

Empirische Sonderpädagogik, 2016, Nr. 4, S. 346-366
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Konstruktion und Überprüfung eines curriculumbasierten Testverfahrens im Fach Mathematik für die vierte Klasse

Julia Rensing, Carolina Käter, Tobias Käter & Clemens Hillenbrand

Universität Oldenburg

Zusammenfassung

Schülerinnen und Schüler sollten arithmetische Operationen am Ende der Grundschulzeit sicher beherrschen (Humbach, 2009; Moser Opitz, 2009), damit eine gesicherte Grundlage für den weiteren mathematischen Kompetenzerwerb in der Sekundarstufe I besteht (Bos et al., 2007). An Lehrkräfte wird die Forderung gestellt, Lernverläufe sensibel zu dokumentieren (Strathmann & Klauer, 2010). Diese Dokumentation stellt jedoch für viele Grundschullehrkräfte eine große Herausforderung dar (Knopp & Hartke, 2010). Eine evidenzbasierte Möglichkeit des Monitorings von Lernverläufen sind dabei curriculumbasierte Messverfahren (Hartke & Diehl, 2013). In Form von Kurztests, die den Standards der wissenschaftlichen Güte entsprechen, kann die Leistungsentwicklung in einer spezifischen schulischen Domäne erhoben und dokumentiert werden. Im vorliegenden Beitrag wird ein solches curriculumbasiertes Messverfahren für das Fach Mathematik für die Jahrgangsstufe 4, das anhand eines Konstruktionsleitfadens entwickelt wurde, vorgestellt und der Überprüfung der Güte unterzogen. Die Ergebnisse zeigen, dass es sich um ein valides und reliables Instrument zur Messung des Lernverlaufs handelt.

Schlüsselwörter: Curriculumbasiertes Messen, CBM, formative Leistungsevaluation, Transition, Inklusion, Mathematik

Construction and Analysis of a cbm for School Mathematics in Grade four

Abstract

When leaving primary school, pupils should be able to master arithmetical operations (Humbach, 2009; Moser Opitz, 2009), a skill that will serve as a basis for additional mathematical skills acquisition (Bos et al., 2007). Teachers need to monitor children's progress sensitively (Strathmann & Klauer, 2010). However, this is very challenging for many teachers at primary schools (Knopp & Hartke, 2010). One evidence-based method to monitor children's progress is curriculum-based measurement (cbm) (Hartke & Diehl, 2013). These short tests that meet scientific standards document and assess school student's learning process in a specific scholastic domain. In this paper a cbm that was developed by using a guideline for the school subject maths for grade four, and its psychometric criteria are presented. Analyses confirm that the developed cbm is a valid and reliable instrument to document learning process.

Key words: Curriculum-based measurement, cbm, formative assessment, transition, inclusion, mathematics

Schülerinnen und Schüler sollen am Ende der Grundschulzeit arithmetische Operationen zum halbschriftlichen, schriftlichen und zum Kopf-Rechnen als „grundsätzliche Methoden zur Bewältigung von Rechenanforderungen“ (Krauthausen & Scherer, 2007, S. 43) mit ausreichender Sicherheit beherrschen und durchführen können, um darauf aufbauend in der Sekundarstufe I Kompetenzen in weiteren mathematischen Bereichen zu erwerben (Humbach, 2008, 2009; Moser Opitz, 2009; Stern, 2008). Das Nichtbeherrschen dieser grundlegenden arithmetischen Fertigkeiten hat schwerwiegende Folgen (Humbach, 2009; Moser Opitz, 2009). Besorgniserregend ist, dass etwa ein Fünftel der Schülerinnen und Schüler die Grundschule mit Defiziten im Bereich Mathematik verlässt (Bos et al., 2007). Die Ergebnisse der PISA-Untersuchung zeigen zudem, dass sich Probleme beim Erwerb grundlegender mathematischer Kompetenzen in der Sekundarstufe I weiter verschärfen (Bos et al., 2007). Dabei wird die Einführung eines inklusiven Bildungssystems innerhalb einer einzelnen Schulklasse zu weitaus größeren Differenzen individueller Lernvoraussetzungen und Lernleistungen von Schülerinnen und Schülern führen (Wilbert, 2014). Daher kommt der frühen Prävention von Schwierigkeiten im Erwerb mathematischer Kompetenzen in der Sekundarstufe I eine hohe Relevanz zu. Diesen kann z. B. durch eine gelungene Transition vorgebeugt werden, in der die individuelle Leistungsentwicklung in den zentralen mathematischen Inhalten schon weit vor dem Übergang dokumentiert und überprüft werden (van Ophuysen & Harazd, 2011; Peter-Koop, Hasemann & Klep, 2006).

Bezüglich der Erhebung von Lernständen in der Grundschule fordert auch die Kultusministerkonferenz (2015) die Anerkennung des individuellen Lernfortschritts: „An den individuellen Stärken orientierte, lernprozessbegleitende Rückmeldungen zeigen den Kindern ihre Lernentwicklung auf und machen Lernfortschritte und Kom-

petenzen bewusst. Dadurch gewinnen sie eine positive Einstellung zum Lernen sowie Selbstvertrauen und können Verantwortung für ihr eigenes Lernen übernehmen“ (S. 6) (s. auch Gebhardt, Diehl & Mühling, 2016).

Für Lehrkräfte ist es jedoch eine Herausforderung, im Schulalltag den individuellen Lernverlauf eines jeden Lernenden objektiv, reliabel, valide und sensibel zu erheben (Diehl, Hartke & Knopp, 2009; Knopp & Hartke, 2010; Wilbert, 2014). Eine mögliche Unterstützung darin stellen curriculumbasierte Messverfahren (CBM) dar.

Curriculumbasierte Messverfahren

CBM gelten als formative Leistungsmessungen, die sich inhaltlich am Unterricht orientieren und deren Durchführung im Regelfall eine bis drei Minuten dauert. Diese zeitliche Ökonomie stellt gegenüber herkömmlichen standardisierten Messverfahren, wie beispielsweise die Reihe des Deutschen Mathematiktests (Gölitz, Roick & Hasselhorn, 2006), einen Vorteil für einen regelmäßigen Einsatz im Unterricht dar (im Gruppentest nimmt die Durchführung dieser Verfahren eine Gesamtzeit von 45 Minuten in Anspruch). CBM dokumentieren den Leistungsverlauf von Schülerinnen und Schülern über einen längeren Zeitraum, der grafisch abgebildet wird. Sie versprechen ein systematisches Monitoring von Lern- und Entwicklungsverläufen (Limbach-Reich, 2015). Besonderes Merkmal curriculumbasierter Messverfahren ist dabei eine enge Bindung an wissenschaftliche und diagnostische Standards (Voß & Hartke, 2014). In der Konstruktion curriculumbasierter Messverfahren ist die Erfüllung der hohen wissenschaftlichen Güte in Anbetracht der versprochenen Unterrichtsnähe eine große Herausforderung. Die deutsche Forschung zu formativer Leistungsdiagnostik ist gegenüber der in den USA noch nicht weit fortgeschritten (Klauer, 2014; Voß & Hartke, 2014).

Formative Leistungsdiagnostik kann auch innerhalb des Response-to-Interven-

tion-Ansatz (Rtl) verortet werden (Hartke & Diehl, 2013). Dabei tauchen unterschiedliche Begrifflichkeiten auf. Der oftmals verwendete Terminus des curriculum-based assessment (CBA) gilt als „übergeordneter Begriff jegliche[r] Maßnahmen des Informationsgewinns zur pädagogischen Entscheidungsfindung“ (Voß & Hartke, 2014, S. 85). Diesem Oberbegriff sind CBM unterzuordnen, wobei sich ebendiese durch ihre enge Bindung an wissenschaftliche diagnostische Standards von anderen CBA-Instrumenten unterscheiden (Voß & Hartke, 2014). Innerhalb der Literatur werden drei Methoden genannt, die weiterhin dem Konzept von CBM unterzuordnen sind: general outcome measures (GOM), skills-based measures (SBM) sowie mastery measures (MM) (s. Abbildung 1).

Die drei Methoden unterscheiden sich hinsichtlich des Einsatzes, des Aufbaus sowie der Struktur und weisen somit verschiedene Vor- und Nachteile auf, wie Tabelle 1 zeigt.

Die Güte curriculumbasierter Messverfahren

Ziel der Lernverlaufsdiagnostik als Änderungsmessung ist es, auf Individualebene die Entwicklung der Lernleistung sichtbar zu machen und somit Zustandsänderung in Abhängigkeit von Wirkfaktoren zu finden (Petermann, 2010; Wilbert & Linnemann, 2011). Für das Messen und Testen von Veränderungen im Lernen existiert bis heute keine allgemeingültige methodologische Grundlegung. „Folglich gibt es eine unübersichtliche Vielzahl von Methoden und Maßen auf technischer Ebene“ (Waldmann & Petermann, 2014, S. 86). Die Auseinandersetzung mit der Verortung von CBM in Testtheorien ist jedoch höchst relevant, denn „sicher hat man es nicht mit Tests gemäß der klassischen Testtheorie zu tun“ (Strathmann & Klauer, 2010). Fuchs (2004, S. 189) schlägt für die Erstellung von CBM eine Orientierung auf drei Ebenen vor:

- (1) „Technical features of the static score“,
- (2) „Technical features of slope“ und
- (3) „Instructional utility“.

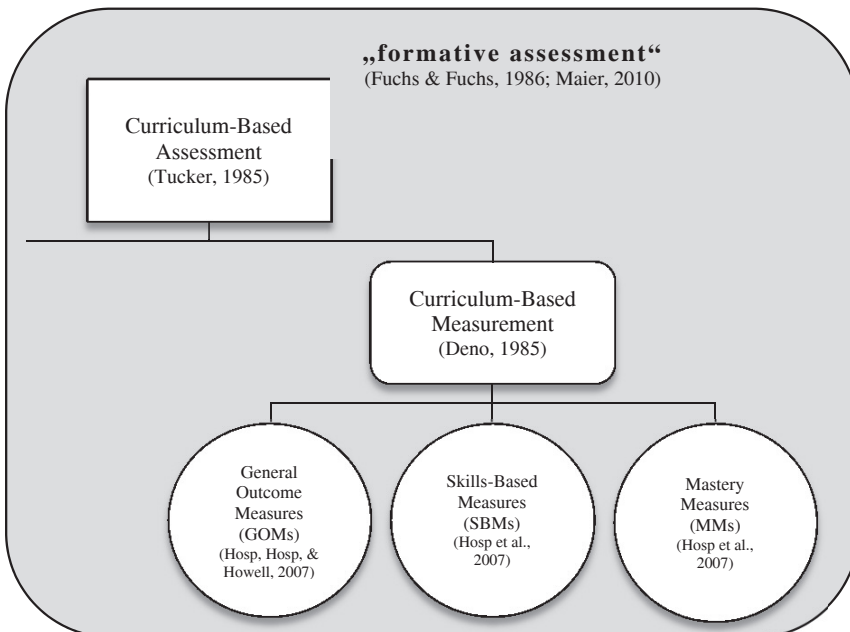


Abbildung 1: Formen des formative assessments (adaptiert nach Voß & Hartke, 2014, S. 86)

Tabelle 1: Vergleich der drei CBM-Typen GOM, SBM und MM (Hosp et al., 2007, S. 15 [eigene Übersetzung])

General outcome measures (GOM)	Skills-based measures (SBM)	Mastery measures (MM)
Beispiel		
Messung der Leseflüssigkeit	Überprüfung der halbschriftlichen Addition und Subtraktion	Überprüfung der Addition im Zahlenraum bis 10
Einsatz		
<ul style="list-style-type: none"> – Screening – Lernverlaufsdokumentation 	<ul style="list-style-type: none"> – Screening – Lernverlaufsdokumentation 	<ul style="list-style-type: none"> – Spezifische Überprüfung eines bestimmten Lernbereichs – Evaluation eher im diagnostischen Sinne
Aufbau und Struktur		
<ul style="list-style-type: none"> – Messung einer „globalen“ Kompetenz – Latente Teilfertigkeiten werden nicht berücksichtigt – Für einen längeren Dokumentationszeitraum angelegt – Enthält zumeist allgemeine „Klassenaufgaben“ 	<ul style="list-style-type: none"> – Gemischte Aufgabentypen auf Grundlage des Curriculums oder des Unterrichts – Aufgaben können übergreifende Kompetenzen überprüfen – Fähigkeiten/ Kompetenzen können isoliert betrachtet werden 	<ul style="list-style-type: none"> – Überprüfung einer bestimmten Teilfertigkeit – Für einen kürzeren Dokumentationszeitraum angelegt – Größerer Aufgabenpool – Items beziehen sich auf bestimmte Fähigkeiten oder Leistungsstufen – Fähigkeiten/ Kompetenzen können isoliert betrachtet werden
Vorteile		
<ul style="list-style-type: none"> – Ermöglicht langfristigen Überblick über die Entwicklung – Gut geeignet für Dokumentation der Entwicklung – Geringe Messverzerrung – Zeigt Stagnationen auf und ermöglicht Aussagen über allgemeine Entwicklung 	<ul style="list-style-type: none"> – Gibt einen Überblick über Kompetenzstand – Gut geeignet für Dokumentation der Entwicklung – Geringe Messverzerrung – Zeigt Stagnationen auf 	<ul style="list-style-type: none"> – Kann als Ergänzung zu einem SBM genutzt werden – Nützlich zur gezielten Hypothesenprüfung – Sehr fokussiert
Nachteile		
<ul style="list-style-type: none"> – Geringe diagnostische Information – Keine Aussagen über Teilkompetenzen – Beinhaltet oftmals Aufgaben, die noch nicht gelöst werden können – Entwicklungsziele sind oftmals per definitionem nicht festgelegt 	<ul style="list-style-type: none"> – Geringe diagnostische Information aufgrund geringer Aufgabenzahl eines Bereichs – Beinhaltet oftmals Aufgaben, die noch nicht gelöst werden können – Verallgemeinerung oder praktisches Ausführen der Fähigkeiten sind nicht erforderlich 	<ul style="list-style-type: none"> – Gibt keinen großen Überblick über allgemeine Kompetenzen – Generalisierung oder Herstellen von Zusammenhängen ist nicht möglich – Verhältnis Teilkompetenz-Kompetenz ist unpassend – Eher nicht für langfristige Verlaufsdokumentation angelegt

Die erste Ebene setzt sich dabei mit Fragen nach traditionellen psychometrischen Kennzeichen hinsichtlich Reliabilität und Objektivität auseinander. Ebene 2 geht der Frage nach, ob zunehmende Werte in den CBM tatsächlich mit der Verbesserung innerhalb der jeweiligen akademischen Domäne einhergehen. Die Auseinandersetzung mit dem praktischen Nutzen der Ergebnisse für Praktikerinnen und Praktiker in Bezug auf die Vermittlung findet auf einer dritten Ebene, „Instructional utility“, statt.

Die Klassische Testtheorie (KTT) zeigt in der Überprüfung der Güte von Verfahren der Lernverlaufsmessung Grenzen auf. Beispielsweise weist Rost (2004) auf ein Validitätsproblem hin, das bei Veränderungsmessungen auftritt, „weil normalerweise mit Hilfe der Korrelation zwischen zwei Variablen geprüft wird, inwieweit beide Variablen dasselbe messen“ (S. 280). Dahingegen ist es gerade das Ziel von Veränderungsmessungen, dass sich Personenwerte auf individuell unterschiedliche Weise ändern sollen (Knopp & Hartke, 2010; Rost, 2004). Zudem ist das Reliabilität-Validitäts-Dilemma zu diskutieren. Denn für änderungssensible Verfahren gilt, dass sie nur mäßige Retest- und Paralleltestreliabilitäten bei einer hohen Split-half Reliabilität aufweisen (Strathmann, 2014). Hohe Retest- oder Paralleltestreliabilitäten „sprechen vielmehr dafür, dass die Probandinnen und Probanden sich in ihren Werten nicht oder alle im gleichen Ausmaß und in der gleichen Richtung verändert haben“ (Waldmann & Petermann, 2014, S. 86). Da zwischen einzelnen Tests Lernfortschritte stattfinden, ist zu erwarten, dass Tests notwendigerweise leichter werden und ihre Mittelwerte somit ansteigen (Strathmann & Klauer, 2010).

Wilbert und Linnemann (2011) schlagen ein Vorgehen vor, das zugleich Aspekte der Klassischen und der Probabilistischen Testtheorie berücksichtigt. Um die Güte zu ermitteln, sind folgende vier Schritte zu gehen:

1. Anwendung einer Itemanalyse gemäß der Klassischen Testtheorie, um Trenn-

schärfe, Schwierigkeit, interne Konsistenz, Homogenität und Retest-Reliabilität des entwickelten Tests zu ermitteln.

2. Untersuchung der Eindimensionalität des Konstrukts anhand einer konfirmatorischen Faktoranalyse.
3. Durchführung der polytomen Raschmodellierung für die Erfassung der Itemschwierigkeit.
4. Berechnung der Differential Item Functioning-Werte anhand ordinal logistischer Regressionsanalysen basierend auf der Item-Response-Theorie zur Bestimmung der Testfairness.

Bisherige deutschsprachige curriculumbasierte Messverfahren im Bereich Mathematik und ihre Grenzen

Aktuell liegen in Deutschland bereits standardisierte Verfahren der Lernverlaufsdagnostik im Bereich Mathematik vor. Sie beziehen sich vornehmlich auf den Grundschulbereich mit dem Schwerpunkt Arithmetik. International ist eine größere Anzahl an Instrumenten der Lernverlaufsdagnostik zu verzeichnen. Die Adaption vorhandener Instrumente aus dem internationalen Sprachraum auf das deutsche Bildungssystem ist jedoch nicht ohne weiteres möglich. Ein Grund dafür sind u. a. die unterschiedlichen Bildungssysteme, unter denen das deutsche mit seiner kurzen, vierjährigen Grundschule eher eine Ausnahme darstellt (Schmitt, 2001). Somit beginnt das Curriculum, das in jedem Land unterschiedlich gestaltet ist, für den Sekundarstufenbereich hierzulande in der Klasse fünf, in anderen Ländern tendenziell eher später. Bekannt ist dabei, dass Transitionen sogar einen negativen Einfluss auf die akademische Leistungsentwicklung der Schülerinnen und Schüler nehmen können (Racherbäumer & Kohnen, 2014).

Weitere Verfahren befinden sich noch in der Entwicklung, wie beispielsweise das gemeinsame Forschungsprojekt LeVuMi (Lern-Verlaufs-Monitoring) von Gebhardt,

Tabelle 2: Bisher in Deutschland veröffentlichte Verfahren bzw. in Projekten implementierte Instrumente zur Lernverlaufsdiagnostik für das Fach Mathematik

Programm	Diagnose- und Förderblätter in drei Bänden	Lernverlaufsdiagnostik Mathematik für zweite bis vierte Klassen (LVD-M 2-4)	Lernfortschrittsdiagnostik Grundrechenarten	Inventar Rechenfische	„quop“-Projekt	Rügener Inklusionsmodell
AutorIn, Jahr	Klauer (1994)	Strathmann & Klauer (2012)	Hartmann & Müller (2014)	Knopp & Hartke (2010)	Souvignier, Förster & Salaschek (seit 2008)	Voß & Hartke (2010–2015)
Einsatz	Klasse 2–4	Klasse 2–4	Klasse 1–4	Klasse 1	Klasse 1–6	Klasse 1–4
Inhalte	Addition, Subtraktion, Multiplikation, Division, Größen	Addition, Subtraktion, Multiplikation, Division	Addition, Subtraktion, Multiplikation, Division	Addition, Subtraktion, Zahlen zerlegen, Zahlreihen ergänzen, Textaufgaben oder Zahlenstrahl	Mathematische Vorläuferkompetenzen, curriculare Kompetenzen	Addition, Subtraktion, Multiplikation, Division
Dauer (Min.)	k. A.	15	3	unbegrenzt	10–15	k. A.
Parallelversionen	30	unbegrenzt	120	3 Versionen (in je A- und B-Form)	8 pro Schuljahr	k. A.
Derzeit öffentlich zugänglich	ja	ja	ja	nein	nein	nein

Diehl und Mühling. Es liegen bereits Lernverlaufstests für die Bereiche Leseflüssigkeit und Buchstabentests vor, die Entwicklung weiterer Reihen, wie auch eine für das Fach Mathematik, befindet sich noch in Planung (Gebhardt & Mühling, in Planung).

Die vorhandenen Verfahren zeigen, dass die Forschung zu und Entwicklung von Instrumenten der Lernverlaufsdiagnostik in Deutschland seit der Jahrtausendwende vorangeschritten ist. Dennoch gibt es Barrieren, weshalb diese CBM bisher nicht in den Unterrichtsalltag implementiert werden. Mögliche Gründe dafür sind eine fehlende

Passung für den individuellen Unterricht, die begrenzte und geringe Anzahl an Testbögen, kein öffentlicher sondern ggf. kostenpflichtiger Zugang zu den Verfahren aufgrund interner Plattformen o. ä. und die z. T. lange Durchführungsdauer, die auch die konzeptionell vorgeschlagene Testdauer für CBM-Kurztests überschreitet (Käter, Käter, Martenstein & Hillenbrand, 2016). Gemäß internationaler Studien lassen sich insbesondere die folgenden Antworten auf die Frage nennen, weshalb Lehrkräfte häufig keine curriculumbasierten Messungen durchführen: Die Gründe sind vorrangig

der zeitliche Aufwand, ein zu geringes Fachwissen, das benötigte Material, die Unwissenheit über Auswertungstechniken, logistische Herausforderungen oder eine angezweifelte Anschauungsvalidität (Deno, 1993; Wesson, King & Deno, 1984; Yell, Deno & Marston, 1992). Križan und Vossen (2016) sehen als weitere Gründe für die nicht vorhandene Implementation von CBM in den schulischen Alltag das wissenschaftliche Eigenleben, in dem Forschungsergebnisse Praktikerinnen und Praktikern nicht zur Verfügung gestellt werden. Zudem nennen sie als weitere Gründe die schulische Tradition, in der sogar oftmals pseudowissenschaftliche Förderverfahren zum Einsatz kommen und letztlich die politische Koordination, in der ein Austausch zwischen Wissenschaft und Schule nicht intensiv genug forciert wird.

Aus diesen Gründen überprüft der vorliegende Beitrag, ob die anhand eines am spezifischen Unterricht orientierten Konstruktionsleitfadens entwickelten CBM den hohen Ansprüchen an die wissenschaftliche Güte entsprechen. Bewährt sich ein solches Vorgehen, eröffnet sich für Lehrkräfte die Möglichkeit einer spezifischen Adaption ihres Handelns: Sie können auf ihren eigenen Unterricht angepasste, individuelle und zuverlässige CBM entwickeln, um die Lernentwicklung ihrer Schülerinnen und Schüler zu dokumentieren und zum Anlass weiterer Unterstützung zu nehmen. Eine solche Orientierung des Unterrichts an der Bedürfnislage der Lernenden entspricht den Forderungen evidenzbasierter sonderpädagogischer Praxis (Casale, Hennemann & Grosche, 2015; Hillenbrand, 2015). Voraussetzung für den Einsatz eines Konstruktionsleitfadens ist jedoch die wissenschaftlich nachgewiesene Güte des Instruments.

Forschungsziel und Forschungsfrage

In dem vorliegenden Beitrag wird ein Instrument der Lernverlaufsdiagnostik für die Jahrgangsstufe 4 im Fach Mathematik vorge-

stellt, das anhand eines siebenschrittigen Leitfadens konstruiert wurde. Die darin beinhalteten arithmetischen Grundoperationen (Addition, Subtraktion, Multiplikation, Division & Stellenwertsystem) stellen als mathematische Basiskompetenzen notwendige Lernvoraussetzungen für weiterführende Inhalte des Mathematikunterrichts in der Sekundarstufe I dar. Das hier vorgestellte Instrument fokussiert sich somit auf den Inhaltsbereich „Zahlen und Operationen“ der Bildungsstandards für das Fach Mathematik für die Grundschule (Kultusministerkonferenz, 2005).

Die zentrale Forschungsfrage des vorliegenden Beitrags lautet: Entspricht ein leitfadengestützt entwickeltes curriculumbasiertes Messverfahren für das Fach Mathematik in der Jahrgangsstufe 4 den hohen, an CBM gestellten, wissenschaftlichen Standards der Güte?

Methode

Design

Zur Überprüfung der Fragestellung wurde ein Parallelgruppen-Design gewählt und mit einem Cross-Over-Design kombiniert. Dabei bearbeiteten Probandinnen und Probanden in zwei Gruppen über insgesamt 12 Messzeitpunkte dieselben entwickelten CBM, jedoch in einer anderen Reihenfolge (s. Tabelle 3: Forschungsdesign). Die Gruppen bearbeiteten dabei jeweils die A- und B-Version eines CBM zum selben Messzeitpunkt, da für die schulische Implementation Parallelversionen, zum Beispiel zur Vermeidung des Abschreibens bei Sitznachbarn, notwendig sind. Die A- und B-Formen eines CBM stellen also Paralleltests dar. Die Gruppe, die zunächst alle A-Formen bearbeitet hat, wird im Folgenden mit „Versuchsgruppe A“ (VG-A) bezeichnet, die andere Gruppe mit „Versuchsgruppe B“ (VG-B).

Ein Vorteil des Cross-Over-Designs liegt darin, dass im Sinne klinischer Studien jede Probandin und jeder Proband seiner eige-

Tabelle 3: Forschungsdesign

	Prä	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	Post
VG-A	HRT 1-4	CBM 1A	CBM 2A	CBM 3A	CBM 4A	CBM 5A	CBM 6A	CBM 1B	CBM 2B	CBM 3B	CBM 4B	CBM 5B	CBM 6B	HRT 1-4
VG-B	HRT 1-4	CBM 1B	CBM 2B	CBM 3B	CBM 4B	CBM 5B	CBM 6B	CBM 1A	CBM 2A	CBM 3A	CBM 4A	CBM 5A	CBM 6A	HRT 1-4

Anmerkung: VG-A = Versuchsgruppe A, VG-B = Versuchsgruppe B

nen Kontrolle dient (Schumacher & Schulgen, 2009). Zudem kann in Anlehnung an Voß (2014) ein möglicher Übungs- bzw. Lerneffekt kontrolliert werden. Vor dem ersten und nach dem letzten Messzeitpunkt erfolgte eine Prä- bzw. Postmessung durch den Heidelberger Rechentest (HRT 1-4) (Subskala Rechenoperationen) (Haffner, Baro, Parzer & Resch, 2005).

Instrumente

Die curriculumbasierten Testbögen. Zur Entwicklung der Testbögen wurde ein Leitfaden zur Konstruktion von CBM hinzugezogen, der bereits für den Bereich Lesen positiv evaluiert wurde (Käter et al., 2016). In diesem Leitfaden wird ein von Fuchs (2004) vorgeschlagener Konstruktionsweg für CBM verfolgt. Sie empfiehlt dabei eine Präzisierung der jeweiligen interessierenden Kompetenz, die dann durch die *Definition* von Aufgabentypen operationalisiert wird. Der hier genutzte Leitfaden ist etwas differenzierter und besteht aus insgesamt sieben Schritten, die in der Konstruktion der Reihe nach durchlaufen werden.

Schritt 1: Auswahl des Lehr-/Lerninhaltes

Schritt 2: Bestimmung des Aufgabentyps

Schritt 3: Definition der Aufgaben- und Schwierigkeitsbereiche

Schritt 4: Bestimmung des Testumfangs

Schritt 5: Generierung der Aufgabenstichprobe

Schritt 6: Erstellung der Testblätter

Schritt 7: Auswertung und Dokumentation des Lernverlaufs

Inhaltlich orientieren sich die hier entwickelten CBM an dem Kompetenzbereich „Zahlen und Operationen“ der Bildungsstandards für das Fach Mathematik für die Grundschule (Jahrgangsstufe 4) (Kultusministerkonferenz, 2005) (Schritt 1). Als Aufgabentyp wurde hier der Typ Skills-based measures (SBM) gewählt, da die Kompetenzen in *verschiedenen* Grundrechenarten zugleich geprüft werden sollen (Müller & Hartmann, 2009) (Schritt 2). Die Definition der Aufgaben- und Schwierigkeitsbereiche ist in Tabelle 4 ersichtlich (Schritt 3). Ein Aufgabenblatt besteht aus insgesamt 15 dichotomen Aufgaben, die in drei Schwierigkeitsblöcke (Niveaus) eingeteilt werden. In jedem Schwierigkeitsblock befindet sich je eine Aufgabe zum Stellenwertsystem, zur Addition, Subtraktion, Multiplikation und Division. Es wurden zwölf Parallelversionen mit insgesamt 180 Aufgaben entwickelt. Davon fallen je 36 Aufgaben auf die jeweiligen Grundrechenarten resp. das Stellenwertsystem (Schritt 4). Die Generierung der Aufgaben erfolgte mittels des Programms „Rechenblatt“, das auf Grundlage vorgegebener Definitionen unendlich viele zufällige Rechenaufgaben generiert (Schritt 5). Anschließend wurden die generierten Aufgaben jeweils parallel in Testbögen festgehalten (Schritt 6) und im vorliegenden Beitrag ausgewertet (Schritt 7).

Der Heidelberger Rechentest 1-4 (HRT 1-4). Der Heidelberger Rechentest 1-4 (HRT 1-4) (Haffner et al., 2005) ist ein standardisiertes Instrument zur Erfassung der, von curricularen Stoffplänen unabhängigen, mathematischen Basiskompetenzen im Grundschulalter. Das Verfahren umfasst die

Tabelle 4: Überblick über Aufgaben und zugrundeliegende Definitionen. Die Bearbeitung konnte halbschriftlich, schriftlich oder im Kopf erfolgen

Aufgabe	Bezeichnung	Definition	Beispielitem												
Aufgabe 1	Stellenwertsystem Niveau 1	$E < 10$ $Z < 10$	<table border="1"> <thead> <tr> <th>Z</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>○○○</td> <td>○○</td> </tr> </tbody> </table>	Z	E	○○○	○○								
Z	E														
○○○	○○														
Aufgabe 2	Addition Niveau 1	$T+H < 10.000$	$8000 + 300 =$												
Aufgabe 3	Subtraktion Niveau 1	$T-H > 0$	$8000 - 400 =$												
Aufgabe 4	Multiplikation Niveau 1	$ZE \cdot E < 1.000$	$24 \cdot 6 =$												
Aufgabe 5	Division Niveau 1	$ZE:E < 100$ ($E \neq 0$)	$63 : 7 =$												
Aufgabe 6	Stellenwertsystem Niveau 2	$10 \geq E \leq 20$ $Z < 10$ $H < 10$ (Übertrag bei Z)	<table border="1"> <thead> <tr> <th>H</th> <th>Z</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>○○○○○</td> <td>○○</td> <td>○○○○○</td> </tr> <tr> <td>○○○</td> <td></td> <td>○</td> </tr> </tbody> </table>	H	Z	E	○○○○○	○○	○○○○○	○○○		○			
H	Z	E													
○○○○○	○○	○○○○○													
○○○		○													
Aufgabe 7	Addition Niveau 2	$THZE + ZE < 10.000$ (Übertrag bei Z)	$2569 + 14 =$												
Aufgabe 8	Subtraktion Niveau 2	$THZE - TH > 0$ (Übertrag bei H)	$9119 - 8800 =$												
Aufgabe 9	Multiplikation Niveau 2	$ZE \cdot ZE < 10.000$	$59 \cdot 57 =$												
Aufgabe 10	Division Niveau 2	$HZE:E < 1.000$ $E \neq 0$	$637 : 7 =$												
Aufgabe 11	Stellenwertsystem Niveau 3	$10 \geq E \leq 20$ $10 \geq Z \leq 20$ $H < 10$ (Übertrag bei Z und H)	<table border="1"> <thead> <tr> <th>H</th> <th>Z</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>○○○○○</td> <td>○○○○○</td> <td>○○○○○</td> </tr> <tr> <td></td> <td>○○○○○</td> <td>○○○○○</td> </tr> <tr> <td></td> <td>○</td> <td>○○○○○</td> </tr> </tbody> </table>	H	Z	E	○○○○○	○○○○○	○○○○○		○○○○○	○○○○○		○	○○○○○
H	Z	E													
○○○○○	○○○○○	○○○○○													
	○○○○○	○○○○○													
	○	○○○○○													
Aufgabe 12	Addition Niveau 3	$THZE + THZE < 20.000$ (Übertrag bei Z und H)	$3656 + 3095 =$												
Aufgabe 13	Subtraktion Niveau 3	$THZE - THZE > 0$ (Übertrag bei Z und H)	$8614 - 4487 =$												
Aufgabe 14	Multiplikation Niveau 3	$HZE \cdot ZE < 100.000$	$567 \cdot 53 =$												
Aufgabe 15	Division Niveau 3	$THZE : E < 10.000$ ($E \neq 0$)	$3395 : 7 =$												

Anmerkung: T = Tausender, H = Hunderter, Z = Zehner, E = Einer

Bereiche „Rechenoperationen“ und „Numerisch-logische und räumlich-visuelle Fähigkeiten“. Normiert wurde das Verfahren anhand einer Stichprobe von $N=2.262$ Grund- und Förderschülerinnen und -schülern der ersten bis vierten Klasse. In der vorliegenden Studie wurde ausschließlich die Skala „Rechenoperationen“ genutzt. Die Reliabilität der einzelnen Untertests beläuft sich auf .68 bis .89 (Haffner et al., 2005). Zusätzlich wurde die Validität über Außenkriterien, wie Schulnoten in den Fächern Deutsch und Mathematik, andere Testverfahren (Aufmerksamkeits-Belastungs-Test (d2) (Brickenkamp, 2002), DEMAT 4 (Gölitz et al., 2006), Würzburger-Leise-Leseprobe (WLLP) (Küspert & Schneider, 1998), Weingartener Grundwortschatz Rechtschreibtest für 3. und 4. Klassen (WRT3+) (Birkel, 2007) sowie das Prüfsystem für Schul- und Bildungsberatung (PSB) (Horn & Lukesch, 2002)) positiv evaluiert.

Stichprobe

Die Größe der Stichprobe beläuft sich auf $N=143$ (s. Tabelle 5). Die Schülerinnen und Schüler stammen dabei aus drei verschiedenen Grundschulen und insgesamt sieben Klassen in Niedersachsen. Die

durchschnittliche Klassengröße liegt bei 20.43 Schülerinnen und Schülern. Es handelt sich um eine anfallende Stichprobe, da die Schulen nicht randomisiert gewählt wurden. Die Schülerinnen und Schüler nahmen an der Untersuchung teil, nachdem eine Zustimmung der Schulleitung und der Klassenlehrkräfte vorlag. Weder Schulleiter, Klassenlehrkräfte noch Schülerinnen und Schüler erhielten eine Entschädigung für die Erhebung.

Dropout. Tabelle 6 ist zu entnehmen, wie viele Schülerinnen und Schüler in den jeweiligen Messzeitpunkten fehlten. Fehlende Werte entstanden hier lediglich durch nicht anwesende Schülerinnen und Schüler. Alle anderen haben die CBM bearbeitet und wurden in die Analyse mit aufgenommen. Daher kann hier von zufälligen fehlenden Werten die Rede sein.

Durchführung

Die Bearbeitung der zwölf CBM-Testbögen durch die Schülerinnen und Schüler erfolgte zwischen den Oster- und Sommerferien im Jahr 2014. Die Prä- und Posterhebung wurden von geschulten Studierenden durchgeführt und fanden in den ersten Stunden des Schultags statt. Sie dauerten in je-

Tabelle 5: Zusammensetzung der Stichprobe

Versuchsgruppe		
$n = 143$		
n Schulen = 3		
n Klassen = 7		
Ø Klassengröße = 20.43 Schülerinnen und Schüler		
männlich	weiblich	k. A.
$n = 77$ (53.8 %)	$n = 63$ (44.1 %)	$n = 3$ (2.1 %)
Ø Alter (in Jahren) \approx 10.52 (k. A.: 18.3 %)		

Tabelle 6: Dropout-Werte zu den einzelnen Messzeitpunkten

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
Fehlend	9.09	10.49	8.05	7.69	5.59	9.09	6.99	5.59	6.29	9.09	5.59	8.39

Anmerkung: Angaben in Prozent

der Klasse ca. eine Schulstunde. Nachdem die Prä-Testung in den Klassen vorgenommen wurde, wurde in der jeweils direkt folgenden Stunde die erste CBM-Testung durchgeführt. Auch diese Durchführung erfolgte durch die Studierenden, damit den anwesenden Mathematiklehrkräften exemplarisch gezeigt werden konnte, wie die weitere Durchführung erfolgen soll. Die Schülerinnen und Schüler lösten in den ersten fünf Minuten die Aufgaben mit einem schwarzen Stift (Speed-Test-Komponente), für die weitere benötigte Zeit schrieben sie mit einem blauen Stift weiter. Dieses Vorgehen wurde gewählt, um im Nachhinein ermitteln zu können, wie viele Aufgaben die Schülerinnen und Schüler in fünf Minuten lösen konnten und ob die Anzahl der gestellten Aufgaben für eine CBM-Messung als *Kurztest* angemessen ist. Die Durchführung der CBM erfolgte folglich nicht ausschließlich in Form eines Speed-Tests. Die Schülerinnen und Schüler hatten ausreichend Zeit, die Bögen zu bearbeiten. Eine Durchführung benötigte insgesamt ca. 20 Minuten. Für diejenigen, die schon früh mit der Bearbeitung der Aufgaben fertig waren, standen für jede der zwölf Durchführungen Knobelaufgaben bereit. Die weiteren elf CBM führten die Lehrkräfte wie beschrieben orientiert am eigenen schulischen Alltag im Mathematikunterricht nach der Vorgabe durch, dass zwei CBM pro Woche durchgeführt werden sollten, wobei die jeweiligen Testungen nicht an aufeinanderfolgenden Tagen stattfanden.

Statistische Analyse

Die Vorgehensweise stützt sich auf den o. g. Vorschlag von Wilbert und Linnemann (2011). Konkret mündet dies in folgenden Erwartungen:

Die CBM stellen ein valides Instrument zur Lernverlaufdiagnostik im Bereich der arithmetischen Kompetenzen dar (Validität).

Die Lösung der CBM ist frei von Fähigkeiten oder Umständen außerhalb des gewählten Inhaltsbereichs (Homogenität).

Die CBM messen die Fähigkeiten der Kinder im Bereich der arithmetischen Kompetenzen zuverlässig und in vergleichbarem Maße (Reliabilität).

Die CBM eignen sich gemäß ihrer angedachten Verwendung als Speed-Test.

Die Items der einzelnen CBM-Tests werden zunehmend, aber zwischen den verschiedenen CBM-Tests in vergleichbarem Maße schwerer.

Ergebnisse

Allgemeine Entwicklung

Bezüglich der allgemeinen Entwicklung zeigt sich, dass sowohl bei den in fünf Minuten gelösten Aufgaben als auch bei den insgesamt richtig gerechneten Aufgaben ein positiver Entwicklungstrend einsetzt (lineare Regression: $R^2=0.66$ bzw. $R^2=0.85$) (s. Abb. 2).

Innerhalb der Messzeitpunkte (1, 2, 3 ... 12) ergaben sich in den meisten Fällen keine Unterschiede zwischen den jeweiligen Formen A und B (siehe Tabelle 7). Ausnahmen bildeten dabei die Messzeitpunkte 2, 6, 9 und 10. Bei der zweiten Bearbeitung dieser Testbögen in den Messzeitpunkten 8, 12, 3 und 4 ließen sich wiederum keine signifikanten Unterschiede feststellen.

Untersuchungen gemäß der Klassischen Testtheorie

Validität. Zur Überprüfung der Kriteriumsvalidität wurden alle Messzeitpunkte der CBM mit den Werten des HRT 1–4 (Haffner et al., 2005) korreliert. Da die vorherige Analyse keine Unterschiede zwischen den A- und B-Formen innerhalb eines Messzeitpunkts aufgezeigte, wurden für die Korrelationen jeweils die CBM in A- und B-Form eines Messzeitpunktes gemeinsam einbezogen. Es ergaben sich stabile signifikante Korrelationen zwischen $r=.36$ und $r=.54$ mit dem Prä-Zeitpunkt und ebenso signifikante Korrelationen zwischen $r=.49$ und $r=.60$

zum Post-Zeitpunkt. Es besteht ein mittlerer Zusammenhang (Diaz-Bone, 2006) der CBM mit den Werten des HRT.

Reliabilität. Zur Betrachtung der Paralleltest-Reliabilität der CBM wurden die jeweils aufeinander folgenden CBM in der Darstellung einer Simplex-Struktur korreliert (siehe Tabelle 9) (s. Schneider & Treiber, 1986). Die CBM benachbarter Zeitpunkte korrelieren in mittlerem bis hohem Maße miteinander ($.62 \leq r \leq .83$ auf der Diagonalen). Es ist erkennbar, dass die Korrelationskoeffizienten tendenziell abneh-

men, je weiter entfernt diese von der Diagonalen liegen (minimal $r = .49$ bei T1/T8).

Interne Konsistenz. Um die interne Konsistenz zu bestimmen, wurde in einem ersten Schritt Cronbachs Alpha der CBM aller Messzeitpunkte anhand der jeweils erreichten Gesamtrohwerte ermittelt. Es ergibt sich ein Alpha von $\alpha_c = .96$.

In einem zweiten Schritt wurde die interne Konsistenz der einzelnen CBM ermittelt. Die Werte liegen bei fast allen CBM im akzeptablen Bereich ($.68 < \alpha_c < .78$ für die gesamte Stichprobe).

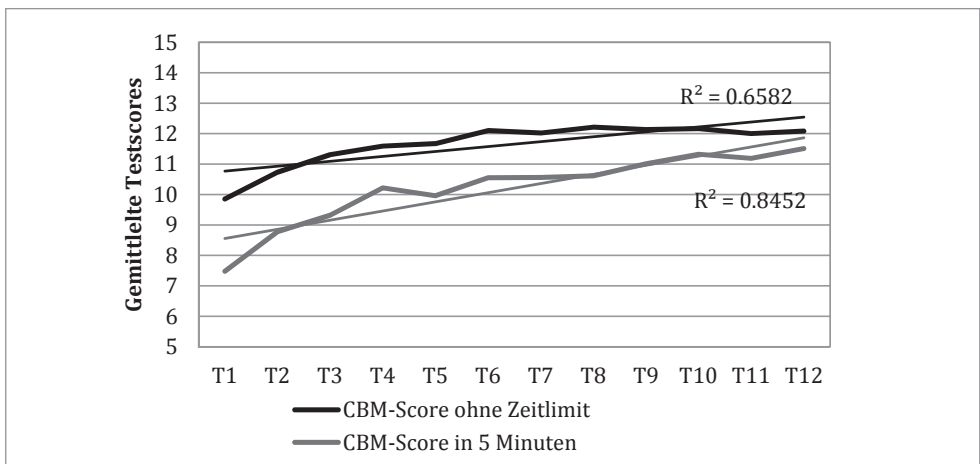


Abbildung 2: CBM-Score mit und ohne Zeitlimit

Tabelle 7: Prüfung von Unterschieden zwischen den Parallelversionen A und B mittels entsprechender t-Tests

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
t	0.72	2.29	0.54	1.68	1.77	2.41	1.92	-0.59	2.26	0.72	2.29	0.54
df	114	126	129	130	110	114	115	122	108	114	126	129
Diff	0.354	1.21*	0.26	0.77	0.78	1.15*	0.89	-0.25	0.98*	0.354	1.21*	0.26

Anmerkung: * – auf 0.05-Niveau signifikant

Tabelle 8: Korrelationen der CBM mit den Werten des HRT 1-4 zum Prä- und Postzeitpunkt. Alle Korrelationen sind signifikant auf dem .001-Niveau

HRT	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
Prä	.51	.54	.36	.44	.42	.44	.45	.43	.51	.51	.48	.50
Post	.53	.55	.49	.53	.52	.56	.54	.52	.54	.60	.57	.51

Um von den Ergebnissen der bisherigen Analysen (CBM *ohne* Zeitvorgabe) auf die CBM *mit* Zeitvorgabe schließen zu können, wurden die Werte der CBM ohne eine Zeitvorgabe mit denen mit Zeitvorgabe korreliert (siehe Tabelle 11). Mit den vorliegenden Werten kann von einem linearen Zusammenhang zwischen den zeitoffenen und -limitierten Werten in allen CBM-Formen ausgegangen werden.

Schwierigkeiten und Trennschärfen. Bei dichotomen Items entspricht der Mittelwert über die Antworten aller Probandinnen und Probanden der Itemschwierigkeit – im Sinne eines Schätzers für die Lösungswahr-

scheinlichkeit. Um für jedes Item eine Aussage treffen zu können, wurden hier alle Lösungen, die ohne Zeitlimit entstanden sind, mit einbezogen. Für jedes der 15 dichotomen Items ergaben sich bei paarweisen Vergleichen zwischen allen zwölf Messzeitpunkten in lediglich 58 Fällen signifikante Mittelwertunterschiede auf dem .05-Niveau. Dies entspricht einer Quote von 5.86%, sodass angenommen werden kann, dass innerhalb eines Items die Schwierigkeiten zwischen den CBM-Formen gleich sind. In Abbildung 3 sind die gemittelten Itemschwierigkeiten für die Aufgabenitems a bis o der 12 CBM-Tests dargestellt. Wie zu se-

Tabelle 9: Inter-Item-Korrelationsmatrix für CBM in den 12 Messzeitpunkten

	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
T1	.62	.61	.56	.56	.55	.56	.49	.51	.53	.55	.57
T2		.71	.71	.65	.67	.65	.59	.65	.67	.62	.57
T3			.79	.76	.77	.74	.75	.73	.73	.72	.66
T4				.75	.75	.76	.67	.67	.76	.72	.72
T5					.78	.79	.75	.75	.77	.74	.72
T6						.79	.72	.75	.78	.75	.72
T7							.67	.77	.79	.72	.71
T8								.74	.80	.78	.74
T9									.83	.78	.77
T10										.79	.78
T11											.80

Tabelle 10: Cronbachs Alpha für jede CBM-Form sowohl für die gesamte Stichprobe als auch für die beiden Versuchsgruppen A und B

Gruppe	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B
Alle	.70	.78	.75	.76	.73	.70	.75	.75	.68	.75	.77	.74
VG A	.48	.66	.70	.60	.66	.52	.62	.77	.47	.68	.71	.61
VG B	.79	.76	.78	.79	.78	.75	.82	.63	.77	.78	.79	.79

Tabelle 11: Korrelation der Ergebnisse nach fünf Minuten mit denen ohne Zeitvorgabe

	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B
r	.82	.87	.85	.82	.86	.84	.83	.93	.83	.87	.89	.85

hen ist, sind die Aufgaben zu Beginn sehr einfach und wurden zum Ende eines jeden CBM-Tests schwieriger. Im Falle der Trennschärfen (korrigierte Item-Skala-Korrelation für das jeweilige Item) ergibt sich ein gegenläufiger Trend. Gemäß den Erwartungen bzgl. der Beziehung zwischen Trennschärfe und Schwierigkeit sind die leichtesten Items am wenigsten trennscharf. Zum Ende eines jeden CBM-Tests steigt die Trennschärfe in einen mittleren Trennschärfebereich.

Aus Tabelle 12 wird ersichtlich, dass sich alle CBM-Formen auf Testebene im leichten bis sehr leichten Bereich bewegen. Alle CBM bewegten sich etwa auf demselben Niveau um $p \approx .75$.

Untersuchung gemäß der Item-Response-Theorie

Da die Ermittlung der Itemparameter auf das Vorhandensein immer gleich vieler beant-

worteter Items aufbaut, konnten für die Analyse lediglich die Fälle mit einbezogen werden, in denen alle Items des jeweiligen CBM beantwortet wurden. Weiter war wegen der gleichzeitigen Berücksichtigung von Itemschwierigkeit und Personenparameter in der Analyse nicht notwendig, zwischen den Versuchsgruppen A und B zu unterscheiden.

In Abbildung 4 sind die Itemparameter aller 12 CBM-Formen zu sehen. Die Itemschwierigkeiten bilden in der Grafik ein enges Band, aus dem in einzelnen Ausnahmefällen Werte herausstechen. Zu erkennen ist, dass die Items in allen CBM schwerer werden. Der Verlauf der Itemparameter verläuft für alle CBM vergleichbar zackig.

In Tabelle 13 sind die Testwerte des Anderson-Likelihood-Ratio-Tests (LRT) (Testwert annähernd χ^2 -verteilt) ersichtlich. Hier zeigt sich in fast allen CBM durch hohe Signifikanzwerte Subgruppeninvarianz.

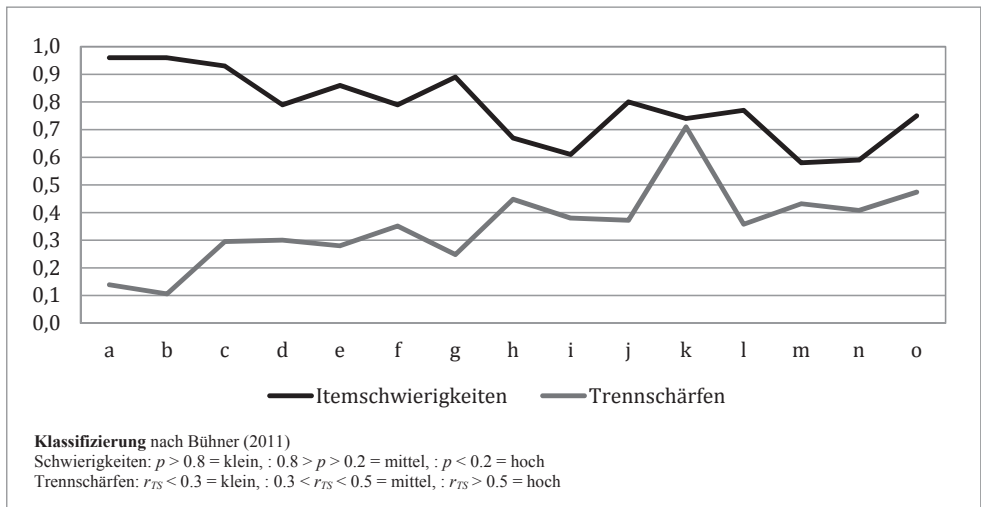


Abbildung 3: Gemittelte Itemschwierigkeiten und Trennschärfen für die Items a bis o über alle 12 CBM-Formen hinweg

Tabelle 12: Schwierigkeitskoeffizienten der 12 CBM-Formen

	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B
Schwierigkeit	.72	.74	.79	.75	.77	.80	.79	.80	.79	.79	.82	.80

Klassifizierung: $p > 0.8$ = klein, $0.8 > p > 0.2$ = mittel, $p < 0.2$ = hoch (Bühner, 2011)

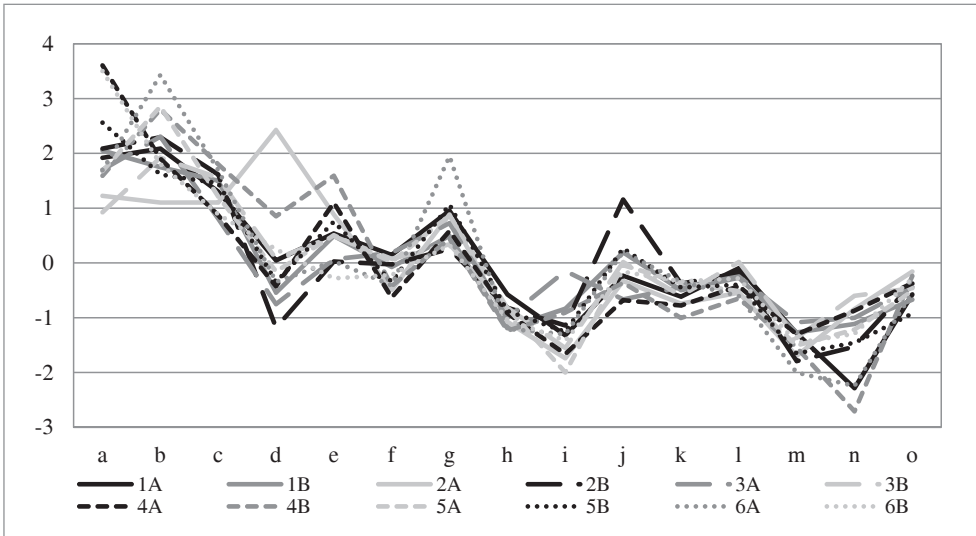


Abbildung 4: Aus der Rasch-Analyse gewonnene Itemparameter β (kleine Werte bedeuten eine hohe Schwierigkeit) für alle 12 CBM

Tabelle 13: Testwerte des LRT zur Bewertung der Subgruppeninvarianz der 12 CBM-Formen

Parameter	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B
Testwert	6.53	7.84	4.33	18.93	10.60	11.78	10.15	8.36	12.96	2.82	10.13	8.47
df	10	12	9	10	14	10	10	7	12	6	8	11
Sig.	.77	.80	.89	.04	.72	.30	.43	.30	.37	.83	.26	.67

Diskussion

Dieser Beitrag verfolgte die Frage, ob das leitfadengestützt entwickelte CBM-Verfahren den hohen Anforderungen wissenschaftlicher Gütekriterien entspricht.

Diskussion der Ergebnisse

Die Auswertung der Ergebnisse anhand des Vorschlags von Wilbert und Linnemann (2011) zeigte, dass es sich bei den entwickelten CBM mit Rückbezug auf die von Fuchs (2004) vorgeschlagene Dreistufung der Überprüfung von CBM auf den ersten beiden Stufen um ein reliables und valides Instrument der Lernverlaufsdiagnostik handelt. Mit Rückbezug auf Fuchs (2004) wurden Überprüfungen auf Ebenen der ersten

beiden Stufen, (1) „Technical features of the static score“ und (2) „Technical features of slope“, vorgenommen. Der Leitfaden zur Erstellung der CBM liefert demzufolge im Rahmen der natürlichen Variabilität gleich schwere und konsistente CBM, die zu unterschiedlichen Zeitpunkten erfolgreich eingesetzt werden konnten. Über die dritte Stufe, (3) „Instructional utility“, kann an dieser Stelle keine Aussage getroffen werden, hier sind unter anderem weiterführende Untersuchungen der Implementationsforschung nötig (vgl. Petermann, 2014).

Allgemein zeigt sich, dass die Schülerinnen und Schüler über die 12 Messzeitpunkte hinweg einen kontinuierlichen Kompetenzzuwachs in den mathematischen Basiskompetenzen verzeichneten. Unterschiede zwischen den jeweiligen Parallelversionen

konnten nicht ausgemacht werden. Bezüglich der Validität konnte zwischen den Werten der CBM und dem standardisierten Außenkriterium ein moderater Zusammenhang festgestellt werden. Hinsichtlich der Reliabilität bilden die 12 eingesetzten CBM mit einem Wert von $\alpha_c = .96$ ein konsistentes Konstrukt zur Lernverlaufsdiagnostik. Die interne Konsistenz der einzelnen CBM beläuft sich auf Werte von $\alpha_c = .68$ bis $.78$. Die Erwartungen hinsichtlich der zunehmenden Schwierigkeit der Items innerhalb der CBM sowie einer adäquaten Trennschärfe der Aufgaben konnten bestätigt werden.

In den Ergebnissen wird deutlich, dass in der Überprüfung der Unterschiede zwischen den Testzeitpunkten Unregelmäßigkeiten aufgetreten sind. Die CBM-Formen 4A, 5A und 6A weisen einen (nicht signifikanten) positiven Unterschied zwischen den Testzeitpunkten auf. Dies ist insofern ungewöhnlich, als dass eigentlich ein negativer Mittelwertunterschied zu erwarten ist. Außerdem ist diese Unregelmäßigkeit in den zeitgleich eingesetzten Formen 4B, 5B und 6B nicht zu verzeichnen. Dieser Sachverhalt erfordert weiterführende Analysen.

Ähnliche Unregelmäßigkeiten konnten auch in der analogen Analyse zu den Messzeitpunkten 2, 6, 9 und 10 verzeichnet werden. Obwohl hier keine Unterschiede vorliegen sollten, konnten zu diesen Zeitpunkten signifikante Unterschiede zwischen den CBM-Formen ermittelt werden. Die redundanten Zeitpunkte 7, 12, 3 und 4 wiesen diese Unterschiede hingegen nicht auf. Ebenso waren entsprechende Mittelwertunterschiede nicht zu verzeichnen, wenn die redundanten Zeitpunkte zusammen auf Unterschiede getestet wurden. Da die CBM selbst innerhalb einer Klasse zufällig nach A- und B-Form verteilt wurden, scheiden hier auch Durchführungseffekte als möglicher Grund aus, womit dessen Ursache bislang nicht weiter ermittelt werden kann.

Die mittleren Korrelationen mit dem HRT 1–4 konnten durch die unterschiedlichen inhaltlichen Konzeptionen der Messinstrumente erklärt werden. Während beim

HRT 1–4 in der Subskala Rechenoperationen auch Aufgaben zu Größenrelationen und Ergänzungsaufgaben erhoben wurden, die in den CBM nicht enthalten sind, prüften die CBM im Gegenzug auch Aufgaben zum Stellenwertsystem, die im HRT 1–4 nicht zu verzeichnen sind. Trotz dieses Unterschieds scheint die Wahl des HRT 1–4 als Bezugsnorm angesichts mangelnder passenderer Instrumente als geeignet, da der HRT 1–4 im Gegensatz zu anderen Instrumenten einerseits entsprechende Rechenoperationen erfasst und zugleich gut im unteren Leistungsbereich differenziert.

Die innere Konsistenz der einzelnen CBM fiel mit $\alpha_c \approx .7$ lediglich in den gerade akzeptablen Bereich. Diese ist nicht als verminderte Konsistenz, sondern vielmehr als konstruktionsbedingt anzusehen. In den CBM wurden unterschiedliche Aufgabentypen behandelt, die alle theoretisch fundiert in dem Bereich der mathematischen Basiskompetenzen zu verorten sind. Dennoch wurden unterschiedliche Rechenarten (z. B. Addition vs. Division) von Schülerinnen und Schülern unterschiedlich gut gelöst und sind damit nicht als untereinander konsistent anzusehen. Der dennoch akzeptable Wert des Cronbachs Alpha deutet darauf hin, dass die CBM insgesamt ein reliables und valides Instrument zur Testung der mathematischen Basiskompetenzen darstellen.

Bezüglich der Itemschwierigkeiten ergaben sich sowohl nach der Analyse in der klassischen Testtheorie als auch im Rasch-Modell sehr hohe Lösungswahrscheinlichkeiten für die ersten beiden Items. Wegen der daraus resultierenden ungünstigen Trennschärfe ist zu prüfen, ob diese Items für ein CBM grundsätzlich geeignet sind. Eine mögliche Ursache kann hier im zur Lerngruppe nicht ganz passend gewählten inhaltlichen Schwierigkeitsgrad der beiden Items liegen.

Diskussion der Methodik

Bei der herangezogenen Stichprobe handelte es sich um eine anfallende, nicht rando-

misierte Stichprobe. Schulen in nahen Einzugsgebieten der Universität sind oftmals mit Forschungsprojekten ausgelastet, sodass gezielt Schulen im Umland angeschrieben wurden. Dabei musste der Faktor „Erreichbarkeit“ gewährleistet bleiben, um eine möglichst problemlose Projektdurchführung zu ermöglichen. Ein weiterer Diskussionspunkt ist die Durchführung der CBM-Erhebungen durch die Mathematiklehrkräfte in den einzelnen Klassen. Zwar erhielten die Lehrkräfte schriftliche Durchführungsanweisungen mit den verbalen Instruktionen, jedoch ist zu berücksichtigen, dass einige Lehrkräfte Unterstützungen oder Hilfen geleistet haben könnten, was jedoch im Rahmen des Projekts nicht kontrolliert wurde. Zudem könnte die Durchführung der Erhebungen durch die Lehrkräfte dazu geführt haben, dass diese gezielt Inhalte noch einmal im Mathematikunterricht wiederholten resp. vertieften, um möglichst gut bei den Erhebungen abzuschneiden. Grafische Darstellungen über die Entwicklungsverläufe erhielten die Lehrkräfte erst nach Projektabschluss, um inhaltliche Reaktionen auf Unterrichtsebene möglichst zu vermeiden.

Um die Zuverlässigkeit und Validität der Prä- und Posterhebungen mittels des HRT 1–4 möglichst hoch zu halten, wurden diese jeweils von denselben Studierenden durchgeführt. Diese wurden im Vorfeld geschult, um eine standardisierte Testung zu gewährleisten. Ebenso positiv hervorzuheben ist, dass die Testungen nicht in Form von Speed-Tests durchgeführt wurden. Die Schülerinnen und Schüler hatten ausreichend Zeit, die CBM auszufüllen; durch den Wechsel der Stiftfarbe konnten die in 5 Minuten gelösten Aufgaben dennoch ermittelt werden. Das Bearbeiten eines Aufgabenblatts beanspruchte ca. 20 Minuten. Der Verzicht auf eine Speed-Testung ermöglichte eine zuverlässige Berechnung von Schwierigkeitsindizes. Schwierige Aufgaben wären vermutlich wegen der zeitlichen Begrenzung nur von wenigen Schülerinnen und Schülern bearbeitet worden. Für den Einsatz in der Praxis wird ein Zeitaufwand

von drei bis fünf Minuten empfohlen, der hier eingehalten wurde.

Forschungsausblick

Der dritten Stufe zur Erstellung von CBM nach Fuchs (2004) (Untersuchung der Einsetzbarkeit und Effektivität) wurde in dem vorliegenden Beitrag nicht nachgegangen.

Es schließen sich somit weiterführende Fragestellungen bezüglich der Untersuchung der Implementationsqualität an, um der Forderung nach evidenzbasierter Praxis zu entsprechen: Wie viel Zeit benötigen Lehrkräfte, um CBM zu erstellen? Wie hoch ist die Hemmschwelle für den Einsatz im eigenen Unterricht? Entsprechen auch die CBM der Lehrkräfte den Ansprüchen der Güte?

Zudem sollte in weiterführenden Analysen oder Replikationen Unregelmäßigkeiten zwischen den Testzeitpunkten (CBM-Formen 4A, 5A, 6A sowie 4B, 5B, 6B) nachgegangen werden. Ebenso sollte die Schwierigkeit der Items überdacht werden, da sowohl die Klassische Testtheorie als auch die Ergebnisse der Rasch-Modellierung hohe Lösungswahrscheinlichkeiten mit ungünstigen Trennschärfen der Items anzeigten. Zukünftig gilt es auch, die prognostische Validität zu überprüfen, um Aussagen darüber treffen zu können, ob es sich bei den vorliegenden CBM um ein Instrument handelt, das dazu geeignet ist, frühzeitig Lernrückstände zu identifizieren. Hierfür gilt es, weitere CBM nach dem vorliegenden Leitfaden zu entwickeln und über einen längeren Zeitraum einzusetzen. Somit können auch Änderungsindizes ermittelt und darauf basierend geeignete Auswertungsstrategien für Lehrkräfte formuliert werden.

Fazit

Die Ergebnisse der Überprüfung von CBM mittels der Klassischen Testtheorie sowie der Rasch-Modellierung zeigten, dass es sich bei den leitfadengestützt konstruierten CBM um ein reliables und valides Instru-

ment der Lernverlaufsdiagnostik für das Fach Mathematik in der vierten Jahrgangsstufe handelt. Nach einer noch anstehenden positiven Evaluation der Implementation in die Praxis (Schritt 3 nach Fuchs, 2004), könnten Lehrkräfte zukünftig leitfadengestützt CBM für ihren eigenen Mathematikunterricht erstellen. Aus der Überprüfung des curriculumbasierten Messverfahrens in der vorliegenden Untersuchung ergeben sich jedoch vorerst weitere Forschungsnotwendigkeiten.

Bei dem vorliegenden Verfahren zur Lernverlaufsdiagnostik handelt es sich zuletzt um einen weiteren wichtigen Beitrag zur deutschen CBM-Forschung, die gegenüber dem Forschungsstand in den USA bislang noch in den Kinderschuhen steckt (Voß & Hartke, 2014).

Literaturverzeichnis

- Birkel, P. (2007). *Weingartener Grundwortschatz Rechtschreib-Test für dritte und vierte Klassen : WRT 3+ (2., neu normierte Auflage)*. PSYTEST, Göttingen [u.a.]: Hogrefe.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. & Walther, G. (2007). Erste Ergebnisse aus IGLU : Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. *Sozialwissenschaftlicher Fachinformationsdienst soFid*, (1), 9–46.
- Brickenkamp, R. (2002). *Test d2 : Aufmerksamkeits-Belastungs-Test : Manual*. PSYTEST, Göttingen [u.a.]: Hogrefe.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. (3., aktualisierte Auflage). München: Pearson
- Casale, G., Hennemann, T. & Grosche, M. (2015). Zum Beitrag der Verlaufsdiagnostik für eine evidenzbasierte sonderpädagogische Praxis am Beispiel des Förderschwerpunkts der emotionalen und sozialen Entwicklung. *Zeitschrift für Heilpädagogik*, 66(7), 325–334.
- Deno, S. L. (1985). Curriculum-Based Measurement: The Emerging Alternative. *Exceptional Children*, 52(3), 219–232.
- Deno, S. L. (1993). Curriculum-Based Measurement. In J. J. Kramer (Hrsg.), *Curriculum-based measurement* (S. 1–23). Lincoln, Neb: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.
- Diaz-Bone, R. (2006). *Statistik für Soziologen*. Konstanz: UVK Verlagsgesellschaft.
- Diehl, K., Hartke, B. & Knopp, E. (2009). Curriculum-Based Measurement & Leerlingonderwijsvolgysteem – Konzepte zur theoriegeleiteten Lernfortschrittsmessung im Anfangsunterricht Deutsch und Mathematik? *Zeitschrift für Heilpädagogik*, (4), 122–130.
- Fuchs, L. S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research. *School Psychology Review*, 33(2), 188–192.
- Fuchs, L. S. & Fuchs, D. (1986). Effects of Systematic Formative Evaluation: A Meta-Analysis. *Exceptional Children*, 53(3), 199–208.
- Gebhardt, M., Diehl, K. & Mühling, A. (2016). Online-Lernverlaufsmessung für alle Schülerinnen und Schüler in inklusiven Klassen. *Zeitschrift für Heilpädagogik*, 10, 444–453.
- Gebhardt, M. & Mühling, A. (in Planung). Über LeVuMi. Zugriff am 16.01.2017. Verfügbar unter <http://www.levumi.de/?page=about>.
- Gölitz, D., Roick, T. & Hasselhorn, M. (2006). *Deutscher Mathematiktest für vierte Klassen (DEMAT 4)*. Göttingen [u.a.]: Hogrefe.
- Haffner, J., Baro, K., Parzer, P. & Resch, F. (2005). *HRT 1-4. Heidelberger Rechen-test. Erfassung mathematischer Basiskompetenzen im Grundschulalter*. Göttingen [u.a.]: Hogrefe.
- Hartke, B. & Diehl, K. (2013). *Schulische Prävention im Bereich Lernen. Problemlösungen mit dem RTI-Ansatz*. Stuttgart: Kohlhammer.

- Hartmann, E. & Müller, C. (2014). *Lernfortschrittsdiagnostik Grundrechenarten*. Hamburg: Persen.
- Hillenbrand, C. (2015). Evidenzbasierung sonderpädagogischer Praxis: Widerspruch oder Gelingensbedingung. *Zeitschrift für Heilpädagogik*, 66(7), 312–324.
- Horn, W. & Lukesch, H. (2002). *Prüfsystem für Schul- und Bildungsberatung für 4. Bis 6. Klassen : PSB-R 4-6* (Rev. Fassu.). PSYTEST, Göttingen [u.a.]: Hogrefe.
- Hosp, M. K., Hosp, J. & Howell, K. W. (2007). *The ABCs of CBM. A practical guide to curriculum-based measurement*. New York: Guilford Press.
- Humbach, M. (2008). *Arithmetische Basis-kompetenzen in der Klasse 10*. Berlin: Dr. Köster.
- Humbach, M. (2009). Arithmetisches Basiswissen in der Jahrgangsstufe 10. In A. Fritz & S. Schmidt (Hrsg.), *Fördernder Mathematikunterricht in der Sek. I* (S. 58–72). Weinheim, Basel: Beltz Verlag.
- Käter, C., Käter, T., Martenstein, R. & Hillenbrand, C. (2016). Leitfadengestützte Konstruktion eines Instruments der Lernverlaufdiagnostik (CBM) im Bereich Lesen. *Zeitschrift für Heilpädagogik*, 66, 168–179.
- Klauer, K. J. (1994). *Diagnose- und Förderblätter. Rechenfertigkeiten im 4. Schuljahr*. Berlin: Cornelsen.
- Klauer, K. J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufdiagnostik: Neue Folge Band 12* (1. Auflage, S. 1–17). Göttingen: Hogrefe Verlag.
- Knopp, E. & Hartke, B. (2010). Das Inventar Rechenfische – Anwendung, Reliabilität und Validität eines Verfahrens zur Erfassung des Leistungsstandes von Erstklässlern in Mathematik. *Empirische Sonderpädagogik*, 2(3), 5–25.
- Krauthausen, G. & Scherer, P. (2007). *Einführung in die Mathematikdidaktik* (3. Auflage.). Heidelberg, Berlin: Spektrum.
- Križan, A. & Vossen, A. (2016). Zeitschrift für Heilpädagogik. *Zeitschrift für Heilpädagogik*, 62, 419–431.
- Kultusministerkonferenz (2005). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4)*. (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, Hrsg.). München [u.a.]: Luchterhand.
- Kultusministerkonferenz (2015). *Empfehlungen zur Arbeit in der Grundschule*. Abgerufen von http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1970/1970_07_02_Empfehlungen_Grundschule.pdf
- Küspert, P. & Schneider, W. (1998). *Würzburger Leise-Leseprobe : (WLLP)*. PSYTEST, Göttingen [u.a.]: Hogrefe.
- Limbach-Reich, A. (2015). „Response to Intervention“ (RTI) im Spannungsfeld Inklusiver Diagnostik. In H. Schäfer & C. Rittmeier (Hrsg.), *Handbuch Inklusive Diagnostik* (S. 478–495). Weinheim, Basel: Beltz.
- Maier, U. (2010). Formative Assessment – Ein erfolgsversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13, 293–308.
- Moser Opitz, E. (2009). Erwerb grundlegender Konzepte der Grundschulmathematik als Voraussetzung für das Mathematiklernen in der Sekundarstufe I. In A. Fritz & S. Schmidt (Hrsg.), *Fördernder Mathematikunterricht in der Sek. I. Rechenschwierigkeiten erkennen und überwinden* (S. 29–45). Weinheim, Basel: Beltz Verlag.
- Müller, C. M. & Hartmann, E. (2009). Lernfortschritte im Unterricht erheben – Möglichkeiten und Grenzen des curriculum-basierten Messens. *Schweizerische Zeitschrift für Heilpädagogik*, (10), 36–42.
- Ophuysen, S. van & Harazd, B. (2011). Der Übergang von der Grundschule zur weiterführenden Schule – Gestaltung, Beratung, Diagnostik. *Handreichungen des Programms SINUS an Grundschulen*.

- Peter-Koop, A., Hasemann, K. & Klep, J. (2006). *SINUS-Transfer Grundschule Mathematik Modul G10: Übergänge gestalten*.
- Petermann, F. (2010). Veränderungsmessung. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4., überarbeitete Auflage, S. 896–902). Weinheim: Beltz.
- Petermann, F. (2014). Implementationsforschung: Grundbegriffe und Konzepte. *Psychologische Rundschau*, 65(3), 122–128.
- Racherbäumer, K. & Kohnen, M. (2014). Schulstufenübergreifende Kooperationen von Lehrkräften der Primar- und Sekundarstufe zur Gestaltung anschlussfähiger Bildungsprozesse am Beispiel Naturwissenschaften und Mathematik. In A. Liegmann, I. Mammes & K. Racherbäumer (Hrsg.), *Facetten von Übergängen im Bildungssystem* (S. 95–109). Münster: Waxmann.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Verlag Hans Huber.
- Schmitt, R. (2001). *Grundlegende Bildung in und für Europa*. Frankfurt am Main: Grundschulverband – Arbeitskreis Grundschule e. V.
- Schneider W. & Treiber B. (1986). Determinanten der Leistungsveränderung im Mathematikunterricht: Ein Modellvergleich in unterschiedlichen Schulklassentypen. In C. Möbius & W. Schneider (Hrsg.), *Strukturmodelle für Längsschnittdaten und Zeitreihen* (S. 209–220) Stuttgart: Huber.
- Schumacher, M. & Schulgen, G. (2009). *Methodik klinischer Studien. Methodische Grundlagen der Planung, Durchführung und Auswertung* (3., überarbeitete Auflage). Berlin, Heidelberg: Springer.
- Souvignier, E., Förster, N. & Salaschek, M. (2014). quop: Ein Ansatz internetbasierter Lernverlaufsdiagnostik mit Testkonzepten für Lesen und Mathematik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik: Neue Folge Band 12* (1. Auflage, S. 239–256). Göttingen: Hogrefe Verlag.
- Stern, E. (2008). Verpasste Chancen? Was wir aus der LOGIK-Studie über den Mathematikunterricht lernen können. In W. Schneider (Hrsg.), *Entwicklung von der Kindheit bis zum Erwachsenenalter: Befunde der Münchner Längsschnittstudie LOGIK* (1. Auflage, S. 187–202). Weinheim, Basel: Beltz Verlag.
- Strathmann, A. M. (2014). Lernverlaufsdiagnostik Mathematik für zweite bis vierte Klassen (LVD-M). In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik: Neue Folge Band 12* (1. Auflage, S. 203–220). Göttingen: Hogrefe Verlag.
- Strathmann, A. M. & Klauer, K. J. (2010). Lernverlaufsdiagnostik: Ein Ansatz zur längerfristigen Lernfortschrittsmessung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42(2), 111–122.
- Strathmann, A. M. & Klauer, K. J. (2012). *LVD-M2-4. Lernverlaufsdiagnostik – Mathematik für zweite bis vierte Klassen. Manual*. Göttingen [u.a.]: Hogrefe.
- Tucker, J. A. (1985). Curriculum-Based Assessment: An Introduction. *Exceptional Children*, 52(3), 199–204.
- Voß, S. (2014). *Curriculumbasierte Messverfahren im mathematischen Erstunterricht. Zur Güte und Anwendbarkeit einer Adaption US-amerikanischer Verfahren im deutschen Schulsystem*. Saarbrücken: SHV.
- Voß, S. & Hartke, B. (2014). Curriculumbasierte Messverfahren (CBM) als Methode der formativen Leistungsdiagnostik im RTI-Ansatz. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik: Neue Folge Band 12* (1. Auflage, S. 83–100). Göttingen: Hogrefe Verlag.
- Waldmann, H.-C. & Petermann, F. (2014). Veränderungsmessung – Methodische Vorschläge für Forschung und klinische Praxis. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 62(2), 85–92.

- Wesson, C. L., King, R. P. & Deno, S. L. (1984). Direct and Frequent Measurement of Student Performance: If It's Good for Us, Why Don't we Do It? *Learning Disability Quarterly*, (7), 45–48.
- Wilbert, J. (2014). Vermittlung von Basiskompetenzen zum Rechnen. In G. W. Lauth, M. Grünke & J. C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen: Förderung, Training und Therapie in der Praxis* (2., überarbeitete Auflage, S. 209–2199). Göttingen, Bern, Toronto, Seattle: Hogrefe.
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik*, (3), 225–242.
- Yell, M., Deno, S. L. & Marston, D. E. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique*, 18, 99–105.
- Julia Rensing (M. Ed.)**
Carl von Ossietzky Universität
Oldenburg
Institut für Sonder- und
Rehabilitationspädagogik
26129 Oldenburg
j.rensing@uni-oldenburg.de
- Carolina Käter (M. Ed.)**
Carl von Ossietzky Universität
Oldenburg
Institut für Sonder- und
Rehabilitationspädagogik
26129 Oldenburg
- Tobias Käter (M. Ed.)**
Carl von Ossietzky Universität
Oldenburg
Institut für Sonder- und
Rehabilitationspädagogik
26129 Oldenburg
- Clemens Hillenbrand**
Carl von Ossietzky Universität
Oldenburg
Institut für Sonder- und
Rehabilitationspädagogik
26129 Oldenburg

Erstmalig eingereicht: 09.08.2016

Überarbeitung eingereicht: 17.01.2017

Angenommen: 25.01.2017