

Empirische Sonderpädagogik, 2018, Nr. 1, S. 100-118
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit

Jana Jungjohann¹, Andreas Gegenfurtner² & Markus Gebhardt¹

¹Technische Universität Dortmund, ²Technische Hochschule Deggendorf

Zusammenfassung

Mit Hilfe von Lernverlaufsmessungen können Lehrkräfte Schwierigkeiten im Leseerwerb identifizieren und eine passende Förderung initiieren. Die Leseflüssigkeit ist ein robuster Indikator der Lesekompetenz, für die mehrere formative Testverfahren konstruiert wurden. In diesem systematischen Review werden empirisch belegte formative Testverfahren für den Leseunterricht der ersten drei Schuljahre dargestellt und die aktuelle Entwicklung für die Schulpraxis sowie wissenschaftliche Forschung abgeleitet. Insgesamt wurden acht Testverfahren zur Leseflüssigkeit aus begutachteten nationalen und internationalen Zeitschriften identifiziert. Die Entwicklungen, Gestalt und Evaluationen dieser Verfahren werden im Hinblick auf zukünftige Forschungsbemühungen ausgewertet, synthetisiert und diskutiert. Die Testverfahren sind überwiegend unabhängig von Testbatterien oder Förderkonzepten, als pen&paper Version verfügbar und nach klassischer Testtheorie konstruiert. Besonders die Überprüfung mittels der Item-Response-Theorie und die Messinvarianz sowie die Nutzung digitaler Versionen stehen für diese Testverfahren noch aus.

Schlagwörter: Systematisches Review, Leseflüssigkeit, Formative Lernverlaufsmessung, CBM, Grundschule

A systematic literature review of the formative assessment of early reading fluency

Abstract

With the help of formative assessment instruments, teachers can identify and support children who have reading difficulties. Reading fluency is a robust indicator of reading ability in many existing formative assessments. This systematic literature review discusses different tests for the formative assessment of reading fluency in grades 1 to 3 and makes scientific recommendations for their scientific development and practical use. In total, eight formative assessments of reading fluency were identified from national and international peer reviewed journals. This review describes how these eight formative assessment instruments have been developed, how they are used, and how the test outcomes are evaluated. The aim is to offer implications for future research. These assessments are independent from existing test batteries or materials, they are administered by pen and paper, and were constructed based on classical test theory. Future research can test these assessment instruments with item-response-theory, estimate their measurement invariance, and validate computer-based test versions.

Keywords: CBM, Formative Assessment, Primary School, Reading Fluency, Systematic Review

Der erfolgreiche Leseerwerb ist ein wichtiger Prädiktor für den weiteren schulischen Verlauf (Ennemoser, Marx, Weber & Schneider, 2012; Korhonen, Linnanmäki & Aunio, 2014; Nagler, Linkersdörfer, Lonnemann, Hasselhorn & Lindberg, 2016). Besonders eine präventive Förderung in den ersten Schuljahren ist für die basalen Lesekompetenzen von hoher Bedeutung, damit sich Leseerwerbsprobleme nicht langfristig festsetzen (Helf & Cooke, 2011; Torgesen, 2002). Es existieren bereits zahlreiche Förderkonzepte für die ersten Schuljahre, allerdings mangelt es an Implementationsstudien und einer Verschränkung von Diagnostik und Förderung im schulischen Feld (Groth, Hasko, Bruder, Kunze & Schulte-Körne, 2013). Die Verschmelzung von theoriegeleiteter Diagnostik mit mehreren Messzeitpunkten und zielgerichteten Interventionen verspricht vor allem für den Anfangsunterricht einen besonders hohen Nutzen im Hinblick auf eine fortschreitende Lernentwicklung (Diehl, Harke & Knopp, 2009).

Engmaschige Lernverlaufsmessungen der intraindividuellen Leseentwicklung, wie die des US-amerikanischen Ansatzes des Curriculum-Based Measurements (CBM), setzen an genau diesem Ziel an. Die Instrumente geben den Lehrkräften Aufschluss über Leseschwierigkeiten und regen die Lehrkräfte an, die Wirksamkeit ihrer unterrichtlichen Maßnahmen fortlaufend zu überprüfen, um die Unterrichtsqualität zu erhöhen (Deno, 2003a; Klauer, 2014; Lankes, Vaccaro & Gegenfurtner, 2013; Stecker, Fuchs & Fuchs, 2005).

Seit einiger Zeit werden CBM-Verfahren auch außerhalb der USA für den Anfangsunterricht im Lesen entwickelt. Trotz ihrer möglichen Potentiale für die Förderung von leistungsschwachen Kindern sind sie deutlich seltener vertreten als Statusverfahren (Organisation for Economic Co-operation and Development, 2005). Nach mehr als einem Jahrzehnt an Forschungsbemühungen in Deutschland ist die Entwicklung und Erforschung von formativen Lernverlaufsmessungen immer noch in ihren Anfängen. Im

Hinblick auf das deutsche Schulsystem könnte ein vermehrter Einsatz von Lernverlaufsmessungen dazu beitragen, dass „wait-to-fail“-Strukturen zu Gunsten von „Response-to-Intervention“ (RTI)-Strukturen abgebaut werden (Huber & Grosche, 2012). Das RTI-Modell zeichnet sich u.a. dadurch aus, dass Lernrückstände frühzeitig identifiziert und Lernentwicklungen fortlaufend überprüft werden (ebd.). Innerhalb von RTI-Strukturen fördern Lehrkräfte individuell und im präventiven Sinn, damit sich keine Lern- und Verhaltensprobleme manifestieren. In inklusiven Klassen arbeiten sonderpädagogische Lehrkräfte und Regelschullehrkräfte gemeinsam für den Lernerfolg ihrer Schülerinnen und Schüler (Gebhardt, Schwab, Krammer & Gegenfurtner, 2015). Besonders unter diesen Bedingungen kann der Einsatz von Lernverlaufsmessungen dazu beitragen, dass Kinder eine sonderpädagogische Förderung erhalten, bevor sie gravierende Schwierigkeiten entwickeln.

Das Ziel dieses Beitrags ist es, einen Überblick zum aktuellen Forschungsstand über lernverlaufsdagnostische Verfahren zur Erhebung der frühen Leseflüssigkeit im Anfangsunterricht darzustellen. Kinder mit einem sonderpädagogischen Förderbedarf oder mit langanhaltenden Leseschwierigkeiten benötigen über das erste Schuljahr hinweg Zeit für die Entwicklung von basalen Lesefähigkeiten (Klicpera, Schabmann, Gastegger-Klicpera & Schmidt, 2017). Besonders in inklusiven Klassen vergrößert sich somit die Heterogenität im Leseerwerb. Um ebenfalls die Leseentwicklung leistungsschwacher Kinder ausreichend berücksichtigen zu können, wird in diesem Review zusätzlich zum Anfangsunterricht das dritte Grundschuljahr eingeschlossen. Somit werden jene Verfahren berücksichtigt, die in den ersten drei Grundschuljahren Anwendung finden. Diese starke Eingrenzung wird mit dem Ziel vorgenommen, detaillierte Implikationen für den Leseanfangsunterricht in leistungsheterogenen Klassen abzuleiten. Von speziellem Interesse sind dafür auch internationale Entwicklungen, die für konsistente Sprachen

wie dem Deutschen konzipiert sind. Der Bedarf eines solchen Reviews liegt darin begründet, dass in den ersten Grundschuljahren wegweisende und altersspezifische Entwicklungen vollzogen werden und diese die Lesekompetenz langfristig beeinflussen. Mit besonderem Blick auf die deutsche Schullandschaft werden die Ergebnisse sowie das weitere Potential dieser Forschung diskutiert.

Entwicklung und Förderung basaler Lesekompetenzen

Lesen ist ein kognitiver, multidimensionaler und aktiver Prozess. Gute Leserinnen und Leser unterscheiden sich von schwachen dadurch, dass sie die benötigten Teilprozesse effizienter bewältigen (Scheerer-Neumann, 2015). In der langen Forschungshistorie zum Leseerwerb sind unterschiedliche Modelle entwickelt worden, die sich hinsichtlich der Terminologie, Ausgestaltung und Anzahl einzelner Phasen unterscheiden. Alle Modelle berücksichtigen das basale (hierarchie-niedrige) Lesen und keine weiterführenden strategischen (hierarchiehohen) Textleseleistungen (Mayer, 2010; Scheerer-Neumann, 2015). Schriftspracherwerbsmodelle, wie die von Uta Frith (1986) oder Klaus Günter (1986), betonen, dass der Leseerwerb kein linearer Vorgang ist und sich das kognitive System kontinuierlich qualitativ umstrukturiert. Klicpera et al. (2017) verweisen darauf, dass sich Teilleistungsprozesse im Deutschen nicht gegenseitig ablösen, sondern parallel als Lesestrategien entwickeln und bei reifen Leserinnen und Lesern verfügbar sind. Nach dem Kompetenzentwicklungsmodell des Worterkennens und des lauten Lesens von Klicpera et al. (2017) bilden Kinder das lexikalische und das nichtlexikalische Lesen ab dem Eintritt in die Grundschule aus. Zu Beginn des Anfangsunterrichts steht das nichtlexikalische Lesen im Vordergrund. Die Kinder erlesen mit dieser Strategie zunächst Wörter sequenziell anhand der Buchstabenreihenfolge (Phonologisches Rekodieren). Ab ca. der zweiten

Klassen greifen Kinder ohne Leseschwierigkeiten beim phonologischen Rekodieren auf immer größere Verarbeitungseinheiten, wie Silben und Morpheme, zurück. Dadurch steigt die Leseflüssigkeit (reading fluency) an. Im Laufe des Anfangsunterrichts werden zunehmend mehr Verarbeitungseinheiten und Wörter im mentalen Lexikon gespeichert. Parallel zum nichtlexikalischen Lesen beginnt die Ausbildung des lexikalischen Lesens. Unter Rückgriff auf gespeicherte Einträge aus dem mentalen Lexikon können Kinder Wörter im Ganzen erkennen. Mit steigender Lesekompetenz wird das lexikalische Lesen zur dominierenden Strategie und erhöht so die Lesegeschwindigkeit (reading speed).

Die Leseflüssigkeit stellt in unserem Leseerwerb eine zentrale literale Zieldimension dar, da sie maßgeblich die gesamte weitere Leseentwicklung beeinflusst (National Institute of Child Health and Human Development, 2000). Um eine hohe Lesekompetenz zu erreichen, müssen dementsprechend Leserinnen und Leser eine gute Leseflüssigkeit ausbilden. Denn je weniger Kapazität die Leseflüssigkeit beansprucht, desto mehr steht für das Leseverstehen (reading comprehension) bereit. Nach Klicpera et al. (2017) wird der Leseerwerb durch individuelle Basiskompetenzen sowie weitere Einflussfaktoren, wie der unterrichtlichen Instruktionen und Förderungen, beeinflusst.

Allerdings erreichen 15% der Schülerinnen und Schüler nur unzureichende Lesefähigkeiten, die das alltäglichen Leben sowie berufliche Chancen beeinträchtigen (Organisation for Economic Co-operation and Development, 2016; Tarelli, Valtin, Bos, Bremerich-Vos & Schwippert, 2012). In der Forschung ist aktueller Konsens, dass sich schon früh Einflussfaktoren für den Leseerwerb erkennen lassen (Ennemoser et al., 2012; Gorecki & Landerl, 2015; Snowling & Melby-Lervag, 2016). In der Längsschnittstudie von Klicpera, Ehgartner, Gasteiger-Klicpera und Schabmann (1993) starteten Kinder mit einem sonderpädagogischen Förderbedarf bereits mit geringen Lernvoraussetzungen

gen in die Schulzeit. Diese Kinder haben im Vergleich zu Kindern ohne Förderbedarf häufig geringere Buchstabenkenntnisse und weniger Einsichten in die Phonemstruktur (ebd.). Ohne passende unterrichtliche Förderung manifestieren sich Leseschwierigkeiten zu Beginn des Leseerwerbs bei einem Großteil der Kinder über die gesamte Schulzeit hinweg (Landerl & Wimmer, 2008). Folglich wird in der Forschung dem Leseanfangsunterricht ein besonders hoher Stellenwert für einen erfolgreichen Leseerwerb zugeschrieben (Heyne, 2014; McDonald Connor et al., 2009). Der Einsatz von formativer Lernverlaufsmessung im Unterricht kann den Leseerwerb von leseschwachen Schülerinnen und Schülern positiv beeinflussen. Ihr Lernzuwachs steigt signifikant an, wenn ihre Lehrkräfte auf der Basis von CBM-Daten ihren Unterricht individuell adaptieren (Stecker et al., 2005). Souvignier und Förster (2011) weisen erstmalig für den deutschsprachigen Raum nach, dass die Zuwächse von leseschwachen Schülerinnen und Schülern im Leseerwerb signifikant höher sind als in Klassen ohne eingesetzte formative Lernverlaufsmessung.

Formative Lernverlaufsmessung

Formative Lernverlaufsmessungen dienen dem Zweck, Lernprozesse mittels Messungen kontinuierlich zu begleiten und sie mit Hilfe der Messergebnisse positiv zu beeinflussen (Bennett, 2011; Reschly, Busch, Betts, Deno & Long, 2009). Einen wegweisenden Grundstein entwickelte die Forschergruppe um Stanley Deno aus der Disziplin der Sonderpädagogik mit dem Ansatz des CBM (Deno, 1985). CBM dient der Visualisierung von Leistungsveränderungen während eines Lernprozesses, die eng an das Curriculum gebunden sind. Dieses Aufzeigen werden durch wiederholte Messungen in zeitlich nahen Abständen (bis zu einmal wöchentlich) mittels paralleler Testversionen realisiert (Gebhardt, Diehl & Mühling, 2016). Des Weiteren werden eine ökonomische Anwendung und Auswertung sowie die

mögliche Berücksichtigung lokaler Bedingungen zu ihren Stärken in der schulischen Praxis gezählt (Walter, 2009a). Die Messverfahren konzentrieren sich auf stellvertretende Aufgaben, die repräsentativ die geforderten Fähigkeiten am Ende eines Schuljahres widerspiegeln (Fuchs, 2004). Es wird zwischen zwei Gestaltungsarten der Aufgaben unterschieden. Zum einen wird ein robuster Indikator stellvertretend für eine umfassende Kompetenz identifiziert, welcher hoch mit der fokussierten Fähigkeit am Ende eines Schuljahres korreliert. Beispielsweise gilt dies im Bereich der allgemeinen Lesekompetenz für die Leseflüssigkeit (Deno, Mirkin & Chiang, 1982; Fuchs, Fuchs, Hosp & Jenkins, 2001). Zum anderen werden systematische Beispielaufgaben entwickelt, welche für mehrere Teilfähigkeiten in unterschiedlichen Schwierigkeitsstufen verfügbar sind. Diese werden typischerweise für mathematisch orientierte CBM gewählt (Fuchs, 2004). Anhand der erhobenen Daten können die Lehrkräfte Rückschlüsse über ihr unterrichtliches Handeln ziehen und Förderungen im präventiven Sinn entsprechend anpassen. Besonders dieser Aspekt charakterisiert CBM maßgeblich, da sie nicht das Ziel der Klassifikation eines Leistungsstandes oder der Selektion verfolgen. Sie grenzen sich so deutlich von Statusverfahren ab. Inhaltlich orientieren sich CBM traditionell unmittelbar an den aktuellen Unterrichtsinhalten, die die Lehrkräfte vermitteln. Forschungen zeigen allerdings, dass zu einem effektiven Einsatz ebenfalls schulübergreifende bzw. überregionale Aufgabenstellungen mit ähnlichen Schwierigkeitsmerkmalen valide sind (Deno, 2003b). Verfahren mit general outcomes measures oder dynamic indicators of basic skills bezeichnet (Fuchs und Deno 1994, Shinn 1998 zitiert nach Deno 2003b).

Für die Erhebung der Lesekompetenz von Kindern mit Lernschwierigkeiten setzten sich im Laufe der Entwicklung drei CBM-Arten durch (Walter, 2009b). Die erste Testform oral reading fluency (ORF) stellt Schü-

lerinnen und Schülern die Aufgabe, in einem kurzen Zeitraum (1-3 Minuten) einen Text möglichst ohne Fehler laut vorzulesen. Das Kompetenzmaß ist die Anzahl richtig gelesener Wörter pro Minute (WPM). Auslassungen, langes Zögern oder Fehlaussprache zählen dabei als Lesefehler. Hierbei ist die Leseflüssigkeit der robuste Indikator. Bei der zweiten Testform *word identification fluency* (WIF) werden Listen mit unzusammenhängenden Wörter, Silben oder Pseudowörter laut vorgelesen. Diese WIF-Verfahren fokussieren ebenfalls die Leseflüssigkeit und stützen sich auf dasselbe Kompetenzmaß. Sowohl ORF- als auch WIF-Verfahren erfassen basale Lesekompetenzen und werden für Messungen unmittelbar ab dem Schuleintritt eingesetzt. Ob ein Verständnis des Gelesenen vorliegt, wird nicht weiter überprüft. Die dritte Testform *maze selection* überprüft das Leseverständnis. Dabei wird den Schülerinnen und Schülern ein Text präsentiert, indem jedes siebte Wort nur als Auswahl gemeinsam mit zwei Distraktoren gezeigt wird. Die Distraktoren ähneln dem Zielwort meistens phonologisch oder semantisch. Die Kinder lesen diese Texte leise und wählen das passende Wort aus. Die Anzahl richtig ausgewählter Wörter stellt hier das Kompetenzmaß dar.

Die Messinstrumente der formativen Lernverlaufsmessung sollten dabei die Standards der klassischen Gütekriterien wie Objektivität, Reliabilität und Validität sowie die Nebengütekriterien der Änderungssensibilität, Ökonomie und Eindimensionalität erfüllen (Good & Jefferson, 1998; Klauer, 2006). Unter der Reliabilität wird im Allgemeinen verstanden, dass die zugrundeliegende Fähigkeit das Messergebnis abbildet und nicht zu stark durch einen Messfehler beeinflusst wird (Wilbert & Linnemann, 2011). Die Reliabilität eines Lernverlaufsinstrumentes kann neben der klassischen Testtheorie auch mittels der Generalisierbarkeitstheorie (GT) geprüft werden. Die GT bietet den Vorteil, dass mehrere einflussnehmende Faktoren (z.B. Rater und Items) im Entwicklungsverlauf so-

wohl inter- als auch intraindividuell analysiert werden können (Hintze, Owen, Shapiro & Daly, 2000). Die Validität beschreibt unterschiedliche mögliche Interpretationen und Verwendungen von Testergebnissen (Hartig, Frey & Jude, 2012; Kane, 2013). Besonders im Hinblick auf die Interpretationen und die Verwendung diagnostischer Ergebnisse für den pädagogischen Kontext muss bei der Testkonstruktion auf die Testfairness geachtet werden. Es dürfen keine Benachteiligungen bestimmter Personen entstehen (Moosbrugger & Kelava, 2012). Um sicherzustellen, dass die einzelnen Tests zu den einzelnen Messzeitpunkten gleich schwer und zu allen Kindern fair sind, empfehlen einige Autorengruppen eine Überprüfung der Testgüte mittels der *Item-Response-Theorie* (IRT) (Gebhardt, Heine, Zeuch & Förster, 2015; Klauer, 2011; Wilbert & Linnemann, 2011). Da Lernverlaufsinstrumente das Ziel verfolgen Veränderungen von Fähigkeiten oder Kompetenzen sensibel zu erfassen, müssen sie über die Zeit hinweg bereits kleine Änderungen in den Fähigkeiten zuverlässig messen (Klauer, 2011). Damit Tests änderungssensibel messen, müssen die Ergebnisse eines Tests eindimensional sein. Eindimensionalität bedeutet in diesem Fall, dass die Testergebnisse in lediglich einem latenten Faktor begründet sind (Wilbert & Linnemann, 2011).

Fuchs (2004) klassifiziert die Evaluation von CBM-Verfahren in drei aufeinander aufbauende Stadien der Güte. Diese Stadien der Güte können ebenfalls als Prüfung der Testgütekriterien aufgefasst werden. Das erste Stadium umfasst eine einmalige Messung mit einer großen Stichprobe eines aktuellen Leistungsstandes (Überprüfung der Reliabilität, Stabilität, Objektivität). Das zweite Stadium prüft die Messinvarianz sowie die Änderungssensibilität bei mehreren Messzeitpunkten und zeigt den Lernverlauf grafisch (Überprüfung der Veränderungssensibilität). Im dritten Stadium sollen die Studien den praktischen Nutzen der Lernverlaufsmessung im Feld, die Wirksamkeit sowie die

Verschänkungen mit den Fördermaßnahmen der Lehrkräfte prüfen (Überprüfung der sozialen Validität).

Forschungsfragen zu formativen Testverfahren

Für weitere Forschungsaktivitäten im Bereich der formativen Lernverlaufsmessung für basale Lesefähigkeiten ist von besonderem Interesse, welche Entwicklungen in den letzten Dekaden national und international stattgefunden haben. Dabei liegt ein Fokus auf der Entwicklung, der Gestalt und der Evaluation formativer Testinstrumente, welche die frühe Leseflüssigkeit erfassen. Die folgenden Forschungsfragen rahmen das vorliegende systematische Review.

1. Wie sind die einzelnen Aufgaben der Testverfahren konzipiert? Für welche Zielgruppe wurde welche Art von Materialien entworfen?
2. Welche theoriegeleiteten (Teil-)Aspekte der basalen Lesekompetenzen werden fokussiert?
3. Wie werden die Testverfahren im Unterrichtsalltag realisiert?
4. Inwiefern sind die Testverfahren flexibel und adaptierbar? Bieten die Testverfahren Fördermaterialien oder Schulungsunterlagen an?
5. Welche Arten von Feedback integrieren die Instrumente?
6. Inwiefern sind die Testverfahren in Bezug auf Gütekriterien und Testtheorien evaluiert? Welche weiteren Forschungsanliegen wurden in Abgrenzung zu den klassischen Gütekriterien diskutiert?

Methodik des systematischen Reviews

Die Methodik dieses systematischen Reviews entspricht dem Leitfaden nach Petticrew und Roberts (2012). Zentrale Schritte eines systematischen Reviews beinhalten die Definition relevanter Forschungsfragen, die Festschreibung von Ein- und Ausschlusskriterien, eine detaillierte Dokumentation des gesamten Reviews, den Suchvorgang selbst, die Sichtung der Suchergebnisse unter kritischer Prüfung, den endgültigen Ein- bzw. Ausschluss relevanter Studien sowie die Verschriftlichung des Reviews unter Beachtung eines möglichen Publikationsbias. Der Ein- bzw. Ausschluss der Studien bzw. Testinstrumente ist nach folgenden Kriterien selektiert worden. Studien sind nicht in das Review aufgenommen worden, wenn sie nicht alle vier Einschlusskriterien erfüllen.

1. Die Studien beschreiben ein formatives Verfahren. Daraus resultiert die vorwiegende Eigenschaft, dass die erhobenen Daten zur Ausgestaltung bzw. Adaption von Unterricht und Förderung genutzt werden sollen.
2. Die identifizierten Testverfahren erheben die Leseflüssigkeit.
3. Die relevante Kohorte bezieht sich auf Schülerinnen und Schüler der ersten drei Klassenstufen. Es ist zulässig, dass sowohl Kinder mit und ohne sonderpädagogischen Förderbedarf in den Stichproben eingeschlossen sind. Zudem dürfen die Testverfahren neben den ersten drei Schuljahren zusätzlich in höheren Klassenstufen eingesetzt werden.
4. Die Studien sind in Zeitschriften mit wissenschaftlichem Begutachtungsverfahren veröffentlicht.

Die Suche wurde von Oktober 2016 bis Februar 2017 durchgeführt. Zuerst wurde die Suche in der deutschsprachigen Literaturdatenbank FIS-Bildung vorgenommen. Danach in der internationalen Literaturdatenbank

EBSCOhost Online Research Databases, die die Datenbanken PsycInfo und ERIC einschloss. Peer Reviewed Artikel aller Sprachen galten als zulässig, die Suchbegriffe wurden als Freitext eingetragen und doppelte Einträge wurden automatisch gelöscht. In der deutschsprachigen Datenbank wurden folgende Suchbegriffe verwendet: „Lesekompetenz“, „Lesefähigkeit“, „Leseflüssigkeit“, „Leseverstehen“, „Leseengeschwindigkeit“ UND „Messung“, „Diagnostik“, „Test“, „Überprüfung“, „Beurteilung“, „Lernverlauf“, „Lernfortschritt“, „Lernentwicklung“, „formativ“, „curriculum basiert“, „cbm“, „Prozess“, „Verlauf“. Insgesamt ergab diese Suche 409 Suchtreffer. Für die internationale Suche wurden analog die folgenden Begriffe verwendet: „reading competence“, „reading ability“, „reading fluency“, „reading speed“ AND „curriculum based“, „formative“, „formative evaluation“, „formative assessment“, „assessment for learning“, „feedback“. Die Suchbegriffe „cbm“ sowie „curriculum-based“ ergaben dieselben Suchtreffer wie Kombinationen mit „curriculum based“. Die Suchbegriffe „measurement“, „diagnostic“, „test“, „assessment“ sowie „process“ lieferten in anfänglichen, explorativen Suchanfragen zu viele Treffer und stellten sich folglich als zu unspezifisch heraus. Daher wurden sie aus Gründen der Effizienz und Praktikabilität zugunsten der letztlich verwendeten Suchbegriffe exkludiert. Insgesamt wurden mit der internationalen Suche 284 Treffer generiert.

Von allen Suchtreffern ($n=693$) wurden die Titel sowie die Abstracts von der Erstautorin gesichtet. Alle als potentiell relevanten Artikel konnten beschafft werden und sind im Vier-Augen-Prinzip mit einem Co-Autor anhand der Selektionskriterien gemeinsam geprüft worden. Das Ergebnis war der Einschluss von sieben Testverfahren auf der Basis von acht Artikeln.

Ein Großteil der für dieses Review ungeeigneten Studien stellen Interventionsstudien und Förderprogramme zu unterschiedlichen Teilleistungen der Lesefähigkeit und -motivation dar. Diese Studien gehen auf spezifi-

sche Vorhersage- oder Einflussindikatoren der Lesekompetenz wie z.B. Geschlecht, Alter, Lesemotivation, spezifische Förderbedarfe oder soziale Disparitäten ein. Beispielsweise werden Elterntrainings und kooperative Leseaktivitäten validiert. Weitere ausgeschlossene Studien vergleichen Messergebnisse von statusorientierten Schulleistungstests, wie beispielsweise dem Ekwall Reading Inventory oder Woodcock Reading Mastery Test, mit CBM-ORF-Werten. Bei diesen Studien wurden zwar CBM-Werte als Referenz herangezogen, allerdings wurden keine ausreichenden Hinweise zu einzelnen CBM-Verfahren angegeben. Weitere Analysen beziehen sich auf großangelegte Schulleistungstests und Vergleichsarbeiten sowie ihre Konsequenzen für den sprachlichen Bereich. Bezug wurde auf beispielsweise PISA, IGLU, PIRLS, BISS und TMMSS genommen. Weitere Artikel thematisieren den Fachsprachenunterricht. Analysiert wurden Unterrichtsmaterialien, Bildungsstandards und die diagnostische Kompetenz von Lehrkräften. Vorgestellte formative Testverfahren erheben andere Teilfähigkeiten der Lesekompetenz wie das Leseverstehen, die PB oder die lexikalische Repräsentation.

Als dritter Schritt wurden die Querverweise und Quellen der identifizierten Studien näher betrachtet, um weitere relevante begutachtete Publikationen ausfindig zu machen. Dadurch wurde eine weitere Studie in das systematische Review aufgenommen. Somit analysiert das Review insgesamt acht Testverfahren auf der Basis von neun Studien.

Ergebnisse der Recherche

Die acht identifizierten Testverfahren wurden anhand von neun Publikationen analysiert und sind in Tab. 1 dargestellt. Die Analyse erfolgt anhand der zuvor aufgestellten Forschungsfragen.

Tabelle 1: Eingeschlossene Studien formativer Lernverlaufsmessungen im Bereich Lesen

Autoren (Jahr, Name)	Konzeption (CBM Art ¹ , Einb. ² , Schul- stufe)	Testmaterial (Anz. Versionen ³ , Referenzquelle, Struktur)	Realisierung im UN (Durchf. ¹⁰ , KM ¹¹ , MZP ¹² , Medienart)	Flexibilität & add. Materialien (Anpassung ans Curriculum, Fördermaterialien)	Land	Feedback (Dokumentation, Übertragung auf UN)	Evaluation (Design, Stichprobe pro Klassenstufe, MZP, Ergebnisse)
Abu-Hamour (2014b)	WIF einzelstehend 1	30 UN ⁴ -Material Gl. Schw. ⁵ N ⁶ &V ⁷	Standardisiert WPM ¹³ - 14 - Pen&Paper	- - - -	JOR	- -	Erprobung: N=75, MZP=15, Retest-R.: 0.97-0.99, Krit.-V ¹⁷ : 0.47 (p < .01)
Diehl & Hartke (2011) IEL-1	WIF Subtest 1	1 Sichtwortschatz Anst. Schw. ⁸ N&F ⁹	Standardisiert WPM 3 - Pen&Paper	- -	DEU	Graphisch -	Erprobung inkl. interne Validität; N = 1642, MZP = 3, Retest-R.: 0.70-0.86, Krit.-V: 0.66-0.70 (p < .01)
Fuchs et al. (2004)	WIF einzelstehend 1	30 Häufigkeitswörter Gl. Schw. F&V	Standardisiert WPM - - Pen&Paper	- -	USA	- -	Erprobung, Zuwachsraten, Vergleich unterschiedlicher Testkonstruktionen; N = 151, MZP = 33, Paralleltest-R.: 0.88, Krit.-V: 0.27-0.93 (p < keine Angabe)
Schaughency et al. (2015) NZWIF	WIF einzelstehend 2 (entspricht 1)	16 Häufigkeitswörter Gl. Schw. Alle Wortarten	Standardisiert WPM - - Pen&Paper	- -	NZ	- -	Erprobung, Zuwachsraten; N = 120, MZP = 3, Retest-R.: 0.90-0.99, Krit.-V: 0.71-0.89 (p < .05)
Zumenta et al. (2012)	WIF einzelstehend 1	20 Häufigkeitswörter Anst. Vorkommen F&V	- WPM - - Pen&Paper	- -	USA	- -	Erprobung, Zuwachsraten, Vergleich unterschiedlicher Testkonstruktionen; N = 699, MZP = 2, Paralleltest-R.: 0.95-0.97, Krit.-V: 0.65-0.88 (p < .01)
Abu-Hamour (2014a)	ORF einzelstehend 1-3	18 Mehrere Quellen Gl. Schw. abgeschlossen	Standardisiert WPM - - Pen&Paper	- -	JOR	Graphisch -	Erprobung: N = 50-100, MZP = 1, Retest-R.: 0.90-0.94, Krit.-V: 0.44-0.57 (p < .01)
Walter (2011) LDL	ORF einzelstehend 1-9	28 1 Quelle Gl. Schw. zusammenhängend	Standardisiert WPM - - Pen&Paper	- -	DEU	- -	Erprobung inkl. interne Validität, Zuwachsraten; N = 98-162, MZP = 5, Paralleltest-R.: 0.91, Krit.-V: 0.68-0.94 (p < keine Angabe)
Wheldall & Madelaine (2000) WRAP	ORF einzelstehend 3-8	14 Neue Texte Gl. Schw. abgeschlossen	Standardisiert WPM - - Pen&Paper	- -	AUS	- -	Erprobung, Zuwachsraten; N = 20-65; MZP = 1; Paralleltest-R: 0.95-0.96 (p < .001), Krit.-V: 0.49-0.80 (p < .01)

Anmerkungen: ¹Art des Curriculum-Based Measurement. ²Einbettung des Testverfahrens. ³Anzahl Parallelversionen. ⁴Unterricht. ⁵Gleiche Schwierigkeit. ⁶Nomen. ⁷Verben. ⁸Ansteigende Schwierigkeit. ⁹Funktionswörter. ¹⁰Durchführung. ¹¹Kompetenzmaß. ¹²Messzeitpunkt. ¹³Words per Minute. ¹⁴trifft nicht zu bzw. keine Angaben vorhanden. ¹⁵additiv. ¹⁶Reliabilität. ¹⁷Kriteriumsvalidität.

Testkonzeptionen

Die gefundenen Testinstrumente sind alle laute Leseverfahren und können den typischen Arten von CBM zugeordnet werden. Es ist keine grundlegende neu konzipierte Art identifiziert worden. Fünf Instrumente sind WIF-Verfahren und erheben die Leseflüssigkeit mithilfe von Wortlisten (Abu-Hamour, 2014b; Diehl & Hartke, 2011; Fuchs, Fuchs & Compton, 2004; Schaughency, McLennan & McDowall, 2015; Zumeta, Compton & Fuchs, 2012). Die anderen drei Instrumente sind als ORF-Verfahren konstruiert (Abu-Hamour, 2014a; Walter, 2011; Wheldall & Madelaine, 2000). Fast alle vorgestellten Testinstrumente ($n=7$) erheben ausschließlich die Leseflüssigkeit und sind nicht Bestandteil einer übergreifenden Lesekompetenzerfassung. Nur bei Diehl und Hartke (2011) ist der Leseflüßigkeitstest Teil des Inventars zur Erfassung der Lesekompetenz im 1. Schuljahr (IEL-1). Alle Verfahren präsentieren ein Erhebungsformat der Leseflüssigkeit, welches als eindimensional betrachtet und mittels des Summenscores ausgewertet wird.

Die WIF-Verfahren ($n=5$) erheben die Leseflüssigkeit ausschließlich in der ersten Jahrgangsstufe und somit bereits zu Beginn des Leseerwerbs. Die neuseeländische Version der WIF (NZWIF) ist für Kinder des zweiten Schulbesuchsjahrs entworfen, welches allerdings dem Curriculum der ersten Klasse im amerikanischen und europäischen Schulsystem entspricht. Die Autoren weisen darauf hin, dass die erste Klasse eher dem letzten Vorschuljahr entspricht (Schaughency et al., 2015). Die Altersspanne der Zielgruppe variiert bei den ORF-Verfahren ($n=3$) hingegen stark. Zwei ORF-Verfahren geben die Einsetzbarkeit ihrer Texte jahrgangsübergreifend bis zum Ende der Sekundarstufe I (Klassenstufe 8 bzw. 9) an (Walter, 2011; Wheldall & Madelaine, 2000). Ab der fünften Jahrgangsstufe beobachteten Wheldall und Madelaine (2000) im Wheldall Assessment of Reading Passages (WARP) Deckeneffekte und diskutierten die große Al-

tersspanne von acht Jahren. Ein Testverfahren stellt hingegen Texte speziell für jede Jahrgangsstufe einzeln zur Verfügung (Abu-Hamour, 2014a).

Die Instrumente unterscheiden sich hinsichtlich a) ihrer Itemrekrutierung, b) der Anzahl der zur Verfügung gestellten Items und Parallelversionen sowie c) ihrer Struktur. Im Hinblick auf die Itemrekrutierung haben sich die Autoren der WIF-Verfahren ($n=5$) hauptsächlich auf fest definierte Wortschätze gestützt. Die spezifische Itemauswahl und der Bezug zum Curriculum werden über eine altersgemäße Repräsentativität des Wortmaterials begründet. Meistens wurde auf Häufigkeitswortlisten zurückgegriffen, wie z.B. der Dolch-word-list, den 500 häufigsten geschriebenen englischen Wörtern oder einem altersentsprechenden Sichtwortschatz ($n=4$, Diehl & Hartke, 2011; Fuchs et al., 2004; Schaughency et al., 2015; Zumeta et al., 2012). Lediglich ein Wörterpool wurde durch die Sichtung von Unterrichtslesematerialien aus den entsprechenden Klassenstufen neu erstellt (Abu-Hamour, 2014b). Die generierten Wörterpools weisen einen erheblichen unterschiedlichen Umfang auf. Die Itemanzahl variiert zwischen 60 und 1500. Die Listen der meisten Testinstrumente ($n=7$) umfassen 50 Wörter, allerdings werden zwischen 16 und 30 Parallelversionen mit gleichem Schwierigkeitsgrad zur Verfügung gestellt. Das IEL-1 von Diehl und Hartke (2011) enthält nur eine Liste mit 60 Items. Entweder sind alle Wortarten innerhalb einer Liste vertreten oder es wurden Kombinationen aus Nomen und Verben, Nomen und Funktionswörter sowie Funktionswörter und Verben integriert. Keine der Listen enthält einzelne Silben oder Pseudowörter. Somit arbeiten alle mit realen Wörtern. Die Struktur der Wortlisten unterscheidet sich darin, ob die Wörter zufällig oder bewusst angeordnet wurden. Zumeta et al. (2012) ordneten die Wörter nach ihrem ansteigenden Vorkommen. Die ersten Wörter auf der Liste kommen im Sprachgebrauch häufiger vor als die letzten. Das IEL-1 präsentiert die Wörter nach Schwierigkeitsgrad

und das leichtere Wortmaterial ist zuoberst abgebildet (Diehl & Hartke, 2011). Die Wörter auf den zufällig angeordneten Listen weisen laut Autoren die gleiche Schwierigkeit auf.

Die Texte der ORF-Testverfahren ($n=3$) differieren ebenfalls in den oben genannten drei Punkten. Alle Testverfahren stellen zwischen 14 und 28 Parallelversionen zur Verfügung, die nicht an eine bestimmte Reihenfolge gebunden sind. Jeder Text umfasst exakt 200 Wörter. Manche Texte wurden aus bereits publizierter Kinder- und Jugendliteratur übernommen (Abu-Hamour, 2014a; Walter, 2011). Für den WARP-Test haben die Autoren die Texte selbst verfasst, um Lerneffekte durch bekannte Inhalte zu vermeiden (Wheldall & Madelaine, 2000). Abu-Hamour (2014a) verzichtete bewusst auf direkte Rede, Gedichts- und Theaterformen sowie untypische Eigennamen oder Wörter. Die Schwierigkeit aller Parallelversionen innerhalb eines Tests wurden als gleich angenommen und mittels des Flesh-Index oder Expertenmeinungen überprüft. Die Struktur der Passagen unterscheidet sich darin, ob die Texte inhaltlich zusammenhängend sind oder nicht. In zwei Verfahren sind die Tests in sich geschlossen und einzelstehend. Nur die Texte der Lernfortschrittsdiagnostik Lesen (LDL) von Walter (2011) bauen inhaltlich auf einander auf und ergeben in Summe eine vollständige Geschichte.

Theoriegeleitete Leseerwerbskonzepte

Bei den meisten Instrumenten ($n=7$) wird kein Bezug zu einem expliziten theoriegeleiteten Leseerwerbsmodell hergestellt. Es lassen sich hauptsächlich unspezifische Verweise in Anlehnungen an gängige Modelle finden (z.B. Schaughency et al., 2015; Walter, 2011). Im Gesamtzusammenhang des IEL-1 wird das Erwerbsmodell nach Uta Frith als Konstruktionsbasis herangezogen (Diehl, 2011). Die Leseflüssigkeit wird bei allen Testverfahren als robuster Indikator für die Lesekompetenz genannt und es wird die Be-

deutung einer guten Lesekompetenz verdeutlicht. In den Argumentationsketten wird entweder das Konstrukt der Leseflüssigkeit definiert oder sie im Zusammenhang der Testung als Voraussetzung für höhere kognitive Leistungen festgeschrieben. Beispielsweise kann die Leseflüssigkeit als ein Zusammenspiel aus Lesegenauigkeit und Lesegeschwindigkeit definiert werden (Wheldall & Madelaine, 2000). Eine weitere Modellvorstellung schließt die Prosodie als relevanten Faktor der Leseflüssigkeit mit ein (Abu-Hamour, 2014a). Alternativ wird das flüssige Lesen als Prädikator für die Automatisierung im Leseprozess und für das Leseverstehen von Texten aufgezeigt (Abu-Hamour, 2014b; Zumeta et al., 2012). Die Berücksichtigung sprachspezifischer Besonderheiten werden nur selten angegeben ($n=3$). Im Arabischen werden die Besonderheiten der Orthographie sowie der Einschluss von Vokalen in der Schriftsprache erörtert. (Abu-Hamour, 2014a; Abu-Hamour, 2014b). Der kulturabhängige mündliche Sprachgebrauch und sein Einfluss auf die Lesegeschwindigkeit wird in Bezug auf das NZWIF thematisiert (Schaughency et al., 2015).

Realisierung im Unterricht

Um ein Testverfahren gut im Unterricht einsetzen zu können, muss es möglichst ökonomisch und nützlich sein (Souvignier & Förster, 2011). Einflussnehmende Aspekte sind a) die Durchführung der Testungen, b) das gewählte Kompetenzmaß, c) die Variation der Messzeitpunkte sowie d) die Medienart der Testmaterialien.

Alle acht Tests sind laute Leseverfahren und als Einzeltests konstruiert. Somit muss jeweils eine Person mit hoher Lesekompetenz über die Lesefehler urteilen. Die meisten Verfahren ($n=7$) geben eine standardisierte Durchführung mit vorgegebener Arbeits- und Auswertungsanleitung vor. Sollten die Kinder nicht innerhalb von 3 Sekunden anfangen vorzulesen, hilft die begleitende Person mit dem ersten Wort und be-

ginnt mit der Zeitmessung (Schaughency et al., 2015; Wheldall & Madelaine, 2000). Bei zu langem Stocken nennt sie das nächste Wort (Fuchs et al., 2004). Die Kinder sollen während der Testung das Dokumentationspapier über Lesefehler nicht einsehen (Abu-Hamour, 2014a).

Alle Verfahren ziehen als Kompetenzmaß WPM heran und dauern genau eine Minute. In einer Studie wurden Überlegungen über ein zweites Kompetenzmaß für die Leseflüssigkeit angestellt. Dabei mussten die Kinder die 200-Wörter-Passage bis zum Ende lesen und die insgesamt richtig gelesenen Wörter pro Passage (WPP) wurden dokumentiert. Die angenommene Hypothese bestand darin, dass ein schnell lesendes Kind anderes Wortmaterial erliest als ein langsam lesendes Kind. Aufgrund einer sehr hohen Korrelation zwischen den beiden Kompetenzmaßen von 0.96 bis 0.99 wurde von WPP aufgrund der schlechteren Ökonomie Abstand genommen (Wheldall & Madelaine, 2000).

Einige Studien ($n = 5$) verweisen auf mehrere mögliche Messzeitpunkte innerhalb einer Klassenstufe. Implizit lassen sich Hinweise darauf finden, dass wöchentliche Messungen möglich sind (Fuchs et al., 2004; Zumeta et al., 2012). Diehl (2011) benennt als einzige drei deutlich strukturierte Messzeitpunkte (Ende erstes Halbjahr, Mitte und Ende zweites Halbjahr) innerhalb des ersten Schuljahres.

Keine der Studien thematisiert explizit, ob die Tests physisch oder digital durchgeführt werden. An dieser Stelle können nur indirekte Schlussfolgerungen gezogen werden. Zum einen sind Diagnoseinstrumente traditionell als Papierversion verfügbar und auf computergestützte Versionen wird in der Regel hingewiesen. Zum anderen kann durch die methodische Beschreibung der Evaluationsstudien darauf geschlossen werden, dass alle Instrumente lediglich als pen&paper Version existieren. Beispielsweise liefern die Durchführungen Hinweise, wenn die Kinder den begleitenden Personen gegenüber sitzen und die Dokumentations-

unterlagen nicht einsehen dürfen (Abu-Hamour, 2014a; Abu-Hamour, 2014b; Zumeta et al., 2012). Fuchs et al. (2004) sprechen davon, dass die Wortlisten als Arbeitsblatt konstruiert sind.

Flexibilität und additive Materialien

Bei allen acht Instrumenten besteht keine Möglichkeit die vorgegebenen Inhalte an das Curriculum oder an den individuellen Lernstand der Kinder anzupassen. Es werden in sich geschlossene Materialien vorgestellt. Punktuell wird eingeräumt, dass die Entwicklung von statischen Materialien eine Einschränkung im ursprünglichen CBM-Sinn darstellt (Wheldall & Madelaine, 2000). Allerdings wird fortgeführt, dass jeglicher kind- und altersgerechte Lesestoff dem aktuellen Curriculum entspricht.

Kein Verfahren steht mit Fördermaterialien im Zusammenhang. Es wird weder auf passende externe noch auf selbst entwickelte Fördermaterialien verwiesen. Grundsätzlich wird angestoßen, dass die Instrumente den Lehrkräften dazu verhelfen sollen, eigene Fördermaterialien passend zu den individuellen Lernständen zu entwickeln (Abu-Hamour, 2014a; Abu-Hamour, 2014b). Ebenso werden mehr Forschungsaktivitäten im Bereich der Fördermaterialien gefordert, die an formativ ausgelegte Lernverlaufsmessungen geknüpft sind (Diehl, 2011). Eine enge Verknüpfung zwischen Testinstrumenten und Fördermaterialien sollte hergestellt werden, damit Kinder mit Schwierigkeiten im Lesen noch gezieltere Unterstützung erhalten können.

Feedback

Die acht Verfahren sprechen die Auswertungen nur indirekt an. Teilweise wird darauf verwiesen, dass die Lernergebnisse in eine Lernkurve übertragen werden sollen (Abu-Hamour, 2014a; Diehl, 2011). Dies impliziert eine reine quantitative Auswertung mit individueller Bezugsnorm. Hilfestellungen zur Interpretation existieren ausschließ-

lich in Form von Normwerten (Abu-Hamour, 2014a; Diehl, 2011; Fuchs et al., 2004; Schaughency et al., 2015; Walter, 2011; Wheldall & Madelaine, 2000; Zumeta et al., 2012).

Alle Autoren fordern, dass die Erkenntnisse aus den formativen Messungen auf die Unterrichtspraxis übertragen werden. Diesbezüglich werden nur instrumentenunspezifische Aussagen getätigt. Beispielsweise trägt der Einsatz einer formativen Lernverlaufsmessung zu einem besseren Unterricht bei und Informationen werden über den Lernprozess transparent abgebildet (Schaughency et al., 2015; Zumeta et al., 2012). Gänzlich werden keine konkreten Hinweise, wie Interpretationshilfen von Lernverläufen oder abgeleitete Konsequenzen für den Unterricht, genannt. Ebenso wenig wird über ein aufgaben- oder lernverlaufsbezogenes Feedback für die Lernenden berichtet.

Evaluation

Für alle Testverfahren werden mindestens die klassischen Testgütekriterien überprüft. Diese Überprüfungen erfolgen alle nach der klassischen Testtheorie. Es konnten keine Itemanalysen im Sinne der Item-Response-Theorie, der Generalisierbarkeitstheorie oder eine Überprüfung der Messinvarianz ausfindig gemacht werden. Diehl und Hartke (2011) verweisen darauf, dass eine Itemanalyse für das WIF-Instrument des IEL-1 nicht forschungsökonomisch sei. Dieses Verfahren umfasst von allen eingeschlossenen Tests die geringste Anzahl von Items. Der Fokus der Evaluationsstudien liegt auf der Reliabilität und der externen Validität. Die Objektivität wird zumeist kurz durch eine standardisierte Durchführung und Auswertung begründet (z.B. Abu-Hamour, 2014a; Diehl & Hartke, 2011). Die anderen Gütekriterien werden durch Studien mit stark unterschiedlicher Qualität untersucht. Es gibt Unterschiede hinsichtlich a) der Stichproben, b) der Anzahl der Erhebungszeitpunkte und c) der verwendeten Parallelversionen. Die Stichproben variieren zwischen 20 und 1642 Kin-

dern pro Klassenstufe. Einige Studien berichten über Einschlusskriterien, wie beispielsweise dem vorliegenden Einverständnis der Erziehungsberechtigten, die Beherrschung der Untersuchungssprache als Muttersprache oder eine durchschnittliche Intelligenz ($n=4$, Abu-Hamour, 2014a; Abu-Hamour, 2014b; Schaughency et al., 2015) oder über Randomisierungsverfahren ($n=1$, Fuchs et al., 2004). Partiiell wurden die Teilnehmenden bereits vor den Messungen nach ihrer aktuellen Lesefähigkeit klassifiziert und in Subgruppen unterteilt. Zur Erhebung der Daten wurden zwischen einem und 33 Messzeitpunkte gewählt. Entweder wurden nicht exakt bestimmbare Zeitpunkte wie Ende des ersten Schulhalbjahrs oder Frühling (Diehl & Hartke, 2011; Schaughency et al., 2015; Zumeta et al., 2012), eine rhythmisierte Erhebung innerhalb eines definierten Zeitraums (Abu-Hamour, 2014b; Schaughency et al., 2015) oder eine Kombination aus beiden Designs (Fuchs et al., 2004) gewählt. Zur Erhebung der Reliabilität wurden nicht immer alle Parallelversionen verwendet. Wenn eine Auswahl stattgefunden hat, wurde dies allerdings schlüssig durch Randomisierung oder hohe Korrelationen zwischen den Parallelversionen begründet ($n=2$, Schaughency et al., 2015; Wheldall & Madelaine, 2000). Alle Testverfahren berichten von einer mittleren bis sehr hohen Retest- oder Paralleltest-Reliabilität. Zur Überprüfung der Kriteriumsvalidität wurden Korrelationsberechnungen zwischen den Rohwerten der Lernverlaufsinstrumente und den Werten von ein bis fünf landesspezifischen standardisierten Diagnoseverfahren durchgeführt (siehe Tabelle 1). Zusätzlich wurden beim IEL-1 und NZWIF nicht standardisierte Lehrerurteile berücksichtigt (Diehl & Hartke, 2011; Schaughency et al., 2015). Eine genauere Erklärung der Lehrerurteile erfolgt in keiner Studie. Alle Testverfahren eignen sich dazu schwache Leserinnen und Leser zu identifizieren, da sich ihre Testwerte im Vergleich zu starken Leserinnen und Lesern unterscheiden. Nicht flüssig lesende Kinder grenzen sich durch

deutlich niedrigere Mittelwerte mit parallelen Verläufen in den entsprechenden Kompetenzmaßen auf allen Altersstufen ab.

Weiterführende Evaluationen wurden bisher nur in Bezug auf einzelne Testverfahren gewonnen. Für die Verfahren IEL-1 und LDL konnte mittels Gruppen- bzw. Extremgruppenvergleich von Kindern mit und ohne Probleme beim Leselerwerb die interne Validität bestätigt werden (Diehl & Hartke, 2011; Walter, 2011). Für das IEL-1 wurden die Werte von Regelschulklassen mit denen von Förderschulklassen und für das LDL die Werte von vier Niveaugruppen innerhalb einer Klasse abgeschätzt. Über mehrere Messzeitpunkte zeigten sich für beide Instrumente parallele Lernverläufe. Deckeneffekte sind hinsichtlich der Zuwachsraten bei Kindern höheren Alters für ORF-Verfahren zu finden. Ab einem bestimmten Alter nimmt ihre Leistung nicht mehr im gleichen Verhältnis wie bei jüngeren Probandinnen und Probanden zu (Walter, 2011; Wheldall & Madelaine, 2000). Eine eindeutige Altersgrenze konnte allerdings aus den eingeschlossenen Studien nicht ausfindig gemacht werden. Schaughency et al. (2015) erhoben, dass Kinder ohne Schwierigkeiten im Lesen grundsätzlich niedrigere Zuwachsraten für das WIF-Verfahren vorweisen als Kinder mit Schwierigkeiten. Zumeta et al. (2012) fanden zudem heraus, dass sich unterschiedlich strukturierte Arten von Wortlisten für unterschiedliche Zwecke vorteilhaft eignen. Wortlisten, die aus einem kleinen Wörterpool generiert worden sind, eignen sich eher zum Screening mit größeren zeitlichen Abständen zwischen den Messungen. Wortlisten, die aus einem weiteren repräsentativen Wörterpool entstanden sind, bieten Vorteile bei engmaschigen Messungen. Diese weiter gefassten Listen sind zudem änderungssensibler gegenüber schwächeren Schülerinnen und Schülern. Fuchs et al. (2004) wiesen eine eindeutige Überlegenheit von Wortlisten nach, die aus realen Wörtern bestehen. Sie wurden im Hinblick auf Zuwachsraten und der prognostischen Validität mit Pseu-

dowortlisten an derselben Stichprobe verglichen.

Zusammenfassung und Diskussion

Dieses systematische Review untersuchte aktuelle Entwicklungen internationaler und nationaler formativer Lernverlaufsmessungen zur Leseflüssigkeit von Schülerinnen und Schülern in den ersten drei Grundschuljahren. Dafür sind ausschließlich Instrumente herangezogen worden, deren Entwicklung sowie Evaluation in Zeitschriften mit wissenschaftlichem Begutachtungsverfahren publiziert sind. Dies ist eine Limitation der Studie, da nicht alle Testverfahren mit einbezogen wurden und ein Publikationsbias vorliegen könnte. Dieses Forschungsgebiet ist noch sehr jung und dementsprechend könnten beispielsweise unbeachtete Dissertationen vorliegen. Allerdings ist die wissenschaftliche Testentwicklung ein sehr komplexer Vorgang, der mit hohen Standards verbunden ist. Aus Gründen der Qualitätssicherung wurde sich dementsprechend für den ausschließlichen Einbezug von Beiträgen, die durch ein Peer-Review-Verfahren begutachtet worden sind, entschieden. Es gibt Instrumente, welche entweder ohne psychometrische Kennwerte oder ohne Möglichkeit der Nachvollziehbarkeit zur Konstruktion im Internet oder in Printform veröffentlicht sind. Beispielsweise gibt es die kostenpflichtige Testbatterie Dynamic Indicators of Basic Early Literacy Skills (DIBELS), welche ebenfalls WIF- und ORF-Verfahren für die ersten Schuljahre zur Verfügung stellen (Good & Kaminski, 2002). Zur Konstruktion und zur Skalierung einzelner DIBELS, Tests sowie deren Evaluation konnten keine wissenschaftlichen Artikel in dieser systematischen Suche gefunden werden. Ebenso gibt es keine Veröffentlichung der einzelnen Testitems. In einigen begutachteten Artikeln wurden z.B. nur kostenlose DIBELS-Tests zur Überprüfung der Akzeptanz und Umsetzung von CBM in der Schule verwendet. Hierbei

wurden keine Informationen über die Skalierung und die psychometrischen Eigenschaften der Tests benannt (Grima-Farrell, 2014). In einer weiteren Studie wird beispielsweise die Anwendungsgenauigkeit von Lehrkräften in der Ausbildung mittels DIBELS-Tests überprüft (Coulter, Shavin & Gichuru, 2009).

Fast alle vorgestellten Testverfahren ($n=7$) sind weder Bestandteil einer kombinierten Testbatterie noch mit Fördermaterialien verbunden. Nur das IEL-1 ist ein Inventar, welches weitere Aspekte der Lesekompetenz erfasst und einen umfassenderen Einblick in die individuellen Lesefähigkeiten ermöglicht (Diehl & Hartke, 2011). Die Testverfahren haben eine unterschiedliche Anzahl von einer bis zu dreißig Parallelversionen. Somit können sie den Lernverlauf unterschiedlich genau beobachten. Nur Verfahren mit einer ausreichenden Anzahl von Parallelversionen können sicherstellen, dass keine Erinnerungseffekte das Ergebnis verzerren. Zudem bleibt offen, welche Anzahl von Parallelversionen für welches Setting als ausreichend eingestuft werden kann.

Die Leseflüssigkeit wird bei allen Tests als robuster Indikator für die Lesekompetenz angesehen. Die einbezogenen WIF-Verfahren ($n=5$) messen ausschließlich in der ersten Klassenstufe und die ORF-Verfahren ($n=3$) zwischen der ersten und neunten Jahrgangsstufe. Alle Verfahren geben klassenstufenbezogene Normen an. Die verwendeten Items wurden anhand unterschiedlicher Materialien ausgewählt. Beispielsweise wurden Grundwortschätze, Häufigkeitswortlisten oder Wörter aus Unterrichtsmaterialien verwendet. Dabei wurde in den meisten Tests angenommen, dass alle Items (Wörter/ Texte) gleich schwierig sind. Das IEL-1 von Diehl und Hartke (2011) nimmt anhand des Kieler Leseaufbaus kürzere Wörter mit einer Silbe als leichter an als mehrsilbige komplexere Wörter. Zumeta et al. (2012) fragen hingegen zuerst die häufigsten Wörter und seltenere Wörter erst am Ende ab. Die Interpretation der Ergebnisse ist durch die Konstruktion bedingt. Fuchs et al. (2004) und

Zumeta et al. (2012) haben herausgefunden, dass je nach Zusammensetzung der Zielgruppe unterschiedliche Konstruktionsverfahren reliabler und sensibler sind, um den Status und den Lernverlauf zu erfassen. Ein eindeutiges Vorgehen bei der Konstruktion dieser Tests ist aktuell in der Forschung nicht erkennbar. Der Einsatz von Lesetexten ist nach aktuellem Stand in der englischsprachigen Forschung nicht sinnvoll, da aufgrund der Überforderung von schwachen Leserinnen und Lesern Bodeneffekte entstehen können (Fuchs et al., 2004; Good, Simmons & Kame'enui, 2001). Berichte über ähnliche Effekte in Schriftsprachen mit konsequenten Graphem-Phonem-Verbindungen wurden nicht angeführt und bedürfen zukünftiger Überprüfung. Nur so können gezielte Implikationen für deutschsprachige CBM-Verfahren abgeleitet werden.

Als Kompetenzmaß ziehen alle Tests die Anzahl korrekt gelesener Wörter pro Minute heran, die händisch erfasst werden muss. Dafür wird überwiegend eine standardisierte Durchführung und Auswertung vorgegeben, was eine objektive Umsetzung gewährleistet. Konkrete Hilfestellungen zur Implementation im Unterricht, zur spezifischen Förderung und zu einem schülerbezogenen Feedback wurden zu keinem der Verfahren ausgearbeitet. Dies ist jedoch für den Erfolg der Anwendung und Umsetzung im Unterricht nach Fuchs (2004) ein wesentliches Kriterium. Besonders die Verschränkung der Messergebnisse mit konkreten pädagogischen Zielen bleibt den Lehrkräften selbst überlassen. Trotz eindringlicher wissenschaftlicher Forderung (Ardoin, Christ, Morena, Cormier & Klingbeil, 2013) werden keine begleitenden Materialien zum Umgang mit den Lernverlaufdaten angeboten. Die Lehrkräfte müssen die Daten eigeninitiativ interpretieren und Konsequenzen für die einzelnen Schülerinnen und Schüler ziehen. Für das Deutsche ist es wichtig, die individuellen Lesestrategien bei der Förderung im Anfangsunterricht einzubeziehen (Klicpera et al., 2017). Hier liegt es nahe, begleitende Schulungsmaterialien für die Erkennung

nung von sprachspezifischen Strategien und Stolpersteine in den Lernverlaufdaten zu konstruieren und zu evaluieren.

Alle Testverfahren werden als pen&paper Version angeboten. Dies ist für die Durchführung leicht praktikabel. Für die Auswertung, Dokumentation und graphische Darstellung des Lernverlaufs jedoch eher weniger, da alle Werte von Hand verrechnet oder in den Computer übertragen werden müssen. Computergestützte Testverfahren haben hier hohes Potential und eröffnen mehrere Möglichkeiten sowie ein ökonomisches Vorgehen im schulischen Einsatz, wie die Metaanalyse von Maier (2014) vorstellt. In Deutschland stehen inzwischen den meisten Grundschulen Computer mit Internetverbindung zur Verfügung und die Kinder sind den Umgang mit dem PC gewöhnt (Tarelli, Lankes, Drossel & Gegenfurtner, 2012). An dieser Schnittstelle könnten Verbindungen geschaffen werden, um die Ökonomie neuer und vorhandener Instrumente zu verbessern.

Die einbezogenen Instrumente erfassen alle reliabel nach klassischer Testtheorie die Leseflüssigkeit ($n_{\text{Retest-Reliabilität}}=4$, $n_{\text{Paralleltestreliabilität}}=4$). Hierbei ist kritisch anzumerken, dass weitere Überprüfungen der Fairness und Schwierigkeit der Aufgaben mittels IRT, GT oder Strukturgleichungsmodellierung (z.B. Messinvarianzanalysen) bisher nicht nachgewiesen wurden. Dies ist insbesondere für Verfahren mit mehreren Parallelversionen kritisch zu beurteilen, da diese für die Fairness der Messung annehmen, dass alle Versionen gleich schwer sind. Zudem wurden die Instrumente meist nur einmalig in einem Artikel überprüft und die Stichproben wiesen erhebliche unterschiedliche Umfänge ($n=20-1500$) auf. Außerdem wurde meist nur im Querschnitt und nicht im Längsschnitt gemessen. Zu vier Testverfahren wird eine Überprüfung der Retest-Reliabilität und somit der Änderungssensibilität angeführt. In Bezug auf die Anforderungen nach Fuchs (2004) hat kein Verfahren die dritte Stufe erreicht. Eine Überprüfung der Verfahren im Feld fehlt gänzlich (Stu-

fe 3). Dafür sollten vermehrt Interventionsstudien mit Experimentalgruppendesign gewählt werden, die den Nutzen der diagnostischen Informationen in der Schullandschaft überprüfen. Zusätzlich eignet sich die Erhebung des Einsatzes und der Effektivität der Instrumente mittels Befragungen von Lehrkräften sowie Schülerinnen und Schülern. Um eine flexible und realistische Einsetzbarkeit im deutschen Schulsystem zu sichern, sollten in diesen Ansätzen die Begebenheiten inklusiver Klassen sowie Förderschulklassen Berücksichtigung finden. Zusammenfassend lässt sich festhalten, dass weiterführende Forschungsbemühungen sowohl in Bezug auf die bestehenden Instrumente als auch auf die Erweiterung der Praktikabilität in diesem in Deutschland noch jungen Forschungsgebiet noch ausstehen.

Literatur

- Abu-Hamour, B. (2014a). A Pilot Study for Standardizing Curriculum-Based Measurement Oral Reading Fluency (CBM ORF) in Arabic. *Journal of the International Association of Special Education*, 15, 16–26.
- Abu-Hamour, B. (2014b). Using Arabic word identification fluency to monitor first-grade reading progress. *Dyslexia: An International Journal of Research and Practice*, 20, 167–174.
- Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C. & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, 51, 1–18.
- Bennett, R. E. (2011). Formative assessment. A critical review. *Assessment in Education: Principles, Policy & Practice*, 18, 5–25.
- Coulter, G., Shavin, K. & Gichuru, M. (2009). Oral Reading Fluency. Accuracy of Assessing Errors and Classification of Readers Using a 1-Min Timed Reading Sample. *Pre-*

- venting School Failure: Alternative Education for Children and Youth, 54, 71–76.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- Deno, S. L. (2003a). Curriculum-Based Measures: Development and Perspectives. *Curriculum-Based Measures: Development and Perspectives*, 28, 3–12.
- Deno, S. L. (2003b). Developments in Curriculum-Based Measurement. *Developments in Curriculum-Based Measurement*, 37, 184–192.
- Deno, S. L., Mirkin, P.K. & Chiang, B. (1982). Identifying a valid measure of reading. *Exceptional Children*, 49, 36–45.
- Diehl, K. (2011). Innovative Lesediagnostik – Ein Schlüssel zur Prävention von Leserechtschreibschwierigkeiten. *Zeitschrift für Heilpädagogik*, 62, 164–172.
- Diehl, K., Harke, B. & Knopp, E. (2009). Curriculum-Based Measurement & Leeringonderwijsvolgysteem – Konzepte zur theoriegeleiteten Lernfortschrittsmessung im Anfangsunterricht Deutsch und Mathematik? *Zeitschrift für Heilpädagogik*, 60, 122–130.
- Diehl, K. & Hartke, B. (2011). Zur Reliabilität und Validität des formativen Bewertungssystems IEL-1. Inventar zur Erfassung der Lesekompetenz von Erstklässlern. *Empirische Sonderpädagogik*, 3, 121–146.
- Ennemoser, M., Marx, P., Weber, J. & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44, 53–67.
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, 36, 67–81.
- Fuchs, L. S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research. *School Psychology Review*, 33, 188–192.
- Fuchs, L. S., Fuchs, D. & Compton, D. L. (2004). Monitoring Early Reading Development in First Grade: Word Identification Fluency Versus Nonsense Word Fluency. *Exceptional Children*, 71, 7–21.
- Fuchs, L. S., Fuchs, D., Hosp, M. K. & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence. A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, 5, 239–256.
- Gebhardt, M., Diehl, K. & Mühlhling, A. (2016). Online-Lernverlaufsmessung für alle Schülerinnen und Schüler in inklusiven Klassen. *Zeitschrift für Heilpädagogik*, 66, 444–453.
- Gebhardt, M., Heine, J.-H., Zeuch, N. & Förster, N. (2015). Lernverlaufsdagnostik im Mathematikunterricht der zweiten Klasse: Raschanalysen und Empfehlungen zur Adaption eines Testverfahrens für den Einsatz in inklusiven Klassen. *Empirische Sonderpädagogik*, 3, 206–222.
- Gebhardt, M., Schwab, S., Krammer, M. & Gegenfurtner, A. (2015). General and special education teachers' perceptions of teamwork in inclusive classrooms at elementary and secondary schools. *Journal for Educational Research Online / Journal für Bildungsforschung Online*, 7, 129–146.
- Good, R. & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Hrsg.), *Advanced applications of curriculum-based measurement* (The Guilford school practitioner series, S. 61–88). New York, NY: Guilford Press.
- Good, R. H. & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills (6th edition)*. New Milford, NJ: Institute for the Development of Educational Achievement.
- Good, R. H., Simmons, D. C. & Kame'enui, E. J. (2001). The Importance and Decision-Making Utility of a Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third-Grade High-Stakes Outcomes. *Scientific Studies of Reading*, 5, 257–288.
- Gorecki, B. & Landerl, K. (2015). Phonologische Bewusstheit. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47, 139–146.

- Grima-Farrell, C. (2014). Curriculum-Based Measurement of Oral Reading fluency (CBM-R). An objective orientated evaluation study. *Support for Learning, 29*, 370–393.
- Groth, K., Hasko, S., Bruder, J., Kunze, S. & Schulte-Körne, G. (2013). Interventionseffekte bei Lese-Rechtschreibstörung. Evaluation von zwei Förderkonzepten unter besonderer Betrachtung methodischer Aspekte. *Lernen und Lernstörungen, 2*, 161–175.
- Günter, K. B. (1986). Ein Stufenmodell der Entwicklung kindlicher Lese- und Schreibstrategien. In H. Brügelmann (Hrsg.), *ABC und Schriftsprache. Rätsel für Kinder, Lehrer und Forscher* (Lesen und Schreiben, Bd. 1, S. 32–54). Konstanz: Faude.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage, S. 143–171). Berlin: Springer Berlin Heidelberg.
- Helf, S. & Cooke, N. L. (2011). Reading Specialist. Key to a Systematic Schoolwide Reading Model. *Preventing School Failure: Alternative Education for Children and Youth, 55*, 140–147.
- Heyne, N. (2014). *Merkmale des Leseunterrichts im vierten Schuljahr im Fokus videobasierter Analysen*. Münster: Waxmann.
- Hintze, J. M., Owen, S. V., Shapiro, E. S. & Daly, E. J. (2000). Generalizability of oral reading fluency measures. Application of G theory to curriculum-based measurement. *School Psychology Quarterly, 15*, 52–68.
- Huber, C. & Grosche, M. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik, 8*, 312–322.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement, 20*, 1–73.
- Klauer, K. J. (2006). Erfassung des Lernfortschritts durch curriculumbasierte Messung. *Heilpädagogische Forschung, 32*, 16–26.
- Klauer, K. J. (2011). Lernverlaufsdiagnostik - Konzepte, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik, 3*, 207–224.
- Klauer, K. J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (Tests und Trends, Bd. 12, 1. Aufl., S. 1–17). Göttingen, Niedersachsen: Hogrefe Verlag.
- Klicpera, C., Ehgartner, M., Gasteiger-Klicpera, B. & Schabmann, A. (1993). Voraussetzungen für das Leselernen bei lernbehinderten Kindern in der Sonderschule und bei guten und schwachen Lesern in der Grundschule: Eine Längsschnittuntersuchung zur Entwicklung des phonematischen Bewußtseins in der ersten Schulstufe. *Heilpädagogische Forschung, 3*, 104–108.
- Klicpera, C., Schabmann, A., Gasteiger-Klicpera, B. & Schmidt, B. (2017). *Legasthenie – LRS. Modelle, Diagnose, Therapie und Förderung: mit 20 Abbildungen und 100 Übungsfragen* (5. überarbeitete und erweiterte Auflage). München: Ernst Reinhardt Verlag.
- Korhonen, J., Linnanmäki, K. & Aunio, P. (2014). Learning difficulties, academic well-being and educational dropout: A person-centred approach. *Learning and Individual Differences, 31*, 1–10.
- Landerl, K. & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography. An 8-year follow-up. *Journal of Educational Psychology, 100*, 150–161.
- Lankes, E.-M., Vaccaro, D. & Gegenfurtner, A. (2013). Wie kommen Evaluationsteams zu ihrer Einschätzung der Unterrichtsqualität bei externen Evaluationen? *Unterrichtswissenschaft, 41*, 197–215.
- Maier, U. (2014). Computergestützte, formative Leistungsdiagnostik in Primar- und Sekundarschulen. Ein Forschungsüberblick zu Entwicklung, Implementation und Effekten. *Unterrichtswissenschaft, 42*, 69–86.

- Mayer, A. (2010). *Gezielte Förderung bei Les- und Rechtschreibstörungen. Mit 11 Tabellen*. München: Reinhardt.
- McDonald Connor, C., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E. et al. (2009). Individualizing student instruction precisely: effects of Child x Instruction interactions on first graders' literacy development. *Child Development*, 80, 77–100.
- Moosbrugger, H. & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage). Berlin: Springer Berlin Heidelberg.
- Nagler, T., Linkersdörfer, J., Lonnemann, J., Hasselhorn, M. & Lindberg, S. (2016). The impact of text fading on reading in children with reading difficulties. *Journal for Educational Research Online*, 8, 26–41.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel - Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Washington, DC: U.S. Government Printing Office.
- Organisation for Economic Co-operation and Development. (2005). *Formative Assessment. Improving Learning in Secondary Classrooms*. Paris, FR: OECD Publishing and Centre for Educational Research and Innovation.
- Organisation for Economic Co-operation and Development. (2016). *PISA 2015 Ergebnisse. Deutschland*. Zugriff am 06.02.2017. Verfügbar unter <https://www.oecd.org/pisa/PISA-2015-Germany-DEU.pdf>
- Petticrew, M. & Roberts, H. (2012). *Systematic reviews in the social sciences. A practical guide*. Malden, MA.: Blackwell Publ.
- Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L. & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: a meta-analysis of the correlational evidence. *Journal of School Psychology*, 47, 427–469.
- Schaughency, E., McLennan, K. M. & McDowall, P. S. (2015). Development and preliminary evaluation of an adaptation of Word Identification Fluency for beginning readers in New Zealand. *Assessment for Effective Intervention*, 40, 67–80.
- Scheerer-Neumann, G. (2015). *Les-Rechtschreib-Schwäche und Legasthenie. Grundlagen, Diagnostik und Förderung* (1. Aufl.). Stuttgart: Kohlhammer Verlag.
- Snowling, M. J. & Melby-Lervag, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, 142, 498–545.
- Souvignier, E. & Förster, N. (2011). Effekte prozessorientierter Diagnostik auf die Entwicklung der Lesekompetenz leseschwacher Viertklässler. *Empirische Sonderpädagogik*, 3, 243–255.
- Stecker, P. M., Fuchs, L. S. & Fuchs, D. (2005). Using Curriculum-Based Measurement to Improve Student Achievement. Review of Research. *Psychology in the Schools*, 42, 795–819.
- Tarelli, I., Lankes, E.-M., Drossel, K. & Gegenfurtner, A. (2012). Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 137–173). Münster: Waxmann.
- Tarelli, I., Valtin, R., Bos, W., Bremerich-Vos, A. & Schwippert, K. (2012). IGLU 2011: Wichtige Ergebnisse im Überblick. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 11–26). Münster: Waxmann.
- Torgesen, J. K. (2002). The Prevention of Reading Difficulties. *Journal of School Psychology*, 40, 7–26.
- Walter, J. (2009a). Theorie und Praxis Curriculumbasierter Messens (CBM) in Unterricht und Förderung. *Zeitschrift für Heilpädagogik*, 60, 162–170.

- Walter, J. (2009b). Eignet sich die Messtechnik "MAZE" zur Erfassung von Lesekompetenzen als lernprozessbegleitende Diagnostik? *Heilpädagogische Forschung*, XXXV, 62-75.
- Walter, J. (2011). Die Messung der Entwicklung der Lesekompetenz im Dienste der systematischen formativen Evaluation von Lehr- und Lernprozessen. *Zeitschrift für Heilpädagogik*, 6, 204–217.
- Wheldall, K. & Madelaine, A. (2000). A curriculum-based passage reading test for monitoring the performance of low-progress readers: The development of the WARP. *International Journal of Disability, Development and Education*, 47, 371–382.
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik*, 3, 225–242.
- Zumeta, R. O., Compton, D. L. & Fuchs, L. S. (2012). Using Word Identification Fluency to Monitor First-Grade Reading Development. *Exceptional Children*, 78, 201–22

Jana Jungjohann

Technische Universität Dortmund
Entwicklung und Erforschung inklusiver
Bildungsprozesse
Emil-Figge-Straße 50
D-44227 Dortmund
jana.jungjohann@tu-dortmund.de

Erstmalig eingereicht: 06.09.2017

Überarbeitung eingereicht: 02.11.2017

Angenommen: 20.11.2017