rückbezug auf überwiegend sprach- und mathematikdidaktische Konzeptualisierungen (u. a. Ahrenholz, 2013; Akinwunmi, 2012; Blum und Leiß, 2005; de Villiers, 1990; Eckhardt, 2008; Feilke, 2012; Gogolin et al., 2004; Heppt, Dragon, Berendes, Stanat & Weinert, 2012; Jeuk, 2010; Koch & Oesterreicher, 1985; Ortner, 2009; Quehl & Scheffler, 2008; Prediger, 2013; Prediger & Wessel, 2011; Rasch & Sturm, 2018; Rösch, 2005; Schmölzer-Eibinger, 2013; Schnotz, Baadte, Müller & Rasch, 2011). Für jeden der sieben kodierten Teilbereiche (1. Nomen/Pronomen/Numerale, 2. Verben, 3. Adjektive, 4. Präpositionen/Adverbien, 5. Satzkonstruktionen, 6. nonverbale Darstellungen und 7. Sprachhandlungen) wurde auf der Grundlage von quantitativen Analysen sowie in Rückbezug zur Theorie ein Cut-Off-Kriterium zur Erlangung eines Niveaus festgesetzt (siehe dazu ausführlich Merkert, 2022).

Zur Gewährleistung der Durchführungsobjektivität wurden Instruktionen zum Ablauf des Tests erstellt, in denen neben den organisatorischen Bestimmungen sämtliche Er-

klärungen für die Testleitung wörtlich vorformuliert und somit standardisiert sind.

Zur Auswertung und Interpretation des Tests wurde ein detailliertes Kodiermanual mit Ankerbeispielen und Grenzfällen entwickelt, anhand dessen die Bewertung einer Äußerung erfolgt. Im Rahmen der Auswertung wird für jeden Teilbereich analysiert, ob genügend Indikatoren für ein bestimmtes Niveau entsprechend der Cut-Off-Kriterien vorliegen. Betrachtet wird dabei jeweils das gesamte verbal und nonverbal gebildete Material. Die Antworten zu den sechs Aufgaben werden damit nicht einzeln, sondern in ihrer Gesamtheit hinsichtlich der sieben Teilbereiche ausgewertet. Werden die Kriterien der fünften und damit höchsten Stufe nicht erreicht, werden die der darunterliegenden geprüft. Im Sinne der Unterscheidung zwischen Types und Tokens (Bußmann, 2008) gehen Doppelungen nicht in die Zählung ein. Wenn also bspw. zweimal derselbe Begriff genutzt wurde (wie z. B. „Dreieckspyramide“ oder „Einmaleins-Aufgabe“), dann wird dieser nur einmal gezählt. Berücksichtigt wird durch den Test nicht nur, ob bestimmte Begriffe bekannt sind bzw. verwendet wurden, sondern auch welche alternativen, wenn auch nicht fachsprachlichen, Ausdrücke zum Einsatz kommen. Orthografische Aspekte spielen nur insofern eine Rolle, als dass geschriebene Wörter phonologisch nachvollziehbar sein müssen, um eindeutig interpretiert werden zu können. Entsprechend der sieben Teilbereiche betreffen die Indikatoren zur Vergabe eines Teilniveaus nicht nur einzelne Begriffe, sondern bspw. im Teilbereich nonverbale Darstellungen auch das Aufstellen von Termen und Gleichungen oder das Anfertigen einer lösungsunterstützenden bzw. lösungsdarstellenden Tabelle oder Zeichnung. Die Indikatoren sind in dem zum Test dazugehörigen Kodiermanual ausführlich sowie auf dem Auswertungsbogen noch einmal in knapper Form beschrieben und mit Beispiel aufgeführt (siehe Merkert, 2022). Aus den Niveaus der sieben Teilbereiche kann im letzten Schritt der Auswertung ein Gesamtniveau der sprachlichen Ausdrucksfähigkeit in Mathematik bestimmt werden. Dieses reicht von 1 = *rudimentäre sprachliche Ausdrucksfähigkeit, d. h. die sprachliche Darstellung ist entweder nicht bzw. nicht eindeutig oder nur im konkreten Kontext zu verstehen* bis 5 = *entfaltete sprachliche Ausdrucksfähigkeit, d. h. die sprachliche Darstellung ist präzise und dem fachlichen Kontext angemessen; sie umfasst auch komplexe Sprachhandlungen sowie verbale und nonverbale Darstellungsformen, die funktional genutzt werden*.

Pilotierung der Skala (dritte Teilstudie)

Im Fokus der dritten Teilstudie standen die Überprüfung der Reliabilität (Cronbachs α) und der Auswertungsobjektivität auf Basis des entwickelten Kodiermanuals. Die Reliabilität lag bei $\alpha = .77$ und kann damit in Anlehnung an Lienert und Raatz (1998) als zufriedenstellend bezeichnet werden. Die Interraterreliabilität (ermittelt durch eine Doppelkodierung der gesamten Daten durch zwei unabhängig arbeitende Raterinnen bzw. Rater) betrug $r_s = .93$, was auf eine hohe Auswertungsobjektivität schließen lässt.

Testvalidierung

Validiert wurde der SAMT-Test im Rahmen des Projekts Eva-Prim: Evaluation im Primarbereich – Sprachförderung in alltäglichen und fachlichen Kontexten¹, das Teil der bundesweiten BiSS-Initiative war (Henschel, Gentrup & Stanat, 2018). Der Fokus des Projekts lag auf dem Mathematikunterricht der Grundschule.

Erhebungsdesign und Erhebungsinstrumente

Der SAMT-Test sowie der DEMAT (= Deutscher Mathematiktest) in den Versionen 2+ (Krajewski, Liehm & Schneider, 2004) und 3+ (Roick, Gölitz & Hasselhorn, 2004) wurden zu jeweils zwei der insgesamt drei Messzeitpunkte der in das Projekt Eva-Prim (= Evaluation im Primarbereich – Sprachförderung in alltäglichen und fachlichen Kontexten) eingebetteten Validierungsstudie (Ende Klasse 2, Mitte Klasse 3 und Mitte Klasse 4) eingesetzt. Im Rahmen des Projekts wurden außerdem weitere Leistungen wie die Lesefertigkeit erhoben (siehe dazu Rank et al., 2021). Die Erhebungen mittels DEMAT 2+ und 3+ fanden dabei zum ersten und dritten Messzeitpunkt und anhand von SAMT zum zweiten und dritten Messzeitpunkt statt (siehe ESM 3).

Der SAMT-Test ist als Paper-Pencil-Verfahren konzipiert, das gemäß der standardisierten Instruktionen einzeln oder als Gruppentestung durchführbar ist. Für die Durchführung des Tests mit einer ganzen Schulklasse einschließlich aller organisatorischen Aspekte wurden im Rahmen der Erhebungen zwei Unterrichtsstunden von jeweils 45 Minuten eingeplant. Nach der Hälfte der Test-

¹ Geleitet wurde das vom BMBF (= Bundesministerium für Bildung und Forschung) geförderte Projekt von Prof. Dr. Astrid Rank, Prof. Dr. Anita Schilcher, Prof. Dr. Stefan Krauss sowie Prof. Dr. Anja Wildemann und Prof. Dr. Gerlinde Lenske.

aufgaben, was auch nur knapp mehr als der Hälfte der Testzeit entspricht, fand stets eine Pause von fünf Minuten statt. Die reine Bearbeitungsdauer der Aufgaben beträgt 55 Minuten. An den Erhebungstagen von SAMT wurden keine weiteren Tests durchgeführt, um möglichen Müdigkeitserscheinungen aufgrund der Testlänge vorzubeugen.

Die demografischen Daten der Schülerinnen und Schüler, wie der Migrationshintergrund (Kind selbst oder beide Elternteile nicht in Deutschland geboren) und der Erwerbskontext des Deutschen (Deutsch als Erst- oder Zweitsprache), wurden mittels eines Elternfragebogens erfasst.

Stichprobe

Der Validierung liegt eine Stichprobe von Schülerinnen und Schülern aus drei deutschen Bundesländern zugrunde. Insgesamt nahmen 384 Kinder (50 % weiblich und 50 % männlich) zu beiden Erhebungen am SAMT-Test teil. Die Kinder verteilten sich auf insgesamt 14 Grundschulen und 31 Klassen. Der Anteil der Kinder, die erst nach ihrem dritten Lebensjahr Kontakt zur deutschen Sprache hatten, lag bei 27 %. Der Anteil der Kinder, die entweder selbst nicht in Deutschland geboren wurden oder beide Elternteile, betrug 33 %. Von diesen liegen für 358 Schülerinnen und Schüler die Daten zu allen zu betrachtenden DEMAT- und SAMT-Erhebungen vor. Die entsprechende Substichprobe ist hinsichtlich ihrer Zusammensetzung mit der Gesamtstichprobe vergleichbar und wird zu den Ergebnisberichten angegeben. Die Analysen, die sich ausschließlich auf die SAMT-Ergebnisse beziehen, schließen alle Schülerinnen und Schüler ein, die zu beiden Messzeitpunkten teilgenommen haben.

Analysemethoden

Die interne Konsistenz der Skala als Maß ihrer Reliabilität wurde anhand von Cronbachs α bestimmt. Zur Überprüfung der faktoriellen Validität wurde zu beiden Messzeitpunkten, zu denen SAMT eingesetzt wurde, eine konfirmatorische Faktorenanalyse durchgeführt. Betrachtet wurde außerdem, ob der SAMT-Test über die Geschlechter hinweg in gleicher Weise misst, wozu eine Untersuchung der Messinvarianz zwischen Jungen und Mädchen vorgenommen wurde. Zur Prüfung der Änderungssensitivität sowie der Vergleichbarkeit von Messwerten zu unterschiedlichen Messzeitpunkten wurde eine longitudinale Messinvarianzanalyse durchgeführt und zur Hypothesenprüfung ein Cross-Lagged-Panel-Modell gewählt. Durchgeführt wurden die Analysen mit der Statistik-Software R (packages lavaan, tidySEM und magrittr).

Die Kennwerte der Faktorenanalysen zu beiden Messzeitpunkten verdeutlichen in Anlehnung an Hu und Bentler (1999) eine gute Modellpassung (erster Messzeitpunkt: $N = 384$; $\chi^2(14) = 21.66$, $p = .086$; CFI = .98; TLI = .97; RMSEA = .04; SRMR = .04; zweiter Messzeitpunkt: $N = 384$; $\chi^2(14) = 18.54$, $p = .183$; CFI = .99; TLI = .98; RMSEA = .03; SRMR = .03).

Ergebnisse

Reliabilität (interne Konsistenz)

Für den ersten Messzeitpunkt mit SAMT ergibt sich ein Wert von $\alpha = .71$ und für den zweiten von $\alpha = .70$ ($N = 384$). Die Werte liegen damit knapp über dem von Lienert und Raatz (1998) für individualdiagnostische Zwecke empfohlenen Schwellenwert von .70.

Faktorielle Validität

Die Kennwerte der Faktorenanalysen zu beiden Messzeitpunkten verdeutlichen in Anlehnung an Hu und Bentler (1999) eine gute Modellpassung (erster Messzeitpunkt: $N = 384$; $\chi^2(14) = 21.66$, $p = .086$; CFI = .98; TLI = .97; RMSEA = .04; SRMR = .04; zweiter Messzeitpunkt: $N = 384$; $\chi^2(14) = 18.54$, $p = .183$; CFI = .99; TLI = .98; RMSEA = .03; SRMR = .03).

Messinvarianz zwischen Jungen und Mädchen

Geprüft wurde die Messinvarianz zwischen den Geschlechtern anhand der Daten des ersten Messzeitpunkts der Validierungsstudie, zu dem der SAMT-Test zum Einsatz kam. Bei der Analyse wurde die jeweils nächste restriktivere Stufe nur getestet, wenn zusätzlich zu einer in Orientierung an Hu und Bentler (1999) guten Modellpassung keine im Vergleich zum Vorgängermodell bedeutsame Verschlechterung vorlag. In Anlehnung an Chen (2007) wurde bei der Interpretation außerdem der CFI, RMSEA und SRMR betrachtet. Da RMSEA und SRMR leicht durch Stichprobengröße und Modellkomplexität beeinflusst werden, gilt der CFI als Hauptkriterium (Chen, 2007; Cheung & Rensvold, 2002)². Entsprechend der Richtwerte wurde die höchste Stufe (Strict) erreicht. Die

² Nach Chen (2007) indiziert ein Absinken des CFI um $\geq .01$ in Kombination mit einer Steigerung des RMSEA um $\geq .02$, dass nicht mehr von Invarianz in Bezug auf Faktorladungen, Intercepts und Residualvarianzen auszugehen ist. Berücksichtigt werden sollte auch der SRMR, der bei größeren Stichproben ($N > 300$) zusätzlich zu einem Absinken des CFI um $\geq .01$ bei faktorieller Invarianz (Weak) nicht um $\geq .03$ und bei Invarianz

Tabelle 1. Messinvarianzprüfung zwischen den Geschlechtern, erste SAMT-Erhebung ($N = 384$)

	χ^2 p -value	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$ p -value	Δ CFI
Configural	.574	1.000	1.008	< .001	.033	/	/
Weak	.420	.997	.997	.012	.049	.172	-.003
Strong	.379	.994	.994	.017	.053	.310	-.003
Strict	.520	1.000	1.003	< .001	.059	.811	.006

Anmerkung: Die Werte wurden auf die dritte Nachkommastelle gerundet.

Tabelle 2. Longitudinale Messinvarianzmodelle zur Überprüfung der Änderungssensitivität ($N = 384$)

	χ^2 p -value	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$ p -value	Δ CFI
Configural	.064	.984	.975	.034	.029	/	/
Weak	.085	.984	.980	.030	.035	.466	< .001
Strong	< .001	.902	.897	.069	.056	< .001	.082
Partial Strong	.088	.984	.982	.029	.037	.351	-.001
Partial Strong (Gleichsetzung der Varianzen)	.007	.964	.966	.039	.055	.007	-.019
Partial Strong (Gleichsetzung der latenten Intercepts)	< .001	.883	.871	.077	.096	< .001	-.101

Anmerkungen: Die Werte wurden auf die dritte Nachkommastelle gerundet; abgelehnte Modelle sind grau hinterlegt.

Messinvarianzprüfung (siehe Tabelle 1) indiziert somit gleiche Strukturen, Faktorladungen, Intercepts und Residualvarianzen zwischen den Geschlechtern (Chen, 2008; Cheung & Rensvold, 2002).

Longitudinale Messinvarianz und Änderungssensitivität

Die longitudinale Messinvarianzprüfung ($N = 384$) erreichte in Anlehnung an Chen (2007) sowie Cheung und Rensvold (2002) die Stufe Partial Strong (siehe Tabelle 2). Da zwei der sieben Intercepts einen uniformen Bias aufwiesen, wurden anstelle der manifesten die latenten (messfehlerbereinigten) Mittelwerte verglichen, um eine Über- bzw. Unterschätzung der Effektstärke zu vermeiden.

Durch das Gleichsetzen der latenten Mittelwerte im Partial-Strong-Modell, wurde der p -Wert des χ^2 -Tests signifikant³. Die Annahme einer Gleichheit der Mittelwerte muss daher verworfen werden. Dies deutet auf einen si-

gnifikanten Anstieg der Mittelwerte vom ersten bis zum zweiten Messzeitpunkt hin. Vor der Ermittlung der Effektstärke wurden ebenfalls im Partial-Strong-Modell die Varianzen gleichgesetzt. Dies führte erneut dazu, dass der p -Wert des χ^2 -Tests die gesetzte Signifikanzschwelle unterschritt. Bei der Berechnung der Effektstärke musste daher die Ungleichheit der Varianzen berücksichtigt werden. Es zeigte sich ein Cohens d von 0.83 und damit ein großer Effekt (Cohen, 1988).

Kriteriale Validität

Zur Prüfung der aufgestellten Hypothese wurde im Rahmen eines Cross-Lagged-Panel-Modells untersucht, ob die sprachliche Ausdrucksfähigkeit in Mathematik die mathematische Kompetenz in einem späteren Schuljahr stärker beeinflusst als umgekehrt (siehe Abbildung 1).

Erwartungsgemäß sind SAMT⁴ ($\beta = .72, p < .001$) und DEMAT ($\beta = .54, p < .001$) stabil. Betrachtet man die

hinsichtlich Intercepts (Strong) und Residualvarianzen (Strict) um $\geq .01$ ansteigen sollte (Chen, 2007). Cheung und Rensvold (2002) legen nahe, dass auch bei einer CFI-Differenz $\leq .01$ von Messinvarianz ausgegangen werden kann.

³ Gewählt wurde ein α -Risiko von $\alpha = .05$. Im Sinne einer möglichst transparenten Ergebnisdarstellung werden im Folgenden die genauen p -Werte berichtet.

⁴ Im Modell berücksichtigt wurde die Autokorrelation der Residuen in Bezug auf dieselben Indikatorvariablen des latenten Konstrukts zu zwei Messzeitpunkten.

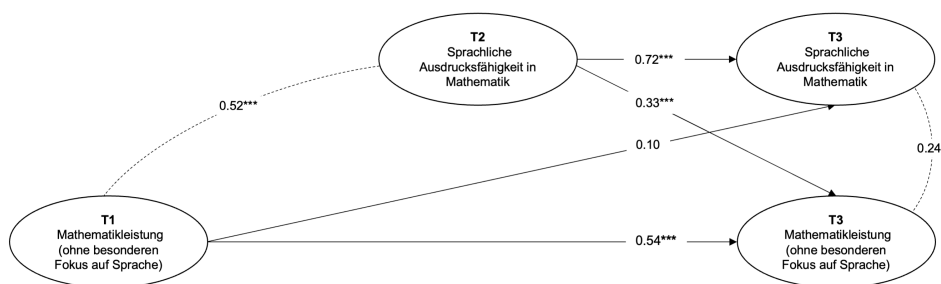


Abbildung 1. Cross-Lagged-Panel-Model zum Einfluss der sprachlichen Ausdrucksfähigkeit in Mathematik auf die Mathematikleistung (ohne besonderen Fokus auf Sprache).

Cross-Lagged-Pfade, fällt auf, dass die SAMT-Ergebnisse zum zweiten Messzeitpunkt einen signifikanten Einfluss auf die DEMAT 3+-Ergebnisse zum dritten Messzeitpunkt haben ($\beta = .33, p < .001$). Der Pfad von den DEMAT 2+-Ergebnissen zum ersten Messzeitpunkt zu den SAMT-Ergebnissen zum dritten Messzeitpunkt kann jedoch nicht als signifikant interpretiert werden ($\beta = .10, p = .133$). Auch nach Aufnahme des sozioökonomischen Status (gemessen über die Bildungsabschlüsse der Eltern) ins Modell bleibt ein signifikanter Einfluss der SAMT-Ergebnisse auf die des DEMAT 3+ bestehen ($\beta = .23, p = .009$), wobei sich allerdings die Stichprobe auf $n = 292$ verringert.

Diskussion

Ziel der vorliegenden Studie war die Prüfung der Gütekriterien des entwickelten Testinstruments im Hinblick auf Reliabilität, faktorielle Validität, Messinvarianz und kriteriale Validität. Diesbezüglich lässt sich abschließend auf Basis der Befunde festhalten, dass alle erfassten psychometrischen Kennwerte des SAMT-Tests als zufriedenstellend bezeichnet werden können.

Cronbachs α lässt auf eine hinreichende Reliabilität des Verfahrens schließen, sodass der SAMT-Test auch für die Individualdiagnostik verwendet werden kann (Lienert & Raatz, 1998). Die konfirmatorischen Faktorenanalysen geben jeweils Hinweise auf die faktorielle Validität zu beiden Messzeitpunkten. Die strikte Messinvarianz zwischen den Geschlechtern ermöglicht den direkten Vergleich der manifesten Mittelwerte zwischen diesen Gruppen (Chen, 2008; Cheung & Rensvold, 2002), was für die Interpretation der Testwerte bedeutsam ist. In Bezug auf die longitudinale Messinvarianz wurde die Stufe Partial Strong erreicht, was einen Vergleich der latenten statt der manifesten Mittelwerte über mehrere Messzeitpunkte hinweg indiziert. Hinsichtlich der Änderungssensitivität ist ein signifikanter Anstieg der latenten Mittelwerte zwischen den beiden Messzeitpunkten festzustellen. Nach Cohen (1988) kann von einem großen Effekt gesprochen werden, was auf die Eignung der Skala zu Evalua-

tionszwecken auch im Rahmen längsschnittlicher Designs verweist. Die Änderungssensitivität deutet gleichzeitig auf die Validität der Skala hin, da sie in der Lage ist, erwartete Leistungsunterschiede entsprechend abzubilden. Die formulierte Hypothese zur kriterialen Validitätsprüfung konnte durch das Cross-Lagged-Panel-Model bestätigt werden: Es zeigte sich nicht nur, dass die sprachliche Ausdrucksfähigkeit von Primarstufenschülerinnen und -schülern in Mathematik einen größeren Einfluss auf die Mathematikleistung (ohne besonderen Fokus auf Sprache) im darauffolgenden Schuljahr besitzt als umgekehrt, sondern auch, dass die mathematische Kompetenz auf die spätere sprachliche Ausdrucksfähigkeit in Mathematik im Gegensatz zum umgekehrten Fall keinen signifikanten Einfluss hat. Dies spricht für die Erfassung einer differenzierbaren sprachlichen Leistung sowie deren Bedeutung für den mathematischen Kompetenzerwerb. Resümierend lassen sich, unter Berücksichtigung der Objektivität, die bei der Pilotierung überprüft wurde, alle untersuchten Gütekriterien als hinreichend bis sehr gut bezeichnen, was für den Einsatz von SAMT in Forschung und Praxis – sowohl zur Erfassung des Leistungsniveaus als auch zur Skizzierung der Leistungsentwicklung spricht.

Trotz der zufriedenstellenden psychometrischen Kennwerte und der Maßnahmen zur Erhöhung der Testökonomie sollen abschließend mögliche Limitationen des SAMT-Tests sowie der Validierungsstudie betrachtet werden. Kritisch zu diskutieren ist, dass die erste Erhebung der Mathematikleistung (Ende Klasse 2) und der sprachlichen Ausdrucksfähigkeit in Mathematik (Mitte Klasse 3) nicht zeitgleich stattfanden. Zwischen der ersten Messung der Mathematikleistung und der zweiten Messung der sprachlichen Ausdrucksfähigkeit in Mathematik ist damit mehr Zeit vergangen als zwischen der ersten Messung der sprachlichen Ausdrucksfähigkeit in Mathematik und der zweiten Messung der Mathematikleistung. Längere Zeiträume zwischen den Messungen zweier Variablen bergen dabei potenziell die Gefahr einer Unterschätzung des Einflusses. Demnach sollte in Bezug auf die kriteriale Validität eine Replikation der Befunde erfolgen.

Limitierend ist darüber hinaus die Testökonomie in Bezug auf die Auswertung sowie auch die Durchführungs-

dauer zu nennen. Insbesondere wenn SAMT lediglich als Kontrollvariable fungieren soll, wäre eine höhere Testökonomie wünschenswert.

Erprobt werden soll zukünftig der Einsatz und Nutzen des SAMT-Tests zur fachbezogenen Sprachdiagnostik an Grundschulen. Zur weiteren Verbesserung der Handhabbarkeit erscheint eine Vereinfachung der Auswertung bzw. die Konzeption einer (ggf. computergestützten) Kurzversion sinnvoll. In Bezug auf die Unterstützung der fachbezogenen Sprachförderung im Primarbereich ist außerdem die Entwicklung von auf den SAMT-Test abgestimmten Sprachfördermaterialien von Mehrwert.

Elektronische Supplemente (ESM)

Die elektronischen Supplemente sind zusammen mit der Online-Version des Artikels verfügbar unter <https://doi.org/10.1026/0012-1924/a000315>

ESM 1. Studien zur Testentwicklung

ESM 2. Testaufgaben

ESM 3. Messzeitpunkte der Validierungsstudie

Literatur

- Abedi, J. & Lord, C. (2010). The language factor in mathematics tests. *Applied Measurement in Education*, 14, 219–243. https://doi.org/10.1207/S15324818AME1403_2
- Ahrenholz, B. (2010). Bildungssprache im Sachunterricht der Grundschule. In B. Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache* (S. 15–35). Tübingen: Narr.
- Ahrenholz, B. (2013). Sprache im Fachunterricht untersuchen. In C. Röhner & B. Hövelbrinks (Hrsg.), *Fachbezogene Sprachförderung in Deutsch als Zweitsprache. Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen* (S. 87–98). Weinheim: Beltz; Juventa.
- Ahrenholz, B. (2017). Sprache in der Wissensvermittlung und Wissensaneignung im schulischen Fachunterricht. Empirische Einblicke. In B. Lütke, B., Petersen, I., Tajmel, T. (Hrsg.), *Fachintegrierte Sprachbildung. Forschung, Theoriebildung und Konzepte für die Unterrichtspraxis* (S. 1–31). Berlin, Boston: De Gruyter.
- Akinwunmi, K. (2012). *Zur Entwicklung von Variablenkonzepten beim Verallgemeinern mathematischer Muster*. Heidelberg: Springer Spektrum.
- Bailey, A. L. (2007). *The language demands of school. Putting academic English to the test*. New Haven, CT: Yale.
- Berendes, K., Dragon, N., Weinert, S., Heppt, B. & Stanat, P. (2013). Hürde Bildungssprache? Eine Annäherung an das Konzept „Bildungssprache“ unter Einbezug aktueller empirischer Forschungsergebnisse. In A. Redder, A. & S. Weinert (Hrsg.), *Sprachförderung und Sprachdiagnostik. Interdisziplinäre Perspektiven* (S. 17–41). Münster: Waxmann.
- Blum, W. & Leiß, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *mathematik lehren*, 128, 18–22.
- Bochnik, K. & Ufer, S. (2016). Die Rolle fachsprachlicher Kompetenzen zur Erklärung mathematischer Kompetenzunterschiede zwischen Kindern mit deutscher und nicht-deutscher Familiensprache. *Zeitschrift für Grundschulforschung*, 9(1), 135–147.
- Bochnik, K. & Ufer, S. (2017). Fachsprachliche Kompetenzen im sprachsensiblen Mathematikunterricht der Grundschule. Implikationen einer Studie zur sprachbezogenen Analyse mathematischer Leistungsunterschiede zwischen Lernenden mit deutscher und nicht-deutscher Familiensprache. In D. Leiß, M. Hagen, A. Neumann & K. Schwippert (Hrsg.), *Mathematik und Sprache. Empirischer Forschungsstand und unterrichtliche Herausforderungen* (S. 81–98). Münster: Waxmann.
- Brown, C. L. (2005). Equity of literacy-based math performance assessments for English language learners. *Bilingual Research Journal*, 29, 337–363.
- Bußmann, H. (2008). *Lexikon der Sprachwissenschaft* (4. Aufl.). Stuttgart: Kröner.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95, 1 005–1 018.
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modelling*, 9, 233–255.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hoboken: Taylor and Francis.
- Daroczy, G., Wolska, M., Meurers, W. D. & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, 6(348), 1–13.
- de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras*, 24, 17–24.
- Dröse, J. & Prediger, S. (2020). Enhancing fifth graders' awareness of syntactic features in mathematical word problems: A design research study on the variation principle. *Journal für Mathematikdidaktik*, 41, 391–422.
- Duarte, J., Gogolin, I. & Kaiser, G. (2011). Sprachlich bedingte Schwierigkeiten von mehrsprachigen Schülerinnen und Schülern bei Textaufgaben. In S. Prediger & E. Özdil (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit* (S. 35–53). Münster: Waxmann.
- Eckhardt, A. G. (2008). *Sprache als Barriere für den schulischen Erfolg. Potenzielle Schwierigkeiten beim Erwerb schulbezogener Sprache für Kinder mit Migrationshintergrund*. Münster: Waxmann.
- Feilke, H. (2012). Bildungssprachliche Kompetenzen – fördern und entwickeln. *Praxis Deutsch*, 39(233), 4–13.
- Fornol (2020). Bildungssprachliche Mittel. Eine Analyse von Schülertexten aus dem Sachunterricht der Primarstufe (*Empirische Forschung im Elementar- und Primarbereich; Bd. 6*). Bad Heilbrunn: Klinkhardt
- Franke, M. & Ruwisch, S. (2010). *Didaktik des Sachrechnens in der Grundschule. Mathematik Primarstufe und Sekundarstufe I + II* (2. Aufl.). Heidelberg: Spektrum Akademischer Verlag.
- Gebhardt, M., Rauch, D., Mang, J., Sälzer, C. & Stanat, P. (2013). Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 275–307). Münster: Waxmann.
- Gogolin, I., Neumann, U. & Roth, H.-J. (2007). *Schulversuch bilinguale Grundschulklassen in Hamburg – Wissenschaftliche Begleitung*. Abschlussbericht über die italienisch-deutschen, por-

- tugiesisch-deutschen und spanisch-deutschen Modellklassen unter Mitarbeit von Annette Grevé und Thorsten Klinger. Hamburg. <https://www.ew.uni-hamburg.de/ueber-die-fakultaet/personen/gogolin/pdf-dokumente/bericht2007.pdf>
- Gogolin, I., Kaiser, G., Roth, H.-J., Deseniss, A., Hawighorst, B. & Schwarz, I. (2004). *Mathematiklernen im Kontext sprachlich-kultureller Diversität. Abschlussbericht 2004*. Hamburg. <https://www.ew.uni-hamburg.de/ueber-die-fakultaet/personen/gogolin/pdf-dokumente/mathe-bericht.pdf>
- Greisen, M., Georges, C., Hornung, C., Sonnleitner, P. & Schiltz, C. (2021). Learning mathematics with shackles: How lower reading comprehension in the language of mathematics instruction accounts for lower mathematics achievement in speakers of different home languages. *Acta Psychologica*, 221. <https://doi.org/10.1016/j.actpsy.2021.103456>
- Gürsoy, E., Benholz, C., Renk, N., Prediger, S. & Büchter, A. (2013). Erlös = Erlösung? Sprachliche und konzeptuelle Hürden in Prüfungsaufgaben zur Mathematik. *Deutsch als Zweitsprache*, 13(1), 14–24.
- Häsel-Weide, U. & Prediger, S. (2017). Förderung und Diagnose im Mathematikunterricht – Begriffe, Planungsfragen und Ansätze. In M. Abshagen, B. Barzel, J. Kramer, T. Riecke-Baulecke, B. Rösken-Winter & C. Selter (Hrsg.), *Basiswissen Lehrerbildung: Mathematik unterrichten mit Beiträgen für den Primar- und Sekundarstufenbereich* (S. 167–181). Seelze: Friedrich, Klett Kallmeyer.
- Heinze, A., Herwartz-Emden, L. & Reiss, K. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik*, 53, 562–581.
- Henschel, S., Gentrup, S. & Stanat, P. (2018). Evaluation im BiSS-Programm: Anlage und Zielsetzungen der zehn Evaluationsprojekte. In BiSS-Trägerkonsortium (Hrsg.), *Projektatlas Evaluation. Erste Ergebnisse aus den BiSS-Evaluationsprojekten* (S. 4–8). Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache.
- Hept, B., Dragon, N., Berendes, K., Stanat, P. & Weinert, S. (2012). Beherrschung von Bildungssprache bei Kindern im Grundschulalter. *Diskurs Kindheits- und Jugendforschung/Discourse. Journal of Childhood and Adolescence Research*, 7, 349–356.
- Hopf, M. & Eckhardt, A. G. (2013). Bildungssprache im Kindergarten. Ein Blick in die konzeptionellen Umsetzungen der Rahmenvorgaben für die frühe Bildung. In C. Röhner & B. Hövelbrinks (Hrsg.), *Fachbezogene Sprachförderung in Deutsch als Zweitsprache. Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen* (S. 118–130). Weinheim, Basel: Beltz Juventa.
- 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.
- Jeuk, S. (2010). Analyse der diskursiven Basisqualifikation bei mehrsprachigen und ein-sprachigen Kindern zum Zeitpunkt der Einschulung. In M. Rost-Roth (Hrsg.), *DaZ Spracherwerb und Sprachförderung Deutsch als Zweitsprache. Beiträge aus dem 5. Workshop „Kinder mit Migrationshintergrund“* (S. 123–139). Freiburg: Filibach.
- Jordan, R. & Stein, M. (2012). *TeMaTex - Test zum mathematischen Textverständnis. Ein Testverfahren zur Ermittlung des mathematischen Textverständnisses am Ende der Sekundarstufe I bzw. zu Beginn der beruflichen Ausbildung*. Münster: WTM.
- Jude, N. & Klieme, E. (2007). Sprachliche Kompetenz aus Sicht der pädagogisch-psychologischen Diagnostik. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistung International)* (S. 9–22). Weinheim: Beltz.
- Kempert, S., Schalk, L. & Saalbach, H. (2019). Sprache als Werkzeug des Lernens. Ein Überblick zu den kommunikativen und kognitiven Funktionen der Sprache und deren Bedeutung für den fachlichen Wissenserwerb. *Psychologie in Erziehung und Unterricht*, 66, 176–195.
- Krajewski, K., Liehm, S. & Schneider, W. (2004). *DEMAT 2+. Deutscher Mathematiktest für zweite Klassen*. Göttingen: Beltz.
- Krummheuer, G. (2010). *Wie begründen Kinder im Mathematikunterricht der Grundschule?* Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) an der Universität Kiel.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* (Bildungsreform Bd. 1, 2. Aufl.). Berlin: BMBF.
- Kniffka, G., Linnemann, M. & Thesen, S. (2007). *C-Test für den Förderunterricht: Handbuch. Universität zu Köln. Kooperationsprojekt Sprachförderung/ Stiftung Mercator*.
- Koch, P. & Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In O. Deutschmann, H. Flasche, B. König, M. Kruse, W. Pabst & W.-D. Stempel (Hrsg.), *Romanistisches Jahrbuch* (Bd. 36, S. 15–43). Berlin: De Gruyter.
- Kultusministerkonferenz (2004). *Bildungsstandards im Fach Mathematik für den Primarbereich*. München: Luchterhand.
- Kultusministerkonferenz (2009). Den Übergang von der Tageseinrichtung für Kinder in die Grundschule sinnvoll und wirksam gestalten – Das Zusammenwirken von Elementarbereich und Primarstufe optimieren. Beschluss der Jugend- und Familienministerkonferenz vom 05.06.2009 und Beschluss der Kultusministerkonferenz vom 18.06.2009. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2009/2009_06_18-uebergang-Tageseinrichtungen-Grundschule.pdf
- Kultusministerkonferenz (2022). *Bildungsstandards im Fach Mathematik. Primarbereich*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2022/2022_06_23-Bista-Primarbereich-Mathe.pdf
- Leiss, D., Plath, J. & Schwippert, K. (2019). Language and mathematics – Key factors influencing the comprehension process in reality-based tasks. *Mathematical Thinking and Learning*, 21(1), 131–153.
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lisker, A. (2013). *Sprachstandsfeststellung und Sprachförderung vor der Einschulung – Eine Bestandsaufnahme in den Ländern. Expertise im Auftrag des Deutschen Jugendinstituts*. München: Jugendinstitut e.V.
- Lühe, J., Becker, M., Neumann, M. & Maaz, K. (2017). Zur Bedeutung der sozialen Herkunft für Geschlechterunterschiede im Bildungserfolg. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 37, 416–436.
- Maier, H. & Schweiger, F. (1999). *Mathematik und Sprache. Zum Verstehen von Fachsprache im Mathematikunterricht*. Wien: öbv.
- Mark, W. & Dowker, A. (2015). Linguistic influence on mathematical development is specific rather than pervasive: Revisiting the Chinese Number Advantage in Chinese and English children. *Frontiers in Psychology*, 6(203), 1–9.
- May, P. (2006). Orthographische Kompetenz und ihre Bedingungen am Ende der vierten Jahrgangsstufe. In W. Bos & M. Pietsch (Hrsg.), *HANSE – Hamburger Schriften zur Qualität im Bildungswesen. Vol. 1. KESS 4 – Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 4 in Hamburger Grundschulen* (S. 111–141). Münster: Waxmann.

- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.
- Mayringer, H. & Wimmer, H. (2014). *SLS 2-9. Salzburger Lesescreening für die Schulstufen 2-9*. Göttingen: Hogrefe.
- Meyer, M. & Prediger, S. (2012). Sprachenvielfalt im Mathematikunterricht – Herausforderungen, Chancen und Förderansätze. *Praxis der Mathematik in der Schule*, 54(45), 2–9.
- Merkert, A. (2022). *Sprachdiagnostik im Mathematikunterricht der Grundschule. Konzeption eines Testinstruments* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 102). Münster: Waxmann.
- Möhn, D. (1981). Fach- und Gemeinsprache – zur Emanzipation und Isolation der Sprache. In W. von Hahn (Hrsg.), *Fachsprachen* (S. 172–217). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Morek, M. & Heller, V. (2012). Bildungssprache - Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für angewandte Linguistik*, 57(1), 67–101.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. & Kennedy, A. M. (2003). PIRLS International Report. *IEA's study of reading literacy achievement in primary school in 35 countries*. Chestnut Hill: PIRLS International Study Center.
- Niederhaus, C., Pöhler, B. & Prediger, S. (2016). Relevante Sprachmittel für mathematische Textaufgaben – Korpuslinguistische Annäherung am Beispiel Prozentrechnung. In E. Tschirner, O. Bärenfänger & J. Möhring (Hrsg.), *Deutsch als fremde Bildungssprache: Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik* (Deutsch als Fremd- und Zweitsprache, Bd. 7, S. 135–162). Tübingen: Stauffenburg.
- OECD (2007). *Science competencies for tomorrow's world. Vol. 2 Data (PISA 2006)*. Paris: OECD.
- Ortner, H. (2009). *Rhetorisch-stilistische Eigenschaften der Bildungssprache*. In U. Fix, A. Gardt & J. Knappe (Hrsg.), *Rhetorik und Stilistik* (S. 2 227–2 240). Berlin: De Gruyter.
- Paetsch, J., Felbrich, A. & Stanat, P. (2015). Der Zusammenhang von sprachlichen und mathematischen Kompetenzen bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Pädagogische Psychologie*, 29(1), 19–29.
- Paetsch, J., Radmann, S., Felbrich, A., Lehmann, R. & Stanat, P. (2016). Sprachkompetenz als Prädiktor mathematischer Kompetenzentwicklung von Kindern deutscher und nicht-deutscher Familiensprache. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 48, 27–41.
- Pietsch, M. & Krauthausen, G. (2006). Mathematisches Grundverständnis von Kindern am Ende der vierten Jahrgangsstufe. In W. Bos & M. Pietsch (Hrsg.), *HANSE – Hamburger Schriften zur Qualität im Bildungswesen. Vol. 1. KESS 4 – Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 4 in Hamburger Grundschulen* (S. 143–163). Münster: Waxmann.
- Prediger, S. (2013). Darstellungen, Register und mentale Konstruktion von Bedeutungen und Beziehungen – Mathematikspezifische sprachliche Herausforderungen identifizieren und überwinden. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach – Sprachlichkeit und fachliches Lernen* (S. 167–183). Münster: Waxmann.
- Prediger, S. (2017). Auf sprachliche Heterogenität im Mathematikunterricht vorbereiten. Fokussierte Problemdiagnose und Förderansätze. In J. Leuders, T. Leuders, S. Prediger & S. Ruwisch (Hrsg.), *Mit Heterogenität im Mathematikunterricht umgehen lernen. Konzepte und Studien zur Hochschuldidaktik und Lehrerbildung Mathematik* (S. 29–40). Wiesbaden: Springer VS.
- Prediger, S., Clarkson, P. & Bose, A. (2016). Purposefully relating multilingual registers: Building theory and teaching strategies for bilingual learners based on an integration of three traditions. In R. Barwell, P. Clarkson, A. Halai, M. Kazima, J. Moschkovich, N. Planas, M. Setati-Phakeng, P. Valero & M. Villavicencio (Hrsg.), *Mathematics education and language diversity: The 21st ICM Study* (S. 193–215). Cham: Springer.
- Prediger, S., Erath, K. & Moser Opitz, E. (2019). The language dimension of mathematical difficulties. In A. Fritz, V. Haase & P. Räsänen (Eds.), *International Handbook of math learning difficulties: From the laboratory to the classroom* (pp. 437–455). Cham: Springer.
- Prediger, S. & Wessel, L. (2011). Darstellen – Deuten – Darstellungen vernetzen. Ein fach- und sprachintegrierter Förderansatz für mehrsprachige Lernende im Mathematikunterricht. In S. Prediger & E. Özdil (Hrsg.), *Mathematiklernen und Bedingungen der Mehrsprachigkeit - Stand und Perspektiven der Forschung und Entwicklung* (S. 163–184). Münster: Waxmann.
- Prediger, S., Wilhelm, N., Büchter, A., Benholz, C. & Gürsoy, E. (2015). Sprachkompetenz und Mathematikleistung – Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10. *Journal für Mathematik-Didaktik*, 36(1), 77–104.
- Prior, A., Katz, M., Mahajna, I. & Rubinsten, O. (2015). Number word structure in first and second language influences arithmetic skills. *Frontiers in Psychology*, 6(266), 1–10.
- Quasthoff, U., Heller, V., Prediger, S. & Erath, K. (2022). Learning in and through classroom interaction: On the convergence of language and content learning opportunities in subject-matter learning. *European Journal of Applied Linguistics*, 10(1), 57–85.
- Quehl, T. & Scheffler, U. (2008). Möglichkeit fortlaufender Sprachförderung im Sachunterricht. In C. Bainski & M. Kürger-Potratz (Hrsg.), *Handbuch Sprachförderung* (S. 66–79). Essen: Neue Deutsche Schule Verlagsgesellschaft.
- Rank, A., Deml, I., Lenske, G., Merkert, A., Binder, K., Schulte, M., Schilcher, A., Wildemann, A., Bien-Miller, L. & Krauss, S. (2021). Eva-Prim: Evaluation von Sprachförderkompetenz und (bildungs)sprachlichen Leistungen von Schülerinnen und Schülern in Mathematik. In S. Gentrup, S. Henschel, K. Schotte, L. Beck, & P. Stanat (Hrsg.), *Sprach- und Schriftsprachförderung wirksam gestalten: Evaluation umgesetzter Konzepte* (S. 105–124). Stuttgart: Kohlhammer.
- Rasch, R. & Sturm, N. (2018). Modellierungspotenzial problemhaltiger Textaufgaben. In K. Eilerts & K. Skutella (Hrsg.), *Neue Materialien für einen realitätsbezogenen Mathematikunterricht 5* (S. 99–112). Wiesbaden: Springer Spektrum.
- Riccomini, P. J., Smith, G. W., Hughes, E. M. & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 31, 235–252.
- Riebling, L. (2013). Heuristik der Bildungssprache. In I. Gogolin, I. Lange, U. Michel & H. Reich (Hrsg.), *Herausforderung Bildungssprache – und wie man sie meistert* (S. 106–153). Münster: Waxmann.
- Rincke, K. (2010). Alltagssprache, Fachsprache und ihre besonderen Bedeutungen für das Lernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 235–260.
- Roelcke, T. (2019). Fachsprache. *(Verbale und nonverbale) Kommunikation in spezialisierten menschlichen Tätigkeitsbereichen*. https://epub.ub.uni-muenchen.de/61967/1/Roelcke_Fachsprache.pdf
- Roick, T., Göllitz, D. & Hasselhorn, M. (2004). *DEMAT 3+. Deutscher Mathematiktest für die dritten Klassen*. Göttingen: Beltz.
- Röscher, H. (2005). Nahtstelle Übergang vom Primar- zum Sekundarbereich. In H. Bartnitzky & A. Speck-Hamdan (Hrsg.), *Deutsch als Zweitsprache lernen* (S. 110–120). Frankfurt am Main: Grundschulverband – Arbeitskreis Grundschule.
- Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education* 12, 431–459.

- Schlepppegrell, M. J. (2004). *The language of schooling. A functional linguistic perspective*. Mahwah, New Jersey: Lawrence Erlbaum.
- Schmölzer-Eibinger, S. (2013). Sprache als Medium des Lernens im Fach. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach – Sprachlichkeit und fachliches Lernen* (S. 25–40). Münster: Waxmann.
- Schnotz, W., Baadte, C., Müller, A. & Rasch, R. (2011). Kreatives Problemlösen mit bildlichen und beschreibenden Repräsentationen. In K. Sachs-Hombach & R. Trotzke (Hrsg.), *Bilder sehen – denken* (S. 204–252). Köln: Halem.
- Sikora, S. & Hartke, B. (2020). Zur Konstruktion und Güte eines formativen Schulleistungstests für das Fach Mathematik in dritten Klassen. *Diagnostica*, 66, 223–234.
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S. & Henschel, S. (2021). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe: Erste Ergebnisse nach über einem Jahr Schulbetrieb unter Pandemiebedingungen*. Münster: Waxmann.
- Steinbring, H. (2017). Von Dingen, Worten und mathematischen Symbolen. In A. S. Steinweg (Hrsg.), *Mathematik und Sprache. Tagungsband des AK Grundschule in der GDM 2017* (Mathematikdidaktik Grundschule, Bd. 7, S. 25–40). Bamberg: University of Bamberg Press.
- Tajmel, T. (2017). *Naturwissenschaftliche Bildung in der Migrationsgesellschaft. Grundzüge einer Reflexion Physikdidaktik und kritisch-sprachbewussten Praxis*. Wiesbaden: Springer.
- Thürmann, E. (2013). Deutsch in allen Fächern: Überlegungen zur Förderung bildungssprachlicher Kompetenzen. In S.-I. Beutel, W. Bos & R. Porsch, (Hrsg.), *Lernen in Vielfalt. Chance und Herausforderung für Schul- und Unterrichtsentwicklung* (S. 133–156). Münster: Waxmann.
- Ufer, S. & Bochnik, K. (2020). The role of general and subject specific language skills when learning mathematics in elementary school. *Journal für Mathematikdidaktik*, 41(1), 81–117.
- Vollmer, H. J. (2011). *Schulsprachliche Kompetenzen: Zentrale Diskursfunktionen*. <http://www.home.uni-osnabrueck.de/hvollmer/VollmerDF-Kurzdefinitionen.pdf>
- Vollmer, H. J. & Thürmann, E. (2010). Zur Sprachlichkeit des Fachlernens. Modellierung eines Referenzrahmens für Deutsch als Zweitsprache. In B. Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache* (S. 107–132). Tübingen: Narr.
- Walther, G., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert & G. Walther (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 189–226). Münster: Waxmann.
- Weis, I. (2013). *Wie viel Sprache hat Mathematik in der Grundschule. Über die Notwendigkeit der Verbindung von sprachlichem und fachlichem Lernen im Mathematikunterricht der Grundschule*. https://www.unidue.de/imperia/md/content/prodaz/wie_viel_sprache_mathematik_grundschule.pdf
- Wessel, L. (2015). *Fach- und sprachintegrierte Förderung durch Darstellungsvernetzung und Scaffolding. Ein Entwicklungsforschungsprojekt zum Anteilbegriff*. Wiesbaden: Springer VS.
- Wildemann, A. & Vach, K. (2018). *Deutsch unterrichten in der Grundschule. Kompetenzen fördern, Lernumgebungen gestalten*. Seelze: Klett, Kallmeyer.
- Zifonun, G., Hoffmann, L. & Strecker, B. (1997). *Grammatik der deutschen Sprache. Teil 1* (Schriften des Instituts für Deutsche Sprache, 7.1). Berlin: De Gruyter.
- 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 capacity. *Journal of Experimental Child Psychology*. 102(1), 60–77.

Historie

Onlineveröffentlichung: 30.03.2023

Förderung

Das Projekt Eva-Prim: Evaluation im Primarbereich – Sprachförderung in alltäglichen und fachlichen Kontexten, auf das im Text Bezug genommen wird, fand im Rahmen der bundesweiten BiSS-Initiative (Bildung durch Sprache und Schrift) statt und wurde durch das Bundesministerium für Bildung und Forschung gefördert.

Open Access-Veröffentlichung ermöglicht durch die Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU).

Dr. Alexandra Merkert

Fachbereich 5: Erziehungswissenschaften
 Institut für Allgemeine Erziehungswissenschaft
 Rheinland-Pfälzische Technische Universität
 Kaiserslautern-Landau (RPTU)
 Arbeitsbereich Professionalisierung und Organisationsentwicklung
 Institut für Allgem. Erziehungswissenschaft
 Bürgerstraße 23
 76829 Landau
 Deutschland
alexandra.merkert@rptu.de