

*Empirische Sonderpädagogik*, 2017, Nr. 2, S. 184-194  
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

## *Kurzbeitrag*

# Vorschlag zur Konzeption eines curriculumbasierten Messverfahrens zur Erfassung der Rechtschreibleistungen im Grundschulbereich

*Stefan Voß<sup>1</sup>, Simon Sikora<sup>1</sup> & Kathrin Mahlau<sup>2</sup>*

<sup>1</sup> *Universität Rostock*

<sup>2</sup> *Universität Greifswald*

## **Zusammenfassung**

Dass die Konzeption und Evaluation von Verfahren zur Lernverlaufsdiagnostik, insbesondere auch curriculumbasierten Messverfahren (CBM; curriculum-based measurements; Deno, 1985, 2003), nicht einfach ist, zeigen die im Fachdiskurs aufgestellten Forschungskriterien (u. a. Fuchs, 2004; Wilbert & Linnemann, 2011) sowie die vergleichsweise geringe Dichte an existierenden Instrumenten. Für den Kompetenzbereich Rechtschreiben liegen derzeit noch keine deutschsprachigen CBM vor. Vorarbeiten wie die von Strathmann und Klauer (2008) und Strathmann, Klauer und Greisbach (2010) haben auf verschiedene Herausforderungen bei der Entwicklung von CBM für das Feld der Rechtschreibung aufmerksam gemacht.

Im vorliegenden Beitrag wird eine Möglichkeit vorgestellt, curriculumbasierte Messverfahren zur formativen Evaluation rechtschriftlicher Kompetenzen von Grundschülerinnen und -schülern zu konstruieren. Dabei werden sowohl inhaltliche als auch psychometrische Kriterien berücksichtigt. Der Beitrag versteht sich als eine Art „Grobskizze“ zur Visualisierung einer Möglichkeit zur Itemgewinnung und -pilotierung und kann – in angepasster Form – auch auf weitere schulische Inhaltsfelder übertragen werden.

Schlüsselwörter: Formative Evaluation, CBM, Rechtschreibung, Item-Response-Theory

## **A suggestion for developing curriculum-based measurements to assess spelling skills at elementary schools**

### **Abstract**

The conceptualization and evaluation of progress monitoring instruments regarding learning outcomes – especially curriculum-based measurements (CBM, Deno, 1985, 2003) – is not easy. This is recognizable by the extensive research criteria set out in the discourse as well as the comparatively low density of existing instruments. There is currently no German CBM available for spelling. Preliminary work such as that of Strathmann and Klauer (2008) and Strathmann, Klauer and Greisbach (2010) have drawn attention to various challenges in the development of CBM in this field.

In this paper we present a possibility to construct curriculum-based measurement methods for the formative evaluation of the spelling competences of primary school students. Both contentual and psychometric criteria are considered. The study shows a possible way of item construction and can be transferred - in an adapted form - to other areas of schooling.

Key words: formative evaluation, CBM, spelling, Item-Response-Theory

## Einleitung

Ziel von Verfahren zur formativen Evaluation von Unterricht ist es, durch regelmäßig wiederholte Leistungserfassungen festzustellen, ob die Lernentwicklung der Schülerinnen und Schüler erwartungsgemäß verläuft bzw. die gewählten pädagogischen Maßnahmen in ausreichendem Maße erfolgreich sind. Dies ermöglicht eine Steigerung der schulischen Leistungen (u. a. Hattie, 2013), ein zeitnahes Erkennen von Lernschwierigkeiten sowie eine produktive Kommunikation zwischen Lehrkräften, Eltern sowie Schülerinnen und Schülern. Der Lernverlaufsdiagnostik ist im deutschsprachigen Forschungsraum im vergangenen Jahrzehnt viel Beachtung geschenkt worden. Im Fokus stehen v. a. die Konzeption und Überprüfung von Verfahren zur formativen Evaluation im Bereich des Lernens (Hasselhorn, Schneider & Trautwein, 2014). Insbesondere die Entwicklung von sog. curriculumbasierten Messverfahren (CBM; curriculum-based measurements; Deno, 1985, 2003; Fuchs, 2004), die sich durch ihre Orientierung an wissenschaftlichen Gütekriterien auszeichnen, wurde vorangetrieben (z. B. Souvignier, Förster & Salaschek, 2014; Strathmann & Klauer 2010; Strathmann & Klauer, 2012; Voß, 2014; Walter, 2010).

Dass der Ansatz der Lernverlaufsdiagnostik auch im Bereich der Rechtschreibung von großer Bedeutung ist, zeigen unterschiedliche Befunde:

1. Erkenntnisse einer interdisziplinär ausgerichteten Schriftspracherwerbsforschung verweisen darauf, dass das Erlernen orthografischer Kompetenzen ein aktiver, kognitiv basierter Konstruktionsprozess ist. Dabei zeigen sich sehr un-

terschiedliche, von verschiedenen Faktoren abhängige Entwicklungsverläufe zwischen den Kindern einer Altersgruppe (Hanke & Schwippert, 2005). In deren Konsequenz sehen sich Lehrkräfte einer großen Bandbreite von Lernausgangslagen sowie -verläufen gegenüber, denen es im Unterricht zu entsprechen gilt.

2. Vor dem Hintergrund von Studien zur Urteilsgenauigkeit von Lehrkräften (zusammenfassend: Hoge & Coladarci, 1989; Südkamp, Kaiser & Möller, 2012) muss angezweifelt werden, dass es allen Praktikerinnen und Praktikern hinreichend gut gelingt, die individuellen Lernvoraussetzungen ihrer Schülerinnen und Schüler präzise einzuschätzen – und dies vor allem im zeitlichen Verlauf. Diese Fähigkeit kann jedoch als eine notwendige Voraussetzung für eine passgenaue, individualisierte Förderung rechtschriftlicher Kompetenzen angesehen werden.

In diesem Spannungsfeld wird das Potenzial einer engmaschigen Lernverlaufsdiagnostik mittels curriculumbasierter Messverfahren betont (Deno, 1985; Klauer, 2006; Voß & Hartke, 2014; Walter, 2008). Allerdings liegen bisher keine entsprechenden Verfahren für den Bereich der Rechtschreibung im deutschen Sprachraum vor. Der Grund dafür liegt sicherlich darin, dass die Konzeption und Evaluation von Verfahren zur Lernverlaufsdiagnostik generell nicht einfach ist, wie verschiedene im Fachdiskurs aufgestellte Forschungskriterien es erahnen lassen (Fuchs, 2004; Klauer, 2011; Wilbert & Linnemann, 2011). Zudem weisen bereits bestehende Vorarbeiten auf Schwierigkeiten bei der Entwicklung von CBM speziell im

Bereich der Rechtschreibung hin (Strathmann & Klauer, 2008; Strathmann, Klauer & Greisbach, 2010). So berichten Strathmann et al. (2010) von unbefriedigenden Reliabilitätswerten und Problemen, Paralleltests zu konstruieren. Insofern liegt hier ein großes Forschungsdesiderat vor.

Im Folgenden wird eine Möglichkeit vorgestellt, die das Potenzial hat, die skizzierten Schwierigkeiten bei der Konstruktion von CBM zur formativen Erfassung rechtsschriftlicher Kompetenzen von Grundschülerinnen und -schülern zu überwinden. In Anlehnung an Deno (2003) sind dabei zwei Fragestellungen handlungsleitend: 1. Was soll gemessen werden? 2. Wie soll gemessen werden? Während sich die erste Frage insbesondere mit dem zu messenden Konstrukt, d. h. mit Aspekten der Validität befasst, zielt die zweite Frage vordergründig auf die Messgenauigkeit des zu prüfenden Instrumentes ab. Insofern wird nachfolgend in einem ersten Schritt die Operationalisierung der Items beschrieben. Dann werden die derart gewonnenen Items in einem zweiten Schritt im Rahmen einer empirischen Studie auf ihre psychometrische Eignung untersucht.

## Was soll gemessen werden?

### Generierung eines validen

### Itempools

Zunächst stellt sich die Aufgabe, einen für das Rechtschreib-Curriculum der Grundschule möglichst geeigneten Wortpool abzubilden. Dies ist eine große Herausforderung, da nach dem Duden der Alltagswortschatz eine halbe Million Wörter umfasst, von denen man ca. 70.000 häufig verwendet (Best, 2000). Schulanfängerinnen und -anfänger verstehen in aller Regel den Alltagswortschatz und verwenden mehrere tausend Wörter verbalsprachlich (Grimm, 2003). Um die notwendige Eingrenzung des Wortumfangs vorzunehmen, erfolgte eine vergleichende Analyse der nachfolgend aufgeführten bildungsministeriellen Emp-

fehlungen, anerkannten fachwissenschaftlichen Modellen und einschlägigen Arbeitsmaterialien. Die Bewertungen der Analysen zeigen, dass es nur wenige Anhaltspunkte für die Definition eines Wortschatzkorpus' für den Bereich der Rechtschreibung in der Primarstufe gibt.

#### a) Empfehlungen der Kultusministerkonferenz (2004)

Die Kultusministerkonferenz (KMK, 2004) hat für alle deutschen Bundesländer verbindliche Bildungsstandards für den Grundschulbereich im Fach Deutsch vorgelegt, die seit dem Schuljahr 2005/2006 sowohl die Basis für den Deutschunterricht der Grundschulstufe als auch für bundesweite Vergleichsarbeiten (VERA) bilden (Bremerich-Vos, Granzer, Behrens & Köller, 2011). Sie beinhalten die Ziele im Fach Deutsch, die bis zum Ende der 4. Jahrgangsstufe erreicht werden sollen. Am Ende der vierten Klasse sollen die Schülerinnen und Schüler geübte, rechtschreibwichtige Wörter normgerecht schreiben, grundlegende Rechtschreibstrategien verwenden und die Zeichensetzung beachten. Die Bildungsstandards explizieren jedoch nicht, welche Wörter „rechtschreibwichtig“ oder welche Strategien „grundlegend“ sind.

#### b) Rahmenrichtlinien Deutsch verschiedener Bundesländer

Gemäß den Rahmenrichtlinien für den Deutschunterricht der Grundschule verschiedener Bundesländer (LISUM, 2004) sollen die Schülerinnen und Schüler bis zum Ende der Grundschulzeit folgende Kompetenzen im Bereich der Rechtschreibung erreichen. Sie sollen u. a.

- Wörter des Übungswortschatzes vollständig sowie eigene Texte überwiegend richtig schreiben,
- die Rechtschreibstrategien mitsprechen, ableiten und einprägen und
- Arbeitstechniken und Rechtschreibhilfen (Duden, Lexika) nutzen.

Es ist ersichtlich, dass die Inhalte der Kompetenzbereiche der Rahmenrichtlinien sehr offen gehalten werden und keine konkreten Anhaltspunkte dafür geben, in welcher Klassenstufe welche Rechtschreibkompetenz erworben sein sollte.

### c) Grundwortschatzlisten

Grundwortschatzlisten weisen diejenigen Wörter aus, die in der Alltagssprache häufig genutzt bzw. geschrieben werden. Von den Bildungsministerien der Bundesländer Bayern, Berlin, Brandenburg und Mecklenburg-Vorpommern wurden entsprechende Aufstellungen von 700 bis 752 Wörtern veröffentlicht (LISUM, 2011; Ministerium für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern, 2016; Senatsverwaltung für Bildung, Jugend und Wissenschaft, 2011; Staatsinstitut für Schulqualität und Bildungsforschung München, 2012). Es werden die ca. 100 häufigsten Wörter jahrgangsstufenübergreifend herausgestellt, zudem weitere 300 bis 400 orthografisch bedeutsame Wörter jeweils für die Klassenstufen 1 und 2 sowie 3 und 4.

Dass es für die Konzeption eines CBM im Bereich der Rechtschreibung über die Grundschulzeit nicht ausschließlich genügt, einen validen Wortpool auszuarbeiten, aus welchem jeweils Zufallsstichproben gezogen werden, zeigen die bereits veröffentlichten Vorarbeiten von Strathmann und Klauer (2008) oder Strathmann et al. (2010). Ein derartiges Vorgehen resultiert in unterschiedlich schwierigen und nicht hinreichend reliablen Testvorlagen. Zufallsstichproben aus einem sehr großen Wortschatzpool ohne weitere Systematisierung – z. B. nach Klassenstufen oder auch Rechtschreibstrategien und -kompetenzen – zu ziehen, lässt für eine curriculumbasierte Messung vermutlich zu viel Spielraum. Die Möglichkeit, dass bestimmte Rechtschreibstrategien noch nicht im Unterricht behandelt und gefestigt wurden, zeigt sich in der Analyse verschiedener Lehrwerke. Die Stratifizierung, also eine Unterteilung der gegebenen Wör-

ter in Kategorien nach unterschiedlichen Kriterien, könnte hierbei Abhilfe schaffen. Entsprechend ist zu überlegen, welche Kriterien zur Systematisierung eines Wortpools geeignet erscheinen.

Eine Unterteilung der Wörter nach ihrer zeitlichen Verortung in der Grundschulzeit stellt diesbezüglich ein sinnvolles Kriterium dar. Nicht nur, dass hierbei ebenfalls inhärent nach unterschiedlichen Rechtschreibphänomenen (eher lautgetreue Wörter in Klassenstufe 1, phonologische und morphologische Strategien wie *st*, *sp* oder Komposita in Klasse 2 etc.) sortiert wird, zudem wird ein hieran orientiertes CBM auch eher dem Anspruch der Curriculumbasierung gerecht. Für die notwendige Eingrenzung des schriftsprachlichen Wortschatzkorpus' nach Klassenstufen ist eine Orientierung an aktuell gültigen Curricula – wie oben ausgeführt – nicht zielführend. Möglichkeiten zur Verortung der Wörter auf die Schuljahre stellen 1) eine Analyse einschlägiger Rechtschreib-Lehrgänge und 2) eine Zuordnung zu Kompetenzprofilen dar, die sich an anerkannten Entwicklungsmodellen orientieren (z. B. das „Kompetenzprofil Rechtschreiben“ von Reber und Kirch, 2013).

#### 1) Analyse ausgewählter Sprachbücher für die Grundschule

Um genauere Hinweise für eine sinnvolle Abfolge der eingeführten Rechtschreibphänomene zu erhalten, können die in einschlägigen Lehrwerken aufgeführten Rechtschreibstrategien und -phänomene den jeweiligen Klassenstufen zugeordnet werden. Eine solche Analyse macht deutlich, dass im Wesentlichen ein stufenweiser Aufbau in den Lehrwerken zu erkennen ist, aber auch, dass kein einheitliches Vorgehen beim Einführen und Festigen von Rechtschreibstrategien besteht. Die Lehrwerke führen einige der Strategien um bis zu zwei Jahre versetzt ein (z. B. die phonologische Strategie *ng*, *nk*). Es ist also nicht davon auszugehen, dass in allen bundesdeutschen Grundschulen bestimmte Rechtschreibstrategien und

-phänomene in derselben Klassenstufe behandelt werden.

## 2) Einordnung der Items in das „Kompetenzprofil Rechtschreiben“ (Reber & Kirch, 2013)

Das Kompetenzprofil Rechtschreiben gibt ein Raster vor, mithilfe dessen die rechtschriftlichen Merkmale der Wörter sich der alphabetischen, phonologischen, morphologischen, orthografischen oder grammatischen Strategie zuordnen lassen. Diese Strategien differenzieren sich sehr detailliert in 60 weitere Rechtschreibkompetenzen (z. B. Verdopplungen, Dehnungen, Auslautverhärtung usw.) aus. Auf diese Weise können Wörter mit gleichen orthografischen Merkmalen und Graphemanzahlen gefunden werden. Dies folgt der Annahme, dass Wörter mit ähnlicher Anforderungsstruktur bestimmt werden können, was die Grundlage für parallele CBM darstellt.

### Fazit zur Generierung eines Itempools

Zur Definition eines hinsichtlich der Rechtschreibung möglichst vollständigen, in mehreren Bundesländern gültigen Wortschatzes wurde in der vorliegenden Studie ein Wortschatzkorpus aus den oben genannten Grundwortschatzlisten und einigen Wörtern gängiger Lehrwerke gebildet. Die Bildungsstandards und die Rahmenpläne fanden insofern keine Berücksichtigung, als dass sie keine konkreten Hinweise auf Inhalte eines Wortschatzes gaben. Auf diese Weise konnten insgesamt 808 Wörter für den Grundschulbereich bestimmt werden. Diese wurden in das beschriebene Kompetenzprofil von Reber und Kirch (2013) eingeordnet. Dabei erfolgte die Einsortierung aller 808 Wörter hinsichtlich der zur korrekten Schreibung notwendigen Rechtschreibkompetenzen (alphabetische, phonologische, morphematische, orthografische und grammatische Strategien) nach 55 unterschiedlichen Kriterien. Dieser Wortschatz

ist Ausgangspunkt zur Entwicklung der hier vorgestellten Rechtschreib-CBM.

Es soll nicht unerwähnt bleiben, dass bei diesem Vorgehen auch verschiedene Schwierigkeiten auftreten. So kann ein Wort teilweise mehrere Rechtschreibphänomene enthalten, die unterschiedlichen Strategien und Kompetenzen zugeordnet werden können (z. B. „flüssig“ – morphologische Strategie [Auslautverhärtung] und orthografische Strategie [Doppelkonsonanz]). Weiterhin ist auch die Zuordnung einzelner Phänomene diskussionsbedürftig (Ist das Wort „putzen“ lautgetreu, das Wort „Wurzel“ aber nicht?). Zur Lösung dieser Schwierigkeiten müssen klare Regeln festgelegt werden. So sind Wörter mit mehreren Strategien und Kompetenzen grundsätzlich der höherwertigen Strategie zuzuordnen. Wörter, in denen alle gesprochenen Laute eindeutig Graphemen zuzuordnen und zu verschriftlichen sind, sind lautgetreu. Wörter, die Phänomene enthalten, die beim Sprechen zwar hörbar sind, aber nicht geschrieben werden, sind nicht lautgetreu, sondern werden einer höheren Rechtschreibstrategie zugeordnet.

### Wie wird gemessen?

Nachdem geklärt wurde, was gemessen werden soll, geht es in einem nächsten Schritt um die Frage, wie man zu Verfahren kommt, mit denen Lernverläufe abgebildet werden können. Hierzu wurde eine empirische Studie durchgeführt, die nachfolgend kurz beschrieben werden soll.

### Methode

#### Stichprobe

Insgesamt haben 4091 Kinder aus 192 ersten bis vierten Klassen von 24 Schulen Mecklenburg-Vorpommerns an der Erhebung teilgenommen. Aus den in die Untersuchung einbezogenen Grundschulen haben alle Klassen die Wortdiktate durchge-

führt. Es wurden Grundschulen in ländlichen, kleinstädtischen sowie städtischen Regionen berücksichtigt. Die genaue Verteilung in den unterschiedlichen Klassenstufen ist Tabelle 1 zu entnehmen.

### Vorgehen

Die ermittelten Wörter für die zu erstellenden CBM wurden mittels eines Multi-Matrix-Designs (Mislevy, Beaton, Kaplan & Sheehan, 1992) jeweils getrennt nach Klassenstufe auf verschiedene Wortlisten verteilt und pilotiert. Die Durchführung der Wortdiktate in den Klassen erfolgte durch die unterrichtende Lehrkraft zum Ende des Schuljahres 2015/16. Ausgewertet wurden die Arbeitsergebnisse durch geschulte studentische Hilfskräfte, die dokumentierten, ob ein Wort richtig oder falsch geschrieben wurde und welche Fehlschreibweisen auftraten.

Aufgrund des Multi-Matrix-Designs verfügt jede Liste über einen Anteil identi-

scher Wörter (sog. Ankeritems) innerhalb einer Klassenstufe und jeweils aufeinanderfolgender Klassenstufen, über welche später ein übergreifend verlinkter Datensatz generiert wurde. Auf Basis der Item-Response-Theorie (IRT) konnten dann psychometrische Kennwerte für alle Items der Gesamtstichprobe ermittelt werden. Hierzu wurde ein eindimensionales Raschmodell geschätzt. Die statistischen Analysen wurden mit dem Programm R (R Core Team, 2013) und dem Zusatzpaket pairwise (Heine, 2014) durchgeführt.

### Ergebnisse

Durch die Skalierung nach IRT liegen für jedes Wort empirische Kennwerte (Schwierigkeit, Trennschärfe, Modellfit) vor, welche bei der Itemauswahl der zu erstellenden Rechtschreib-CBM berücksichtigt werden. Ein exemplarischer Auszug der Ergebnisse für Wörter mit einem Silbentrennenden-h ist in Tabelle 2 aufgeführt.

Tabelle 1: Charakteristika der Stichprobe getrennt nach Klassenstufe

Klassenstufe	N	Alter	Anteil Mädchen	Anzahl Klassen
1	1066	7;6	48.6 %	48
2	1062	8;7	51.0 %	52
3	984	9;9	46.8 %	48
4	979	11;1	51.7 %	44

Tabelle 2: Auszug der empirisch gewonnenen Kennwerte auf Grundlage einer Skalierung des Gesamtdatensatzes nach IRT (rechtschriftliches Phänomen: Silbentrennendes-h)

Wort	Itemschwierigkeit	Trennschärfe $r_{pbis}$	Modellfit Infit MnSq
blühen	0.04	0.43	0.86
drehen	-1.19	0.25	0.87
gehen	-2.39	0.66	1.21
Kühe	-1.25	0.22	0.82
stehen	-2.26	0.37	1.16
Uhu	-1.17	0.20	1.03
Zehe	-0.05	0.27	1.14
ziehen	-0.06	0.30	0.94



Von den ursprünglich 808 pilotierten Wörtern wurden ein Wort aufgrund einer zu geringen Schwierigkeit (Lösungsrate bei 100 %), 21 Wörter aufgrund einer zu geringen Trennschärfe ( $r_{pbis} < 0.2$ ) und 41 aufgrund einer ungenügenden Passung zum Rasch-Modell ( $0.7 < Infit < 1.3$ ; Bond & Fox, 2015) aus dem Wortpool eliminiert. Insgesamt erwiesen sich unter psychometrischen Gesichtspunkten 745 Wörter für die zu erstellenden CBM als geeignet. Vor der Verteilung einzelner Wörter dieses Wortpools auf CBM-Testvorlagen wurde festgelegt, dass

- nur *tatsächliche Rechtschreibleistungen* (also keine Vorausläuferfähigkeiten), beginnend mit dem Schreiben kurzer, lautgetreuer Wörter, erhoben werden. Die zu entwickelnden Rechtschreib-CBM beziehen sich daher auf den Grundschulunterricht ab dem zweiten Halbjahr der ersten Klasse bis zum Ende der vierten Klassenstufe.
- die Verfahren möglichst zeitsparend einzusetzen sind. Das Aufgabenformat mit der höchsten Durchführungs- und Auswertungsökonomie ist wohl das Wortdiktat. Auf diese Weise können möglichst viele rechtschriftliche Phänomene in möglichst kurzer Zeit abgefragt werden.
- die Anzahl der zu prüfenden orthografischen Phänomene mit zunehmender Schulungsdauer ansteigt und die Wortanzahl der Diktate gestaffelt wird (Klasse 1: 12 Items, Klasse 2: 16 Items, Klasse 3: 20 Items, Klasse 4: 24 Items). Aufgrund der sich verbessernden schriftlichen Fähigkeiten der Kinder ist davon auszugehen, dass die Durchführung in allen Klassenstufen zwischen 8 und 12 Minuten in Anspruch nimmt.

Eine rein inhaltliche Analyse erscheint für die Erzeugung solcher inhaltsparalleler Verfahren nicht hinreichend. Durch den Einbezug der ermittelten psychometrischen Kennwerte konnte eine begründete Itemauswahl erfolgen. In diesem Punkt unter-

scheidet sich das hier vorgeschlagene Vorgehen von dem bisheriger Vorarbeiten. Die Itemzuordnung erfolgt somit nicht zufällig mittels Itemsampling (Strathmann & Klauer, 2008; Strathmann et al., 2010), stattdessen bedarf es eines Abwägungsprozesses zwischen den gesetzten inhaltlichen (Welche Strategien/ Phänomene/ Wörter sind in welcher Klassenstufe aktuell?) und den gewonnenen empirischen Kriterien (Schwierigkeit, Trennschärfe, Fit-Statistik). So wird sichergestellt, dass das Verfahren zum einen inhaltlich bzw. curricular valide und zum anderen hinreichend messgenau ist.

Auf dieser Grundlage wurden die Items aus dem 745 Wörter umfassenden Itempool auf fünf jeweils strukturähnliche CBM pro Schulhalbjahr verteilt. Grundsätzlich wurde versucht, so viele Wörter wie möglich aus dem Wörterpool zu nutzen, um eine hohe inhaltliche Validität zu erreichen. Bei vereinzelt Rechtschreibphänomenen, für die nur eine geringe Auswahlmöglichkeit besteht (bspw. *qu*), wurden a) Wörter mehrfach verwendet oder b) die empirischen Kriterien etwas breiter gefasst. Die resultierenden CBM sind daher nicht auf Itemebene parallel, aber auf Ebene der gesamten Wortlisten (siehe Tabelle 3). Zudem wurde auch bei der Bildung der CBM-Verfahren ein Multi-Matrix-Design (Mislevy et al., 1992) genutzt, sodass eine statistische Verlinkung über die Testzeitpunkte und Schuljahre ermöglicht wird. Diese erfolgt über identische Grapheme, welche sich unter empirischen Kriterien als parallel erweisen. In der nachfolgenden Übersicht werden zwei CBM-Wortlisten für die dritte Klassenstufe samt ihren psychometrischen Kennwerten dargestellt.

## Fazit und Ausblick

Ziel des Beitrags war es, ein Vorgehen zur Itemgewinnung und -analyse zu erproben und auf seine Handhabbarkeit zur Konzeption und Evaluation von CBM zu prüfen. Exemplarisch wurde hier der Bereich Rechtschreibung gewählt. Nach inhaltlichen

Tabelle 3: Exemplarische Darstellung zweier CBM-Wortlisten für die dritte Klassenstufe

CBM-Wortliste 1					CBM-Wortliste 2				
Wort	Schwierigkeit	r <sub>pbis</sub>	Infit	Länge	Wort	Schwierigkeit	r <sub>pbis</sub>	Infit	Länge
du	-1.60	0.35	0.94	2	er	-1.74	0.48	0.88	2
Blut	-1.86	0.37	1.07	4	Hufe	-1.45	0.33	1.10	4
Papier	-0.03	0.43	1.05	6	liest	0.18	0.55	0.86	5
Hexe	-1.00	0.32	0.92	4	hart	-0.16	0.36	1.13	4
Ostern	-1.99	0.38	1.00	6	Garten	-2.80	0.30	1.16	6
springst	0.12	0.34	1.11	8	gestern	0.69	0.40	1.09	7
Würstchen	0.61	0.58	0.87	9	Krankheit	0.35	0.41	1.05	9
heiß	-0.66	0.64	0.82	4	groß	-0.66	0.67	0.76	4
viel	-1.39	0.33	0.90	4	vom	-0.82	0.50	0.94	3
offen	0.69	0.56	0.94	5	grillen	0.05	0.57	0.91	7
Flügel	0.70	0.41	1.09	6	zwölf	0.25	0.47	0.96	5
Gräser	1.51	0.50	0.99	6	älter	0.67	0.32	1.11	5
Wetter	-1.06	0.41	0.90	6	Blätter	-0.25	0.58	0.82	7
Bügeleisen	0.02	0.40	1.03	10	Klassenfahrt	0.08	0.37	0.94	12
gehen	-2.39	0.66	1.21	5	stehen	-2.26	0.37	1.16	6
Zimmer	-1.59	0.42	1.02	6	Tanne	-1.62	0.25	0.94	5
dick	-0.25	0.43	0.96	4	jung	0.56	0.46	0.99	4
pflegen	1.24	0.46	1.04	7	Donnerstag	-0.16	0.50	0.85	10
schmecken	0.17	0.50	0.82	9	drücken	0.30	0.45	0.86	7
Dienstag	0.34	0.50	0.94	8	gesund	0.27	0.43	0.90	6
<b>MW</b>	<b>-0.42</b>	<b>0.45</b>	<b>0.98</b>	<b>5.95</b>	<b>MW</b>	<b>-0.43</b>	<b>0.44</b>	<b>0.97</b>	<b>5.90</b>
<b>SD</b>	<b>1.14</b>	<b>0.10</b>	<b>0.10</b>	<b>2.06</b>	<b>SD</b>	<b>1.02</b>	<b>0.11</b>	<b>0.12</b>	<b>2.40</b>

(Was soll gemessen werden?) sowie methodischen Überlegungen (Wie soll gemessen werden?) konnte ein Pool an Wörtern eingegrenzt werden, auf dessen Grundlage eine Zusammenstellung von CBM für den Bereich der Rechtschreibung für die Klassenstufen 1 bis 4 erfolgte. Diese CBM wurden im Schuljahr 2016/2017 in Stichproben mit mehreren hundert Schülerinnen und Schülern Mecklenburg-Vorpommerns erprobt. Erste Ergebnisse zur Reliabilität der CBM für die Klassenstufen 2 bis 4 konnten hierbei berechnet werden. Die Werte variieren im Bereich zwischen  $\alpha = .81$  und  $\alpha = .89$  und sind damit als gut einzuschätzen (Tabelle

4). Dies erhärtet die Annahme, dass durch das beschriebene Vorgehen homogene CBM erstellt werden konnten.

Im Rahmen der weiterführenden Forschungsarbeit wurden zudem zeitgleich mit den CBM konvergente und diskriminante Verfahren zur Einschätzung der Konstruktvalidität sowie zeitlich versetzt konvergente Messinstrumente zur Beurteilung der prognostischen Validität eingesetzt. Zum Zeitpunkt der Manuskriptvorbereitung konnten diese Daten jedoch noch nicht analysiert werden. Voraussichtlich zum Ende des Jahres 2017 liegen Ergebnisse zur Validität der entwickelten Verfahren vor.



Tabelle 4: Überblick über die Kennwerte zur Homogenität der erarbeiteten CBM über mehrere Messzeitpunkte und Klassenstufen

Schulwoche im Schuljahr 2016/17	Klassenstufe	N	Cronbachs $\alpha$
8	2	449	.83 (16 Items)
	3	456	.82 (20 Items)
	4	457	.88 (24 Items)
12	2	324	.83 (16 Items)
	3	390	.81 (20 Items)
	4	409	.89 (24 Items)
16	2	426	.83 (16 Items)
	3	456	.82 (20 Items)
	4	457	.88 (24 Items)

Eine weitere offene Forschungsfrage bezieht sich auf die Auswertung der CBM-Daten. Neben einer quantitativen Analyse der richtig bzw. falsch gelösten Items ist eine qualitative Auswertung nach Rechtschreibphänomenen denkbar. Dies würde die Förderrelevanz der Ergebnisse enorm erhöhen, bleibt jedoch noch im Detail auszuarbeiten.

Die bisherigen Daten weisen darauf hin, dass das im Rahmen dieses Beitrags beschriebene Vorgehen zur Entwicklung von CBM vielversprechend ist, vermutlich nicht nur für den Bereich der Rechtschreibung. Durch klar definierte inhaltliche sowie psychometrische Kriterien können stratifizierte Itempools erarbeitet werden, welche die Grundlage zur Erstellung paralleler CBM bilden. Inwieweit der vorgelegte Vorschlag zur Konzeption von CBM für andere schulische Bereiche übertragungsfähig ist, sollte zukünftig auf theoretischer sowie empirischer Ebene diskutiert werden.

## Literatur

- Best, K.-H. (2000). Unser Wortschatz. Sprachstatistische Untersuchungen. In R. Hoberg & K. Eichhoff-Cyrus (Hrsg.), *Die deutsche Sprache zur Jahrtausendwende. Sprachkultur oder Sprachverfall?* (S. 35-52). Mannheim / Leipzig / Wien / Zürich: Du-  
denverlag.
- Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3. Aufl.). New York & London: Routledge.
- Bremerich-Vos, A., Granzer, D., Behrens, U. & Köller, O. (2011). *Bildungsstandards für die Grundschule: Deutsch konkret* (3. Aufl.). Berlin: Cornelsen.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- Deno, S. L. (2003). Developments in Curriculum-based Measurement. *Journal of Special Education*, 37, 184–192.
- Fuchs, L. S. (2004). The Past, Present, and Future of Curriculum-based Measurement Research. *School Psychology Review*, 33, 188–192.
- Grimm, H. (2003). *Störungen der Sprachentwicklung* (2. Aufl.). Göttingen: Hogrefe.
- Hanke, P. & Schwippert, K. (2005). Orthographische Lernprozesse im Grundschulbereich. Ergebnisse aus Mehrebenenanalysen. *Zeitschrift Unterrichtswissenschaft*, 1, 72–93.
- Hasselhorn, M., Schneider, W. & Trautwein, U. (Hrsg.) (2014). *Lernverlaufsdiagnostik*. Göttingen: Hogrefe.

- Hattie, J. (2013). *Lernen sichtbar machen*. Baltmannsweiler: Schneider-Verl. Hohengehren.
- Heine, J.-H. (2014). *pairwise: Rasch Model Parameters by Pairwise Algorithm* [Computer software]. Munich. Zugriff am 28.05.2017. Verfügbar unter <http://cran.r-project.org/web/packages/pairwise/index.html> (R package version 0.4.1).
- Hoge, R. D. & Coladarsi, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology*, 59, 297–313.
- Klauer, K. J. (2006). Erfassung des Lernfortschritts durch curriculumbasierte Messung. *Heilpädagogische Forschung*, 32, 16–26.
- Klauer, K. J. (2011). Lernverlaufsdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik*, 3, 207–224.
- Kultusministerkonferenz (2004). *Bildungsstandards im Fach Deutsch für den Primarbereich Beschluss vom 15.10.2004*. München: Wolters.
- LISUM (2004). Rahmenplan Grundschule Deutsch. Roggentin: ariant.
- LISUM (2011). Grundwortschatz für die Grundschule in Brandenburg – Rechtschreiben. Abgerufen am 26.01.2017 unter [http://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/faecher/sprachen/deutsch/Grundschule/Grundwortschatz\\_BB\\_WEB\\_2013.pdf](http://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/faecher/sprachen/deutsch/Grundschule/Grundwortschatz_BB_WEB_2013.pdf)
- Ministerium für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern (2016). Mindestwortschatz für die Jahrgangsstufen 1 bis 4 in Mecklenburg-Vorpommern. Abgerufen am 23.02.2017 unter <http://www.regierung-mv.de/serviceassistent/download?id=1580760>
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29, 133–161.
- Reber, K. & Kirch, M. (2013). Richtig schreiben lernen. Kompetenzorientierter, inklusiver Rechtschreibunterricht. *Praxis Sprache*, 4, 254–257.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing [Computer software]. Vienna, Austria. Retrieved from <http://www.R-project.org>.
- Senatsverwaltung für Bildung, Jugend und Wissenschaft (2011). Grundwortschatz Berlin. Abgerufen am 23.02.2017 unter [https://www.berlin.de/sen/bildung/schule/bildungswege/grundschule/handreichung\\_grundwortschatz.pdf](https://www.berlin.de/sen/bildung/schule/bildungswege/grundschule/handreichung_grundwortschatz.pdf)
- Souvignier, E., Förster, N. & Salaschek, M. (2014). quop: ein Ansatz internet-basierter Lernverlaufsdiagnostik und Testkonzepte für Mathematik und Lesen. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik (Tests und Trends N.F. Band 12, S. 239-256)*. Göttingen: Hogrefe.
- Staatsinstitut für Schulqualität und Bildungsforschung München (2012). Neuerungen im Deutschlehrplan für die Grundschule. Abgerufen am 23.02.2017 unter <https://www.isb.bayern.de/download/1353/neuerungen-deutschlp-gs-07-01-12.pdf>
- Strathmann, A. & Klauer, K. J. (2008). Diagnostik des Lernverlaufs. Eine Pilotstudie am Beispiel der Entwicklung der Rechtschreibkompetenz. *Sonderpädagogik*, 38, 5–24.
- Strathmann, A. & Klauer, K. J. (2010). Lernverlaufsdiagnostik: Ein Ansatz zur längerfristigen Lernfortschrittsmessung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42, 111–122.
- Strathmann, A. & Klauer, K. J. (2012). *LVD-M 2–4. Lernverlaufsdiagnostik Mathematik für die zweiten bis vierten Klassen*. Göttingen: Hogrefe.
- Strathmann, A., Klauer, K. J. & Greisbach, M. (2010). Lernverlaufsdiagnostik - Dargestellt am Beispiel der Entwicklung der Rechtschreibkompetenz in der Grundschule. *Empirische Sonderpädagogik*, 2, 64–77.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of stu-

- dents' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104, 743–762.
- 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* (Tests & Trends, NF Bd. 12, S. 83–99). Göttingen: Hogrefe.
- Voß, S. (2014). *Curriculumbasierte Messverfahren im mathematischen Erstunterricht - Zur Güte und Anwendbarkeit einer Adaption US-amerikanischer Verfahren im deutschen Schulsystem*. Saarbrücken: SVH.
- Walter, J. (2008). Curriculumbasiertes Messen (CBM) als lernprozessbegleitende Diagnostik: Erste deutschsprachige Ergebnisse zur Validität, Reliabilität und Veränderungssensibilität eines robusten Indikators zur Lernfortschrittsmessung beim Lesen. *Heilpädagogische Forschung*, 34, 62–79.
- Walter, J. (2010). *Lernfortschrittsdiagnostik Lesen. Ein curriculumbasiertes Verfahren*. Göttingen: Hogrefe.
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik*, 3, 225–242.

**Stefan Voß**

Universität Rostock

18051 Rostock

stefan.voss3@uni-rostock.de

Erstmalig eingereicht: 01.03.2017

Überarbeitung eingereicht: 30.05.2017

Angenommen: 01.08.2017