

**Theoriegeleitete Lernverlaufsdagnostik zur Erfassung von basalen
Lesekompetenzen in der inklusiven Grundschule – Entwicklung, Evaluation
und Implementation von webbasierten Testverfahren**

vorgelegt von

Jana Jungjohann

als Dissertation zur Erlangung des Grades
einer Doktorin der Philosophie (Dr. phil.)

in der

Fakultät Rehabilitationswissenschaften
der Technischen Universität Dortmund

Dortmund

2019

Betreuer: Prof. Dr. Markus Gebhardt

Betreuer: Prof. Dr. Jan Kuhl

Tag der Disputation: 06.03.2020

Inhaltsverzeichnis

Inhaltsverzeichnis	i
Danksagung	ii
Abkürzungsverzeichnis.....	iii
1. Einleitung	1
2. Theoretische Grundlagen.....	6
2.1 Modelle des Lesens	6
2.2 Leseschwierigkeiten in der inklusiven Grundschule	11
2.3 Lernverlaufsdiagnostik im Bereich Lesen.....	17
3. Fragestellungen der Dissertation	27
4. Vorgelegte Fachpublikationen der Dissertation.....	29
4.1 Artikel 1	29
4.2 Artikel 2	31
4.3 Artikel 3	35
4.4 Artikel 4	40
4.5 Artikel 5	43
5. Diskussion und Limitationen	46
6. Fazit und Implikationen	60
7. Literaturverzeichnis	65
8. Eidesstattliche Versicherung.....	94
9. Anhang.....	95

Danksagung

„Der Mensch wächst mit seinen höheren Zielen.“

Wolfgang P. Schmitter

Die vorliegende Dissertation beschäftigt sich mit der Frage, wie Lernverlaufsdiagnostik den inklusiven Grundschulunterricht bereichert. Sie ist in mehreren Projekten entstanden, in denen mich verschiedenste Persönlichkeiten geprägt und begleitet haben.

Den größten Dank richte ich an Prof. Dr. Markus Gebhardt, der mir während meiner Promotion ununterbrochen mit Rat, Tat und vor allem Zeit zur Seite stand. Seine unzähligen Ideen haben mich inspiriert und mir geholfen, meine Ziele zu erreichen. Mein weiterer Dank gilt Prof. Dr. Jan Kuhl, der mich als Zweitprüfer in allen Fragen konstruktiv beriet. Prof.ⁱⁿ Dr. Kirsten Diehl danke ich für den immer anregenden Austausch. Prof. Dr. Andreas Mühling unterstützte mich nicht nur in Levumi Fragen, sondern hat mir einen neuen Blick auf das wissenschaftliche Leben ermöglicht. Danke.

Für den fachlichen Austausch und die unterstützenden Worte möchte ich von Herzen Sven Anderson, Jeffrey M. DeVries, Claudia Gottwald, Michael Schurig und Sandra Szczecina danken. Liebe Leevke, für all deine Mühen und deine fachliche Kompetenz danke ich dir in ganz besonderer Weise. Für den außeruniversitären Austausch möchte ich zuerst meinen Eltern Eva und Armin sowie Jonas und Hannah danken, die mir doch so oft zuhörten und mich trotzdem uneingeschränkt unterstützten. Meinen Freunden ging es ähnlich. Ich danke euch für die Diskussionen über Wissenschaft, Bildung und deren Zusammenhänge. Ein ganz besonderer Dank gilt Tobi, der mich während meiner Promotionszeit immer wieder geerdet hat. Anna E. - ohne Worte. Dank dir für alles!

Darüber hinaus vielen Dank an meine Hilfskräfte Rebecca Reimering und Leonie Bredtmann, die in jeder Notsituation bei Testungen eingesprungen sind. Nicht zuletzt danke ich allen Schulleitungen, Lehrkräften, Schülerinnen und Schülern sowie Studierenden. Denn erst sie haben meine Forschung ermöglicht.

Dortmund, November 2019

Jana Jungjohann

Abkürzungsverzeichnis

Abkürzung	Bedeutung
DIF	Differential item functioning
ICC	Intra-Klassenkorrelation
IEL-1	Inventar zur Erfassung von Lesekompetenzen im 1ten Schuljahr
KMK	Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland
KLA	Kieler Leseaufbau
RtI-Modell	Response-to-Intervention-Modell
RWpM	Richtig gelesenen Wörter pro Minute
SiL-Levumi	Testverfahren „Silben lesen“ der Onlineplattform Levumi
SinnL-Levumi	Testverfahren „Sinnkonstruierendes Satzlesen“ der Onlineplattform Levumi
SPF	Bedarf an sonderpädagogischer Unterstützung / Sonderpädagogischer Förderbedarf
SPF-L	Sonderpädagogischer Förderbedarf Lernen
SPF-S	Sonderpädagogischer Förderbedarf Sprache
WpM	Anzahl gelesener Wörter pro Minute
ZPID	Leibniz-Zentrum für Psychologische Information und Dokumentation

1. Einleitung

Ein erfolgreicher Lesererwerb ist für alle Schülerinnen und Schüler in den Anfangsjahren der Schulzeit essentiell. Die Qualität der Lesekompetenzen beeinflusst maßgeblich das schulische Lernen und den Berufserfolg (Esser, Wyschkon & Schmidt, 2002; Korhonen, Linnanmäki & Aunio, 2014; Taylor, 2012). Aktuell erreichen fast 20% der deutschen Grundschülerinnen und Grundschüler keine ausreichenden Lesekompetenzen für das Lernen in der Sekundarstufe I, wie die Internationale Grundschul-Lese-Untersuchung (IGLU) 2016 erneut aufzeigt (Hußmann et al., 2017). Verschiedenen Gruppen von Schülerinnen und Schülern wird ein besonderes Risiko zur Ausbildung von Leseschwierigkeiten in der Grundschulzeit zugeschrieben. Zu ihnen zählen Schülerinnen und Schüler mit Sprachentwicklungsstörungen, mit Schwierigkeiten im Verhalten und Empfinden sowie mit sozialen und zuwanderungsbezogenen Disparitäten (Bental & Tirosh, 2007; Diehl, 2010; Haag, Kocaj, Jansen & Kuhl, 2017; Maitz, Paleczek, Seifert & Gasteiger-Klicpera, 2018). Des Weiteren wirken sich gravierende Leseschwierigkeiten negativ auf die individuelle Schulbiographie aus, da mit ihnen eine Zuschreibung des Bedarfs an sonderpädagogischer Unterstützung im Förderschwerpunkt Lernen (SPF-L) einhergehen kann (Heimlich, 2007; Wember & Greisbach, 2018). Die Folgen von unzureichenden Lesekompetenzen im Grundschulalter erstrecken sich bis in die Sekundarstufe I, da sie sich in vielen Fällen ohne effektive Förderung langfristig manifestieren (Gebhardt, Sälzer, Mang, Müller & Prenzel, 2015; Klicpera & Gasteiger-Klicpera, 1993; Klicpera, Schabmann & Gasteiger-Klicpera, 2006; Landerl & Wimmer, 2008). Ziel sollte es daher sein, Leseschwierigkeiten frühzeitig aufzudecken, damit eine präventive Förderung mit unterschiedlicher Intensität erfolgt (Kuhl & Hecht, 2014). Durch eine frühzeitige adaptive Förderung ist es wahrscheinlich, dass auch Schülerinnen und Schüler mit ungünstigen Lernvoraussetzungen einen positiven Lernverlauf im Lesen haben (Martschinke, 2015; Wember, 2001).

Ein möglicher Ansatz zur frühzeitigen Identifizierung von Schülerinnen und Schülern mit Leseschwierigkeiten ist die Lernverlaufsdagnostik (Ardoin, Christ, Morena, Cormier & Klingbeil, 2013). Die Lernverlaufsdagnostik ermöglicht mittels kurzer und leicht handhabbarer Lesetests eine Evaluation der Erfolge des regulären Unterrichts oder einer spezifischen Fördermaßnahme (Gebhardt, Diehl & Mühling, 2016b; Klauer,

2006, 2014; Walter, 2009b, 2011b). Im Unterricht wurden lernförderliche Effekte auf die Leseentwicklung von Schülerinnen und Schülern mit und ohne gravierenden Leseschwierigkeiten durch den Einsatz einer Lernverlaufsdagnostik gefunden (Stecker, Fuchs & Fuchs, 2005). Die Anwendung der Lernverlaufsdagnostik deckt neben erhofften Lernfortschritten auch nicht erwartungsgemäße Lernentwicklungen auf. Durch eine Visualisierung der Lernverläufe in Lernverlaufsgraphen erhalten die Lehrkräfte eine Rückmeldung, welche Schülerinnen und Schüler vom aktuellen Unterricht wenig oder gar nicht profitieren (Espin, Wayman, Deno, McMaster & de Rooij, 2017; van den Bosch, Espin, Chung & Saab, 2017). Somit können ihre Lehrkräfte auf die individuellen Lernbedürfnisse mit intensiven und/oder adaptiven Fördermaßnahmen reagieren, bevor sich die Leseschwierigkeiten verfestigen. Eine entwicklungsorientierte Diagnostik berücksichtigt theoretisch fundierte Entwicklungsmodelle der zu fördernden Kompetenz in der Testentwicklung (Kuhl & Wittich, 2018). Durch eine solche Modellorientierung in der Lernverlaufsdagnostik sind Lernschritte in unterschiedlichen Lesestrategien zu verorten. Für die Schulpraxis werden daher insbesondere Lernverlaufstests gefordert, welche in Anlehnung an differenzierte (Lese-)Modelle konstruiert sind (Diehl, Hartke & Knopp, 2009; Förster & Souvignier, 2011).

Klauer (2006) stellt in Deutschland als einer der ersten Forscher die Prinzipien und den Bedarf des Forschungsfeldes Lernverlaufsdagnostik vor. Seitdem wurden wenige deutschsprachige Leseverfahren von Forschenden publiziert. Das langsame aber stetige Wachstum in der Veröffentlichung von Lernverlaufstests ist u.a. darin begründet, dass die methodischen Ansprüche bei der Testentwicklung sehr hoch sind (Wilbert, 2014). Bisher waren nur wenige Forschende bereit, diesen Aufwand zu leisten. Die Tests müssen gleichzeitig testtheoretische und schulpraktische Anforderungen erfüllen. Sie müssen sowohl reliabel und änderungssensibel messen (Wilbert & Linnemann, 2011; Wilbert, 2014) als auch durch eine kurze Durchführungszeit für die anwendenden Lehrkräfte praktikabel sein (Deno, 2003a; Fuchs, 2017). Die Nutzerfreundlichkeit und die reliablen Messungen widersprechen sich teilweise gegenseitig und sind daher die große Herausforderung für die Testkonstruktion und -evaluation. Erschwerend kommt hinzu, dass die Lernverlaufstests im optimalen Fall auch verschiedene Leistungsniveaus beachten sollten. Seit der Ratifizierung der UN-Behindertenrechtskonvention (UN-BRK; United

Nations, 2006) werden im deutschen Bildungssystem zunehmend mehr Schülerinnen und Schüler mit Bedarf an sonderpädagogischer Unterstützung (SPF) in inklusiven Settings unterrichtet (Lange, 2017). Durch die gemeinsame inklusive Beschulung verbreitert sich das Leistungsspektrum in den Klassen (Gebhardt, 2015). Dies liegt daran, dass Schülerinnen und Schüler mit SPF mit einem langsameren Tempo lernen und viele Lernhürden bewältigen müssen. Dadurch entstehen Leistungsunterschiede von mehreren Schuljahren zu ihren Mitschülerinnen und Mitschülern ohne SPF (Wocken & Gröhlich, 2009). Damit die Lernentwicklungen von allen Schülerinnen und Schülern einer inklusiven Lerngruppe mittels der Lernverlaufsdagnostik überprüft werden können, muss diese auf allen Leistungsniveaus erfolgreich messen und jahrgangsübergreifend einsetzbar sein. Insbesondere im unteren Leistungsbereich ist ein erfolgreiches Messen notwendig, damit Leseschwierigkeiten frühzeitig und differenziert aufgedeckt werden. Für den inklusiven Grundschulunterricht stehen praktikable Instrumente noch gänzlich aus, was ein Forschungsdesiderat darstellt. Aktuell gibt es für den Primarbereich keine deutschsprachigen Leseverfahren, die der Lernverlaufsdagnostik entsprechen, theoriegeleitet konstruiert, über mehrere Jahrgangsstufen einsetzbar sind und sehr geringe Leseleistungen ohne Bodeneffekte in inklusiven Klassen differenziert messen.

Die vorliegende Dissertation beschäftigt sich daher mit der Entwicklung und Evaluation von theoriegeleiteten Lesetests zur Lernverlaufsdagnostik sowie mit Barrieren und Förderfaktoren zur Implementation dieser in der inklusiven Grundschule. Das Ziel ist es, die Bedarfe und Anforderungen der Schulpraxis zu analysieren und empirisch geprüfte Lernverlaufstests zur Erfassung von basalen Lesekompetenzen für leistungsheterogene Lerngruppen bereitzustellen. In Anlehnung an die Leseforschung wird die Erfassung der Leseflüssigkeit auf mehreren Ebenen sowie das Leseverständnis auf Satzbasis betrachtet (Christmann & Groeben, 1999; Fuchs, Fuchs, Hosp & Jenkins, 2001; Kintsch & Rawson, 2011; Kuhn & Stahl, 2003; Perfetti, 1985; Perfetti, Landi & Oakhill, 2011). Dafür werden theoriegeleitete Instrumente zur Lernverlaufsdagnostik von Lesekompetenzen entwickelt und für verschiedene Gruppen von Schülerinnen und Schülern evaluiert. Außerdem werden die Möglichkeiten und Grenzen der Nutzung und Implementation aufgezeigt. Dabei wird ein besonderer Fokus auf das Messen der Lesekompetenzen von Schülerinnen und Schüler mit SPF sowie von welchen mit lesebezogenen Disparitäten gelegt, um

speziell ihre Bedürfnisse zu adressieren. Dadurch wird ein Beitrag für den inklusiven Leseunterricht in Deutschland geleistet.

Zur Erreichung der gesetzten Ziele werden insgesamt fünf Publikationen vorgestellt. Der erste Beitrag ist ein systematisches Review, in welchem literaturbasiert Lernverlaufsdiagnostikinstrumente zur Erfassung der Leseflüssigkeit im Hinblick auf die Testkonstruktion und -evaluation, den Praxiseinsatz und unterstützende Materialien zur Implementation analysiert werden. Dieser Beitrag gibt einen Überblick über aktuelle Instrumente und leitet Bedarfe für den Leseanfangsunterricht ab. Die zweite Studie stellt die theoriegeleitete Testkonstruktion der Leseflüssigkeitstests der Onlineplattform Levumi (Gebhardt et al., 2016b) vor. Es werden die vorhandenen Niveaustufen der Leseflüssigkeitstests am Beispiel des Tests Silbenlesen (SiL-Levumi; Jungjohann, Diehl & Gebhardt, 2019) beschrieben und die Reliabilität sowie die Einsatzfähigkeit des SiL-Levumi Tests für Schülerinnen und Schüler mit SPF im Quer- und Längsschnitt empirisch geprüft. Inhalte der dritten vorgestellten Studie sind die theoriegeleitete Entwicklung eines neu konstruierten Lernverlaufstests zur Messung des Leseverständnisses auf Satzbasis (SinnL-Levumi; Jungjohann & Gebhardt, 2019) sowie die empirische Evaluierung des Instruments für unterschiedliche Gruppen von Schülerinnen und Schülern der dritten Jahrgangsstufe. In der vierten empirischen Studie werden der dritte und vierte Jahrgang einer inklusiven Grundschule im Längsschnitt über ein gesamtes Schuljahr mit drei Leseflüssigkeitstests (Silben, Wörter, Pseudowörter) und dem SinnL-Levumi Test begleitet, um explorativ mögliche Lernverläufe und den jahrgangsübergreifenden Einsatz der Instrumente für dieses Setting zu prüfen. Mit dem Ziel, die Implementation in der Praxis vorzubereiten, diskutiert der fünfte Beitrag Förderfaktoren und Barrieren für die Lernverlaufsdiagnostik in der deutschen Schulkultur sowie Hürden im Umgang mit Lernverlaufsgraphen.

Erster Artikel

Jungjohann, J., Gegenfurtner, A. & Gebhardt, M. (2018). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit. *Empirische Sonderpädagogik*, 10(1), 100-118.

Zweiter Artikel

Jungjohann, J., DeVries, J. M., Gebhardt, M. & Mühling, A. (2018) Levumi. A web-

based curriculum-based measurement to monitor learning progress in inclusive classrooms. In K. Miesenberger, G. Kouroupetroglou & P. Penaz (Eds.), *Computers helping people with special needs. ICCHP 2018. Lecture Notes in Computer Science*, (pp. 369-378). Wiesbaden: Springer. https://doi.org/10.1007/978-3-319-94277-3_58

Dritter Artikel

Jungjohann, J., DeVries, J.M., Mühling, A. & Gebhardt, M. (2018). Using theory-based test construction to develop a new curriculum-based measurement for sentence reading comprehension. *Frontiers in Education*, 3,1. <https://doi.org/10.3389/feduc.2018.00115>

Vierter Artikel

Jungjohann, J., Schurig, M. & Gebhardt, M. (2019). Konstruktion einer Lernverlaufsdagnostik im Lesen für den Längsschnitt. Messungen mit Leseflüssigkeits- und Leseverständnistests in der 3. und 4. Jahrgangsstufe in einer inklusiven Grundschule. *Manuskript zur Veröffentlichung eingereicht*.

Fünfter Artikel

Jungjohann, J., Diehl, K., Mühling, A. & Gebhardt, M. (2018). Graphen der Lernverlaufsdagnostik interpretieren und anwenden – Leseförderung mit der Onlineverlaufsmessung Levumi. *Forschung Sprache*, 6(2), 84-91. <http://dx.doi.org/10.17877/DE290R-19806>

Im theoretischen Teil der Arbeit werden die zu messenden Leseteilkompetenzen definiert. Für die Klärung des späteren Einsatzzwecks der Lernverlaufsdagnostik werden spezifische Leseschwierigkeiten und die Bedingungen des inklusiven Unterrichts hergeleitet. Anschließend wird das Konzept der Lernverlaufsdagnostik mit seinen spezifischen Charakteristika erörtert und eine testtheoretische Rahmung zur Evaluation geschaffen. Danach werden die Forschungsfragen der kumulativen Dissertation genannt und die zu Grunde liegenden Beiträge im Entstehungskontext zusammengefasst. Die Erkenntnisse aus den einzelnen Studien werden mit theoretischen und empirischen Befunden verbunden und in einer Gesamtdiskussion zusammengeführt. Daraus werden Implikationen für die Testentwicklung extrahiert und Empfehlungen für zukünftige Forschungsaktivitäten im Kontext der Lernverlaufsdagnostik im Bereich Lesen abgeleitet.

2. Theoretische Grundlagen

2.1 Modelle des Lesens

Lesen ist ein hochkomplexer Prozess. Zahlreiche Teilkompetenzen müssen ineinandergreifen, bis das übergeordnete Ziel der Sinnkonstruktion des Gelesenen automatisiert und erfolgreich abläuft. Innerhalb des Leseprozesses wird zwischen basalen (hierarchieniedrigen) und weiterführenden (hierarchiehohen) Prozessen differenziert. Die basalen Leseprozesse fungieren als Grundvoraussetzungen für das weiterführende Lesen (Richter & Müller, 2017). Basale Lesekompetenzen umfassen kognitive Prozesse des Wortlesens, Leseverständnisleistungen auf Wort- und Satzebene sowie die Herstellung einer lokalen Kohärenz auf Satzebene (Lenhard, 2019b; National Institute of Child Health and Human Development, 2000; Rosebrock & Nix, 2017). Zur Beschreibung von Lesekompetenzen und deren Entwicklungen in der Grundschule lassen sich zwei grundsätzliche Perspektiven betrachten, die Entwicklungsmodelle und die Prozessmodelle (Mayer, 2016; Reber, 2017). Entwicklungsmodelle des Lesenlernens beschreiben den Leseerwerb im zeitlichen Verlauf bei Kindern und Jugendlichen. Innerhalb dieser Modelle werden zumindest im ungestörten Schriftspracherwerb ähnliche Erwerbsverläufe und unterscheidbare Phasen mit dominanten Strategien angenommen (Euker, Kuhl & Probst, 2012). An Entwicklungsmodellen werden fehlende Erklärungsansätze zur Art und Weise des Erwerbs der Lesekompetenzen kritisiert. In Abgrenzung dazu beschreiben Prozessmodelle den Leseprozess anhand der zugrundeliegenden kognitiven Kompetenzen, deren Einflussfaktoren und deren Zusammenwirken. Die Prozessmodelle fokussieren sich überwiegend auf das Wortlesen und ermöglichen Erklärungsansätze bei Schwierigkeiten während des Leseprozesses. An ihnen wird kritisiert, dass sie weder auf Entwicklungswege eingehen noch die persönliche Bedeutung des Lesens in einer konkreten Situation berücksichtigen (Rosebrock & Nix, 2017). Im Folgenden wird durch eine modellübergreifende Betrachtung eine theoretische Basis für die Messung von basalen Lesekompetenzen in der Grundschule geschaffen.

Zu Beginn der Grundschulzeit erwerben Schülerinnen und Schüler zunächst die Fähigkeit einzelne Wörter zu erlesen. Entwicklungsmodelle zum Erwerb des

Wortlesens wurden seit den 1980er Jahren unter psycholinguistischen, kognitionspsychologischen und entwicklungspsychologischen Einflüssen entwickelt (Mayer, 2010; Scheerer-Neumann, 2015). Das erste Stufenmodell aus dem angloamerikanischen Raum von Frith (1985) und die darauf basierende Adaption für die deutsche Schriftsprache von Günther (1986) gelten als erste Modellannahmen zum Wortlesen. Diese Modelle gehen von voneinander abgrenzbaren Phasen aus, bei denen sich die kognitiven Abläufe mit zunehmendem Kompetenzzuwachs qualitativ umstrukturieren. Günthers Adaption wird inzwischen angezweifelt, da für die beschriebene logographische Phase keine ausreichend hohe Bedeutung im deutschsprachigen Leseerwerb nachgewiesen wurde (Klicpera & Gasteiger-Klicpera, 1993; Wimmer, Hartl & Moser, 1990). Das Kompetenzentwicklungsmodell des Lesens von Klicpera, Schabmann, Gasteiger-Klicpera und Schmidt (2017) ist ein Entwicklungsmodell, welches ebenfalls den Erwerb von basalen Wortlesekompetenzen in einer regulären Schriftsprache, wie dem Deutschen, beschreibt. Diesem Modell kommt eine besondere Bedeutung zu, da es Forschungsbefunde aus dem deutschsprachigen Raum zum Schriftspracherwerb sowie den Einfluss der (schulischen) Instruktionen berücksichtigt. Zusammengefasst beschreibt das Kompetenzentwicklungsmodell in Anlehnung an Prozessmodelle des Wortlesens (Coltheart, 1978; Scheerer-Neumann, 1981), dass Lesende parallele Strategien zum Wortlesen ausbilden. Zu Beginn des Leseerwerbs dominiert das nicht lexikalische Lesen von Wörtern, welches mit zunehmender Leseerfahrung von dem partiell lexikalischen Lesen und dem lexikalischen Lesen als primäre Lesestrategie abgelöst wird. Das nicht lexikalische Lesen umfasst dabei die lautsprachliche Erschließung über die Phonem-Graphem-Struktur einer Graphemfolge. Diese Lesestrategie nutzen Lesende mit elaborierten Lesekompetenzen immer noch bei Pseudowörtern und unbekanntem fremdsprachigen Wörtern. Bei dem partiell lexikalischen Lesen erfolgt ein Abruf von lexikalischen Verarbeitungseinheiten (z.B. Silben, häufige Graphemcluster oder Morpheme) aus dem mentalen Lexikon und beim lexikalischen Lesen werden vollständige Wörter abgerufen. Klicpera et al. (2017) unterteilen den Erwerb des Wortlesens in Anlehnung an Ehri (1999) in die präalphabetische Phase und die alphabetische Phase. Beide Phasen werden durch die Einschulung voneinander getrennt. In der alphabetischen Phase bilden die Schülerinnen und Schüler mithilfe von schulischen Leseinstruktionen die zuvor genannten Wortlesestrategien in Abhängigkeit von spezifischen Vorläuferfähigkeiten

aus. Unter Vorläuferfähigkeiten werden internale Fertigkeiten und Funktionen verstanden, die gemeinsam mit externalen Einflussfaktoren die schriftsprachliche Informationsverarbeitung in dieser Phase begünstigen oder hemmen. Je nach didaktischer Ausrichtung der Leseinstruktionen kommen Einflussfaktoren im Erstleseunterricht in unterschiedlicher Ausprägung zum Tragen (Klicpera et al., 2017). Beispielsweise können die phonologische Bewusstheit, das Arbeitsgedächtnis, die Benennungsgeschwindigkeit, der Wortschatz, die allgemeine linguistische Kompetenz, der sozioökonomische Status sowie die häusliche Lernumgebung das Lesenlernen beeinflussen (Bus & von Ijzendoorn, 1999; Duzy, Ehm, Souvignier, Schneider & Gold, 2013; Ennemoser, Marx, Weber & Schneider, 2012; Juska-Bacher, Beckert, Stalder & Schneider, 2016; Mayer, 2018; Nagler, Lindberg & Hasselhorn, 2018; Vellutino, Fletcher, Snowling & Scanlon, 2004). Mit fortschreitender alphabetischer Phase speichern die Lesenden größere lexikalische Verarbeitungseinheiten in ihrem mentalen Lexikon ab, wodurch dieses wächst und sich ausdifferenziert. Mithilfe des ausdifferenzierteren mentalen Lexikons erhöht sich die Automatisierung des Wortlesens und das Lesen wird flüssiger (Perfetti, 1985).

Der Ausbildung einer angemessenen Leseflüssigkeit (engl. reading fluency) wird ein besonderer Stellenwert im Leseerwerb von basalen Kompetenzen zugeschrieben. Die Leseflüssigkeit fungiert als Brücke zwischen Dekodier- und Verständnisleistungen (García & Cain, 2014; Gold, 2018; Hudson, Pullen, Lane & Torgesen, 2008; National Institute of Child Health and Human Development, 2000). Sie ist ebenfalls ein multidimensionales Konstrukt und im Zuge der wissenschaftlichen Auseinandersetzung entwickelten sich unterschiedliche Definitionen und Konzepte (Wolf & Katzir-Cohen, 2001). Stark vereinfacht wird aktuell unter der Leseflüssigkeit die Fähigkeit zum automatisierten und unbewusst ablaufenden Lesen von Texten verstanden (Rosebrock, Nix, Rieckmann & Gold, 2017) und geht über das in den 1970er Jahren vorherrschende enge Verständnis des kompetenten Worterkennens hinaus. Kuhn, Schwanenflugel, Meisinger, Levy und Rasinski (2010) kommen zu folgender Definition:

Fluency combines accuracy, automaticity, and oral reading prosody, which, taken together, facilitate the reader's construction of meaning. It is demonstrated during oral reading through ease of word recognition, appropriate

spacing, phrasing, and intonation. It is a factor in both oral and silent reading that can limit or support comprehension. (Kuhn et al., 2010, S. 240)

Auf dieser kognitionspsychologischen Definition basierend stellt die Leseflüssigkeit ein dreiteiliges Zusammenspiel aus Lesegenauigkeit (engl. reading accuracy), Lesegeschwindigkeit (engl. reading automaticity) und prosodischen Aspekten (engl. prosody) dar. Diese Komponenten interagieren und beeinflussen sich gegenseitig (Gold, 2009; Gold, Nix, Rieckmann & Rosebrock, 2011; Lenhard, 2019a; Rosebrock & Nix, 2017; Wolf & Katzir-Cohen, 2001). Die Lesegenauigkeit beschreibt, wie gut die lautsprachliche Zuordnung von Phonemen zu Graphemen (Rekodierfähigkeit) mit anschließender Erfassung der Wortbedeutung (Dekodierfähigkeit) unter Berücksichtigung einflussnehmender Faktoren abläuft (Scheerer-Neumann, 2006). Zur Beurteilung der Lesegenauigkeit wird die Anzahl der richtig laut vorgelesenen Wörter pro Minute (RWpM) gemessen (Holle, 2010). Bei dieser Messung wird häufig nicht kontrolliert, ob die Wortbedeutung konstruiert wurde. Die Lesegenauigkeit beeinflusst die Geschwindigkeit des Lesens, da die Lesegeschwindigkeit als Produkt zwischen der Lesegenauigkeit und der automatisierten Worterkennung verstanden wird (Moors & de Houwer, 2006; Samuels, 2006). Operationalisiert wird nach Holle (2010) die Lesegeschwindigkeit als Anzahl gelesener Wörter pro Minute (WpM). Prosodische Aspekte werden als das selbstständige Betonen und Segmentieren in semantisch und syntaktisch angemessene Einheiten während des Lesevorgangs beschrieben (Kuhn et al., 2010; Rasinski, 2004). Diese greifen erst, wenn Wörter in einem Zusammenhang gelesen werden, also auf Phrasen-, Satz- und Textebene.

Wenn Grundschülerinnen und Grundschüler flüssig Wörter rekodieren und ihre Bedeutung dekodieren, sind diese Teilkompetenzen für das Verständnis auf Satzebene noch nicht ausreichend. Sätze sind mehr als die reine Summe der einzelnen Wortbedeutungen. Für ein eindeutiges Verständnis verarbeiten die Lesenden die semantische Oberflächenstruktur und die syntaktische Tiefenstruktur eines Satzes inkrementell (Christmann & Groeben, 1999). Diese Verarbeitung wird unter dem Konzept der lokalen Kohärenzbildung zusammengefasst (Lenhard, 2019b). In Anlehnung an die Kasusgrammatik nach Fillmore (1968) und Chafe (1970) und die bis heute anhaltende Vorstellung der Sprachverarbeitung unterteilen Lesende Sätze in kognitive Bedeutungseinheiten. Dabei wird der Bedeutungsgehalt eines Satzes durch die Prädikat-Argument-Struktur repräsentiert (Christmann & Groeben, 1999).

Bei der Verarbeitung der semantischen Oberflächenstruktur eines Satzes werden mehrere Wörter zu bedeutungstragenden Propositionen gruppiert. Dabei assoziieren die Lesenden bereits während des Leseprozesses eine möglichst wahrscheinliche Satzkonstruktion und weisen den einzelnen Wörtern eine bestimmte Position zu. In jedem Satz gibt es ein eindeutiges Prädikat (z.B. Zustände, Ereignisse, Eigenschaften), welches mehrere mögliche Argumente (z.B. Objekte, Personen, Sachverhalte) impliziert. Dabei spezifizieren stets die Argumente eines Satzes das Prädikat. Je nach Kontext sind die vorhandenen Propositionen eines Satzes allerdings nicht eindeutig (z.B. Passivkonstruktionen), sodass die assoziierte semantische Satzstruktur revidiert werden muss. Für eine eindeutige Bedeutungszuweisung wird zusätzlich die syntaktische Tiefenstruktur eines Satzes verarbeitet. Dieser Vorgang, also die Analyse der Syntax, wird im Deutschen Parsing genannt (Klicpera et al., 2017). Während des Parsens werden Wörter und Wortgruppen einzelnen Phrasen (z.B. Nominal- oder Verbalphrasen) sowie syntaktischen Funktionen (z.B. Subjekt, Prädikat, Objekt, Präpositionen, Konjunktionen) zugeordnet und mit der semantischen Bedeutung in Beziehung gesetzt (Christmann & Groeben, 1999). Dabei nutzen die Lesenden unterschiedliche Grundprinzipien (z.B. „minimal-attachment“-Prinzip, „late-closure“-Prinzip), um die syntaktischen und semantischen Informationen eindeutig und effizient zu verarbeiten (Meng & Bader, 2000). Es bestehen unterschiedliche Ansichten, ob die semantischen und syntaktischen Analysen parallel oder modular ablaufen oder ob eine Analyse bevorzugt genutzt wird (Kennison, 2009; Taraban & McClelland, 1990).

Zusammengefasst bilden Grundschülerinnen und Grundschüler zunächst das Wortlesen mit dem Ziel einer angemessenen Leseflüssigkeit aus. Dabei entwickeln sie unterschiedliche Lesestrategien, um das Wortmaterial sowohl phonologisch zu rekodieren als auch automatisiert aus ihrem mentalen Lexikon abzurufen. Zusätzlich stehen basale Verständnisleistungen auf Wort- und Satzebene unter Berücksichtigung syntaktischer und semantischer Strukturen im Fokus des Leseerwerbs. Diese Kompetenzen gelten als Grundvoraussetzung zum Entnehmen und Verstehen von Informationen aus Texten, welche in der Sekundarstufe I vorausgesetzt werden (Beste et al., 2012; Scheerer-Neumann, 2015). Die Ausbildung von basalen Lesekompetenzen ist ein langandauernder Prozess. Bei Schülerinnen und Schülern ohne Schwierigkeiten erstreckt sich dieser Prozess mindestens über die

ersten beiden Grundschuljahre (Klicpera et al., 2017). Lesende mit lesespezifischen Lernhürden benötigen deutlich mehr Zeit, um die notwendigen Lesekompetenzen zu erwerben. Bei ihnen kann der Leselernprozess über das Ende der Grundschule hinaus andauern (Klicpera et al., 2017).

2.2 Leseschwierigkeiten in der inklusiven Grundschule

Lesen fordert von vielen Schülerinnen und Schülern eine große Anstrengung und gelingt unterschiedlich gut. Der Unterschied zwischen guten und schlechten Leserinnen und Lesern besteht darin, wie effizient die Lesenden die einzelnen kognitiven Teilprozesse bewältigen (Perfetti, 2007; Richter & Christmann, 2009; Scheerer-Neumann, 2015). Bereits zu Beginn des Schuleintritts werden individuelle Unterschiede im Lesen durch lesebezogene Prädiktoren und korrelierende Leseteilkompetenzen bedingt (Bowey, 2011; Cromley & Azevedo, 2007; Ennemoser et al., 2012). Je nach Ausprägung der Prädiktoren und didaktischer Ausrichtung des Unterrichts bereiten diese Unterschiede Schwierigkeiten im Lesen. Leseschwierigkeiten treten in allen Teilkompetenzen des basalen Lesens auf. Sowohl die Leseflüssigkeit als auch das Leseverständnis kann ineffizient ablaufen. In der Leseflüssigkeit können Schülerinnen und Schüler isolierte Defizite beim Wortlesen, in der Lesegeschwindigkeit sowie in der Lesegenauigkeit entwickeln (Catts, Adlof & Weismer, 2006; Schabmann & Kabicher, 2007; Stothard & Hulme, 1995; Wimmer, Mayringer & Landerl, 2000). Untersuchungen von prosodischen Aspekten im Kontext von Leseschwierigkeiten sind in der aktuellen Forschung noch rar und eine differenzierte Befundlage kann nicht angeführt werden (Nix, 2011; Sappok & Fay, 2018). Konsequenzen einer unzureichenden Leseflüssigkeit sind, dass die kognitiven Verarbeitungskapazitäten für das Lesen und Verstehen von Wörtern gebündelt werden und sie für das basale Leseverständnis folglich fehlen (Lenhard, 2019b; Perfetti, 2007). Diese Lesenden dekodieren fehlerhaft, wodurch ihnen dann das Verstehen des Inhalts schwerfällt. In der Grundschule treten Schwierigkeiten in der Leseflüssigkeit und im Leseverständnis häufig gemeinsam auf (Catts, Hogan & Fey, 2003). Dies wird darüber erklärt, dass in der ersten Hälfte der Grundschule das Wortlesen und die Leseflüssigkeit das Leseverständnis maßgeblich beeinflussen (García & Cain, 2014; Juska-Bacher et al., 2016; Schabmann & Kabicher, 2007). Seltener hingegen werden Lesende mit isolierten Schwierigkeiten in der Leseflüssigkeit ohne Einbußen des Verständnisses oder mit ausreichender

Leseflüssigkeit ohne adäquates Leseverständnis identifiziert (Catts et al., 2006; Hamilton & Shinn, 2003; Hoover & Gough, 1990; Leach, Scarborough & Rescorla, 2003; Meisinger, Bradley, Schwanenflugel, Kuhn & Morris, 2009; Spencer, Quinn & Wagner, 2014; Stothard & Hulme, 1995). Erklärungsansätze für isolierte Verständnisprobleme auf Wortebene liegen beispielsweise in den Bereichen Sprachverständnis und -verarbeitung, Wortschatz und/oder grammatikalisches Wissen (Catts, Compton, Tomblin & Bridges, 2012; Juska-Bacher et al., 2016; Kendeou, Papadopoulos & Spanoudis, 2012; Martohardjono et al., 2005; Nation, 2011; Richter, Isberner, Naumann & Neeb, 2013; Waltzman & Cairns, 2000). Auf Satzbasis wird das Verständnis zusätzlich durch die verwendeten Wörter bzw. Wortarten sowie durch die syntaktischen und semantischen Satzstrukturen beeinflusst (Ecalte, Bouchafa, Potocki & Magnan, 2013; Oakhill, Cain & Bryant, 2003; Waltzman & Cairns, 2000). Für den schulischen Kontext resultieren aus den zuvor genannten internationalen Forschungsergebnissen mehrere Leseschwierigkeitstypen, die in der Grundschule auftreten können. In der Leseflüssigkeit gibt es Schülerinnen und Schüler, die a) langsam ohne viele Fehler lesen, b) eine angemessene Lesegeschwindigkeit bei einer geringen Lesegenauigkeit haben oder c) in allen Bereichen der Leseflüssigkeit Schwierigkeiten zeigen (Scheerer-Neumann, Schnitzler & Ritter, 2008). Das Leseverständnis scheitert hingegen, wenn die Schülerinnen und Schüler a) ein ausreichendes Sprachverständnis haben, aber nicht flüssig lesen, b) sie trotz guter Dekodierleistungen keine inhaltlichen Zusammenhänge zwischen den Komponenten eines Satzes herstellen können oder c) sowohl das Dekodieren als auch die Bildung der lokalen Kohärenz erfolglos verlaufen (Knoepke, Richter, Isberner, Neeb & Naumann, 2013; Nix, 2011).

Kompetenzvergleichsstudien zeigen stark variierende Lesekompetenzen bei deutschen Schülerinnen und Schülern am Ende der Grundschule. In der IGLU-Studie 2016 wurde im Vergleich zu den anderen teilnehmenden Staaten eine besonders hohe Streuung in den Leseleistungen der deutschen Viertklässlerinnen und Viertklässler festgestellt (Hußmann et al., 2017). Knapp über 11% der deutschen Schülerinnen und Schüler erreichen in dieser Lesestudie die höchste Kompetenzstufe. Sie können Aussagen aus zusammenhängenden Texten eigenständig konstruieren, interpretieren und kombinieren. Ihre sehr guten Leseleistungen werden das weitere Lernen in Sekundarstufe wahrscheinlich positiv

beeinflussen. Annähernd doppelt so groß wie die Gruppe der sehr gut Lesenden ist die Gruppe der Schülerinnen und Schüler mit unzureichenden Leseleistungen. Diese Gruppe umfasst 18.9% der Gesamtstichprobe. Ihr wird ein großes Risiko vorher gesagt, „dass sie [die Gruppe der Schülerinnen und Schüler mit unzureichenden Leseleistungen] in der Sekundarstufe I mit erheblichen Schwierigkeiten beim Lernen in allen Fächern konfrontiert sein wird“ (Bos, Valtin, Hußmann, Wendt & Goy, 2017, S. 15). Zu dieser Gruppe zählen Schülerinnen und Schüler mit unterschiedlich ausgeprägten Leseschwierigkeiten. Je nach Umfang, Dauer und Ausprägung der Leseschwierigkeiten wird in Deutschland zwischen einer inhaltlich begrenzten, aber manifesten Lernstörungen einer Teilleistung (z.B. Lese-Rechtschreib-Schwäche) oder einer milderen Lernschwäche mit Minderleistungen im Lesen unterschieden (Al-Yagon et al., 2013; Gold, 2018). Beide Begriffe schließen eine beeinträchtigte intellektuelle Befähigung als Erklärungsansatz für die Leseschwierigkeiten aus. Der verallgemeinernde Begriff Lernschwierigkeiten fasst hingegen alle Formen einer beeinträchtigten Leistungsentwicklung im Lesen, in der Rechtschreibung oder beim Rechnen ein (Gold, 2018). In Anlehnung an Gold (2018) werden in der vorliegenden Arbeit unter Leseschwierigkeiten ebenfalls alle Formen einer beeinträchtigten Leseentwicklung verstanden.

Zu der Gruppe von Schülerinnen und Schülern mit Leseschwierigkeiten zählen jene mit einem SPF-L. Im Zuschreibungsprozess des sonderpädagogischen Förderbedarfs im Förderschwerpunkt Lernen sind manifeste Leseschwierigkeiten ein entscheidender Faktor (Grünke & Cavendish, 2016; Kölm, Gresch & Kuhl, 2019). Für die Zuschreibung eines SPF-L und der damit einhergehenden sonderpädagogischen Förderung muss auch unter Ausschöpfung aller pädagogischer und unterrichtsfachlichen Unterstützungsressourcen ein Nicht-Erreichen der Mindeststandards und der Lernziele der allgemeinen Schule über einen längeren Zeitraum in mehreren schulischen Kompetenzen zu erwarten sein (Kultusministerkonferenz, 2019). Konkret genannt werden in der aktuellen Empfehlung der Ständigen Konferenz der Kultusminister (KMK; 2019) des sonderpädagogischen Schwerpunkts Lernen erhebliche Schwierigkeiten in den Kompetenzbereichen Lesen, Schreiben und Rechnen. Folglich ist es konsequent, dass Schülerinnen und Schüler mit SPF-L in allen Phasen der Schulzeit von Leseschwierigkeiten betroffen sind. Zu Beginn der Schullaufbahn zeigen sie bereits geringe lesebezogene Vorläuferfähigkeiten (Klicpera

& Gasteiger-Klicpera, 1993; Klicpera, Ehgartner, Gasteiger-Klicpera & Schabmann, 1993). In der vierten Klassenstufe erreichen die Schülerinnen und Schüler mit SPF-L im IQB-Ländervergleich 2016 im Mittel die geringste Kompetenzstufe I (Kocaj, Kuhl, Haag, Kohrt & Stanat, 2017). Und in der siebten Klassenstufe liegen ihre Leseleistungen mehr als drei Schuljahre hinter denen von Schülerinnen und Schülern ohne SPF (Wocken & Gröhlich, 2009).

Weitere Gruppen von Schülerinnen und Schülern sind von der Ausbildung von Leseschwierigkeiten während der Schulzeit aufgrund ihrer individuellen und situativen Merkmale bedroht (sog. Risikokinder). Schülerinnen und Schüler mit dem Förderschwerpunkt Sprache (SPF-S) weisen entwicklungsbedingte oder erworbene Sprachentwicklungsstörungen auf. Sie sind in ihren lautsprachlichen, grammatikalischen und lexikalischen Fähigkeiten beeinträchtigt, wodurch ihr Leseerwerb mit einer hohen Wahrscheinlichkeit benachteiligt wird (Diehl, 2010; Lindsay & Strand, 2016). Außerdem haben Schülerinnen und Schüler mit Schwierigkeiten im emotionalen und sozialen Erleben oder in der Aufmerksamkeit (Bental & Tirosh, 2007; Jones, Greenberg & Crowley, 2015), mit einem Migrationshintergrund oder mit Deutsch als Zweitsprache (Maitz et al., 2018; Schründer-Lenzen & Merkens, 2006; Taylor, 2012) und sozialen Disparitäten (Haag et al., 2017; Rjosk et al., 2014) aufgrund ihrer spezifischen Personenmerkmale ein erhöhtes Risiko für unzureichende Lesekompetenzen. Des Weiteren ist eine Beeinflussung der Leseleistungen durch die Beschulungsart denkbar. Eine Analyse der PISA-Kohorte 2012 zeigt signifikant höhere Lesekompetenzen bei 15-Jährigen mit SPF in inklusiven Settings als an Förderschulen (Gebhardt, Sälzer et al., 2015). Im Grundschulalter fanden Kocaj, Kuhl, Kroth, Pant und Stanat (2014) vergleichbare Ergebnisse zugunsten der inklusiven Beschulung. Durch ein Propensity Score Matching wurden in dieser Studie mögliche Selektionseffekte vor der Einschulung in den Vergleichsgruppen (Regelschule vs. Förderschule) kontrolliert. Der positive Einfluss zu Gunsten der inklusiven Regelschule fiel für die Schülerinnen und Schüler mit SPF-L größer aus als bei der Gruppe mit SPF-S (Kocaj et al., 2014).

Der Bedarf an einem lernwirksamen Unterricht und an gezielter Leseförderung wird durch den großen Anteil von Schülerinnen und Schülern mit unzureichenden Lesekompetenzen sowie durch die Dauer und Tragweite der Leseschwierigkeiten ersichtlich. Weis et al. (2016) deklarieren, dass insbesondere die schulische

Förderung für Schülerinnen und Schüler mit einem geringen Lesenniveau in Deutschland noch verbessert werden müsse. Seit der Ratifizierung der UN-BRK 2009 ist das deutsche Schulsystem im Wandel und zunehmend mehr Schülerinnen und Schüler werden in inklusiven Settings unterrichtet (Lange, 2017). Nach Piezunka, Schaffus und Grosche (2017) kann der inklusive Unterricht als Lernen am gemeinsamen Lerngegenstand von Schülerinnen und Schülern mit unterschiedlichen Lernvoraussetzungen, Lernbedürfnissen und Lernhürden, bei dem die bestmögliche Leistungsförderung aller Schülerinnen und Schüler im Zentrum des Unterrichts steht, verstanden werden. Aufgrund der breiten Heterogenität der Schülerinnen und Schüler stellt der inklusive Unterricht methodische und didaktische Herausforderungen an die unterrichtenden Lehrkräfte. Kuhl und Hecht (2014) konstatieren, dass eine schulische Prävention das Lernen positiv beeinflusst und einen wichtigen Baustein zum Gelingen der schulischen Inklusion darstellt. Zur Rahmung von präventiver Förderung in inklusiven Settings verweisen Kuhl und Hecht (2014) auf das von Kretschmann (2007) übertragene dreistufige Präventionsmodell der entwicklungsförderlichen Gestaltungen von Systemen der Weltgesundheitsorganisation. Die primäre Prävention umfasst in diesem Modell eine adäquate vorschulische und schulische Ausstattung und Gestaltung der Lebens- und Lernbedingungen für alle Schülerinnen und Schüler (Kretschmann, 2007). Alle unterrichtlichen Maßnahmen dieser Stufe sollen möglichen Lernschwierigkeiten vorbeugen und werden allen Schülerinnen und Schülern angeboten (Kuhl et al., 2012). Im Kontext von Leseschwierigkeiten kann die primäre Prävention durch ein vorschulisches Training von lesebezogenen Vorläuferfähigkeiten realisiert werden (Kuhl & Hecht, 2014). Die sekundäre Prävention umfasst zeitlich begrenzte additive Förderangebote für Schülerinnen und Schüler mit einem Risiko zur Ausbildung von oder solchen mit milden Lernschwierigkeiten. Kuhl und Hecht (2014) nennen Sprachkurse für Schülerinnen und Schüler mit Deutsch als Zweitsprache oder Förderkurse für Schülerinnen und Schüler mit geringen Leseleistungen als Umsetzungsbeispiele der zweiten Präventionsstufe. Explizite Therapien, Förderungen und Unterstützungsangebote bei manifesten Lernschwierigkeiten werden in der Stufe der tertiären Prävention realisiert (Kretschmann, 2007). Kuhl und Hecht (2014) spezifizieren dies durch eine intensive sonderpädagogische Lernförderung.

Innerhalb der drei Präventionsstufen benötigen Lehrkräfte unterschiedliche Förderkonzepte zur Unterstützung des Lesenlernens. Es liegt in ihrer Hand zu entscheiden, welche Schülerinnen und Schüler welche Art der Unterstützung erhalten. Viele Leseförderkonzepte sind für spezifische Leseschwierigkeiten konzipiert (Galuschka & Schulte-Körne, 2015; Ise, Engel & Schulte-Körne, 2012) und greifen damit erst in der tertiären Präventionsstufe. Im Sinne der Wait-to-Fail-Problematik (Huber & Grosche, 2012) müssen die Lernenden erst manifeste Lernschwierigkeiten ausbilden, bevor sie eine ausreichende (sonder-)pädagogische und intensive Leseförderung erhalten. Die meisten Leseförderkonzepte weisen zudem keine empirische Prüfung vor und können daher nicht als evidenzbasiert bezeichnet werden (zur Diskussion der Evidenzbasierung in der Sonderpädagogik siehe Hillenbrand, 2015; Kuhl & Euker, 2016; Kuhl et al., 2017). Der Kieler Leseaufbau (KLA; Dummer-Smoch & Hackethal, 2016) ist ein wirksames Beispiel für den Leseanfangsunterricht, welches in der sekundären und tertiären Präventionsstufe Anwendung findet. Nach dem Prinzip der Vermeidung von Schwierigkeiten werden zu Beginn des Leseerwerbs zunächst nur eine reduzierte Auswahl von Graphemen (z.B. dehbare Konsonanten, alle Vokale) sowie leichte Silben- und Wortstrukturen (z.B. Vokal-Konsonant-Vokal oder Konsonant-Vokal-Konsonant) verwendet. Schülerinnen und Schüler mit milden Leseschwierigkeiten arbeiten mit zunehmend komplexerem linguistischem Material und Schülerinnen und Schüler mit manifesten Leseschwierigkeiten festigen ihre Kompetenzen zunächst mit leichterem Material ohne Überforderung.

Insbesondere für den inklusiven Leseunterricht in der Grundschule wird gefordert, dass förderbedürftige Schülerinnen und Schüler mit einem Risiko zur Ausbildung von Leseschwierigkeiten frühzeitig identifiziert werden, damit sie eine präventive und auf ihre Bedürfnisse abgestimmte Fördermaßnahme erhalten (Euker et al., 2012; Fischer & Gasteiger-Klicpera, 2013; Gasteiger-Klicpera & Fischer, 2008; Torgesen, 2005). Allerdings gelingt eine intuitive Identifikation von Schülerinnen und Schülern mit Leseschwierigkeiten im Unterricht oft nicht rechtzeitig (Artelt & Rausch, 2014; Feinberg & Shapiro, 2009; Madelaine & Wheldall, 2005). In solchen Fällen bleiben Schülerinnen und Schüler mit einem Risiko zur Ausbildung von Leseschwierigkeiten oft unentdeckt. Diagnostische Informationen über individuelle Entwicklungsverläufe und Leseschwierigkeiten aus standardisierten Verfahren können diesen Identifikationsprozess absichern. Standardisierte Statusdiagnostiken haben meist

eine hohe Durchführungsdauer, eine begrenzte Einsatzhäufigkeit und lassen direkte Vergleiche mit den Unterrichtszielen nicht zu, sodass sie zur Beobachtung von Entwicklungsverläufen im Unterrichtsalltag nicht praktikabel sind (Scheer, 2014). Als vorteilhafte Alternative zu statusdiagnostischen Verfahren nennen Kuhl und Wittich (2018): „Eine zumindest in Deutschland relativ neue Form der Modifikationsdiagnostik ist die Lernverlaufsdagnostik oder das Curriculumbasierte Messen (CBM)“ (S.388).

2.3 Lernverlaufsdagnostik im Bereich Lesen

Die Grundidee der Lernverlaufsdagnostik¹ ist eine kontinuierliche Messung der individuellen Lernentwicklungen von Schülerinnen und Schülern mithilfe von standardisierten Tests mit kurzer Durchführungsdauer und hoher Praktikabilität im Unterricht. Im Vordergrund steht dabei die Nutzung der Lernverlaufdaten als Grundlage für das unterrichtliche Handeln der Lehrkräfte und nicht die Bewertung der kindlichen Leistungen (Espin et al., 2017). Mithilfe der Lernverlaufsdagnostik wird eine Identifizierung von Schülerinnen und Schülern mit einem Risiko für die Ausbildung von Leseschwierigkeiten ermöglicht, sodass bei ausbleibendem Lernerfolg eine adaptive Intervention im präventiven Sinn initiiert werden kann.

Das Konzept der Lernverlaufsdagnostik ist in Deutschland seit der Jahrtausendwende unter starkem Einfluss des US-amerikanischen Konzepts des curriculumbasierten Messens (CBM; Deno, 1985, 2003a) gewachsen (Klauer, 2006; Walter, 2009b). Lernverlaufstests stammen aus dem Kontext der Sonderpädagogik und werden vorrangig in leistungsheterogenen Lerngruppen mit Schülerinnen und Schülern mit benachteiligenden Lernfaktoren eingesetzt (Deno, Fuchs, Marston & Shin, 2001). Um die Lernentwicklung der Schülerinnen und Schüler über die Zeit hinweg zu beobachten, werden die Lernverlaufstests hochfrequent (bis zu ein Mal wöchentlich) im Unterricht eingesetzt (Deno, 2003b). Die Ergebnisse der Tests werden anschließend in Lernverlaufsgraphen dargestellt, um die Lernentwicklungen zu

¹ Neben dem Begriff Lernverlaufsdagnostik finden sich in der deutschsprachigen Literatur die Begriffe Lernverlaufsmessung bei Gebhardt, Diehl und Mühling (2016b), Lernfortschrittsdiagnostik bei Walter (2010a), prozessbegleitende Diagnostik bei Liebers, Landwehr, Marquardt und Schlotter (2015) und Lernfortschrittsmessung bei Klauer (2006). Trotz Unterschieden in den Testkonstruktionen liegt ein einheitliches Verständnis des Grundkonzepts vor, welches durch diagnostische Tests den Lernverlauf für Evaluationszwecke sichtbar macht.

visualisieren und für diagnostische Prozesse zu nutzen (Walter, 2009b). Die Lehrkräfte evaluieren anhand der kindlichen Lernentwicklungen, ob die Schülerinnen und Schüler von dem regulären Unterricht profitieren (primäre Präventionsstufe) oder ob eine spezifische (Lese-)Förderung (sekundäre und tertiäre Präventionsstufe) erforderlich ist (Ardoin et al., 2013). Zur Identifikation von Schülerinnen und Schülern mit besonderem Förderbedarf im Lesen werden die individuellen Lernverlaufsdaten entweder mit Daten aus einer sozialen Bezugsgruppe oder mit welchen aus Normierungsstudien verglichen (Förster, Kuhn & Souvignier, 2017). Für einen solchen Vergleich beurteilen die Lehrkräfte vorrangig die numerischen Lernverlaufsdaten (z.B. Werte zu einzelnen Messzeitpunkten, Steigung der Lernkurven). Für die Beurteilung der Effektivität des regulären Unterrichts oder einer intensiven Fördermaßnahme müssen zusätzlich die fachspezifischen Lesekompetenzen der Schülerinnen und Schüler analysiert (Stecker, 2017) sowie mit den gesetzten Zielen und realisierten Förderbemühungen verknüpft werden (Datnow & Hubbard, 2016; Mandinach, 2012; Stecker et al., 2005). Dieser Prozess wird als datenbasierte Förderplanung bezeichnet (van den Bosch et al., 2017). Über diese Verknüpfung sind Rückschlüsse auf die Verbesserungen schulischer Kompetenzen unter dem Einfluss einer spezifischen Fördermaßnahme herstellbar. Diesem Analyseprozess liegt die Annahme zu Grunde, dass durch effektive Fördermaßnahmen ein höherer Lernanstieg beobachtbar ist als ohne spezifische Intervention. In der primären Präventionsstufe kann die Lernverlaufsdagnostik als Screening mit größeren Messabständen eingesetzt werden (Huber & Grosche, 2012). Je nachdem, ob in dieser Stufe erwartete oder langsame Fortschritte, Stagnationen oder gar Rückschritte beobachtet werden, wählen Lehrkräfte Alternativen in der Leseförderung. Schülerinnen und Schüler mit einem Risiko für die Ausbildung von Leseschwierigkeiten fallen bereits in dieser Stufe durch ihre nicht erwartungsgemäßen Lernfortschritte auf (Silberglitt & Hintze, 2007). Daraufhin können die Lehrkräfte auf noch geringe Lernlücken und individuelle Lernhürden frühzeitig reagieren, ohne dass wertvolle Lernzeit verstreicht. Wenn die initiierten Maßnahmen keinen gewünschten Lernerfolg bewirken, kann mit der Lernverlaufsdagnostik die Effektivität von zusätzlichen Fördermaßnahmen der sekundären und tertiären Präventionsstufe mit engmaschigen Messungen evaluiert werden. Das übergeordnete Ziel der Lernverlaufsdagnostik ist eine positive Wirkung auf die Lernentwicklung von Schülerinnen und Schülern (Bennett, 2011).

Die Lesetests der Lernverlaufsdagnostik messen Leistungen in verschiedenen Kompetenzbereichen, wie Leseflüssigkeit oder Leseverständnis. Dafür bearbeiten die Schülerinnen und Schüler in einer vorgegebenen Zeit von meist wenigen Minuten, so viele Aufgaben wie möglich (sog. Speedtests; Rost, 2004). Die amerikanischen CBM Aufgaben wurden ursprünglich direkt aus dem aktuellen Curriculum abgeleitet, um eine sehr enge Verbindung zum Unterricht herzustellen. Diese Aufgaben erheben Kompetenzen, die am Ende eines Schuljahres in dem Curriculum für diese Jahrgangsstufe gefordert sind (Fuchs, 2004). Valide Messungen sind aber ebenfalls über Aufgaben realisierbar, die nur eng an das Curriculum angelehnt sind und nicht unmittelbar aus dem tatsächlichen Unterricht entstammen (Deno, 2003b). Zur Erfassung von basalen Lesekompetenzen haben sich zwei grundsätzliche Arten von Lernverlaufstests durchgesetzt: Leseflüssigkeitstests und Maze-Aufgaben (Graney, Martínez, Missall & Aricak, 2010; Shin, Deno & Espin, 2000). Leseflüssigkeitstests erheben die Leseflüssigkeit durch lautes Vorlesen. Das laute Vorlesen gilt als robuster Indikator für die allgemeine Lesekompetenz (Deno, Mirkin & Chiang, 1982; Fuchs et al., 2001; Hudson et al., 2008). Das zu lesende Material sind entweder zusammenhängende Texte oder Listen mit unzusammenhängenden Silben, Wörtern oder Pseudowörtern (sog. Itemlisten; Fuchs, Fuchs & Compton, 2004; Walter, 2009a; Zumeta, Compton & Fuchs, 2012). Itemlisten unterscheiden sich dahingehend, ob Silben, reale Wörter oder Pseudowörter zum Vorlesen angeboten werden (Fuchs et al., 2004). Während den Messungen bewertet eine lesekompetente Person (z.B. die Lehrkraft), ob Lesefehler gemacht wurden. Bei der Beurteilung von Lesefehlern in der Leseflüssigkeit werden prosodische Aspekte traditionell ausgeschlossen, da sowohl die Messung von prosodischen Aspekten als auch der Mehrwert als noch unzureichend erforscht gilt (Biancarosa & Cummings, 2015; Schwanenflugel, Westmoreland & Benjamin, 2015). Als Kompetenzmaß wird die Anzahl der richtig gelesenen Wörter pro Minute genutzt (Holle, 2010). Leseflüssigkeitstests werden als Einzeltests administriert, wodurch die Praktikabilität im Unterrichtsalltag eingeschränkt ist. Um die Praktikabilität von Lernverlaufstests im Unterricht zu erhöhen, wurden im US-amerikanischen Raum Maze-Aufgaben zur Messung des Leseverständnisses (dt. Labyrinth-Aufgaben; Ardoin et al., 2004) entwickelt. Sie sind in Anlehnung an Lückentextaufgaben ohne Zeitlimit und Auswahlmöglichkeiten (sog. Cloze-Tests) entstanden (Cunningham & Cunningham, 1978; Gellert & Elbro, 2013; Louthan, 1965). Bei Maze-Aufgaben werden den Lesenden Textpassagen von ca. 250 Wörtern

vorgelegt, bei denen einzelne Wörter nach festgesetzten Regeln fehlen. Die literarische Art der Texte ist höchst unterschiedlich (Förster & Souvignier, 2011; Tichá, Espin & Wayman, 2009; Walter, 2011a). Abwandlungen der ursprünglichen Maze-Aufgaben messen durch ihre Testkonstruktion zusätzlich die semantischen und syntaktischen Leseprozesse bei der lokalen Kohärenzbildung (Ecalte et al., 2013; Forster, 2011). In diesen Abwandlungen werden nur einzelne unzusammenhängende Sätze anstatt vollständige Texte verwendet (Witzel & Witzel, 2016). Unter der Angabe eines richtigen Targetwortes und mehreren falschen Distraktoren konstruieren die Lesenden dann den Sinn der Sätze und füllen die Lücken. Die Lücken werden entweder über numerische Regeln (z.B. jedes siebte Wort) oder über linguistische Bedingungen (z.B. bestimme Wortarten) bestimmt (January & Ardoin, 2012; Kingston & Weaver, 1970). Die verwendeten Distraktoren unterscheiden sich dahingehend, ob sie entweder syntaktisch möglich, semantisch sinnvoll oder phonologisch und phonetisch dem Targetwort ähnlich sind (Fuchs & Fuchs, 1992; Gillingham & Garner, 1992; McKenna & Miller, 1980). In der Grundschule werden Maze-Aufgaben als leise Leseaufgaben bearbeitet und sowohl als Gruppen- sowie als Einzeltest administriert (García & Cain, 2014; Hale et al., 2011). Diese Tests haben meist ein festgesetztes Zeitlimit, was je nach Testkonstruktion bis zu zehn Minuten beträgt (Brown-Chidsey, Davis & Maya, 2003; Fuchs & Fuchs, 1992; Wiley & Deno, 2005). Unter Berücksichtigung der Zeitkomponente gelten Maze-Aufgaben als robuste Indikatoren für das Leseverständnis (Hale et al., 2011). Als Kompetenzmaß wird bei Maze-Aufgaben die Anzahl der richtig gelösten Lücken pro Zeiteinheit zur Beurteilung der Kompetenzveränderung herangezogen. Die überwiegend amerikanischen Forschungsergebnisse wurden größtenteils von deutschen Forschenden adaptiert, sodass ähnliche Testarten in deutschsprachigen Lernverlaufstests wiederzufinden sind (Diehl & Hartke, 2012; Walter, 2010a, 2013).

Aus lesetheoretischer Perspektive sind die zuvor genannten Unterschiede innerhalb der Testkonstruktionen nicht zu vernachlässigen. Denn je nach Konstruktion messen die Tests zwar reliabel, beanspruchen aber unterschiedliche Teilkompetenzen bzw. Strategien während des Leseprozesses. Wenn die Testkonstruktion aktuelle Leserwerbsmodelle berücksichtigt, können Leseschwierigkeiten in spezifischen Teilkompetenzen bzw. Strategien differenziert aufgedeckt werden. Werden bei der Leseflüssigkeit beispielsweise einzelne Silben, Wörter oder Pseudowörter verwendet,

wird das linguistische Material entweder nicht lexikalisch, partiell lexikalisch oder vollständig lexikalisch erlesen. Werden als linguistisches Material zusammenhängende Texte verwendet, nutzen die Lesenden den Satzkontext sowohl für das korrekte Vorlesen als auch für das Verständnis in unterschiedlichem Maße (Martohardjono et al., 2005; West & Stanovich, 1978). Durch eine lesetheoretische Testkonstruktion resultiert, dass die Lernverlaufsdagnostik neben der numerischen Lernentwicklung ebenfalls lesespezifische Hinweise für die Förderplanung offenlegt.

Lernverlaufstests sind diagnostische Instrumente zur Beurteilung von individuellen Leistungsveränderungen über einen längeren Zeitraum, die einen hohen Anwendungsbezug im Schulalltag aufweisen. Daher müssen sie den Standards der psychometrischen Gütekriterien (Objektivität, Reliabilität und Validität) sowie den Nebengütekriterien Änderungssensibilität und Ökonomie entsprechen (R. Good & Jefferson, 1998; Klauer, 2011). Die psychometrischen Gütekriterien sowie die Testkonstruktion der Lernverlaufstests haben diverse Bedingungen zu berücksichtigen, die teilweise in hoher Abhängigkeit zu einander stehen. Einerseits müssen die parallelen Testformen zu allen Messzeitpunkten gleich schwer sein und stets dasselbe Konstrukt messen (Gebhardt, Heine, Zeuch & Förster, 2015; Wilbert & Linnemann, 2011). Andererseits müssen die Tests bzw. die Testergebnisse auch in der Schulpraxis für Lehrkräfte leicht interpretierbar und zeitlich ökonomisch sein. Für die Skalierung von Lernverlaufstests werden Item-Response-Modelle aus der probabilistischen Testtheorie empfohlen (Gebhardt, Heine et al., 2015; Klauer, 2011; Wilbert & Linnemann, 2011; Wilbert, 2014). Mittels des Rasch-Modells wird die Wahrscheinlichkeit einer Person beschrieben, Aufgaben mit unterschiedlichen Schwierigkeitsgraden zu lösen. Meist wird im Rasch-Modell angenommen, dass alle Items ein einziges latentes Konstrukt messen und sich die Schwierigkeitsgrade der Items nur in einer Dimension unterscheiden (Rasch, 1960). Eine notwendige Voraussetzung für die Gültigkeit des Rasch-Modells ist im Hinblick auf die Kompetenzerfassung die Eindimensionalität der Skala (Wilbert & Linnemann, 2011). Für die Lernverlaufsdagnostik bedeutet eine testtheoretische Eindimensionalität, dass die Testergebnisse ausschließlich aus einem latenten Faktor resultieren und nicht auf weiteren, konfundierenden Faktoren (z.B. Anstrengungsbereitschaft, Vertrautheit mit dem Aufgabentyp) beruhen (Wilbert, 2014). Als weitere Voraussetzung für die Gültigkeit des Rasch-Modells muss eine essentielle

stochastische bzw. lokale Unabhängigkeit der einzelnen Items bestehen (Stout et al., 1996). Die einzelnen Items dürfen sich also in ihrer Lösungswahrscheinlichkeit nicht gegenseitig beeinflussen und nur durch Fähigkeitsunterschiede der Personen bedingt sein.

Wenn alle Annahmen eines solchen Modells gültig sind, kann der Summenwert aller richtig gelösten Items als Approximation der Kompetenz der Schülerinnen und Schüler verwendet werden (Rost, 2004). Die Skalierung nach der Item-Response-Theorie hat den Vorteil, dass nach erfolgreicher Prüfung des Modells eine zufällige Auswahl von Items unter gleichbleibender Reliabilität vorgenommen werden kann (Mühling, Gebhardt & Diehl, 2017). Die Tests der Lernverlaufsdagnostik werden für eine Vergleichbarkeit der Testergebnisse zu mehreren Messzeitpunkten konstruiert. Damit dies möglich ist, muss die Itemschwierigkeit aller zur Verfügung stehenden Parallelformen möglichst gleich schwer mit gleicher Trennschärfe sein (Wilbert & Linnemann, 2011). Ist dies nicht der Fall, muss die unterschiedliche Itemschwierigkeit bei der Testkonstruktion und -interpretation berücksichtigt werden. Damit die Tests in leistungsheterogenen Klassen sinnvoll eingesetzt werden können, müssen sie von unterschiedlichen Gruppen von Schülerinnen und Schülern (z.B. Geschlecht, Migrationshintergrund, SPF) gleich gut lösbar sein. Unter Testfairness wird der Ausschluss einer systematischen Benachteiligung unterschiedlicher Personengruppen, die durch einen Item-Bias verursacht wird, verstanden (Moosbrugger & Kelava, 2012). Die Testfairness kann geprüft werden, indem die Schwierigkeiten von Testitems für verschiedene Gruppen (z.B. Geschlecht, SPF, Migrationshintergrund) bestimmt und geprüft werden (Differential Item Functioning (DIF); Holland & Wainer, 1993). Alternativ kann die Fairness des gesamten Tests mittels Messinvarianzanalysen geprüft werden (Steenkamp & Baumgartner, 1998). Dabei werden schrittweise Trennschärfen und Fehleranteile zwischen Gruppen verglichen. Lernverlaufstests sind für einen regelmäßigen Einsatz während des Lernprozesses ausgelegt. Daher ist es notwendig, dass die Tests Veränderungen im Lernen sensibel messen bzw. das Kriterium der Änderungssensibilität erfüllt wird (Klauer, 2011).

Wilbert und Linnemann (2011) empfehlen für die psychometrische Prüfung vier Schritte. Zuerst sollen die Items nach der Klassischen Testtheorie sowie in Bezug auf die Test-Retest-Reliabilität geprüft werden. Als zweites soll mit einer

konfirmatorischen Faktorenanalyse die Dimensionalität des Konstrukts getestet werden. Drittens wird die Itemschwierigkeit über eine Anwendung des Rasch-Modells bestimmt. Abschließend kann die Testfairness mittels DIF für relevante Subgruppen getestet werden. Lernverlaufstests werden vorrangig für die Beurteilung von Lernentwicklungen eingesetzt. Dafür empfiehlt Wilbert (2014) die visuelle Inspektion der Lernverlaufsdaten (z.B. anhand der Lernverlaufsgraphen) auf der Basis deskriptiver Analysen. Ergänzend können Regressionsanalysen zur Modellierung der Lernverlaufsdaten eingesetzt werden. Lineare Regressionsanalysen haben bei internationalen und nationalen Testautorinnen und Testautoren eine lange Tradition (Deno et al., 2001; Fuchs, Fuchs, Hamlett, Walz & Germann, 1993; Walter, 2010b). Dabei werden die Lernverläufe mittels des Regressionskoeffizienten (Steigungsmaß, b -Wert) und des Intercepts (Ausgangsniveau) beschrieben. Die Lernverläufe werden überwiegend als lineare Steigung angenommen und erklärt. Dadurch ist eine Vorhersage zukünftiger Werte des Lernverlaufs möglich. Für eine reliable Einschätzung des Lernverlaufs empfehlen Christ, Zopluoglu, Monaghan und Van Norman (2013) zwischen sechs bis acht Messzeitpunkte. Walter (2014) weist zusätzlich daraufhin, dass bei Lernverlaufstests Boden- und Deckeneffekte entstehen können. Diese Effekte treten auf, wenn ein Test im unteren oder oberen Merkmalsbereich die Kompetenzen der Personen nicht abbildet. Entstehen diese Effekte ist der Test nicht fähig, die entsprechenden Leistungen zu messen.

Fuchs (2004) betrachtet die Forschung zur Lernverlaufsdagnostik in drei Gütestadien. Ziel ist es dabei die Qualität der Lernverlaufsdagnostik für die schulische Praxis zu sichern. Im ersten Stadium der Studien geht es um die Frage, ob der entwickelte Test überhaupt reliabel ist. Daher wird der Test mit einer großen Stichprobe einmalig im jeweiligen Leistungsbereich gemessen. In diesen Studien konzentrieren sich die Analysen auf Fragen zur Dimensionalität und Reliabilität. Im zweiten Stadium geht es um die Frage, ob der Test auch Veränderungen in der Lernentwicklung im Längsschnitt reliabel misst. Hier wird mit mehreren Messungen in einem zeitlich gleichen Abstand die Leistungsentwicklung untersucht. Zusätzlich wird die Messinvarianz sowie die Änderungssensibilität anhand der Daten geprüft. Durch solche Studien wird gewährleistet, dass Lernverläufe über einen längeren Zeitraum grafisch dokumentiert und interpretiert werden können. Für die wiederholten Messzeitpunkte muss sichergestellt werden, dass die einzelnen Parallelversionen auf

Itemebene miteinander verbunden sind. Die Stichproben für das erste und zweite Stadium stammen überwiegend aus kontrollierten Studien. Im dritten Stadium nach Fuchs (2004) wird die Implementation, der praktische Nutzen und die Wirksamkeit der Lernverlaufsdagnostik im Feld geprüft. Dazu zählen auch Studien, die die Verknüpfung zwischen beobachteten Lernverläufen und ausgewählten Interventionen untersuchen. Eine Voraussetzung für Studien des dritten Gütestadiums ist, dass Lehrkräfte die Lernverlaufsdagnostik bereits im Unterricht eigenständig anwenden. Diese Studien sind aktuell noch sehr selten und in Deutschland aufgrund der geringen Verbreitung der Lernverlaufsdagnostik kaum möglich. Die Gütestadien nach Fuchs (2004) bieten eine gute Orientierung für die Entwicklung und Durchführung von Evaluationsstudien. Ergänzend machen Förster et al. (2017) auf die Notwendigkeit von Normierungsstudien aufmerksam. Für Normierungsstudien muss insbesondere auf die Repräsentativität der Stichproben geachtet werden. Dazu zählen die Unterrichtsformen und -themen sowie die Vergleichbarkeit der Untersuchungssituationen. Damit Vergleichs- und Richtwerte für spezifische Verfahren angeboten werden können, müssen die Lernverläufe von Kontrollgruppen im regulärem Unterricht und von Interventionsgruppen unter dem Einfluss effektiver Leseförderung untersucht werden (Förster et al., 2017).

Für eine Implementation der Lernverlaufsdagnostik in der Schulpraxis müssen die Tests praktikabel, ökonomisch und zeiteffizient sein (Deno, 2003a; Fuchs, 2017). Daher sollten die Tests leicht in den Unterricht integrierbar, schnell durchführbar, einfach zu handhaben sein und als gewinnbringend wahrgenommen werden. Je nach konkreter Ausgestaltung der Testsysteme wird die Implementation begünstigt oder gehemmt. Fuchs und Fuchs (1992) erkennen schon früh die Potentiale von digitalen Testsystemen. Digitale Testsysteme nehmen Lehrkräften organisatorische Aufgaben bei der Vorbereitung, Dokumentation und Nachbereitung ab, ermöglichen adaptive Testungen oder stellen automatisierte Auswertungen der Verlaufsdaten bereit (Fuchs & Fuchs, 2001; Gebhardt & Jungjohann, im Druck; Seifert, 2019). Diese Eigenschaften reduzieren die zu investierende Zeit für die Lehrkräfte. Gleichzeitig konfrontieren digitale Instrumente der Lernverlaufsdagnostik die Schullandschaft und Lehrkräfte mit neuen administrativen Aufgaben (Maier, 2014). Außerdem werden Lehrkräfte ohne Zugang zu digitalen Endgeräten oder dem Internet von der Nutzung von Testsystemen mit einem rein digitalen bzw. webbasierten Angebot ausgeschlossen.

Allerdings nimmt in Deutschland die Bedeutung von digitalen Testsystemen durch die schulische Digitalisierung zu, da zunehmend mehr internetfähige Endgeräte verfügbar sind (Seifert, 2019). Mühling, Jungjohann und Gebhardt (2019) haben gezeigt, dass sich die Anzahl von registrierten Lehrkräften in der webbasierten Lernverlaufsplattform www.levumi.de (Gebhardt et al., 2016b) zwischen 2015 und 2018 mehr als versechsfacht hat. Obwohl die Gründe der steigenden Zahlen an Nutzerinnen und Nutzern bisher nicht geklärt sind, kann dieser enorme Zuwachs als Hinweis für den Bedarf und die Akzeptanz von digitalen Testsystemen bei Lehrkräften gesehen werden. Die Handhabung von Lernverlaufstests in der Schulpraxis sollte durch erklärende Materialien und gezielte Fortbildungen unterstützt werden (Ardoin et al., 2013). Zu Unterstützungsmaterialien zählen beispielsweise Manuale, die die Anwendung des eigentlichen Instruments, die Interpretation der resultierenden Verlaufsdaten sowie weiterführende Informationen zur Förderung bereitstellen. Wie bei jedem diagnostischen Instrument müssen Lehrkräfte zunächst Einarbeitungszeit investieren. Da sich Leseteilkompetenzen bei Grundschülerinnen und Grundschulern parallel und unterschiedlich schnell entwickeln, verbessert ein Testsystem für mehrere Lesekompetenzen und einen jahrgangsübergreifenden Einsatz die Ökonomie. Für eine Einbettung der Tests in den Unterrichtsalltag schlagen Jungjohann und Gebhardt (2018) beispielsweise eine rhythmisierte Anwendung während individuellen Lernzeiten vor. Dort haben Lehrkräfte Zeit die Leseflüssigkeitstests in Einzelsituationen durchzuführen und ältere Schülerinnen und Schüler können digitale schülerzentrierte Tests zum Leseverständnis mit automatischer Zeitkontrolle eigenständig bearbeiten. Bereits in frühen Studien wurde die Wirksamkeit der Lernverlaufsdagnostik auf die Lernentwicklung nachgewiesen (Fuchs, Fuchs, Hamlett & Ferguson, 1992). Nach der Interventionsstudie von Graney und Shinn (2005) hat die alleinige Bereitstellung von Lernverlaufsdaten keine positiven Effekte auf die Lernentwicklung der Schülerinnen und Schüler. Stecker et al. (2005) schlussfolgern in ihrem Review zur Lernwirksamkeit von CBM Lesetests, dass der kombinierte Einsatz von Lernverlaufsdagnostik mit Interpretationshilfen für die praktizierenden Lehrkräfte die größten Effekte hinsichtlich der kindlichen Lernentwicklung bewirken.

Erst seit jüngster Zeit ist die Erforschung des Verständnisses und der Interpretation von Lernverlaufsdaten und -graphen Teil der Forschungsbemühungen. Fuchs (2017) beschreibt, dass die technische Anwendung der Lernverlaufsdagnostik geringere

Anforderungen an die Lehrkräfte stellt als die Interpretation und Übertragung der Lernverlaufsdaten auf die unterrichtliche Förderung im Sinne der datenbasierten Förderplanung. Aus mehreren Untersuchungen ist bekannt, dass die Dateninterpretation einerseits durch diagnostische und fachbezogene förderpädagogische Kompetenz der Lehrkräfte (van den Bosch et al., 2017; Wagner, Hammerschmidt-Snidarich, Espin, Seifert & McMaster, 2017; Zeuch, Förster & Souvignier, 2017) und andererseits durch die Struktur (z.B. Steigung, Extremwerte) und die Aufbereitung (z.B. optische Hilfsmittel, Vergleichswerte) der Daten selbst (Klapproth, 2018; Newell & Christ, 2017) beeinflusst wird. Nach der Interventionsstudie von van den Bosch, Espin, Pat-El und Saab (2019) wird das Verständnis von Lernverlaufsgraphen bei Grundschullehrkräften durch videobasierte Interventionen verbessert. Bei den zuvor genannten Studien zum Umgang mit Lernverlaufsdaten ist übergreifend Konsens, dass Lehrkräfte insbesondere bei der Verknüpfung der Lernverlaufsdaten mit leredidaktischen Aspekten Schwierigkeiten haben.

3. Fragestellungen der Dissertation

In Deutschland erreichen nicht alle Grundschülerinnen und Grundschülern ausreichend gute Lesekompetenzen (Bos et al., 2017; Gebhardt, Sälzer et al., 2015; Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017). Zur Prävention von Leseschwierigkeiten ist eine frühzeitige Identifikation und Förderung von förderbedürftigen Schülerinnen und Schülern notwendig (Diehl, 2010). Der lernwirksame Ansatz der Lernverlaufsdagnostik verspricht insbesondere für den inklusiven Unterricht Potentiale zur Vermeidung unzureichender Lesekompetenzen, wenn Entwicklungsmodellen in der Testkonstruktion berücksichtigt sind. Allerdings existiert für die deutsche Primarstufe noch keine theoriegeleitete Lernverlaufsdagnostik, die über mehrere Jahrgangsstufen auch die Lernentwicklungen von Lesenden mit niedrigen Leseleistungen misst. Diese Lücke schließt die vorliegende Dissertation.

Das Ziel der vorliegenden Dissertation ist die Entwicklung, die Evaluation und die praxisnahe Aufbereitung von theoriegeleiteten Lernverlaufstests zur Erfassung von basalen Lesekompetenzen in leistungsheterogenen Grundschulklassen. Mithilfe der neuen Verfahren wird eine Früherkennung von Leseschwierigkeiten und die Beobachtung von Lernentwicklungen im Längsschnitt ermöglicht. Im Zentrum der Dissertation stehen drei Forschungsschwerpunkte. Der erste Schwerpunkt thematisiert die Testentwicklung von Lernverlaufstests zur Messung basaler Lesekompetenzen in der inklusiven Grundschule in Deutschland. Lernverlaufstests für den deutschen Unterricht müssen testtheoretische Bedingungen sowie Bedingungen, die aus der gemeinsamen Beschulung von Schülerinnen und Schülern mit und ohne SPF resultieren, erfüllen. Daher werden die psychometrischen und schulbezogenen Bedingungen in der Testkonstruktion unter Berücksichtigung relevanter Lesemodelle vereint, um eine wirksame Lernverlaufsdagnostik für die Schullandschaft bereitzustellen. Die leitende Forschungsfrage lautet:

1. Wie können theoriegeleitete Lernverlaufstests zur Messung basaler Lesekompetenzen konstruiert sein, damit sie den Ansprüchen des inklusiven Unterrichts entsprechen?

Der zweite Forschungsschwerpunkt umfasst die Überprüfung der psychometrischen Güte der theoriegeleiteten Lernverlaufstests. In Anlehnung an Fuchs (2004) werden in diesem Forschungsschwerpunkt Studien des ersten und zweiten Gütestadiums realisiert. Lernverlaufstests für den inklusiven Unterricht müssen Leseleistungen fair und änderungssensibel auf unterschiedlichen Kompetenzniveaus differenziert erfassen. Insbesondere bei der Messung von Schülerinnen und Schülern mit geringen Leseleistungen und manifesten Leseschwierigkeiten dürfen keine Bodeneffekte (Walter, 2014) entstehen, da sonst die Kompetenzentwicklungen nicht ausreichend erfasst werden. Um eine ausreichende Kompetenzabbildung zu prüfen, werden in den Studien extreme Leseleistungen (besonders hohes und niedriges Ausgangsniveau) untersucht. Zur Prüfung der Testfairness werden unterschiedliche Gruppen von Schülerinnen und Schülern mit bestimmten Personenmerkmalen (Geschlecht, Migrationshintergrund, SFP) explizit mit einbezogen. Zusätzlich wird die Einsatzfähigkeit in und über mehreren Jahrgangsstufen der Grundschule hinweg geprüft. Die leitende Fragestellung zum zweiten Forschungsschwerpunkt lautet:

2. Eignen sich die theoriegeleiteten Lernverlaufstests zur langfristigen Messung von Leseleistungen auf unterschiedlichen Kompetenzniveaus?

Der dritte Forschungsschwerpunkt thematisiert die praxisnahe Aufbereitung der Lernverlaufstests für eine Implementation in den inklusiven Grundschulunterricht. In Deutschland ist der Ansatz der Lernverlaufsdagnostik für viele Lehrkräfte neu und die Tests sind unbekannt. Für eine barrierefreie Implementation in der Praxis benötigen Lehrkräfte unterstützende Materialien und spezifische diagnostische Kompetenzen (Ardoin et al., 2013). Die Testsysteme müssen zu den Bedingungen in der Schule passen und von den Lehrkräften akzeptiert werden. Dieser Schwerpunkt schafft den Grundstein für wissenschaftliche Studien zum dritten Gütestadium der Erforschung von Lernverlaufsdagnostik nach Fuchs (2004). Die zentrale Forschungsfrage lautet:

3. Welche Anforderungen aus der Schulpraxis müssen adressiert werden, um eine theoriegeleitete Lernverlaufsdagnostik in deutschen Grundschulen zu implementieren?

4. Vorgelegte Fachpublikationen der Dissertation

Im folgenden Kapitel werden die fünf Fachpublikationen der kumulativen Dissertation zusammengefasst sowie ihr Erkenntnisgewinn erläutert. Zusätzlich werden die Eigenleistung der Erstautorin und bei Bedarf Vorleistungen der Publikationen thematisiert. Abschließend werden Erkenntnisse aus zusammenhängenden wissenschaftlichen Arbeiten angeführt. Die fünf angeführten Publikationen der Dissertation befinden sich in gleicher Reihenfolge im Anhang.

4.1 Artikel 1

Inhaltliche Zusammenfassung

Artikel 1 (Jungjohann, Gegenfurtner & Gebhardt, 2018) untersucht und beschreibt die aktuellen Entwicklungstrends von Lernverlaufstests, welche die Leseflüssigkeit im Anfangsunterricht messen. Als Anfangsunterricht sind im Kontext des Artikels 1 die ersten drei Klassenstufen definiert, um den Einschluss von Instrumenten für Schülerinnen und Schüler mit milden und manifesten Leseschwierigkeiten zu gewährleisten. Im Rahmen des systematischen Reviews werden Forschungsfragen zu sechs Inhaltsbereichen analysiert. Zu den Inhaltsbereichen zählen (1) die allgemeine Testkonstruktion, (2) die Einbindung von Lesemodellen in die Testkonstruktion, (3) die Anwendung der Testverfahren im Unterricht, (4) die Bereitstellung und Verknüpfung von weiterführenden Unterstützungsmaterialien für die praktizierenden Lehrkräfte, (5) die Darstellung und Aufbereitung der Testergebnisse in Lernverlaufgraphen sowie (6) die psychometrische Prüfung der Instrumente. Dieser Überblicksartikel ist ein literaturbasiertes systematisches Review nach dem Leitfaden von Petticrew und Roberts (2012). Der Reviewprozess umfasst folgende Schritte: Definition von Ein- und Ausschlusskriterien der gesuchten Studien, die schriftliche Dokumentation des Reviewprozesses, die systematische Suche nach geeigneten Studien, die Eignungsprüfung der gefundenen Studien, die Entscheidung über den Ein- oder Ausschluss der Studien sowie die Verschriftlichung des Reviews. Als Grundlage für diese Übersichtsarbeit dienen ausschließlich wissenschaftliche Studien, die in einem Journal mit Begutachtungsverfahren veröffentlicht wurden. Außerdem müssen die Studien ein Testverfahren im Sinne der Lernverlaufdiagnostik zur Messung der Leseflüssigkeit vorstellen. Die Stichproben der empirischen Studien

dürfen neben den ersten drei Klassenstufen ebenfalls weitere Schulstufen sowie Schülerinnen und Schüler mit und ohne SPF berücksichtigen. Durch die festgeschriebenen Ein- und Ausschlusskriterien wurden in nationalen und internationalen Publikationsorganen insgesamt 693 Suchtreffer generiert und auf ihre Eignung geprüft. Auf der Basis von neun Veröffentlichungen wurden acht Testverfahren in das Review aufgenommen. Die Ergebnisse zeigen, dass im Anfangsunterricht sowohl Lesetexte als auch Itemlisten als ausschließlich analoge Lernverlaufsdagnostik angeboten werden. Die Testautorinnen und Testautoren nehmen zwar teilweise eine unterschiedliche Schwierigkeit der Items an, prüfen diese aber nicht explizit. Ebenfalls fehlt eine Prüfung im Sinne der Item-Response-Theorie gänzlich. Außerdem erfolgt bei keinem Verfahren eine Prüfung der Testfairness für Schülerinnen und Schüler mit SPF. Daher ist die Einsatzfähigkeit für den inklusiven Unterricht nicht hinreichend geklärt. Die bestehenden Verfahren sind weder mit weiterführenden erklärenden noch mit lesedidaktischen Materialien verknüpft. Aus den Ergebnissen leiten sich die aktuellen Bedarfe für den inklusiven Grundschulunterricht in Deutschland ab. Für die deutsche Schullandschaft fehlen digitale Verfahren, die in einer leistungsheterogenen Lerngruppe mit Schülerinnen und Schülern mit SPF einsetzbar sind und die die Lehrkräfte in ihrer Anwendung durch Unterstützungsmaterialien begleiten.

Eigenleistungen der Erstautorin

Die Idee zu Artikel 1 (Jungjohann, Gegenfurtner et al., 2018) entstand aus dem Diskurs mit Markus Gebhardt. Gemeinsam wurde das Ziel der Studie erarbeitet: eine Übersichtsarbeit zu internationalen und nationalen Entwicklungstrends von Lernverlaufstests zur Messung der Leseflüssigkeit im Anfangsunterricht. Die Motivation zu diesem Beitrag lag darin, anderen Forschenden Entwicklungslücken und Bedarfe aufzuzeigen und Anregungen für weitere Forschungsbemühungen zu geben. Die Erstautorin war federführend für diesen Beitrag zuständig. Sie führte die theoretische Aufarbeitung, die Festlegung der Forschungsfragen, die Festschreibung der Ein- und Ausschlusskriterien sowie den gesamten Suchprozess eigenständig durch. Die Begutachtung der potentiellen Studien fand gemeinsam mit Markus Gebhardt statt. Die Analyse der Studien sowie das Verfassen des Reviews lagen in der Verantwortung der Erstautorin. Die Diskussion entstand in Kooperation mit beiden Ko-Autoren, wobei die Erstautorin den maßgeblichen Anteil übernahm. Beide Ko-

Autoren waren in den gesamten Prozess eingebunden. Andreas Gegenfurtner lieferte Anregungen zur Ergebnisdarstellung. Die Erstautorin berücksichtigte die Anregungen der Ko-Autoren bei der Erstellung des Manuskripts und pflegte sie ein. Die Überarbeitung nach dem Review leistete vollständig die Erstautorin.

4.2 Artikel 2

Inhaltliche Zusammenfassung

Artikel 2 (Jungjohann, DeVries, Gebhardt & Mühling, 2018) stellt die Onlineplattform Levumi sowie die theoriegeleitete Testkonstruktion der Leseflüssigkeitstests vor und evaluiert die psychometrische Güte des Silbentests SiL-Levumi im Quer- und Längsschnitt unter Berücksichtigung von Schülerinnen und Schülern mit SPF. Im ersten Teil des Beitrags wird Levumi als universitäre Forschungsplattform (Gebhardt et al., 2016b) vorgestellt, die webbasierte Lernverlaufstests kostenlos für Lehrkräfte und Forschende anbietet. Nach einer Registrierung können alle Tests direkt im Browser durchgeführt werden. Bei jedem Messzeitpunkt kreiert die Plattform eine individuelle Paralleform durch eine zufällige Itemziehung aus einem fest definierten Itempool. Die Plattform speichert die Ergebnisse anonymisiert für Forschungszwecke und bietet Lehrkräften, Schülerinnen und Schülern automatisierte Rückmeldungen an. Die Lehrkräfte erhalten Lernverlaufsgraphen für die gesamte Lerngruppe und Individualgraphen mit einer qualitativen Analyse für jede einzelne Schülerin und jeden einzelnen Schüler. Die Plattform kann in leistungsheterogenen Settings an Grundschulen und Förderschulen eingesetzt werden. Die Leseflüssigkeit wird mittels unterschiedlicher Testarten und Niveaustufen erhoben. Die verfügbaren Testarten sind das Silbenlesen, das Wörterlesen sowie das Pseudowörterlesen. Jede Testart misst unterschiedliche Teilkompetenzen des Lesens (für mehr Informationen siehe Jungjohann & Gebhardt, 2018). Jede Testart kann auf unterschiedlichen Niveaustufen administriert werden. Die Konstruktion der Niveaustufen erfolgt in Anlehnung an den KLA (Dummer-Smoch & Hackethal, 2016). Die Niveaustufen beinhalten eine begrenzte Graphenauswahl und unterschiedlich komplexe Lautstrukturen, um die Schwierigkeit der Items einer Niveaustufe zu kontrollieren.

Im empirischen Abschnitt von Artikel 2 werden die drei Niveaustufen (N2b, N3, N4) zur Evaluation des Silbentests SiL-Levumi herangezogen. Der Beitrag adressiert vier

Forschungsfragen. Zuerst wird untersucht, ob das Rasch-Modell angenommen werden kann. Dies dient der Prüfung der theoriegeleiteten Itemkonstruktion innerhalb der drei Niveaustufen. Zweitens wird gefragt, ob der Vergleich der Summenwerte der Leseflüssigkeitstests gute Test-Retest-Reliabilitätswerte über zwei Messzeitpunkte ergeben. Diese Forschungsfrage dient der Prüfung, ob geeignete Parallelförmungen über den zufälligen Ziehprozess von Items durch die Plattform erzeugt werden. Drittens wird untersucht, ob die Leseflüssigkeitstests die Leistungen von Schülerinnen und Schülern mit SPF über mehrere Messzeitpunkte änderungssensibel erfassen. Viertens wird geprüft, ob die Leseflüssigkeitstests die Leistungen von Schülerinnen und Schülern mit und ohne SPF fair messen. Zur Beantwortung der Forschungsfragen werden fünf verschiedene große Stichproben ($N = 8 - 334$) gebildet, die zwischen ein und 14 mal die Leseflüssigkeitstests bearbeiteten. In der Itemfit-Statistik hat keins der untersuchten Items einen kritischen Itemfit aufgewiesen (Wright & Linacre, 1994). Daher sind alle theoriegeleiteten Items für den Itempool zulässig und kein Item musste entfernt werden. Innerhalb des Rasch-Modells weisen alle drei Niveaustufen eine mittlere bis hohe Reliabilität ($WLE - Reliabilität_{N2b-N4} = .883 - .919$) auf, wodurch die Passung des Rasch-Modells als gut bewertet wird. Als Konsequenz werden die Summenwerte für weitere Berechnungen herangezogen. Auf der Basis der Summenwerte fällt die Test-Retest-Reliabilität geringer aus als bei den Raschanalysen ($r_{N2b-N4} = .76 - .85$), ist aber mit der $WLE - Reliabilität$ im Einklang. Folglich ist die Zusammenstellung der individuellen Parallelförmungen durch die zufällige Itemziehung geeignet. Der Vergleich der Summenwerte der abhängigen Stichprobe von Schülerinnen und Schülern mit SPF im Längsschnitt zeigt innerhalb von zwei Niveaustufen (N2b, N4) signifikante Änderungen in den Leseleistungen. Für die Niveaustufe N3 sind keine signifikanten Änderungen gefunden worden. Trotzdem wird geschlussfolgert, dass die Leseflüssigkeitstests individuelle Änderungen in den Leseleistungen von Schülerinnen und Schülern mit SPF sensibel erfassen. Ein Erklärungsansatz für das Ausbleiben signifikanter Änderungen bei Niveaustufe N3 liegt in der Jahreszeit während der Erhebung. Die Messzeitpunkte lagen zwischen März und Mai und schlossen somit die Osterferien ein. Durch die Osterferien könnte ein Lernplateau entstanden sein, was zu geringeren Lernzuwächsen führt (Coelen & Siewert, 2008). Sowohl der Einfluss von Ferien als auch eine nähere Prüfung der Änderungssensibilität der Niveaustufe N3 steht noch aus. Die Untersuchung der Testfairness ergab keinen signifikanten Unterschied zwischen den Summenwerten

von Schülerinnen und Schülern mit und ohne SPF in der Niveaustufe N4 ($n_{SPF} = 46, n_{kein\ SPF} = 254, Welch's\ t(83.4) = 1.08, p > .25$). Folglich misst der Silbenlesetest N4 gleichermaßen gut die Leistungen von Schülerinnen und Schülern mit und ohne SPF und ist somit in leistungsheterogenen Klassen einsetzbar. Artikel 2 bekräftigt wichtige Aspekte in der Testentwicklung von Lernverlaufstests für den inklusiven Unterricht. Es wird bestätigt, dass die Levumi Leseflüssigkeitstests gleichermaßen gut für Schülerinnen und Schüler mit und ohne SPF nutzbar sind. Durch diese Eigenschaft arbeiten sich die Lehrkräfte nur in eine Testsystematik ein und alle Schülerinnen und Schüler einer inklusiven Lerngruppe können auf ihrem individuellen Kompetenzniveau die Leseflüssigkeitstests nutzen. Außerdem ermöglicht das Levumi Testsystem die Evaluierung unterschiedlicher Leseteilkompetenzen (z.B. Phonologisches Rekodieren beim Silbenlesen). Ein weiterer Beitrag für die Testentwicklung von Lernverlaufstests des Artikels 2 liegt in der Prüfung der digitalen Testadministration, da die Erstellung einer individuellen Parallelfarm zu jedem Messzeitpunkt durch die zufällige Ziehung der Items erfolgreich funktioniert. Dies reduziert die organisatorische Arbeit von Lehrkräften bei der Nutzung von Lernverlaufstests und erhöht somit die Praktikabilität im Unterricht.

Eigenleistung der Erstautorin

Im Austausch zwischen Markus Gebhardt, Jeffrey M. DeVries und der Erstautorin entstand die Idee und Struktur des Artikels 2 (Jungjohann, DeVries, Gebhardt et al., 2018). Gemeinsam wurde als Ziel die Prüfung des Rasch-Modells, der Änderungssensibilität sowie der Testfairness festgelegt. Die theoretische Aufarbeitung erstand vollständig durch die Erstautorin. Sie sichtete eigenständig die benötigten Forschungsdaten und bereitete sie auf. Die statistischen Auswertungen sowie die Formulierung der Ergebnisse wurden zwischen der Erstautorin und dem zweiten Ko-Autor aufgeteilt. Herr Jeffrey M. DeVries übernahm die Raschanalysen im Sinne der Item-Response-Theorie und die Prüfung der Änderungssensibilität. Die Erstautorin übernahm die Berechnung der Testfairness und prüfte die Analysen von Herrn Jeffrey M. DeVries. Die Ausführungen zur Diskussion sowie zukünftigen Forschungsarbeiten strukturierte und formulierte die Erstautorin mit Anregungen der Ko-Autoren eigenständig. Das gesamte Manuskript editierte Jeffrey M. DeVries als englischsprachiger Native Speaker. Markus Gebhardt beriet und editierte das Manuskript während des Entstehungsprozesses. Die Anregungen der gutachtenden

Personen pflegte die Erstautorin ein. Jeffrey M. DeVries übernahm die Formatierung des Beitrags.

Nachfolgende Leistungen

Das Ziel der Prüfung der psychometrischen Güte von Lernverlaufstests ist die Bereitstellung von wissenschaftlich gesicherten Tests in der Schulpraxis (Deno, 2003b). Als Vorbereitung für die Testveröffentlichung wurden während und nach der Veröffentlichung des Artikels 2 weitere Forschungsdaten zur Prüfung des Silbentests SiL-Levumi in allen Niveaustufen erhoben. Des Weiteren wurde die Akzeptanz der Tests in der Praxis geprüft. Dafür holte die Erstautorin die Meinungen von Lehrkräften aus Kooperationsschulen ein und wertete sie aus. Die Expertinnen und Experten aus der Schulpraxis deklarierten einige Items als untauglich. Auf Basis aller Forschungsdaten prüfte die Erstautorin mehrfach die Itempools aller Niveaustufen. Bei diesen Prüfungen wurden ungeeignete, doppelte sowie fehlerkonstruierte Items aus den Itempools entfernt. Zusätzlich wurde die Kriteriumsvalidität der Leseflüssigkeitstests mit der Kurzversion des standardisierten Leseverständnistests ELFE II (Lenhard, Lenhard & Schneider, 2017) geprüft (Anderson, Jungjohann & Gebhardt, angenommen). Der Test Silbenlesen korreliert mit dem Gesamtwert des ELFE II zu Beginn des zweiten Schulhalbjahres positiv ($r_{\text{Pearson}} = .71, p < .001$). Abschließend wurden die Itempools aller Niveaustufen veröffentlicht.

Der Silbentest SiL-Levumi wurde durch die Erstautorin im Testarchiv (<https://www.testarchiv.eu>) des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) unter der Open Access Creative-Commons-Lizenz CC-BY-NC-SA veröffentlicht (Jungjohann, Diehl et al., 2019). In dieser Veröffentlichung sind die Test- und Itemkonstruktion, vollständige Itemlisten, Beschreibungen der Testanwendung, Durchführung und Auswertung als digitale und analoge Testversion dokumentiert. Zusätzlich wurden für jede Niveaustufe vier feste Parallelversionen veröffentlicht, die eine Testanwendung als Papierversion ermöglichen. Die Erstautorin entschied sich für vier Parallelversionen, damit die Tests mindestens als Screening mit größeren Messabständen in der ersten Präventionsstufe eingesetzt werden können. Diese Veröffentlichung dient zum einen als weitere Informationsquelle für Lehrkräfte und Forschende und zum anderen als Bereitstellung einer analogen Testdurchführung. Für Forschende wird die Testkonstruktion und Testanwendung

transparent publiziert, sodass der SiL-Levumi Test für weitere Forschungsaktivitäten zur Verfügung steht.

4.3 Artikel 3

Vorbereitende Leistungen

Artikel 3 (Jungjohann, DeVries, Mühling & Gebhardt, 2018) stellt eine neu entwickelte Lernverlaufsdiagnostik vor, die das basale Leseverständnis auf Satzbasis unter Berücksichtigung semantischer und syntaktischer Satzstrukturen erhebt. Als Einsatzgebiet ist die Grundschule vorgesehen. Die neue Lernverlaufsdiagnostik fungiert als fachliche Weiterführung der Leseflüssigkeitstests der Onlineplattform Levumi, da zu den basalen Lesekompetenzen neben der Leseflüssigkeit Verständnisleistungen auf Satzebene zählen (National Institute of Child Health and Human Development, 2000). Für einen kompetenzübergreifenden Einsatz wurde der Leseverständnistest SinnL-Levumi auf zwei Niveaustufen (N2, N4) konstruiert. Die Testkonstruktion berücksichtigt maßgeblich drei lesetheoretische Aspekte. Erstens bildet die Graphemauswahl der Niveaustufen der Onlineplattform Levumi die Grundlage der Testitems. Die Regeln zu den Lautstrukturen wurden modifiziert, damit vollständige Sätze bildbar sind. Festgelegt wurde, dass die Wörter der Niveaustufe N2 ebenfalls Konsonantencluster enthalten. Zweitens wurden altersgerechte Wörter verwendet, damit das Wortwissen das Leseverständnis nicht beeinträchtigt. Diese Bedingung wurde über den Abgleich mit Wortlisten (z.B. Grundwortschätze, häufige Wörterlisten) aus mehreren Bundesländern gewährleistet. Es wurden nur Wörter verwendet, die von mindestens zwei Bundesländern empfohlen wurden, um regionsspezifische Wörter (z.B. Boje) auszuschließen. Drittens wurde eine altersgerechte Satzgrammatik (z.B. Vermeidung von Passiv-Konstruktionen und mehreren Nebensätzen) genutzt. Die Niveaustufe N2 enthält 66 Items in zwei Dimensionen und die Niveaustufe N4 enthält 60 Items in drei Dimensionen.

Als Vorbereitung für Artikel 3 fand eine Pilotierung einer Kurzversion der Niveaustufe N4 im Juli 2017 in einer dritten inklusiven Grundschulklasse im westlichen Ruhrgebiet statt. Das Ziel der Pilotierung war die Prüfung der Verständlichkeit der Items und des Aufgabenformats für die anvisierte Zielgruppe. Die teilnehmenden Schülerinnen und Schüler ($N = 20$, $M_{Alter} = 8;8 \text{ Jahre}$, $n_{Mädchen} = 10$,

$n_{Förderbedarf} = 3$, $n_{Migrationshintergrund} = 5$) bearbeiteten eine von zwei papierbasierten Parallelversionen, die 24 zufällig ausgewählte Items enthielt. In beiden Parallelversionen wurden unterschiedliche Items verwendet und randomisiert an die Schülerinnen und Schüler verteilt. Die Schülerinnen und Schüler hatten im Sinne eines Powertests (Rost, 2004) 15 Minuten Bearbeitungszeit. Sie beantworteten zwischen acht und 24 Items korrekt. Alle Items wurden von mindestens der Hälfte der Lerngruppe korrekt beantwortet. Die Lehrkraft stellte in Bezug auf das Aufgabenformat keine Verständnisschwierigkeiten fest. Bei zwei Items waren verwendete Wörter unbekannt, sodass Nachfragen aufkamen. Im Anschluss an die Pilotierung wurden die theoriegeleiteten Items überarbeitet und die missverständlichen Wörter ersetzt. Außerdem wurde die Testzeit unter testtheoretischen Aspekten eines Speedtests (Rost, 2004) und der Expertenmeinung der Lehrkraft auf acht Minuten gekürzt. Nach Abschluss der Pilotierung wurden die Tests für die Onlineplattform Levumi von einer wissenschaftlichen Hilfskraft unter Anleitung der Erstautorin programmiert. Neben der Optimierung des Itempools wurde die Onlineplattform durch die Option der Tabletnutzung erweitert. Außerdem dienen neu installierte einführende Beispiele der Vermeidung vorzeitiger Teststarts.

Inhaltliche Zusammenfassung

Artikel 3 (Jungjohann, DeVries, Mühling et al., 2018) erläutert die theoriegeleitete Testkonstruktion der neuen Lernverlaufsdagnostik SinnL-Levumi auf Niveaustufe N4 zum Leseverständnis auf Satzbasis und prüft die psychometrische Güte des Tests mit einer abhängigen Stichprobe zu zwei Messzeitpunkten. Die Testkonstruktion orientiert sich an den US-amerikanischen Maze-Tests und modifiziert diese unter Berücksichtigung verschiedener Leseverständnismodelle (Cain, Patson & Andrews, 2005; Christmann & Groeben, 1999; Kennison, 2009). Die markantesten Modifikationen stellen sich folgendermaßen dar. Es werden einzelne Sätze verwendet, um das Satzverständnis nicht durch eine zusammenhängende Textstruktur zu beeinflussen und Lesende mit geringen Verständnisleistungen dadurch nicht zu benachteiligen. Des Weiteren werden die zu füllenden Lücken in den Sätzen anhand spezifischer Wortarten bestimmt. In Anlehnung an bisherige Forschungen zum Verständnis einzelner Wortarten innerhalb von Sätzen (Cain et al., 2005; Frisson, Rayner & Pickering, 2005; Kennison, 2009) wurden drei inhaltliche Dimensionen festgelegt. Die Dimensionen sind Auslassung von (1) Argumenten

(Subjekt oder Objekt), (2) Prädikaten (Verb oder Adjektiv) oder (3) Konnektiva (Konjunktion oder Präposition). Durch diese Systematik können Lehrkräfte spezifische Schwierigkeiten im Satzverständnis aus den Lernverlaufsdaten ableiten. Ähnlich wie die Leseflüssigkeitstests wird der SinnL-Levumi Test webbasiert administriert und die Parallelformen werden durch eine zufällige Ziehung bei jeder Messung erstellt.

Artikel 3 beantwortet fünf Forschungsfragen im Sinne der ersten beiden Gütestadien nach Fuchs (2004). Die erste Frage untersucht die Itemschwierigkeit und die Schwierigkeitsunterschiede zwischen den drei Dimensionen. Die zweite Forschungsfrage untersucht die Faktorenstruktur des Tests und fragt, ob die Testergebnisse eindimensional interpretiert werden können. Mit der dritten Forschungsfrage wurde die Messinvarianz einerseits gegenüber verschiedenen Gruppen von Schülerinnen und Schülern (Geschlecht, SPF, Migrationshintergrund) und andererseits zu den zwei Messzeitpunkten geprüft. Den Einfluss der Klassenzugehörigkeit auf die Testergebnisse untersucht die vierte Forschungsfrage. Abschließend wird verglichen, ob die Leistungen der unterschiedlichen Gruppen untereinander und im Längsschnitt unterschiedlich ausfallen. Die Stichprobe umfasst $N = 761$ Drittklässlerinnen und Drittklässler ($n_{Mädchen} = 354$, $n_{SPF} = 106$, $n_{Migrationshintergrund} = 344$) aus dem nordwestlichen Ruhrgebiet. Die Schülerinnen und Schüler haben zu zwei Messzeitpunkten im Abstand von drei Wochen den Leseverständnistest SinnL-Levumi bearbeitet. Die Itemanalyse mittels der einfaktoriellen ANOVA zeigt signifikant unterschiedlich schwierige Auslassungsdimensionen ($F_{(2,57)} = 25.215, p < .001$). Der Tukey post-hoc Test zeigt keinen Unterschied zwischen der ersten (Subjekt oder Objekt) und der zweiten (Verb oder Adjektiv) Auslassungsdimension ($p > .05$). Die dritte Dimension (Konjunktionen oder Präpositionen) ist hingegen schwieriger als die ersten beiden Dimensionen ($p < .05$). Dieses Ergebnis spiegelt die angenommenen theoretischen Schwierigkeiten in der Testkonstruktion wieder. In der Testanwendung wird dieses Ergebnis durch die zufällige Ziehung aufgefangen, da aus allen drei Dimensionen gleichmäßig zufällig Items ausgewählt werden. Folglich erhalten alle Schülerinnen und Schüler gleich viele leichte und schwierige Items. Die konfirmatorischen Faktorenanalysen haben gute Fitwerte für alle drei getesteten Modelle (Hu & Bentler, 1998) ergeben, wodurch eine eindimensionale Testinterpretation möglich ist. Alle drei getesteten Faktorenmodelle sind messinvariant gegenüber den drei untersuchten

Gruppen von Schülerinnen und Schülern (Geschlecht, SPF, Migrationshintergrund) sowie gegenüber beiden Messzeitpunkten. Diese Ergebnisse bestärken die eindimensionale Testinterpretation für alle getesteten Gruppen. Der Einfluss der Klassenzusammensetzungen beträgt ca. 15% ($ICC = .15$). Trotz der hohen Varianz (Hedges & Hedberg, 2007) funktioniert die Lernverlaufsdagnostik in allen Klassen ähnlich gut. Der SinnL-Levumi Test misst über den Zeitraum von drei Wochen änderungssensibel, da die gesamte Stichprobe zum zweiten Messzeitpunkt signifikant höhere Summenwerte erreicht ($F_{(1,658)} = 93.32, p < .001$). Schülerinnen und Schüler mit SPF erreichen signifikant geringere Leistungen als die ohne SPF ($F_{(1,658)} = 89.01, p < .001$). Diese Ergebnisse zeigen eine gute Einsatzfähigkeit des Tests in leistungsheterogenen Klassen und die Erfüllung der Identifikationsfunktion. Ähnlich wie Artikel 2, bestärkt Artikel 3 die Eignung der Levumi Lernverlaufstests für den Einsatz in leistungsheterogenen Klassen. Lehrkräfte von inklusiven Klassen können den Leseverständnistest für die gesamte Lerngruppe nutzen. Zusätzlich unterstützt der Test die Identifikation von Schülerinnen und Schülern mit geringen Leseleistungen. Insbesondere diese Schülerinnen und Schüler profitieren von der Verwendung von einzelnen, unzusammenhängenden Sätzen. Roland H. Good, Simmons und Kame'enui (2001) und Fuchs et al. (2004) beobachten bei Schülerinnen und Schülern mit manifesten Leseschwierigkeiten Überforderungen durch vollständige Texte. Eine solche Überforderung wird durch die neue Testkonstruktion ausgeschlossen. Artikel 3 zeigt zusätzlich, dass die aus den Lesemodellen abgeleiteten schwierigkeitsgenerierenden Regeln in der theoretischen Testkonstruktion abgebildet werden. Dies stellt einen großen Schritt in der Testentwicklung der Lernverlaufsdagnostik dar, weil der didaktische Nutzen der Diagnostik dadurch erhöht wird. Mithilfe des Levumi Lesetests können Lehrkräfte nun auch in Bezug auf das Leseverständnis spezifische Verständnisschwierigkeiten bei einzelnen Schülerinnen und Schülern erkennen und in ihrer Förderung berücksichtigen. Für Forschende stellt Artikel 3 ebenfalls einen Mehrwert dar, da die verwendeten Analysen (Intra-Klassenkorrelation (ICC), konfirmatorische Faktorenanalyse, Prüfung der Messinvarianz) in das Forschungsfeld der Lernverlaufsdagnostik zur psychometrischen Prüfung erfolgreich übertragen wurden.

Eigenleistung der Erstautorin

Den Bedarf einer theoriegeleiteten Lernverlaufsdagnostik zur Erfassung des Leseverständnisses für die Onlineplattform Levumi formulierte Markus Gebhardt. Die Erstautorin des Artikels 3 übernahm vollständig die Ausgestaltung der Testkonstruktion, die Festlegung der schwierigkeitsgenerierenden Regeln (Auslassungsdimensionen), die Einbettung der Niveaustufen (Graphemauswahl und Lautstruktur der verwendeten Wörter) in die Items sowie die Auswahl der verwendeten Wörter. Als Vorarbeit erstellte eine studentische Hilfskraft eine Wörterliste aus mehreren Grundwortschätzen und häufigen Wörterlisten unter Anleitung der Erstautorin. Mit dieser Wörterliste formulierte die Erstautorin die Items beider Niveaustufen (N2, N4) aus. Die Erstautorin koordinierte die Pilotierung sowie die Hauptstudie ohne fremde Hilfe. Die Datenerhebung fand durch die Erstautorin und durch von ihr geschulten Studierenden statt. Die theoretische Fundierung, die Beschreibung der Studie sowie die Diskussion verschriftlichte die Erstautorin eigenständig. Die statistischen Auswertungen wurden zwischen Jeffrey M. DeVries und der Erstautorin mit unterschiedlichen Anteilen aufgeteilt. Die Datenaufbereitung übernahm die Erstautorin und überprüfte alle Rechnungen. Jeffrey M. DeVries und die Erstautorin formulierten die Forschungsergebnisse gemeinsam aus. Andreas Mühling übernahm die Einbettung des Testverfahrens in die Onlineplattform. Markus Gebhardt empfahl das Studiendesign und supervisierte während des gesamten Entstehungs- und Forschungsprozesses. Abschließend editierte Markus Gebhardt das Manuskript. Jeffrey M. DeVries korrigierte als englischsprachiger Native Speaker das finale Manuskript sprachlich. Die Überarbeitungsvorschläge der gutachtenden Personen pflegte die Erstautorin in das Manuskript ein.

Nachfolgende Leistungen

Die Intention von Artikel 3 (Jungjohann, DeVries, Mühling et al., 2018) war ebenfalls die Bereitstellung eines praktikablen Instruments der Lernverlaufsdagnostik für die Schulpraxis. Damit der Test möglichst wenig Unterrichtszeit in Anspruch nimmt, wurde die Testdauer anhand der Stichprobe aus Artikel 3 unter Berücksichtigung der Reliabilität ($WLE - Reliabilität = .86 - .90$) auf fünf Minuten gekürzt (Schurig, Jungjohann & Gebhardt, 2019). Wie bei den Leseflüssigkeitstests (siehe Artikel 2) wurde zur Sicherung der Kriteriumsvalidität die Pearson Korrelation zwischen dem

Levumi Leseverständnistest (N4) und dem standardisierten Leseverständnistest ELFE II (Lenhard et al., 2017) berechnet (Anderson et al., angenommen). Der Levumi Leseverständnistest auf Satzbasis korreliert hoch mit dem ELFE II Subtest Satzverständnis ($r = .81, p < .001$).

Mit der gleichen Intention wie bei der Testveröffentlichung der Leseflüssigkeitstests, publizierte die Erstautorin den SinnL-Levumi Test zum Leseverständnis auf Satzbasis im Testarchiv des ZPID unter der Open Access Creative-Commons-Lizenz CC-BY-NC-SA (Jungjohann & Gebhardt, 2019). In der Testveröffentlichung werden die Test- und Itemkonstruktion, die vollständigen Itemlisten, die Beschreibungen der Testanwendung, Durchführung und Auswertung als digitale und analoge Testversion beschrieben. Ebenfalls wurden in der Testveröffentlichung pro Niveaustufe vier feste Parallelversionen für die analoge Testanwendung abgedruckt.

4.4 Artikel 4

Inhaltliche Zusammenfassung

Artikel 4 (Jungjohann, Schurig & Gebhardt, 2019) analysiert explorativ mögliche Lernentwicklungen in den basalen Lesekompetenzen (Leseflüssigkeit und Leseverständnis auf Satzbasis) über ein Schuljahr hinweg in der dritten und vierten Jahrgangsstufe einer inklusiven Grundschule. Das Ziel der Studie ist die Untersuchung von Lernverläufen unterschiedlich starker Lesenden im regulären Grundschulunterricht, um die jahrgangsübergreifende Anwendung der Levumi Tests zu prüfen und Implikationen für die Anwendung der Lernverlaufsdagnostik abzuleiten. Die Lehrkräfte der teilgenommenen Klassen haben auf freiwilliger Basis die Lernverlaufsdaten während der Studienlaufzeit nicht für die Förderplanung eingesetzt und erst nach Abschluss der Studie Fortbildungen erhalten. Diese Studie ist dem zweiten Gütestadium nach Fuchs (2004) zuzuordnen, da die Forschungsdaten durch geschulte Studierende im Längsschnitt erhoben und die formative Wirkung der Lernverlaufsdagnostik durch das Studiendesign ausgeschlossen wurde. Folgende Forschungsfragen leiten die Studie in Artikel 4. (1) Wie entwickeln sich die Leseleistungen über ein gesamtes Schuljahr hinweg? (2) Lassen sich die Lernverläufe von Lesenden mit unterschiedlichem Ausgangsniveau einer Klassenstufe voneinander abgrenzen und (3) weisen die gebildeten Subgruppen differente

Lernverlaufskurven auf? Die Stichprobe umfasst $N = 90$ Schülerinnen und Schüler ($M_{\text{Alter zu } t_1} = 9; 1 \text{ Jahre}$, $n_{\text{Mädchen}} = 43$, $n_{\text{SPF}} = 2$, $n_{\text{Migrationshintergrund}} = 11$, $n_{\text{Fluchtgeschichte}} = 2$) aus jeweils zwei Klassenverbänden ($n_{\text{Stufe3}} = 49$, $n_{\text{Stufe4}} = 41$) einer inklusiven Grundschule aus dem westlichen Ruhrgebiet. Die Leseleistungen wurden jeweils vor den großen Schulferien (Herbst, Winter, Ostern, Sommer) durch drei Leseflüssigkeitstests (Silbenlesen, Wörterlesen, Pseudowörterlesen) sowie durch den Leseverständnistest SinnL-Levumi der Onlineplattform Levumi erhoben. Alle Schülerinnen und Schüler bearbeiteten die Tests auf derselben Niveaustufe N4, in derselben Reihenfolge und an den gleichen technischen Endgeräten (Computer im schuleigenen PC-Raum). Die Ergebnisse zeigen bis auf eine Ausnahme (Pseudowörterlesen in Stufe 3 zwischen Frühling und Sommer) zu allen Messzeitpunkten und in allen Tests beobachtbare Lernzuwächse. Die Effektstärken zeigen in allen Bedingungen starke Effekte ($d_{\text{Cohen}} = 1.05 - 1.57$) und zwischen dem ersten und letzten Messzeitpunkt steigen in beiden Schulstufen die Lernentwicklungen in allen Tests signifikant. Gemessen an linearen Regressionen erreichen die Schülerinnen und Schüler der dritten Klassenstufe zu Schuljahresbeginn in allen Tests geringere Intercepts als die der vierten Stufe bei ähnlichen Regressionskoeffizienten (b -Werte). In der Untersuchung der Schülerinnen und Schüler mit dem geringsten Leseausgangsniveau verlaufen ihre Lernentwicklungen ähnlich, wie die der durchschnittlich Lesenden. Dies trifft für beide exemplarisch untersuchten Lernverlaufstests Wörterlesen und SinnL-Levumi zu. In allen Bedingungen sind bei den Schülerinnen und Schülern mit dem höchsten Leseausgangsniveau Deckeneffekte entstanden, da sie bereits bei der ersten Testung zu Schuljahresbeginn die meisten oder alle Items der Tests lösten. Das letzte Studienergebnis ist die Anschlussfähigkeit der Tests zwischen beiden Jahrgangsstufen, da sich die Mittelwerte und Konfidenzintervalle der dritten Stufe beim Schuljahresende der vierten Stufe zu Schuljahresbeginn sehr stark ähneln. Die Ergebnisse aus Artikel 4 zeigen, dass die eingesetzten Testverfahren unter realen Unterrichtsbedingungen die Lernverläufe der Lesenden mit einem geringen und durchschnittlichen Ausgangsniveau in beiden Jahrgangsstufen erfolgreich erfassen. Dies entspricht dem grundsätzlichen Fokus der Lernverlaufsdagnostik, die Messung von Lernverläufen von förderbedürftigen Schülerinnen und Schülern (Fuchs, 2017). Artikel 4 bestätigt den gewinnbringenden Einsatz einer theoriegeleiteten Lernverlaufsdagnostik mit mehreren Niveaustufen und Testarten. Denn die Lernentwicklungen der besonders

lesestarken Schülerinnen und Schüler werden aufgrund von Deckeneffekten nicht abgebildet. In der Unterrichtspraxis können die Lehrkräfte die Leseleistungen dieser Schülerinnen und Schüler auf einen schwierigeren Test mit einer höheren Niveaustufe zurückgreifen. Dies trägt insbesondere zu einer einfachen Anwendung in leistungsheterogenen Klassen bei, da alle Schülerinnen und Schüler mit demselben Testsystem arbeiten können. Artikel 4 deckt außerdem potentielle Hürden bei der Interpretation der Lernverlaufsdaten auf, deren Erforschung noch aussteht. Je nach überprüfter Leseteilkompetenz und Leseausgangsniveau der Schülerinnen und Schüler entwickeln sich die Lernverläufe unterschiedlich. Damit Lehrkräfte testtheoretische Deckeneffekte nicht als ausbleibenden Kompetenzzuwachs interpretieren, benötigen sie testspezifische Interpretationshilfen sowie anwendungsbezogenes, diagnostisches und fachdidaktisches Wissen.

Eigenleistung der Erstautorin

Die Idee, in einer Langzeitstudie die basalen Lesekompetenzen über mehrere Jahrgangsstufen gemeinsam zu untersuchen, stammte von der Erstautorin des Artikels 4 (Jungjohann, Schurig et al., 2019). Markus Gebhardt beriet bei der Ausgestaltung des Studiendesigns. Die Koordination und Durchführung der Studie initiierte und realisierte die Erstautorin eigenständig. Die theoretische Fundierung, die statistischen Analysen, die Ergebnisdarstellung sowie die Diskussion des Beitrags konzipierte und verfasste die Erstautorin. Michael Schurig berechnete die Intra-Klassenkorrelation und pflegte diese Ergebnisse in das Manuskript ein. Außerdem beriet Michael Schurig die Erstautorin bei den statistischen Berechnungen. Beide Ko-Autoren editierten das Manuskript und gaben Anregungen zur Überarbeitung. Die Erstautorin pflegte diese Anregungen ein.

Nachfolgende Leistungen

Artikel 4 (Jungjohann, Schurig et al., 2019) bestärkt den Bedarf an unterstützenden Materialien für anwendende Lehrkräfte (Ardoin et al., 2013). Damit sich Lehrkräfte in die Lernverlaufstests der Onlineplattform Levumi möglichst barrierefrei einarbeiten, entwarf die Erstautorin in Kooperation mit weiteren Testautorinnen und Testautoren das „Levumi Handbuch für Lehrkräfte Deutsch“ und veröffentlichte es als Open Access Publikation (Jungjohann, Mau, Diehl & Gebhardt, 2019). Dieses Handbuch verbindet eine fachdidaktische Einführung in die basalen Lese- und Rechtschreibkompetenzen

mit anwendungsbezogenen und diagnostischen Hinweisen der Lernverlaufsdagnostik. Anhand von Beispielen werden (1) die messbaren Teilkompetenzen der einzelnen Tests sowie die differenzierten Niveaustufen, (2) die Zielgruppen der Tests und Niveaustufen, (3) die technische Durchführung, (4) die Möglichkeit strukturierter Implementation in den Unterricht und (5) die Kriterien zur Auswertung der Lernverlaufsdaten beschrieben.

4.5 Artikel 5

Vorbereitende Leistungen

Die Instrumente der Lernverlaufsdagnostik werden dazu entwickelt, dass Lehrkräfte die kindlichen Lernentwicklungen evaluieren, Non-Responder identifizieren und daraufhin geeignete Fördermaßnahmen initiieren (Stecker, 2017). Für eine effektive Nutzung der Lernverlaufsdagnostik wird die Verknüpfung der diagnostischen Daten mit fachdidaktischem Wissen zur Kompetenzförderung gefordert (Ardoin et al., 2013). Als Konsequenz aus Artikel 4 und zur Vorbereitung zu Artikel 5 (Jungjohann, Diehl, Mühling & Gebhardt, 2018) veröffentlichte die Erstautorin das Förderhandbuch „Förderansätze im Lesen mit Levumi“, welches die Anwendung der Lernverlaufsdagnostik Levumi mit konkreten Fördermaßnahmen zur Steigerung der basalen Lesekompetenzen verknüpft und es beinhaltet Arbeitsblätter für den Unterricht (Jungjohann, Gebhardt, Diehl & Mühling, 2017). Der Schwerpunkt des Förderhandbuchs liegt auf fachdidaktischem Wissen zum Leseerwerb im Anfangsunterricht. Es wurde ebenfalls Open Access publiziert, um eine uneingeschränkte Nutzung für unterrichtliche Zwecke zu sichern. Das Handbuch soll Lehrkräfte bei der Auswahl von geeigneten Fördermaßnahmen in Abhängigkeit von den kindlichen Lernbedürfnissen durch fachdidaktisches Wissen unterstützen und damit Überförderung und Demotivation durch ungeeignetes Übungsmaterial verhindern.

Inhaltliche Zusammenfassung

Artikel 5 (Jungjohann, Diehl et al., 2018) thematisiert literaturbasiert Chancen und Grenzen von digitalen Testsystemen der Lernverlaufsdagnostik im Hinblick auf die Implementation in der deutschen Schullandschaft und deckt Barrieren und Förderfaktoren bei der Interpretation von Lernverlaufsgraphen auf. Abschließend wird

das umfassende Testsystem der Onlineplattform Levumi als Praxisbeispiel kurz vorgestellt. Leitende Forschungsfrage ist, welche Anforderungen entstehen aus der neuen Diagnostikart an digitale Testsysteme und Lehrkräfte. Das Ziel des Artikels ist die Aufdeckung neuer Forschungsfelder im Hinblick auf die Implementation der Lernverlaufsdagnostik. Nach einer Einführung in die Lernverlaufsdagnostik werden die Anforderungen an die Testsysteme aufgearbeitet. Die digitalen Testsysteme benötigen eine explorative Erkundung für digitale Natives sowie eine detaillierte Anleitung für technikunerfahrene Nutzerinnen und Nutzer, um allen Lehrkräften einen möglichst barrierefreien Zugang zu ermöglichen. Außerdem müssen die Testsysteme äußerst flexibel einsetzbar sein, da die technische Ausstattung in den Schulen unterschiedlich ist. Es gibt Schulen mit und ohne WLAN und mit Endgeräten unterschiedlicher Standards (Desktop-PCs, Laptops, Tablets). Aufgrund der breiten sprachlichen Heterogenität der Schülerinnen und Schüler (Dirim, Hauenschild & Lütje-Klose, 2008) sind adaptive Testsysteme wünschenswert, die in Abhängigkeit des Kompetenzprofils Tests und Fördermaterialien automatisiert bereitstellen. Die Lehrkräfte profitieren vor allem durch flexibel einsetzbare Systeme. Starre Testzeiträume und eine verpflichtende Anzahl von Messzeitpunkten entsprechen nicht der benötigten Flexibilität in leistungsheterogenen Klassen. Umfassende Systeme mit theoriegeleiteten Tests und passenden Fördermaterialien erhöhen die Bereitschaft von Lehrkräften, die neue Art der Diagnostik im eigenen Unterricht zu etablieren. Die Etablierung hängt ebenfalls davon ab, wie sehr die Nutzung der Lernverlaufsdagnostik in das Unterrichtsgeschehen eingreift. Lehrkräfte mit unterschiedlichen Berufserfahrungen wünschen sich Systeme, die entweder die Unterrichtsplanung leiten oder möglichst wenig in das eigene didaktische Konzept eingreifen. Als zentrale Voraussetzung für die Nutzung der Lernverlaufsdagnostik muss das Verstehen und die Interpretation der Lernverlaufgraphs gesichert werden. Von Vorteil ist es, wenn die Lehrkräfte Sicherheit in der Auswertung und Interpretation der Daten erwerben und die Potentiale und Gefahren von graphischen Hilfsmitteln in der Interpretation präsent sind. Der Lernbereich Lesen der Onlineplattform Levumi bietet ein breites unterstützendes Angebot für Lehrkräfte, das eine Implementation in der Schulpraxis fördern könnte. Neben den Tests zur Messung der Leseflüssigkeit sowie des Leseverständnisses existieren Hilfen zur Anwendung und Interpretation sowie Fördermaterialien, die mit der Systematik der Lernverlaufsdagnostik unmittelbar verknüpft sind. Außerdem orientieren sich die Testkonstruktionen an

Lesetheorien, sodass Hinweise für die Förderplanung aus den Lernverlaufsdaten abgeleitet werden können. Im Sinne des dritten Gütestadiums nach Fuchs (2004) stehen Effektivitätsprüfungen der unterstützenden Materialien und der Dateninterpretation sowie systematische Implementationsuntersuchungen der Levumi Plattform noch aus.

Eigenleistung der Erstautorin

Die Idee, den aktuellen Stand der Forschung als Vorbereitung für weitere Implementationsstudien aufzubereiten, entwickelte die Erstautorin des Artikels 5 (Jungjohann, Diehl et al., 2018). Die Erstautorin strukturierte und formulierte den Beitrag. Kirsten Diehl verfasste anteilig die Einleitung. Markus Gebhardt ergänzte Anforderungsaspekte der Testsysteme. Beide Ko-Autoren und die Ko-Autorin editieren das Manuskript und lieferten Überarbeitungshinweise. Die Erstautorin setzte die Überarbeitungshinweise des Verlags sowie die Formatierung des Beitrags um.

5. Diskussion und Limitationen

Das Ziel der vorliegenden kumulativen Dissertation ist es, eine Früherkennung von Leseschwierigkeiten im präventiven Sinn in inklusiven Klassen durch den Einsatz der Lernverlaufsdagnostik zu ermöglichen. Dafür wurden Lernverlaufstests zur Messung von basalen Lesekompetenzen für den Einsatz in der inklusiven Grundschule entwickelt und evaluiert. Auf der Basis von fünf Fachpublikationen werden zwei theoriegeleitete Lernverlaufsdagnostiken vorgestellt und Überlegungen zur Implementation angeführt. Das erste Verfahren misst die Leseflüssigkeit mit drei Testarten (Silben, Wörter und Pseudowörter) auf fünf Niveaustufen (Jungjohann, Diehl et al., 2019). Das zweite Verfahren SinnL-Levumi erhebt das Leseverständnis auf Satzbasis auf zwei Niveaustufen (Jungjohann & Gebhardt, 2019). Die Testkonstruktion des Leseverständnistests knüpft an die der Leseflüssigkeitstests an, sodass beide Tests zusammen ein Kontinuum zur Messung basaler Lesekompetenzen bilden. Die Evaluation beider Verfahren zeigt, dass die Tests in unterschiedlichen Jahrgängen der Grundschule reliabel und überwiegend bereits im Abstand von drei Wochen änderungssensibel messen. Außerdem messen sie die Lesekompetenzen von unterschiedlichen Schülerinnen und Schülern fair. Geprüft wurde die Testfairness für Schülerinnen und Schüler mit und ohne SPF, beide Geschlechter und solche mit einem Migrationshintergrund. Sowohl der Wortlesetest als auch der Leseverständnistest bilden die Lernverläufe im Längsschnitt über ein Schuljahr hinweg im unteren und mittleren Kompetenzniveau ohne Bodeneffekte ab. Die Tests eignen sich zum Einsatz in inklusiven Klassen und identifizieren förderbedürftige Lesende. Im Folgenden werden die Ergebnisse der einzelnen Artikel studienübergreifend anhand der gestellten Forschungsfragen der kumulativen Dissertation diskutiert. Anschließend werden Limitation und ein Gesamtfazit mit Implikationen für ausstehende Forschungsaktivitäten angeführt.

1. Wie können theoriegeleitete Lernverlaufstests zur Messung basaler Lesekompetenzen konstruiert sein, damit sie den Ansprüchen des inklusiven Unterrichts entsprechen?

Zwei zentrale Ziele des Leseunterrichts in der Grundschule sind die Ausbildung einer angemessenen Leseflüssigkeit und guten Verständnisleistungen auf Satzebene

(National Institute of Child Health and Human Development, 2000). Damit sich diese Kompetenzen erfolgreich ausbilden, müssen Schwierigkeiten in beiden Kompetenzbereichen frühzeitig und während der Lernentwicklung innerhalb der ersten Präventionsstufe aufgedeckt werden (Diehl, 2010; Wember, 2001). Somit benötigt die deutsche Schullandschaft Lernverlaufstests, die diese Kompetenzen in einem Kontinuum messen. Für eine effektive Leseförderung reicht es allerdings nicht aus, dass ausschließlich verlangsamte oder stagnierende Lernverläufe festgestellt werden. Kuhl und Wittich (2018) fordern für den inklusiven Unterricht, dass diagnostische Ergebnisse mit didaktischem Unterrichtshandeln verbunden werden. Diese Verknüpfung kann erreicht werden, wenn eine Lernverlaufsdagnostik Lesemodelle in der Testkonstruktion berücksichtigt und sie mit der Testinterpretation verknüpft. Dadurch werden Implikationen für die anschließende Leseförderung ableitbar. Jungjohann und Gegenfurtner et al. (2018) haben in ihrem systematischen Review festgestellt, dass bisher nahezu kein Lernverlaufstest einen konkreten Bezug zu einem Lesemodell herstellt. Die Ausnahme ist das statusorientierte Inventar zur Erfassung von Lesekompetenzen (IEL-1) von Diehl und Hartke (2012). Das IEL-1 erhebt Kompetenzen im Bereich der phonologischen Bewusstheit, der Phonem-Graphem-Korrespondenzen sowie zum Silbensegmentieren, Wortlesen, Satzlesen und Textlesen. Ergänzend wird im Sinne einer Lernverlaufsdagnostik ein Leseflüssigkeitstest als Eingangsscreening angeboten. Das IEL-1 ist ausschließlich in der ersten Klassenstufe einsetzbar. Zur Erfassung der Leseflüssigkeit werden im IEL-1 in Anlehnung an den KLA (Dummer-Smoch & Hackethal, 2016) Wörter mit unterschiedlichem theoretischen Schwierigkeitsgrad in einer Wortliste gebündelt. Dieses Vorgehen ist für die Leseflüssigkeitstests der Onlineplattform Levumi weiter ausdifferenziert worden (Gebhardt et al., 2016b; Jungjohann, DeVries, Gebhardt et al., 2018). Die unterschiedlichen Testarten (Silben, Wörter, Pseudowörter) repräsentieren in Anlehnung an das Kompetenzentwicklungsmodell von Klicpera et al. (2017) die nicht lexikalische, die partiell lexikalische und die lexikalische Wortlesestrategie (Jungjohann & Gebhardt, 2018). Dadurch wird aufgedeckt, in welcher Lesestrategie aktuell Schwierigkeiten liegen. Die Items der einzelnen Niveaustufen enthalten jeweils nur eine begrenzte Graphemauswahl und unterschiedlich schwierige Silbenstrukturen in Anlehnung an die Schwierigkeitsstufen des KLA (Jungjohann, Diehl et al., 2019). Dies ermöglicht eine Identifikation, ob Schülerinnen und Schüler bereits Konsonantenverbindungen und seltene Grapheme

sicher erlesen. Die Kombination aus verschiedenen Testarten und Niveaustufen ermöglicht ein sehr differenziertes Messen der Leseteilkompetenzen auf unterschiedlichen Kompetenzniveaus. Die Erfassung des Leseverständnisses wird bei dem SinnL-Levumi Test über die Syntaxanalyse der Prädikat-Argumente-Strukturen (Christmann & Groeben, 1999) innerhalb einzelner Sätze realisiert (Jungjohann, DeVries, Mühling et al., 2018). Es wurden drei inhaltliche Dimensionen (Subjekt und Objekt, Verb und Adjektiv, Konjunktion und Präpositionen) konstruiert (Jungjohann & Gebhardt, 2019). Jungjohann, DeVries und Mühling et al. (2018) haben nachgewiesen, dass die dritte Dimension (Konjunktionen und Präpositionen) signifikant schwieriger zu beantworten ist als die ersten beiden. Dadurch werden differenzierte Rückschlüsse auf das Verständnis der syntaktischen Funktionen ermöglicht. Um ein Kontinuum in der Lernverlaufdiagnostik zu schaffen, wurden die Niveaustufen auf die Verständnistests übertragen. Die Übertragung konnte allerdings nur mit Einschränkungen stattfinden, da die begrenzte Graphemauswahl die Anzahl potentieller realer Wörter drastisch minimiert. Der SinnL-Levumi Test bündelt die Graphemauswahl der Leseflüssigkeitstests in zwei Niveaustufen. Konsonantenverbindungen sind für alle Niveaustufen zulässig (Jungjohann & Gebhardt, 2019). Dadurch sind die Leseteilkompetenzen parallel beobachtbar. Die theoriegeleitete Testkonstruktion bedient in mehreren Aspekten die Bedarfe des inklusiven Unterrichts. Erstens ermöglichen die unterschiedlichen Testarten eine parallele Beobachtung mehrerer Leseteilkompetenzen in einem Testsystem. Dies verringert die Wahrscheinlichkeit, dass spezifische Leseschwierigkeiten bei einem oder einer Lernenden im Unterrichtsalltag übersehen werden. Zweitens ermöglichen die differenzierten Niveaustufen die Messung der Lernentwicklung aller Lernenden einer Lerngruppe, selbst wenn die Schülerinnen und Schüler große Kompetenzunterschiede haben. Zusätzlich werden durch die Niveaustufen bei einzelnen Lernenden Lesekompetenzen auf unterschiedlichen Niveaus messbar. Drittens ermöglicht die Orientierung an den Lesemodellen einen jahrgangsübergreifenden Einsatz, da die überprüften Kompetenzen sich nicht auf das Curriculum eines Schuljahres beschränken.

Die US-amerikanischen Leseflüssigkeitstests und Maze-Aufgaben gelten als lernwirksamer Ansatz zur Steigerung von Lesekompetenzen von Schülerinnen und Schülern mit Lernschwierigkeiten (Fuchs et al., 1992; Graney & Shinn, 2005; Stecker

et al., 2005). Damit Lernverlaufstests ebenfalls in Deutschland ihre lernwirksamen Potentiale entfalten können, sollte sich die Testkonstruktion an den etablierten Verfahren orientieren. In großen Teilen wurde die Testkonstruktion der amerikanischen CBM bei den Levumi Tests übernommen. Allerdings haben sich Varianten innerhalb der Testkonstruktionen entwickelt (Ardoin et al., 2013; Shin et al., 2000). Die potentiellen Eigenschaften wurden für die Adaption der neuen Verfahren im Hinblick auf die Bedürfnisse des inklusiven Unterrichts abgewogen. Zur Identifikation von Schülerinnen und Schülern mit einem Risiko zur Ausbildung von Leseschwierigkeiten (primäre Prävention) müssen die Tests vor allem die Lesekompetenzen im unteren Kompetenzbereich ohne Bodeneffekte abbilden (Walter, 2014). Bei der Verwendung von zusammenhängenden Texten als Lesematerial sind sowohl bei Leseflüssigkeitstests als auch bei Maze-Aufgaben Bodeneffekte bei Schülerinnen und Schülern mit milden und manifesten Leseschwierigkeiten beobachtet worden (Fuchs et al., 2004; Roland H. Good et al., 2001). Für den inklusiven Unterricht ist es allerdings unumgänglich, dass die Kompetenzen dieser Lernenden in den Tests abgebildet werden. Denn genau diese Schülerinnen und Schüler benötigen eine effektive Leseförderung. Zur präventiven Vermeidung von Überforderung durch das Lesematerial wurde bei den Levumi Tests vollständig auf den Einsatz zusammenhängender Texte verzichtet. Diese Art der Testkonstruktion wurde bereits in anderen Verfahren erfolgreich erprobt (January & Ardoin, 2012; Jungjohann, Gegenfurtner et al., 2018; Witzel & Witzel, 2016). Bei den Leseflüssigkeitstests bilden somit einzelne Silben, Wörter oder Pseudowörter und bei dem SinnL-Levumi Test einzelne unzusammenhängende Sätze die Items.

Lernverlaufstests können nur positiv auf die Lernentwicklung wirken, wenn sie in der Schule von Lehrkräften eingesetzt werden. Dafür müssen sie praktikabel und in den Unterricht integrierbar sein (Deno, 2003a; Fuchs, 2017). Digitale Testsysteme nehmen Lehrkräften viele organisatorische und diagnostische Aufgaben ab und reduzieren somit die zu investierende Zeit (Fuchs & Fuchs, 1992; Gebhardt & Jungjohann, im Druck; Maier, 2014). Allerdings haben Jungjohann und Gegenfurtner et al. (2018) festgestellt, dass die untersuchten Testverfahren ausschließlich als Papierversion vorliegen. Im Zuge der Digitalisierung werden in Deutschland immer mehr Schulen, wenn auch in unterschiedlichem Maße, mit internetfähigen Endgeräten ausgestattet. Für die Konstruktion der neuen Lernverlaufstests resultieren aus der

neuen Ausstattung mehrere Konsequenzen. Damit die Potentiale der digitalen Testadministration ausgeschöpft werden, sind die Leseflüchtigkeits- und Leseverständnistests in die Onlineplattform Levumi eingebettet. So sind sie flexibel auf unterschiedlichen Endgeräten, wie Smartphones, Tablets oder Desktop-PCS, nutzbar (Mühling et al., 2017). Da beide neuen Lernverlaufstests in eine gemeinsame Plattform eingebettet sind, erfordert dies von den Lehrkräften nur die Einarbeitung in eine Testumgebung. Die Entscheidung für eine webbasierte Anwendung hat den Vorteil, dass die Lehrkräfte keine Programme installieren müssen (Jungjohann, Diehl et al., 2018). Als Einschränkung resultiert daraus allerdings, dass für die Nutzung der Tests stets eine Internetverbindung bestehen muss. Um diese Einschränkung zu entschärfen, wurden die Tests ergänzend frei zugänglich als Papierversion veröffentlicht (Jungjohann, Diehl et al., 2019; Jungjohann & Gebhardt, 2019). Durch die unterschiedlichen Zugänge zu den Lernverlaufstests wird eine möglichst breite Gruppe von Nutzerinnen und Nutzern angesprochen und der Ausschluss aufgrund von fehlenden Ressourcen vermieden.

2. Eignen sich die theoriegeleiteten Lernverläufe zur langfristigen Messung von Leseleistungen auf unterschiedlichen Kompetenzniveaus?

Zur Prüfung psychometrischer und schulpraktischer Eigenschaften von Lernverlaufstests schlägt Fuchs (2004) drei Gütestadien vor. In den ersten beiden Studien stammen die Forschungsdaten aus kontrollierten Schulstudien mit geschulten Testerinnen und Testern. Zuerst müssen die Dimensionalität und die Reliabilität im Querschnitt (Stadium 1) und anschließend die Messinvarianz und die Änderungssensibilität im Längsschnitt (Stadium 2) geprüft werden. Wilbert und Linnemann (2011) konkretisieren die Überprüfung innerhalb der ersten beiden Stadien. Dieses Vorgehen wurde bisher nicht angewendet, wie Jungjohann und Gegenfurtner et al. (2018) bei der systematischen Analyse von Leseflüchtigkeits- und Leseverständnistests feststellen. Ohne dieses Vorgehen wird nicht gesichert, dass die Summenwerte für weitere psychometrische Analysen und die Interpretation in der Schule genutzt werden können. Daher folgt die Eignungsprüfung der neuen Levumi Tests den empfohlenen Richtlinien von Wilbert und Linnemann (2011). Zusammengefasst war die Prüfung des ersten Gütestadiums für alle Tests erfolgreich. Die Schwierigkeit der Items lag stets im unkritischen Bereich und mit der zufälligen Itemziehung wurden reliable Ergebnisse bei der Leseflüchtigkeit und beim Leseverständnis erreicht.

(Gebhardt et al., 2016b; Jungjohann, DeVries, Gebhardt et al., 2018; Jungjohann, DeVries, Mühling et al., 2018). Die Eindimensionalität wurde ebenfalls für beide Verfahren erfolgreich bestätigt (Jungjohann, DeVries, Gebhardt et al., 2018; Jungjohann, DeVries, Mühling et al., 2018). Die Untersuchungen im Längsschnitt (Stadium 2) müssen differenzierter betrachtet werden. Beide Tests sind messinvariant für Schülerinnen und Schüler mit und ohne SPF, für beide Geschlechter und für solche mit einem Migrationshintergrund (Jungjohann, DeVries, Gebhardt et al., 2018; Jungjohann, DeVries, Mühling et al., 2018). Daher können die Tests gut im inklusiven Setting eingesetzt werden. Zusätzlich sind die Tests mit den berücksichtigten Messabständen für verschiedene Subgruppen der einbezogenen Stichproben unterschiedlich änderungssensibel. In beiden Testverfahren haben sich die Leistungen der Schülerinnen und Schüler ohne SPF erwartungsgemäß im Abstand von drei Wochen signifikant verbessert (Jungjohann, DeVries, Gebhardt et al., 2018; Jungjohann, DeVries, Mühling et al., 2018). Keine signifikanten Unterschiede in den Leseleistungen wurden in den vorliegenden Studien bei dem SiL-Levumi Test auf der Niveaustufen N3 und im SinnL-Levumi Test auf Niveaustufe N4 bei Schülerinnen und Schülern mit SPF festgestellt. Dies kann darüber erklärt werden, dass Schülerinnen und Schüler mit SPF mehr Zeit zum Erlernen von Lesekompetenzen benötigen (Klicpera et al., 2017; Wocken & Gröhlich, 2009). Folglich ist es wahrscheinlich, dass es länger dauert bis die Schülerinnen und Schüler einen signifikanten Lernzuwachs erreichen. Hinzu kommt, dass in Artikel 2 (Jungjohann, DeVries, Gebhardt et al., 2018) ein studienbedingter Bias möglich ist. Zwischen den Messungen der Leseflüssigkeit auf Niveaustufe N3 lagen die Osterferien. Die Einflüsse von Ferieneffekten sind für deutsche Schülerinnen und Schüler noch nicht ausreichend geklärt (Coelen & Siewert, 2008). Es ist möglich, dass die stagnierenden Leseleistungen bei den Schülerinnen und Schülern mit SPF durch die fehlende schulische Förderung während der Ferienzeiten bedingt wurde. Daher muss noch in zukünftigen Studien geklärt werden, bei welchem zeitlichen Abstand und unter Berücksichtigung von Ferienzeiten Veränderungen in der Leseleistung bei Schülerinnen und Schülern mit SPF erwartet werden können. Die studienübergreifenden Ergebnisse lassen aber den Schluss zu, dass sich die theoriegeleitete Testkonstruktion zur Messung des Lernverlaufs im Lesen für alle Schülerinnen und Schüler mit den betrachteten Personenmerkmalen eignet.

Die Eignung einer Lernverlaufsdagnostik im inklusiven Unterricht hängt neben den psychometrischen Faktoren von der Anschlussfähigkeit an die Leseleistungen der Schülerinnen und Schüler ab (Deno, 2003a; Fuchs, 2017). Grundsätzlich muss es den Lehrkräften möglich sein, die Entwicklung ihrer Schülerinnen und Schüler aufgrund der beobachteten Lernverläufe einzuschätzen und vorauszusagen (Wilbert, 2014). Dafür dürfen keine Boden- oder Deckeneffekte entstehen (Walter, 2014). Für den speziellen Einsatz im inklusiven Unterricht kommt die breite Leistungsheterogenität der Schülerinnen und Schüler (Gebhardt, 2015) erschwerend hinzu. Decken die Tests die Leistungsheterogenität der anvisierten Lerngruppe nicht ab, werden Lehrkräfte auf mehrere Testsysteme zurückgreifen. Die Praktikabilität und Ökonomie wird beeinträchtigt (Klauer, 2011) und Schülerinnen und Schüler werden von der Lernverlaufsdagnostik ausgeschlossen. Vorrangig muss eine Lernverlaufsdagnostik im unteren Leistungsbereich differenziert messen, da diese Schülerinnen und Schüler ein besonderes Risiko zur Ausbildung von manifesten Leseschwierigkeiten haben und daher insbesondere einer präventiven Leseförderung bedürfen. Die Leseflüssigkeitstests und der SinnL-Levumi Leseverständnistest messen ausreichend gut im unteren und mittleren Leistungsbereich (Jungjohann, Schurig et al., 2019). Bei den Lesenden mit dem höchsten Ausgangsniveau sind in beiden untersuchten Klassenstufen (3. und 4. Jahrgang) Deckeneffekte entstanden. Für diese Gruppe von Schülerinnen und Schülern liefert die eingesetzte Niveaustufe keine verlässlichen Ergebnisse. In der schulischen Anwendung können Lehrkräfte diese Deckeneffekte umgehen, indem sie höhere Niveaustufen oder anspruchsvollere Testarten nutzen. In der Studie wurden für wissenschaftliche Zwecke alle teilnehmenden Schülerinnen und Schüler in derselben Niveaustufe pro Testart (Silben, Wörter, Pseudowörter, Leseverständnis) getestet (Jungjohann, Schurig et al., 2019). Durch das gewählte Studiendesign wurde geprüft, ob die Tests auch über mehrere Jahrgangsstufen hinweg einsetzbar sind und ob sie Leserinnen und Leser mit einem Risiko für Lernschwierigkeiten identifizieren. Die Ergebnisse zeigen, dass beide Kriterien erfüllt werden. Die Leseleistungen der dritten Klasse befinden sich am Ende des Schuljahres auf einem ähnlichen Niveau, wie die Leseleistungen in der vierten Jahrgangsstufe zu Beginn des Schuljahres. Dies stellt einen fließenden Übergang zwischen den Klassenstufen dar. Folglich sind die Tests innerhalb einer Niveaustufe auch jahrgangsübergreifend einsetzbar. Außerdem können die Tests Lernverläufe von Lesenden mit geringen Ausgangsniveaus abbilden und somit identifizieren. Einerseits

wurden bei den schwächsten Lesenden geringere Lernzuwächse als bei den durchschnittlichen Schülerinnen und Schülern festgestellt (Jungjohann, Schurig et al., 2019). Andererseits erreichten die Schülerinnen und Schüler mit SPF signifikant geringere Leistungen in der Leseflüssigkeit als solche ohne SPF (Jungjohann, DeVries, Gebhardt et al., 2018). Abschließend lässt sich festhalten, dass sich sowohl die Leseflüssigkeitstests als auch der SinnL-Levumi Leseverständnistest für den Einsatz in inklusiven Klassen eignen. Sie erfüllen die notwendigen psychometrischen Gütekriterien und messen die Leseleistungen im Quer- und Längsschnitt erfolgreich. Sie können jahrgangsübergreifend eingesetzt werden und messen die Lernverläufe von Schülerinnen und Schülern mit geringen Leseleistungen. Die unterschiedlichen Niveaustufen und Testarten sichern zudem die Nutzung der Lernverlaufdiagnostik in den zentralen Leseteilkompetenzen und ebenfalls im inklusiven Unterricht.

3. Welche Anforderungen aus der Schulpraxis müssen adressiert werden, um eine theoriegeleitete Lernverlaufdiagnostik in deutschen Grundschulen zu implementieren?

Als Voraussetzung für die Implementation müssen Lehrkräfte den Mehrwert der Lernverlaufdiagnostik erkennen. Denn sonst wird keine Lehrkraft eine Lernverlaufdiagnostik eigeninitiativ in ihren Unterricht einbetten und Studien des dritten Gütestadiums nach Fuchs (2004) sind nicht möglich. Jungjohann und Diehl et al. (2018) haben herausgestellt, dass Lehrkräfte unterschiedliche Vorstellungen bezüglich der Implementation einer Lernverlaufdiagnostik haben. Einerseits bestehen Bedenken, dass die Lernverlaufdiagnostik zu sehr in die aktuellen Unterrichtskonzepte eingreift. Andererseits wünschen sich Lehrkräfte durch die Lernverlaufdiagnostik Anregungen zur Gestaltung des Unterrichts. Mithilfe der Lernverlaufdiagnostik prüfen die Lehrkräfte die Effektivität der eigenen Materialien und Fördermaßnahmen. Die Lehrkräfte müssen somit ihr Unterrichtsrepertoire nicht ersetzen, sondern es durch die Tests ergänzen. Des Weiteren wird im Kontext des SPF-L explizit gefordert, dass sonderpädagogische Bildungsangebote durch Prozesse der Diagnostik geprägt und regelmäßig auf ihre Wirksamkeit, auf notwendige Anpassungsbedarfe sowie auf das Erreichen eines Lernziels überprüft werden (Kultusministerkonferenz, 2019). Laut KMK (2019) sind für diese Prüfung standardisierte Testverfahren notwendig, die „die erbrachten Leistungen an den jeweils angestrebten Zielen messen“ (Kultusministerkonferenz, 2019, S. 12). All diese

Forderungen bedient eine standardisierte und erfolgreich geprüfte Lernverlaufsdagnostik, sodass Lehrkräfte durch ihren Einsatz der KMK Empfehlungen zur schulischen Bildung, Beratung und Unterstützung von Schülerinnen und Schülern mit SPF-L nachkommen können. Über die Ergebnisse der Lernverlaufsdagnostik kann die Notwendigkeit einer zusätzlichen und zeitintensiven Leseförderung (tertiäre Präventionsstufe) gegenüber Eltern oder Schulleitern begründet werden. Die Onlineplattform Levumi bietet eine enge Verzahnung zwischen Diagnostik und Förderung als Unterstützungsangebot in der Förderplanung an. In dem Handbuch „Förderansätze im Lesen mit Levumi“ stehen den Lehrkräften didaktische Hinweise und unterrichtsnahe Übungsmaterialien zur Verfügung (Jungjohann et al., 2017). Dieses kann zur Erweiterung des bestehenden Repertoires oder zur Neugestaltung von Lesefördermaßnahmen genutzt werden. Daher können die Nutzerinnen und Nutzer des Levumi Angebots selbst entscheiden, in welchem Ausmaß sie ihren Unterricht mit den Tests und Förderangeboten gestalten (Jungjohann, Diehl et al., 2018). Levumi wird als Erweiterungsangebot zur Steigerung von Leseleistungen verstanden. Dadurch wird eine flexible Implementation gewährleistet.

Für eine erfolgreiche Implementation ist es notwendig, dass Lehrkräfte bei der Anwendung und Interpretation der Tests Unterstützung erhalten (Ardoin et al., 2013). Obwohl die Bereitschaft zur Nutzung einer Lernverlaufsdagnostik in Deutschland in den letzten Jahren gestiegen ist (Mühling et al., 2019), werden Lernverlaufstests bisher nur selten genutzt. Dies kann einerseits damit erklärt werden, dass es bisher nur wenige Verfahren gibt. Andererseits fehlen zumeist passende Unterstützungsangebote sowie eine konkrete Verknüpfung mit Fördermöglichkeiten (Jungjohann, Gegenfurtner et al., 2018). In Bezug auf die technische Anwendung stellen Jungjohann und Diehl et al. (2018) heraus, dass sich die Lehrkräfte eine möglichst hohe Flexibilität für den Einsatz der Tests wünschen. Dazu zählt die Möglichkeit sich eigenständig einzuarbeiten, aber auch Erklärungen zur Anwendung und Beurteilung in schriftlicher Form zu erhalten. Die Leseflüssigkeitstests und die Leseverständnistests der Onlineplattform Levumi streben eine möglichst große Freiheit in der technischen Anwendung an. Die Plattform richtet in jedem Account automatisch eine fiktive Klasse ein, in der explorativ alle Funktionen erkundbar sind (Jungjohann, Diehl et al., 2018). Parallel liefern die Handbücher detaillierte

Informationen über die Plattform, die Anwendung und die Test (Jungjohann, Mau et al., 2019). In Bezug auf die Häufigkeit der Anwendung gibt das Handbuch Deutsch allerdings ausschließlich Anregungen und keine starren Vorgaben. Die Anregungen erhöhen die Flexibilität im Unterricht. Aktuell bleibt es den Lehrkräften überlassen, wie häufig und zu welchen Zeitpunkten die Tests durchgeführt werden. Dadurch wird in Kauf genommen, dass nicht alle Lehrkräfte eine reliable Messung mit den empfohlenen sechs bis acht Messzeitpunkten (Christ et al., 2013) realisieren. Jungjohann und Schurig et al. (2019) zeigen, dass bereits die Umsetzung von vier Messzeitpunkten in einem Schuljahr die Unterrichtsorganisation herausfordert. Die maximale Anzahl von Levumi Messungen wurde von einer geschulten Lehrkraft erhoben und umfasste 14 Messzeitpunkte (Jungjohann, DeVries, Gebhardt et al., 2018). Untersuchungen, wie viele Messzeitpunkte im deutschen Unterrichtsalltag möglich sind, stehen aktuell noch aus. Daher motiviert die Beschreibungen im Handbuch Deutsch (Jungjohann, Mau et al., 2019) bei geringen Leseleistungen engmaschiger zu messen (sekundäre und tertiäre Präventionsstufe) und bei erwarteten Lernzuwächsen die Messabstände zu vergrößern (primäre Präventionsstufe). Folglich sollen die Lehrkräfte die Levumi Tests nach dem individuellen Bedarf in ihren Klassen einsetzen. So wird eine möglichst hohe Ökonomie der Tests erreicht, da nicht immer alle Schülerinnen und Schüler getestet werden müssen.

Die diagnostisch-didaktische Anwendung der Lernverlaufsdagnostik verlangt eine fachspezifische Kompetenzanalyse sowie die Verknüpfung der Daten mit den aktuellen Fördermaßnahmen (Datnow & Hubbard, 2016; Mandinach, 2012; Stecker, 2017). Die Bereitschaft eine Lernverlaufsdagnostik zu nutzen wird gesteigert, wenn Implikationen für die Förderung unmittelbar aus den Testergebnissen ableitbar sind und die Tests didaktisch relevante Informationen liefern (Jungjohann, Diehl et al., 2018). Alle untersuchten Levumi Lernverlaufstests decken Schwierigkeiten in spezifischen Teilkompetenzen auf (Jungjohann & Gebhardt, 2018; Jungjohann, DeVries, Mühling et al., 2018). Diese Ergebnisse werden in den automatisierten Auswertungen der Levumi Plattform transparent aufgeführt. Zu jedem Test werden die Lernverläufe für eine visuelle Inspektion (Wilbert, 2014) in Klassengraphen und Individualgraphen grafisch aufbereitet. Ergänzend werden die Lesegenauigkeit und Lesegeschwindigkeit numerisch sowie die Lernentwicklungen über die Zeit

tabellarisch bereitgestellt (Jungjohann, Diehl et al., 2018). Über die quantitativen Informationen können die Lehrkräfte die Lesenden einem Leseschwierigkeitstyp zuordnen. In der Leseflüssigkeit werden drei Leseschwierigkeitstypen beschrieben (Scheerer-Neumann et al., 2008), die durch die Levumi Tests identifiziert werden können. Lesende des ersten Leseschwierigkeitstyps erreichen in den Levumi Leseflüssigkeitstests insgesamt nur geringe Summenwerte (WpM) bei einer hohen Lösungswahrscheinlichkeit (RWpM). Lesende des zweiten Leseschwierigkeitstyps sind durch geringe Summenwerte in der Lesegenauigkeit (RWpM) bei einer angemessenen Anzahl von Wörtern pro Minute (WpM) identifizierbar. Lesende des dritten Leseschwierigkeitstyps erreichen sowohl in der Lesegeschwindigkeit als auch in der Lesegenauigkeit nur geringe Summenwerte. Somit stellen die gewählten Kompetenzmaße einen unmittelbaren Bezug zur Leseförderung her. Beim Leseverständnis ist eine eindeutige Zuweisung zu einem Leseschwierigkeitstyp hingegen komplexer. Dies liegt an dem hohen Einfluss der Leseflüssigkeit auf das Leseverständnis während der Grundschulzeit (García & Cain, 2014; Schabmann & Kabicher, 2007). In den von Nix (2011) definierten Leseschwierigkeitstypen beim Leseverständnis werden stets die Dekodierleistungen mit einbezogen. Für die Auswertung der Levumi Tests bedeutet dies, dass die Ergebnisse eines Leseverständnistests optimaler Weise gemeinsam mit denen der Leseflüssigkeit betrachtet werden sollten. Schülerinnen und Schüler des ersten Schwierigkeitstyps (gutes Leseverständnis ohne ausreichende Dekodierfähigkeit) zeigen folglich hohe Summenwerte im Leseverständnistests und geringe Summenwerte in der Leseflüssigkeit. Beim zweiten Schwierigkeitstyp drehen sich die Ergebnisse um. Hohe Summenwerte in der Leseflüssigkeit deuten auf gute Dekodierleistungen hin und ein unzureichendes Leseverständnis wird durch geringe Summenwerte im Leseverständnistest aufgedeckt (Schwierigkeitstyp 2). Schülerinnen und Schüler des dritten Schwierigkeitstyps erreichen in beiden Tests nur geringe Summenwerte. Die Langzeitstudie von Jungjohann und Schurig et al. (2019) identifiziert anhand des Leseausgangsniveaus (Intercept) zum ersten Messzeitpunkt im dritten und vierten Jahrgang jeweils die schwächsten Lesenden beim Wortlesen und Satzverständnis. Jedem Schwierigkeitstyp im Leseverständnis nach Nix (2011) konnten in dieser Studie Schülerinnen und Schüler zugeordnet werden. Zusammengefasst bedeuten diese Ergebnisse, dass Auswertungen der theoriegeleiteten Lernverlaufstests unmittelbar am Unterricht anknüpfen und Implikationen für die Förderung auf verschiedenen

Ebenen zulassen. Durch diese Eigenschaften der Lernverlaufstests kann die Bereitschaft zur Implementation nur durch den hohen praktischen Nutzen gesteigert werden.

Zusammenfassung der Diskussion

Im Hinblick auf die erste Fragestellung kann konstatiert werden, dass die theoriegeleitete Testkonstruktion und die vorgenommenen Adaptionen den Ansprüchen der inklusiven Grundschule standhalten. Erstens messen die Tests basale Lesekompetenzen in Anlehnung an aktuelle Lesemodelle, wodurch unmittelbar Implikationen zur Leseförderung ableitbar sind. Zweitens ermöglichen die gemeinsamen schwierigkeitsgenerierenden Regeln in beiden Testarten eine kontinuierliche Messung der basalen Lesekompetenzen über mehrere Jahrgangsstufen hinweg. Drittens sichert die Testkonstruktion eine Messung der Lernentwicklung ohne Bodeneffekte insbesondere im unteren Kompetenzbereich. Viertens schaffen die Einbindung der Tests in die kostenlose Onlineplattform Levumi sowie die Open Access Veröffentlichungen der papierbasierten Parallelformen einen breiten Zugang für möglichst viele Lehrkräfte. Die Ergebnisse zur zweiten Forschungsfrage lassen den Schluss zu, dass beide Testarten im Quer- und im Längsschnitt den Anforderungen der Lernverlaufsdagnostik und des inklusiven Unterrichts entsprechen. Die Tests erfassen die Leseleistungen von unterschiedlichen Gruppen von Schülerinnen und Schülern reliabel und änderungssensibel über mehrere Messzeitpunkte und über ein Schuljahr hinweg. Zusätzlich sichern die verschiedenen Niveaustufen und Testarten die Abbildung verschiedener Leseteilkompetenzen bei leistungsheterogenen Lerngruppen. Auch die praktische Anwendung wird durch die Niveaustufen und Testarten im Unterricht gefördert. Im Hinblick auf die Vorbereitung der Implementation der Lernverlaufstests in den Unterricht (Forschungsfrage 3) wird resümiert, dass die unterstützenden Materialien sowie die Flexibilität der Tests die Lehrkräfte zur Anwendung in eigener Initiative motiviert. Folglich hat die vorliegende Forschung einen Grundstein für die Nutzung der theoriegeleiteten Lernverlaufsdagnostik in der deutschen Schullandschaft geschaffen.

Limitationen

In der vorliegenden Forschung bestehen Limitationen in der Testkonstruktion, im Untersuchungsdesign sowie bei der Vorbereitung zur Implementation der Tests. Im Hinblick auf die Testkonstruktion wurde entschieden, auf zusammenhängende Lesetexte zur Erfassung der Lesekompetenzen gänzlich zu verzichten. Diese Testeigenschaft verhindert die Messung von gewissen Lesekompetenzen. Bei den Leseflüssigkeitstests wird durch die Verwendung von Wortlisten die Messung der prosodischen Aspekte ausgeschlossen. Denn prosodische Aspekte des Vorlesens greifen erst in einem zusammenhängenden literarischen Kontext (Kuhn et al., 2010; Rasinski, 2004). Das vorrangige Ziel der Lernverlaufsdagnostik ist die Messung von Lesekompetenzen zur Identifikation förderbedürftiger Schülerinnen und Schüler. Daher wurde im Kontext der Levumi Tests dieses Merkmal priorisiert berücksichtigt. Für den schulischen Kontext stellt die Messung von prosodischen Aspekten im Kontext der Lesekompetenzen einen Faktor dar, auch wenn der Einfluss der Prosodie auf die Lesekompetenzen bisher noch nicht ausreichend geklärt ist (Schwanenflugel et al., 2015). Folglich sollten Diagnostikverfahren zur expliziten Messung von prosodischen Aspekten vorangetrieben werden. Ein Beispiel stellt die remediale Ratingprozedur mit Audioaufnahmen zur Messung prosodischer Aspekte in der Leseflüssigkeit von Sappok und Fay (2018) dar. Für die Leseverständnistests stellt die Verwendung von einzelnen Sätzen ebenfalls eine Einschränkung in der Kompetenzmessung dar. Nachteilig an Aufgaben mit einzelnen Sätzen ist, dass die Lesenden keine lokale Kohärenz des Gesamtzusammenhangs bilden können und dass Verständnisseleistungen ausschließlich auf Satzebene stattfinden (Lenhard, 2019b). Auch hier überwog die Bedingung, dass die Leseleistungen der Schülerinnen und Schüler mit geringen Verständnisseleistungen ohne Bodeneffekte messbar sind. In der Studie von January und Ardoin (2012) nutzen allerdings die Schülerinnen und Schüler bei der Bearbeitung von Maze-Aufgaben den satzübergreifenden Kontext zu Lösung der Aufgaben nicht. Somit ist es fraglich, ob die Maze-Aufgaben tatsächlich die lokale Kohärenzbildung messen oder ob die Maze-Aufgaben mit zusammenhängenden Texten trotzdem vorrangig das Verständnis auf Satzbasis abfragen. Der SinnL-Levumi Leseverständnistest korreliert hoch mit dem Subtest Satzlesen des ELFE-II Tests (Anderson et al., angenommen). Eine weitere

Validierung des SinnL-Levumi Tests mit standardisierten Leseverständnistests auf Textbasis steht noch aus.

Eine weitere Limitation besteht in der Methodik der Untersuchungen. Zur Untersuchung der Reliabilität der Lernverlaufstests wurden auch die Lesekompetenzen von Schülerinnen und Schülern mit SPF analysiert. In zwei Studien wurden die Lesedaten von Schülerinnen und Schülern aus Förderschulen und inklusiven Settings gemeinsam betrachtet (Jungjohann, DeVries, Gebhardt et al., 2018; Jungjohann, DeVries, Mühling et al., 2018). Dies war notwendig, um eine ausreichend große Stichprobe für die Reliabilitätsprüfung zu generieren. Die Ergebnisse zweier Studien fanden bei Schülerinnen und Schülern mit SPF in inklusiven Settings höhere Leseleistungen als an Förderschulen (Gebhardt, Sälzer et al., 2015; Kocaj et al., 2014). Ungeklärt ist bisher der kausale Zusammenhang zwischen Leseleistungen und Beschulungsart. Trotzdem ist ein Einfluss der Beschulungsart auf die Leseleistungen erwartbar. Eine Reanalyse der genutzten Forschungsdaten würde Aufschluss über mögliche Effekte zwischen der Beschulungsart und den Leseleistungen liefern. Die Untersuchung zu Klasseneffekten innerhalb der Studie von Jungjohann und Schurig et al. (2019) zeigt innerhalb einer inklusiven Grundschule bereits relativ starke Effekte ($ICC = .145 - .287$). Für die Reanalysen ist daher zu erwarten, dass die Einflüsse der Beschulungsarten höher ausfallen als die Klasseneffekte innerhalb einer Schule. Für die vorliegenden Studien stellt das gewählte Vorgehen trotz der angeführten Limitationen eine gute Möglichkeit dar, die grundsätzliche Eignung der Lernverlaufstests für Schülerinnen und Schüler mit SPF im Sinne des ersten und zweiten Gütestadiums nach Fuchs (2004) zu prüfen.

In Bezug auf die Implementation der Lernverlaufstests in der Praxis wäre eine Normierung im Sinne von Förster et al. (2017) wünschenswert. In den bisherigen Studien wurde nicht erhoben, welche didaktischen Maßnahmen die Lehrkräfte der einbezogenen Schülerinnen und Schüler zum Zeitpunkt der Datenerhebung nutzten. Für Studien des ersten und zweiten Gütestadiums sind diese Informationen allerdings nicht notwendig (Fuchs, 2004). Zur individuellen Zielsetzung in der Förderplanung wäre es aber von großem Interesse, welche gesteigerte Lernwicklung in Abhängigkeit der kindlichen Personenmerkmale und Fördermaßnahmen zu erwarten sind. Die aktuellen Studien zeigen bereits, dass die Niveaustufen in Abhängigkeit von der Lesekompetenz unterschiedlich gut die Lernentwicklungen abbilden (Jungjohann,

DeVries, Gebhardt et al., 2018; Jungjohann, DeVries, Mühling et al., 2018; Jungjohann, Schurig et al., 2019). Außerdem wurden in der zweiten Schulstufe unter Einsatz eines adaptiven Lesetrainings größere Lernzuwächse in der Leseflüssigkeit und dem Leseverständnis erreicht (Anderson et al., angenommen). Insbesondere bei der Interpretation von stagnierenden Lernverläufen müssen die Lehrkräfte das Kompetenzniveau der Lesenden mit den Kompetenzansprüchen der Tests in Verbindung setzen (Jungjohann, Schurig et al., 2019). Diese Erkenntnisse müssen in weiteren Jahrgängen rezipiert und anschließend in die Unterstützungsmaterialien zur Dateninterpretation aufgenommen werden.

6. Fazit und Implikationen

„Ultimately, we have to change the system, not just the approach we take to formative assessment, if we want to have maximum impact on learning and instruction.” (Bennett, 2011, S. 19)

Die bisherigen Ausführungen der vorliegenden Dissertation stellen die Lehrkraft und ihr Handeln sowie Effekte auf der Ebene der Schülerinnen und Schüler in den Vordergrund. Alle hiesigen Forschungsbemühungen sind darauf ausgelegt, einer individuellen Lehrkraft im inklusiven Schulsystem die Anwendung der Lernverlaufstests in ihrem Unterricht barrierefrei zu ermöglichen. Dieser Ansatz folgt dem Verständnis, dass die Lehrkräfte ein präventives Arbeiten auf Mikroebene anstreben und dafür die Lernverlaufsdagnostik in ihrem Unterricht implementieren. Eine solche innovative Implementation ist allerdings nur unter besonders guten Voraussetzungen auf der Mikro- und Mesoebene des Schulsystems (Rolff, 2016) möglich. Zunächst muss die Vermittlung fachlicher Inhalte in hohem Maße gesichert und der Bedarf für neue Konzepte erkannt sein. Wenn dann den einzelnen Lehrkräften und dem gesamten Kollegium ausreichend Kapazitäten eingeräumt werden, kann die Lernverlaufsdagnostik ein präventives Handeln im schulinternen Curriculum unterstützen. Ist allerdings eine Schule durch beispielsweise einen hohen Mangel an Lehrkräften, ein inkonsistentes Kollegium durch temporäre Abordnungen oder durch Umstrukturierungsmaßnahmen des Unterrichts stark belastet, ist eine initiative Implementation auf diesen beiden Ebenen wahrscheinlich nicht möglich. Und dies ist

in Deutschland aufgrund der noch nicht abgeschlossenen Umstrukturierung des Schulsystems aktuell oft der Fall. Wie Bennett (2011) es bereits für die formative Diagnostik gefordert hat, muss für die Ausschöpfung des Potentials der Lernverlaufsdagnostik eine Veränderung auf der Makroebene des Schulsystems stattfinden. Und auch dies ist für Deutschland entscheidend. In der Sonderpädagogik herrscht überwiegend noch ein remediales Verständnis. Eine sonderpädagogische Förderung findet zumeist erst statt, nachdem das Lernen in gravierendem Ausmaße nicht erfolgreich verlaufen ist. Die Lernverlaufsdagnostik muss in das schulische Fördersystem integriert werden, bevor das Lernen manifest scheitert. Huber und Grosche (2012) zeigen anhand des Response-to-Intervention-Modells (RtI-Modell; Reschly & Bergstrom, 2009) auf, wie sich das sonderpädagogische Selbstbild in Deutschland hin zum präventiven Fördergedanke wandeln könnte. Das RtI-Modell konkretisiert das mehrstufige Präventionsmodell. Es zeichnet sich dadurch aus, dass alle Schülerinnen und Schüler in drei Förderebenen unterschiedlich häufig und intensiv durch verschiedene Professionen präventiv gefördert werden. Laut Huber und Grosche (2012) stellt die Lernverlaufsdagnostik in jeder Förderebene ein zentrales Fundament für die Begründung von Förderentscheidungen im Unterricht sowie Leistungs- und Entwicklungsrückmeldungen dar. Die Lernverlaufsdagnostik ist dabei Bestandteil eines präventiven Rahmenmodells auf makrosystemischer Ebene. Eine Implementation der Lernverlaufsdagnostik durch systemische Schulstrukturen würden außerdem dazu führen, dass die Verantwortlichkeit der Implementation weg von den einzelnen Lehrkräften hin zu einer systemischen Entscheidung übertragen wird. Laut Huber und Grosche (2012) sind allerdings umfangreiche Änderungen auf der Makroebene notwendig bis eine systemische Verankerung des RtI-Modells hierzulande realisierbar ist. Es bleibt also noch abzuwarten, ob dieses oder ein alternatives Rahmenmodell zur präventiven Förderung systemübergreifend implementiert wird. Bis dahin ist es aber notwendig, dass parallel die Beforschung und Implementation der Lernverlaufsdagnostik auf der Mikro- und Makroebene weiter vorangetrieben werden. Durch eine jetzige Implementation auf Mikroebene können erfolgreiche Einzelfälle und Schulversuche strukturiert untersucht werden. Mit diesem induktiven Vorgehen werden förderliche und hemmende Systembedingungen für eine Implementation durch die Makroebene auf das deutsche Schulsystem übertragbar. Gleichzeitig profitieren einige leseschwache Schülerinnen und Schüler bereits dann durch den lernwirksamen Ansatz der Lernverlaufsdagnostik in Deutschland. Daher

werden im Folgenden Implikationen zur Implementation, der Dateninterpretation und Verknüpfung mit der Förderplanung auf Mikroebene angeführt.

Implikationen

Für die Schulpraxis ergeben sich aus den vorliegenden Befunden folgende Implikationen. Zuerst muss das Konzept der Lernverlaufsdagnostik mit den konkreten Lernverlaufstests in der Schullandschaft weiter Einzug erhalten. Das Forschungsfeld ist in Deutschland immer noch jung. Daher kennen bisher nur wenige Lehrkräfte konkrete Anwendungsbeispiele und die zugrundeliegenden Konzepte. In der Aus- und Weiterbildung von Lehrkräften sollten zum einen die Tests vorgestellt und zum anderen die Potentiale zur Steigerung der Lesekompetenzen verbreitet werden. Die durch Informationen gewonnene Fachkompetenz der Lehrkräfte kann die Akzeptanz und Bereitschaft gegenüber der Lernverlaufsdagnostik erhöhen. Diese Maßnahmen sollten sich an sonderpädagogische Lehrkräfte, Regelschullehrkräfte und Schulpsychologinnen und -psychologen im inklusiven Setting richten. Denn diese Lehrkräfte tragen einen großen Teil der Verantwortung ihre Schülerinnen und Schüler in der Ausbildung der Lesekompetenzen effektiv zu fördern. Des Weiteren benötigen Lehrkräfte die Möglichkeit ihre diagnostischen Kompetenzen im Bereich der Lernverlaufsdagnostik zu schulen. Dafür sollten ebenfalls entsprechende Seminare angeboten werden (van den Bosch et al., 2019). Anhand von Fallbeispielen können die Lehrkräfte üben, leseschwache Schülerinnen und Schüler zu identifizieren und Fördermaßnahmen in Abhängigkeit der linguistischen Schwierigkeiten fachlich zu begründen. Eine Möglichkeit ist es, die fachdidaktischen Leseschwierigkeitstypen in Fallbeispielen abzubilden. Eine Herausforderung ist die Zuordnung der Schülerinnen und Schüler zu den Schwierigkeitstypen. In den Lernverlaufsgraphen wird typischer Weise die Anzahl der richtig gelösten Aufgaben pro Messzeitpunkt abgetragen (van den Bosch et al., 2017). In der Leseflüssigkeit unterscheiden sich die drei Leseschwierigkeitstypen (Scheerer-Neumann et al., 2008) in den Lernverlaufsgraphen auf den ersten Blick nicht. Alle Schwierigkeitstypen zeichnen sich durch niedrigere Summenwerte in der Lesegenauigkeit (RWpM) aus. Erst die qualitative Analyse lässt einen Rückschluss auf die unterschiedlichen Typen zu. Ergänzend zu den Fallbeispielen könnten Übungstools zur Bewertung von Lesefehlern in der Leseflüssigkeit entwickelt und bereitgestellt werden. Denkbar ist beispielsweise ein Leseflüssigkeitstest, indem Antworten durch Audioaufnahmen

simuliert werden. In den simulierten Antworten werden richtig vorgelesene Items und welche mit Lesefehlern zufällig hintereinander präsentiert. Die Lehrkräfte müssten dann bewerten, ob ein Lesefehler zum präsentierten Item vorlag. Ein solches Übungstool würde die Lehrkräfte für typische Lesefehler sensibilisieren und die diagnostische Kompetenz dementsprechend weiter ausbilden. Die Verbreitung der Lernverlaufsdagnostik in der Schulpraxis trägt außerdem zu weiteren schulbezogenen Forschungsaktivitäten bei. Für Studien des dritten Gütestadiums (Fuchs, 2004) werden Daten aus dem Feld benötigt. In diesem Stadium werden die Implementation und der praktische Nutzen erforscht. Insbesondere eine Überprüfung der Wirksamkeit in der Praxis der Levumi Tests steht noch aus. Forschungsaktivitäten in diesem Bereich werden für Lehrkräfte wertvolle Erkenntnisse über eine möglichst praktikable und effektive Anwendung der Lernverlaufsdagnostik bereitstellen.

Für die weitere wissenschaftliche Forschung ergeben sich ebenfalls Implikationen aus den Befunden. Zur Erweiterung des Repertoires an Lernverlaufstests in der deutschen Schulpraxis könnten die hier vorgestellten Adaptionen auf weitere Kompetenzbereiche übertragen werden. Einerseits ist es möglich die Regeln auf Lesekompetenzen der Sekundarstufe zu übertragen. Beispielsweise könnten in Anlehnung an die Syntaxtheorie satzübergreifende Bedeutungseinheiten, wie komplexe Nebensätze, anaphorische Überbrückungsinferenz sowie die kausale, temporale oder örtliche Kohärenz (Brinker, Antos, Heinemann & Sager, 2000; Zimmermann, 2014) mit entsprechenden Items gemessen werden. Andererseits können die schulpraktischen Eigenschaften (mehrere Niveaustufen, fließende Übergänge zwischen Testarten, Handbücher und Fördermaterialien) der Levumi Tests auf Lernverlaufstests im mathematischen Bereich angewendet werden. Als weitere Implikationen für zukünftige Forschungsaktivitäten benennt die vorliegende Dissertation den noch nicht hinreichend erforschten Umgang mit Lernverlaufdaten (Jungjohann, Diehl et al., 2018). Die Identifizierung von Schülerinnen und Schülern mit Leseschwierigkeiten im Unterricht ist häufig fehleranfällig und Schwierigkeiten in der Leseentwicklung bleiben oft unerkannt (Artelt & Rausch, 2014; Feinberg & Shapiro, 2009). Zusätzlich zeigen Erkenntnisse über die Interpretation der Lernverlaufsgraphen, dass Fehlinterpretationen durch die Struktur und Aufbereitung der Daten entstehen (Klapproth, 2018; Newell & Christ, 2017). Eine Verbindung zwischen den diagnostischen Kompetenzen von Lehrkräften sowie der Interpretation

von Lernverlaufsgraphen könnte genauere Erkenntnisse über die Dateninterpretation von theoriegeleiteten Lernverlaufstests generieren. In entsprechenden Studien könnte gefragt werden, wie Lehrkräfte bei der Dateninterpretation von lesetheoriegeleiteten Lernverlaufstests vorgehen, um spezifische Leseschwierigkeiten zu identifizieren. Alternativ könnte gefragt werden, ob anhand von lesetheoriegeleiteten Lernverlaufstests tatsächlich mehr Schülerinnen und Schüler mit Leseschwierigkeiten identifiziert werden. Als nächste Implikation für weitere Forschungsaktivitäten werden Effektivitätsprüfungen der Unterstützungsmaterialien angeführt. Zu Unterstützungsangeboten werden Fortbildungen und schriftliche Informationen zur Testanwendung und -interpretation gezählt (Ardoin et al., 2013; van den Bosch et al., 2019). Im Zusammenhang mit den Levumi Lernverlaufstests wurden bereits drei Handbücher (Gebhardt, Diehl & Mühling, 2016a; Jungjohann et al., 2017; Jungjohann, Mau et al., 2019) und ein praxisnaher Beitrag zur Implementation der Lernverlaufsdagnostik im Unterricht (Jungjohann & Gebhardt, 2018) publiziert. Bisher steht eine Überprüfung der Effektivität der Unterstützungsmaterialien noch weitestgehend aus. Zur Prüfung der Effektivität werden zwei Fragestellungen assoziiert: (1) Nutzen die Lehrkräfte die vorhandenen Publikationen? Hierfür könnte eine Fragebogenstudie in der Onlineplattform Levumi implementiert werden. Durch die Einbindung des Fragebogens in die Plattform würden die direkten Nutzerinnen und Nutzer angesprochen. Der Fragebogen könnte einerseits Fachwissen aus den Handbüchern und andererseits Aktivitäten zur Informationsbeschaffung erfragen. (2) Welche Effekte haben die schriftlichen Informationen auf die Plattformnutzung, die Dateninterpretation und Förderplanung? Hierfür bietet sich die Adaption eines Forschungsdesigns aus der Technikdidaktik an (Stemmann & Lang, 2018). In einer Interventionsstudie könnten die Auswirkungen zweier Fortbildungskonzepte gegenübergestellt werden. In der einen Fortbildung erhalten die Teilnehmenden die schriftlichen Informationen aus den Publikationen. In der Vergleichsgruppe würden die Teilnehmenden zur explorativen Erkundung motiviert. Durch die webbasierte Anwendung werden die Aktivitäten innerhalb der Plattform anhand der Logfiles ausgewertet. Eine Gegenüberstellung der Ergebnisse geben Aufschluss, welche Funktionen der Plattform nach welcher Fortbildung genutzt wurden.

7. Literaturverzeichnis

Al-Yagon, M., Cavendish, W., Cornoldi, C., Fawcett, A. J., Grünke, M., Hung, L.-Y. et al. (2013). The proposed changes for DSM-5 for SLD and ADHD: International perspectives - Australia, Germany, Greece, India, Israel, Italy, Spain, Taiwan, United Kingdom, and United States. *Journal of Learning Disabilities, 46*(1), 58–72. <https://doi.org/10.1177/0022219412464353>

Anderson, S., Jungjohann, J. & Gebhardt, M. (angenommen). Effects of using curriculum-based measurement (CBM) for progress monitoring in reading and an additive reading instruction in second classes. *Zeitschrift für Grundschulforschung*.

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), 1–18. <https://doi.org/10.1016/j.jsp.2012.09.004>

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L. et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33*(2), 218–233.

Artelt, C. & Rausch, T. (2014). Accuracy of teacher judgments. In S. Krolak-Schwerdt, S. Glock & M. Böhmer (Eds.), *Teachers' professional development. Assessment, training, and learning* (pp. 27–43). Rotterdam: Sense Publishers. https://doi.org/10.1007/978-94-6209-536-6_3

Bennett, R. E. (2011). Formative assessment. A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>

Bental, B. & Tirosh, E. (2007). The relationship between attention, executive functions and reading domain abilities in attention deficit hyperactivity disorder and reading disorder: A comparative study. *Journal of Child Psychology and*

Psychiatry and Allied Disciplines, 48(5), 455–463. <https://doi.org/10.1111/j.1469-7610.2006.01710.x>

Beste, G., Franke, C., Gutzmann, M., Hattendorf, E., Hoppe, I., Pilz, A. et al. (2012). *Handreichung zur Förderung von Lesekompetenz in der Schule. Für die Jahrgangsstufen 1 bis 10 in Grundschulen und allen weiterführenden Schulen (Bildungsregion Berlin-Brandenburg)*. Ludwigsfelde-Struveshof: LISUM.

Biancarosa, G. & Cummings, K. D. (2015). New metrics, measures, and uses for fluency data: An introduction to a special issue on the assessment of reading fluency. *Reading and Writing*, 28(1), 1–7. <https://doi.org/10.1007/s11145-014-9516-1>

Bos, W., Valtin, R., Hußmann, A., Wendt, H. & Goy, M. (2017). IGLU 2016: Wichtige Ergebnisse im Überblick. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes et al. (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 13–28). Münster: Waxmann Verlag.

Bowey, J. A. (2011). Predicting individual differences in learning to read. In M. Snowling & C. Hulme (Eds.), *The science of reading. A handbook* (Blackwell handbooks of developmental psychology, pp. 155–172). Malden, Mass.: Blackwell Publishers.

Brinker, K., Antos, G., Heinemann, W. & Sager, S. F. (2000). *Text- und Gesprächslinguistik*. Berlin, New York: De Gruyter.
<https://doi.org/10.1515/9783110135596.1>

Brown-Chidsey, R., Davis, L. & Maya, C. (2003). Sources of variance in curriculum-based measures of silent reading. *Psychology in the Schools*, 40(4), 363–377.
<https://doi.org/10.1002/pits.10095>

Bus, A. G. & von Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91(3), 403-414.

- Cain, K., Patson, N. & Andrews, L. (2005). Age- and ability-related differences in young readers' use of conjunctions. *Journal of Child Language*, 32(4), 877–892. <https://doi.org/10.1017/S0305000905007014>
- Catts, H. W., Adlof, S. M. & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech Language and Hearing Research*, 49(2), 278–293. [https://doi.org/10.1044/1092-4388\(2006/023\)](https://doi.org/10.1044/1092-4388(2006/023))
- Catts, H. W., Compton, D., Tomblin, J. B. & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, 104(1), 166–181. <https://doi.org/10.1037/a0025323>
- Catts, H. W., Hogan, T. P. & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, 36(2), 151–164. <https://doi.org/10.1177/002221940303600208>
- Chafe, W. L. (1970). *Meaning and the structure of language*. Chicago: The University of Chicago Press.
- Christ, T. J., Zopluoglu, C., Monaghan, B. D. & Van Norman, E. R. (2013). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology*, 51(1), 19–57. <https://doi.org/10.1016/j.jsp.2012.11.001>
- Christmann, U. & Groeben, N. (1999). Psychologie des Lesens. In B. Franzmann, K. Hasemann, D. Löffler & E. Schön (Hrsg.), *Handbuch Lesen* (S. 145–223). Berlin: De Gruyter.
- Coelen, H. & Siewert, J. (2008). Der Ferieneffekt - Auch in Deutschland schichtspezifisch? In J. Ramseger & M. Wagener (Hrsg.), *Chancenungleichheit in der Grundschule. Ursachen und Wege aus der Krise* (Bd. 66, S. 87–90). Wiesbaden: VS Verlag für Sozialwissenschaften/ GWV Fachverlage GmbH. https://doi.org/10.1007/978-3-531-91108-3_10

- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Eds.), *Strategies of information processing* (pp. 151–216). London: Academic Press.
- Cromley, J. G. & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311–325. <https://doi.org/10.1037/0022-0663.99.2.311>
- Cunningham, J. W. & Cunningham, P. M. (1978). Validating a limited — Cloze procedure. *Journal of Reading Behavior*, *10*(2), 211–213. <https://doi.org/10.1080/10862967809547272>
- Datnow, A. & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *17*(1), 7–28. <https://doi.org/10.1007/s10833-015-9264-2>
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*(3), 219–232. <https://doi.org/10.1177/001440298505200303>
- Deno, S. L. (2003a). Curriculum-based measures: Development and perspectives. *Curriculum-Based Measures: Development and Perspectives*, *28*(3-4), 3–12. <https://doi.org/10.1177/073724770302800302>
- Deno, S. L. (2003b). Developments in curriculum-based measurement. *The Journal of Special Education*, *37*(3), 184–192. <https://doi.org/10.1177/00224669030370030801>
- Deno, S. L., Fuchs, L. S., Marston, D. & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, *30*(4), 507–524.
- Deno, S. L., Mirkin, P.K. & Chiang, B. (1982). Identifying a valid measure of reading. *Exceptional Children*, *49*(1), 36–45.
- Diehl, K. (2010). Lesenlernen unter erschwerten Bedingungen im Anfangsunterricht - Leselehrwerke im Vergleich. *Zeitschrift für Heilpädagogik*, *61*(3), 109–117.

- Diehl, K. & Hartke, B. (2012). *IEL-1 - Inventar zur Erfassung der Lesekompetenzen von Erstklässlern*. Göttingen: Hogrefe.
- Diehl, K., Hartke, B. & Knopp, E. (2009). Curriculum-based measurement & Leerlingonderwijsvolgsysteem – Konzepte zur theoriegeleiteten Lernfortschrittsmessung im Anfangsunterricht Deutsch und Mathematik? *Zeitschrift für Heilpädagogik*, 60(4), 122–130.
- Dirim, İ., Hauenschild, K. & Lütje-Klose, B. (2008). Ethische Vielfalt und Mehrsprachigkeit an Schulen. Einführung. In İ. Dirim, K. Hauenschild, B. Lütje-Klose, J. Löser & I. Sievers (Hrsg.), *Ethnische Vielfalt und Mehrsprachigkeit an Schulen. Beispiele aus verschiedenen nationalen Kontexten* (S. 9–12). Frankfurt am Main: Brandes & Apsel.
- Dummer-Smoch, L. & Hackethal, R. (2016). *Kieler Leseaufbau. Handbuch und Übungsmaterialien: Ausgabe C, Druckschrift* (9. Auflage). Kiel: Veris-Verlag.
- Duzy, D., Ehm, J.-H., Souvignier, E., Schneider, W. & Gold, A. (2013). Prädiktoren der Lesekompetenz bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45(4), 173–190. <https://doi.org/10.1026/0049-8637/a000093>
- Ecalte, J., Bouchafa, H., Potocki, A. & Magnan, A. (2013). Comprehension of written sentences as a core component of children's reading comprehension. *Journal of Research in Reading*, 36(2), 117–131. <https://doi.org/10.1111/j.1467-9817.2011.01491.x>
- Ehri, L. C. (1999). Phases of development in learning to read words. In J. Oakhill & R. Beard (Eds.), *Reading development and the teaching of reading. A psychological perspective* (pp. 79–108). Oxford: Blackwell Publishers.
- 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(2), 53–67. <https://doi.org/10.1026/0049-8637/a000057>

- Espin, C. A., Wayman, M. M., Deno, S. L., McMaster, K. L. & de Rooij, M. (2017). Data-based decision-making. Developing a method for capturing teachers' understanding of CBM graphs. *Learning Disabilities Research & Practice*, 32(1), 8–21. <https://doi.org/10.1111/ldrp.12123>
- Esser, G., Wyszkon, A. & Schmidt, M. H. (2002). Was wird aus Achtjährigen mit einer Lese- und Rechtschreibstörung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 31(4), 235–242. <https://doi.org/10.1026//1616-3443.31.4.235>
- Euker, N., Kuhl, J. & Probst, H. (2012). Individuelle Förderung des Leseerwerbs im Rahmen Inklusiven Unterrichts. *Gemeinsam leben*, 20(3), 139–150. Zugriff am 04.08.2019. Verfügbar unter https://content-select.com/media/moz_viewer/527fccdb-e050-4a5b-af4b-68cd2efc1343/language:de
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, 102(6), 453–462. <https://doi.org/10.3200/JOER.102.6.453-462>
- Fillmore, C. J. (1968). The case for case. In R. T. Harms & E. W. Bach (Eds.), *Universals in linguistic theory* (pp. 1–25). London: Holt, Rinehart and Winston.
- Fischer, U. & Gasteiger-Klicpera, B. (2013). Prävention von Leseschwierigkeiten. Diagnose und Förderung im Anfangsunterricht. *Didaktik Deutsch*, 19(35), 62–81.
- Forster, K. (2011). Using a maze task to track lexical and sentence processing. *Mental Lexicon*, 5(3), 347–357. <https://doi.org/10.1075/ml.5.3.05for>
- Förster, N., Kuhn, J.-T. & Souvignier, E. (2017). Normierung von Verfahren zur Lernverlaufsdagnostik. *Empirische Sonderpädagogik*, 9(2), 116–122.
- Förster, N. & Souvignier, E. (2011). Curriculum-based measurement. Developing a computer-based assessment instrument for monitoring student reading progress on multiple indicators. *Learning Disabilities: a Contemporary Journal*, 9(2), 65–88.
- Frisson, S., Rayner, K. & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of*

Experimental Psychology. Learning, Memory, and Cognition, 31(5), 862–877.
<https://doi.org/10.1037/0278-7393.31.5.862>

Frith, U. (1985). *Beneath the surface of developmental dyslexia*. Hillsdale, N.J.: Lawrence Erlbaum.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, 33(2), 188–192.

Fuchs, L. S. (2017). Curriculum-based measurement as the emerging alternative. Three decades later. *Learning Disabilities Research & Practice*, 32(1), 5–7.
<https://doi.org/10.1111/ldrp.12127>

Fuchs, L. S. & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, 21(1), 45–58.

Fuchs, L. S. & Fuchs, D. (2001). Computer applications to curriculum-based measurement. *Special Services in the Schools*, 17(1-2), 1–14.

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(1), 7–21.
<https://doi.org/10.1177/001440290407100101>

Fuchs, L. S., Fuchs, D., Hamlett, C. L. & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement, using a reading maze task. *Exceptional Children*, 58(5), 436–450.
<https://doi.org/10.1177/001440299205800507>

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L. & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*, 22(1), 27–48.

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(3), 239–256.
https://doi.org/10.1207/S1532799XSSR0503_3

- Galuschka, K. & Schulte-Körne, G. (2015). Evidenzbasierte Interventionsansätze und forschungsbasierte Programme zur Förderung der Leseleistung bei Kindern und Jugendlichen mit Lesestörung – Ein systematischer Review. *Zeitschrift für Erziehungswissenschaft*, 18(3), 473–487. <https://doi.org/10.1007/s11618-015-0650-6>
- García, J. R. & Cain, K. (2014). Decoding and reading comprehension. *Review of Educational Research*, 84(1), 74–111. <https://doi.org/10.3102/0034654313499616>
- Gasteiger-Klicpera, B. & Fischer, U. (2008). Evidenzbasierte Förderung bei Leserechtschreibschwierigkeiten. In M. Fingerle & S. Ellinger (Hrsg.), *Sonderpädagogische Förderprogramme im Vergleich. Orientierungshilfen für die Praxis* (Heil- und Sonderpädagogik, S. 67–84). Stuttgart: Kohlhammer Verlag.
- Gebhardt, M. (2015). Gemeinsamer Unterricht von Schülerinnen und Schülern mit und ohne sonderpädagogischen Förderbedarf. Ein empirischer Überblick. In E. Kiel (Hrsg.), *Inklusion im Sekundarbereich* (S. 39–52). Stuttgart: Kohlhammer Verlag.
- Gebhardt, M., Diehl, K. & Mühling, A. (2016a). *Lern-Verlaufs-Monitoring LEVUMI Lehrerhandbuch*. Dortmund: Technische Universität Dortmund. <https://doi.org/10.17877/DE290R-17792>
- Gebhardt, M., Diehl, K. & Mühling, A. (2016b). Online Lernverlaufsmessung für alle SchülerInnen in inklusiven Klassen. www.LEVUMI.de. *Zeitschrift für Heilpädagogik*, 67(10), 444–453.
- Gebhardt, M., Heine, J.-H., Zeuch, N. & Förster, N. (2015). Lernverlaufsdagnostik im Mathematikunterricht der zweiten Klasse: Raschanalysen und Empfehlungen zur Adaptation eines Testverfahrens für den Einsatz in inklusiven Klassen. *Empirische Sonderpädagogik*, 7(3), 206–222.
- Gebhardt, M. & Jungjohann, J. (im Druck). Digitale Unterstützung bei der Dokumentation von Verhalts- und Leistungsbeurteilungen. In B. E. Meyer, T. Tretter & U. Englisch (Hrsg.), *Praxisleitfaden auffällige Schüler und Schülerinnen. Basiswissen und Handlungsmöglichkeiten mit Online-Materialien*. Weinheim: Beltz.

- Gebhardt, M., Sälzer, C., Mang, J., Müller, K. & Prenzel, M. (2015). Performance of students with special educational needs in Germany. Findings from Programme for International Student Assessment 2012. *Journal of Cognitive Education and Psychology*, 14(3), 343–356. <https://doi.org/10.1891/1945-8959.14.3.343>
- Gellert, A. S. & Elbro, C. (2013). Do experimental measures of word learning predict vocabulary development over time? A study of children from grade 3 to 4. *Learning and Individual Differences*, 26, 1–8. <https://doi.org/10.1016/j.lindif.2013.04.006>
- Gillingham, M. G. & Garner, R. (1992). Readers' comprehension of mazes embedded in expository texts. *The Journal of Educational Research*, 85(4), 234–241. <https://doi.org/10.1080/00220671.1992.9941121>
- Gold, A. (2009). Leseflüssigkeit. Dimensionen und Bedingungen bei leseschwachen Hauptschülern. In A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität. Bildungsaufgabe und Forschungsfeld* (Lesesozialisation und Medien, S. 151–164). Weinheim: Juventa Verlag.
- Gold, A. (2018). *Lernschwierigkeiten. Ursachen, Diagnostik, Intervention* (Kohlhammer Standards Psychologie, 2. erweiterte und überarbeitete Auflage). Stuttgart: Kohlhammer Verlag.
- Gold, A., Nix, D., Rieckmann, C. & Rosebrock, C. (2011). Leseflüssigkeit und Textverstehen. Vom flüssigen zum verstehenden Lesen. *lernchancen*, 14(81-82), 52–59.
- Good, R. & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Eds.), *Advanced applications of curriculum-based measurement* (The Guilford school practitioner series, pp. 61–88). New York: Guilford Press.
- 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(3), 257–288. https://doi.org/10.1207/S1532799XSSR0503_4

- Graney, S. B., Martínez, R. S., Missall, K. N. & Aricak, O. T. (2010). Universal screening of reading in late elementary school. *Remedial and Special Education*, 31(5), 368–377. <https://doi.org/10.1177/0741932509338371>
- Graney, S. B. & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review*, 34(2), 184–201.
- Grünke, M. & Cavendish, W. (2016). Learning disabilities around the globe: Making sense of the heterogeneity of the different viewpoints. *Learning Disabilities: A Contemporary Journal*, 14(1), 1–8.
- Günther, K. B. (1986). Ein Stufenmodell der Entwicklung kindlicher Lese- und Schreibstrategien. In H. Brügelmann (Hrsg.), *ABC und Schriftsprache. Rätsel für Kinder, Lehrer und Forscher* (S. 32–54). Konstanz: Faude.
- Haag, N., Kocaj, A., Jansen, M. & Kuhl, P. (2017). Soziale Disparitäten. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 213–236). Münster: Waxmann Verlag.
- Hale, A. D., Hawkins, R. O., Sheeley, W., Reynolds, J. R., Jenkins, S., Schmitt, A. J. et al. (2011). An investigation of silent versus aloud reading comprehension of elementary students using maze assessment procedures. *Psychology in the Schools*, 48(1), 4–13. <https://doi.org/10.1002/pits.20543>
- Hamilton, C. & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review*, 32(2), 228–240.
- Hedges, L. V. & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1), 60–87. <https://doi.org/10.3102/0162373707299706>

- Heimlich, U. (Hrsg.). (2007). *Didaktik des Unterrichts im Förderschwerpunkt Lernen. Ein Handbuch für Studium und Praxis* (Heil- und Sonderpädagogik). Stuttgart: Kohlhammer Verlag.
- Hillenbrand, C. (2015). Evidenzbasierung sonderpädagogischer Praxis: Widerspruch oder Gelingensbedingung? *Zeitschrift für Heilpädagogik*, 66(7), 312–324.
- Holland, P. W. & Wainer, H. (1993). *Differential item functioning*. Hillsdale, N.J.: Lawrence Erlbaum.
- Holle, K. (2010). Psychologische Lesemodelle und ihre lesedidaktischen Implikationen. In C. Garbe, K. Holle & T. Jesch (Hrsg.), *Texte lesen. Lesekompetenz - Textverstehen - Lesedidaktik - Lesesozialisation* (2. Auflage, S. 103–165). Stuttgart: UTB.
- Hoover, W. A. & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. <https://doi.org/10.1007/BF00401799>
- Hu, L.-t. & Bentler, P. M. (1998). Fit indices in covariance structure modeling. Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. <https://doi.org/10.1037/1082-989X.3.4.424>
- Huber, C. & Grosche, M. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik*, 63(8), 312–322.
- Hudson, R. F., Pullen, P. C., Lane, H. B. & Torgesen, J. K. (2008). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly*, 25(1), 4–32. <https://doi.org/10.1080/10573560802491208>
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M. et al. (Hrsg.). (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann Verlag. Verfügbar unter <https://content-select.com/de/portal/media/view/5aa259f2-7384-45b8-9192-27cab0dd2d03>

- Ise, E., Engel, R. R. & Schulte-Körne, G. (2012). Was hilft bei der Leserechtschreibstörung? *Kindheit und Entwicklung*, 21(2), 122–136.
<https://doi.org/10.1026/0942-5403/a000077>
- January, S.-A. A. & Ardoin, S. P. (2012). The impact of context and word type on students' maze task accuracy. *School Psychology Review*, 41(3), 262–271.
- Jones, D. E., Greenberg, M. & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, 105(11), 2283–2290.
<https://doi.org/10.2105/AJPH.2015.302630>
- Jungjohann, J., DeVries, J. M., Gebhardt, M. & Mühling, A. (2018). Levumi. A web-based curriculum-based measurement to monitor learning progress in inclusive classrooms. In K. Miesenberger & G. Kouroupetroglou (Eds.), *Computers helping people with special needs. ICCHP 2018. Lecture notes in computer science* (pp. 369–378). Cham: Springer International Publishing.
- Jungjohann, J., DeVries, J. M., Mühling, A. & Gebhardt, M. (2018). Using theory-based test construction to develop a new curriculum-based measurement for sentence reading comprehension. *Frontiers in Education*, 3, 1.
<https://doi.org/10.3389/feduc.2018.00115>
- Jungjohann, J., Diehl, K. & Gebhardt, M. (2019). *SiL-Levumi - Tests der Leseflüssigkeit zur Lernverlaufsdagnostik - "Silben lesen" der Onlineplattform www.levumi.de. [Verfahrensdokumentation aus PSYNDEX Tests-Nr. 9007767 und Silbenlisten]* (Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Hrsg.). Trier: ZPID.
<https://doi.org/10.23668/psycharchives.2462>
- Jungjohann, J., Diehl, K., Mühling, A. & Gebhardt, M. (2018). Graphen der Lernverlaufsdagnostik interpretieren und anwenden – Leseförderung mit der Onlineverlaufsmessung Levumi. *Forschung Sprache*, 6(2), 84–91.
<https://doi.org/10.17877/DE290R-19806>
- Jungjohann, J. & Gebhardt, M. (2018). Lernverlaufsdagnostik im inklusiven Anfangsunterricht Lesen – Verschränkung von Lernverlaufsdagnostik,

- Förderplanung und Wochenplanarbeit. In F. Hellmich, G. Görel & M. F. Löper (Hrsg.), *Inklusive Schul- und Unterrichtsentwicklung. Vom Anspruch zur erfolgreichen Umsetzung* (S. 160–172). Stuttgart: Kohlhammer Verlag.
- Jungjohann, J. & Gebhardt, M. (2019). *SinnL-Levumi - Tests zum sinnkonstruierenden Satzlesen als Lernverlaufsdiagnostik - "Sinnkonstruierendes Satzlesen" der Onlineplattform www.levumi.de. [Verfahrensdokumentation aus PSYNDEX Tests-Nr. 9007837 mit Manuskriptfassung, Itemlisten und Auswertungsbögen]* (Leibniz Institut für Psychologische Information und Dokumentation (ZPID), Hrsg.). Trier: ZPID.
<https://doi.org/10.23668/psycharchives.2463>
- Jungjohann, J., Gebhardt, M., Diehl, K. & Mühlning, A. (2017). *Förderansätze im Lesen mit LEVUMI*. Dortmund: Technische Universität Dortmund.
<https://doi.org/10.17877/DE290R-18042>
- Jungjohann, J., Gegenfurtner, A. & Gebhardt, M. (2018). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit. *Empirische Sonderpädagogik*, 10(1), 100–118.
- Jungjohann, J., Mau, L., Diehl, K. & Gebhardt, M. (2019). *Levumi: Handbuch für Lehrkräfte Deutsch*. Dortmund: Technische Universität Dortmund.
<https://doi.org/10.17877/DE290R-19921>
- Jungjohann, J., Schurig, M. & Gebhardt, M. (2019). Konstruktion einer Lernverlaufsdiagnostik im Lesen für den Längsschnitt. Messungen mit Leseflüchtigkeits- und Leseverständnistests in der 3. und 4. Jahrgangsstufe in einer inklusiven Grundschule. *Manuskript zur Veröffentlichung eingereicht*.
- Juska-Bacher, B., Beckert, C., Stalder, U. & Schneider, H. (2016). Die Bedeutung des Wortschatzes für basale Lesekompetenzen. *Didaktik Deutsch*, 21(40), 20–40.
- Kendeou, P., Papadopoulos, T. C. & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction*, 22(5), 354–367. <https://doi.org/10.1016/j.learninstruc.2012.02.001>

- Kennison, S. M. (2009). The use of verb information in parsing. Different statistical analyses lead to contradictory conclusions. *Journal of Psycholinguistic Research*, 38(4), 363–378. <https://doi.org/10.1007/s10936-008-9096-9>
- Kingston, A. J. & Weaver, W. W. (1970). Feasibility of cloze techniques for teaching and evaluating culturally disadvantaged beginning readers. *The Journal of Social Psychology*, 82(2), 205–214. <https://doi.org/10.1080/00224545.1970.9919952>
- Kintsch, W. & Rawson, K. A. (2011). Comprehension. In M. Snowling & C. Hulme (Eds.), *The science of reading. A handbook* (Blackwell handbooks of developmental psychology, pp. 209–226). Malden, Mass.: Blackwell Publishers.
- Klapproth, F. (2018). Biased predictions of students' future achievement. An experimental study on pre-service teachers' interpretation of curriculum-based measurement graphs. *Studies in Educational Evaluation*, 59, 67–75. <https://doi.org/10.1016/j.stueduc.2018.03.004>
- Klauer, K. J. (2006). Erfassung des Lernfortschritts durch curriculumbasierte Messungen. *Heilpädagogische Forschung*, 32(1), 16–26.
- Klauer, K. J. (2011). Lernverlaufsdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik*, 3(3), 207–224.
- Klauer, K. J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 1–17). Göttingen: Hogrefe.
- Klicpera, C., Ehgartner, M., Gasteiger-Klicpera, B. & Schabmann, A. (1993). Voraussetzungen für das Lesenlernen 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*, 19(3), 104–108.
- Klicpera, C. & Gasteiger-Klicpera, B. (1993). *Lesen und Schreiben. Entwicklung und Schwierigkeiten. Die Wiener Längsschnittuntersuchung über die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit*. Bern: Huber.

- Klicpera, C., Schabmann, A. & Gasteiger-Klicpera, B. (2006). Die mittelfristige Entwicklung von Schülern mit Teilleistungsschwierigkeiten im Bereich der Lese- und Rechtschreibschwierigkeiten. *Kindheit und Entwicklung*, 15(4), 216–227.
<https://doi.org/10.1026/0942-5403.15.4.216>
- Klicpera, C., Schabmann, A., Gasteiger-Klicpera, B. & Schmidt, B. (2017). *Legasthenie - LRS. Modelle, Diagnose, Therapie und Förderung: mit 100 Übungsfragen* (5. überarbeitete und erweiterte Auflage). München: Ernst Reinhardt Verlag. Verfügbar unter <http://www.utb-studi-e-book.de/9783838548166>
- Knoepke, J., Richter, T., Isberner, M.-B., Neeb, Y. & Naumann, J. (2013). Leseverstehen = Hörverstehen X Dekodieren? Ein stringenter Test der Simple View of Reading bei deutschsprachigen Grundschulkindern. In A. Redder & S. Weinert (Hrsg.), *Sprachförderung und Sprachdiagnostik. Interdisziplinäre Perspektiven* (S. 256–276). Münster u.a.: Waxmann Verlag.
- Kocaj, A., Kuhl, P., Haag, N., Kohrt, P. & Stanat, P. (2017). Schulische Kompetenzen und schulische Motivation von Kindern mit sonderpädagogischem Förderbedarf an Förderschulen und an allgemeinen Schulen. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 302–315). Münster: Waxmann Verlag.
- Kocaj, A., Kuhl, P., Kroth, A. J., Pant, H. A. & Stanat, P. (2014). Wo lernen Kinder mit sonderpädagogischem Förderbedarf besser? Ein Vergleich schulischer Kompetenzen zwischen Regel- und Förderschulen in der Primarstufe. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(2), 165–191.
<https://doi.org/10.1007/s11577-014-0253-x>
- Kölm, J., Gresch, C. & Kuhl, P. (2019). Zuwanderungsbezogene Disparitäten bei der Diagnose eines sonderpädagogischen Förderbedarfs Lernen und der besuchten Schulart. *Zeitschrift für Erziehungswissenschaft*, 22(4), 771–789.
<https://doi.org/10.1007/s11618-019-00896-y>

- 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. <https://doi.org/10.1016/j.lindif.2013.12.011>
- Kretschmann, R. (2007). Schulalter. In J. Walter, F. B. Wember, J. Borchert & H. Goetze (Hrsg.), *Sonderpädagogik des Lernens* (S. 245–266). Hogrefe Verlag.
- Kuhl, J. & Euker, N. (2016). Evidenzbasierung von Unterricht und Förderung – Chancen und Grenzen des Konzepts. In J. Kuhl & N. Euker (Hrsg.), *Evidenzbasierte Diagnostik und Förderung von Kindern und Jugendlichen mit intellektueller Beeinträchtigung* (1. Auflage, S. 19–38). Bern: Hogrefe.
- Kuhl, J., Gebhardt, M., Bienstein, P., Käßler, C., Quinten, S., Ritterfeld, U. et al. (2017). Implementationsforschung als Voraussetzung für eine evidenzbasierte sonderpädagogische Praxis. *Sonderpädagogische Förderung*, 62(4), 383–393.
- Kuhl, J. & Hecht, T. (2014). Prävention von Lernschwierigkeiten durch die Implementierung von Diagnostik und Förderung - Ein Praxisbeispiel für das erste Schuljahr. *Zeitschrift für Heilpädagogik*, 65(11), 406–415.
- Kuhl, J., Križan, A., Sinner, D., Probst, H., Hofmann, C. & Ennemoser, M. (2012). Von der sonderpädagogischen Diagnostik zur pädagogisch-psychologischen Diagnostik im Dienst schulischer Prävention. In V. Moser (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online. Behinderten- und Integrationspädagogik: Institutionelle Felder*. Weinheim: Beltz Juventa.
- Kuhl, J. & Wittich, C. (2018). Individuelle Förderung im inklusiven Unterricht. In K. Rathmann & K. Hurrelmann (Hrsg.), *Leistung und Wohlbefinden in der Schule: Herausforderung Inklusion* (S. 381–395). Weinheim: Beltz Juventa.
- Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A. & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, 45(2), 230–251. <https://doi.org/10.1598/RRQ.45.2.4>

Kuhn, M. R. & Stahl, S. A. (2003). Fluency. A review of developmental and remedial practices. *Journal of Educational Psychology*, 95(1), 3–21.
<https://doi.org/10.1037/0022-0663.95.1.3>

Kultusministerkonferenz. (2019). *Empfehlungen zur schulischen Bildung, Beratung und Unterstützung von Kindern und Jugendlichen im sonderpädagogischen Schwerpunkt LERNEN. Beschluss vom 14.03.2019*. Zugriff am 17.10.2019.
Verfügbar unter
https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2019/2019_03_14-FS-Lernen.pdf

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(1), 150–161. <https://doi.org/10.1037/0022-0663.100.1.150>

Lange, V. (2017). *Inklusive Bildung in Deutschland. Ländervergleich*. Berlin: Friedrich-Ebert-Stiftung Abteilung Studienförderung. Verfügbar unter
<http://library.fes.de/pdf-files/studienfoerderung/13834.pdf>

Leach, J. M., Scarborough, H. S. & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95(2), 211–224.
<https://doi.org/10.1037/0022-0663.95.2.211>

Lenhard, W. (2019a). Gegenwärtiger Stand der empirischen Unterrichtsforschung zur Vermittlung von Lesekompetenz. In M. Kämper-van den Boogaart & K. H. Spinner (Hrsg.), *Lese- und Literaturunterricht. Teil 2: Kompetenzen und Unterrichtsziele, Methoden und Unterrichtsmaterialien* (3. Auflage, S. 305–329). Baltmannsweiler: Schneider Verlag.

Lenhard, W. (2019b). *Leseverständnis und Lesekompetenz. Grundlagen - Diagnostik - Förderung* (2. aktualisierte Auflage). Stuttgart: Kohlhammer Verlag.

Lenhard, W., Lenhard, A. & Schneider, W. (2017). *ELFE II - ein Leseverständnistest für Erst- bis Siebtklässler. Version II*. Göttingen: Hogrefe Schultests.

Liebers, K., Landwehr, B., Marquardt, A. & Schlotter, K. (Hrsg.). (2015). *Lernprozessbegleitung und adaptives Lernen in der Grundschule*.

Forschungsbezogene Beiträge. Wiesbaden: Springer VS.

<https://doi.org/10.1007/978-3-658-11346-9>

Lindsay, G. & Strand, S. (2016). Children with language impairment. Prevalence, associated difficulties, and ethnic disproportionality in an English population. *Frontiers in Education*, 2, 1. <https://doi.org/10.3389/feduc.2016.00002>

Louthan, V. (1965). Some systematic grammatical deletions and their effects on reading comprehension. *The English Journal*, 54(4), 295–299. <https://doi.org/10.2307/811113>

Madelaine, A. & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52(1), 33–42. <https://doi.org/10.1080/10349120500071886>

Maier, U. (2014). Computergestützte, formative Leistungsdiagnostik in Primar- und Sekundarschulen. Ein Forschungsüberblick zu Entwicklung, Implementation und Effekten. *Unterrichtswissenschaft*, 42(1), 69–86.

Maitz, K., Paleczek, L., Seifert, S. & Gasteiger-Klicpera, B. (2018). Zusammenhang der Leseverständnisleistungen mit sozialen Herkunftsfaktoren bei SchülerInnen der dritten Schulstufe. *Zeitschrift für Grundschulforschung*, 11(1), 147–160. <https://doi.org/10.1007/s42278-018-0012-6>

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71–85. <https://doi.org/10.1080/00461520.2012.667064>

Martohardjono, G., Otheguy, R., Gabriele, A., Goeas-Malone, M. de, Szupica-Pyrzanowski, M., Rivero, S. et al. (2005). The role of syntax in reading comprehension: A study of bilingual readers. In J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (p. 1522–1544). Somerville, MA: Cascadilla Press.

Martschinke, S. (2015). Facetten adaptiven Unterrichts aus der Sicht der Unterrichtsforschung. In K. Liebers, B. Landwehr, A. Marquardt & K. Schlotter

- (Hrsg.), *Lernprozessbegleitung und adaptives Lernen in der Grundschule. Forschungsbezogene Beiträge* (S. 15–23). Wiesbaden: Springer VS.
- Mayer, A. (2010). *Gezielte Förderung bei Lese- und Rechtschreibstörungen. Mit 11 Tabellen*. München: Ernst Reinhardt Verlag.
- Mayer, A. (2016). *Lese-Rechtschreibstörungen (LRS). Ursachen, Diagnose und Förderung*. München: Ernst Reinhardt Verlag.
- Mayer, A. (2018). Benennungsgeschwindigkeit und Lesen. *Forschung Sprache*, 6(1), 20–42. Verfügbar unter http://www.edu.lmu.de/shp/personen/lehrstuhlinhaber/andreas_mayer/publikationen/mayer_ran-und-lesen.pdf
- McKenna, M. C. & Miller, J. W. (1980). The effects of age and distractors type on maze performance. In K. L. Kamil (Eds.), *Perspectives on reading research and instruction: 29th yearbook of the National Reading Conference* (pp. 288–292). Washington, DC: National Reading Conference.
- Meisinger, E. B., Bradley, B. A., Schwanenflugel, P. J., Kuhn, M. R. & Morris, R. D. (2009). Myth and reality of the word caller: The relation between teacher nominations and prevalence among elementary school children. *School Psychology Quarterly: the Official Journal of the Division of School Psychology, American Psychological Association*, 24(3), 147–150. <https://doi.org/10.1037/a0017191>
- Meng, M. & Bader, M. (2000). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech*, 43(1), 43–74. <https://doi.org/10.1177/00238309000430010201>
- Moors, A. & de Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132(2), 297–326. <https://doi.org/10.1037/0033-2909.132.2.297>
- Moosbrugger, H. & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Auflage). Berlin: Springer VS.

- Mühling, A., Gebhardt, M. & Diehl, K. (2017). Formative Diagnostik durch die Onlineplattform LEVUMI. *Informatik-Spektrum*, 40(6), 556–561. <https://doi.org/10.1007/s00287-017-1069-7>
- Mühling, A., Jungjohann, J. & Gebhardt, M. (2019). Progress monitoring in primary education using Levumi: A case study. In H. Lane, S. Zvacek & J. Uhomoihi (Eds.), *Proceedings of the 11th International Conference on Computer Supported Education* (pp. 137–144). Setúbal: Science and Technology Publications.
- Nagler, T., Lindberg, S. & Hasselhorn, M. (2018). Leseentwicklung in der Kindheit. *Kindheit und Entwicklung*, 27(1), 5–13. <https://doi.org/10.1026/0942-5403/a000240>
- Nation, K. (2011). Children's reading comprehension difficulties. In M. Snowling & C. Hulme (Eds.), *The science of reading. A handbook* (Blackwell handbooks of developmental psychology, pp. 248–265). Malden, Mass.: Blackwell Publishers.
- 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. Zugriff am 30.03.2017.
- Newell, K. W. & Christ, T. J. (2017). Novice interpretations of progress monitoring graphs: Extreme values and graphical aids. *Assessment for Effective Intervention*, 42(4), 224–236. <https://doi.org/10.1177/1534508417694855>
- Nix, D. (2011). *Förderung der Leseflüssigkeit. Theoretische Fundierung und empirische Überprüfung eines kooperativen Lautlese-Verfahrens im Deutschunterricht*. Weinheim: Juventa Verlag.
- Oakhill, J. V., Cain, K. & Bryant, P. E. (2003). The dissociation of word reading and text comprehension. Evidence from component skills. *Language and Cognitive Processes*, 18(4), 443–468. <https://doi.org/10.1080/01690960344000008>
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford Univ. Pr.

- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Perfetti, C. A., Landi, N. & Oakhill, J. (2011). The acquisition of reading comprehension skill. In M. Snowling & C. Hulme (Eds.), *The science of reading. A handbook* (Blackwell handbooks of developmental psychology, pp. 228–247). Malden, Mass.: Blackwell Publishers.
- Petticrew, M. & Roberts, H. (2012). *Systematic reviews in the social sciences. A practical guide*. Malden, Mass.: Blackwell Publishers.
- Piezunka, A., Schaffus, T. & Grosche, M. (2017). Vier Definitionen von schulischer Inklusion und ihr konsensueller Kern. Ergebnisse von Experteninterviews mit Inklusionsforschenden. *Unterrichtswissenschaft*, 45(4), 207–222.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks pædagogiske Institut.
- Rasinski, T. V. (2004). Creating fluent readers. *Educational Leadership*, 61(6), 46–51.
- Reber, K. (2017). *Prävention von Lese- und Rechtschreibstörungen im Unterricht. Systematischer Schriftspracherwerb von Anfang an mit 64 Abbildungen und 27 Tabellen mit Online-Zusatzmaterial* (2. überarbeitete Auflage). München: Ernst Reinhardt Verlag.
- Reschly, D. J. & Bergstrom, M. K. (2009). Response to intervention. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4. Edition, pp. 434–460). Hoboken, NJ: Wiley.
- Richter, T. & Christmann, U. (2009). Lesekompetenz: Prozessebene und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz. Bedingungen, Dimensionen, Funktionen* (3. Auflage, S. 25–58). Weinheim: Juventa Verlag.
- Richter, T., Isberner, M.-B., Naumann, J. & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, 17(6), 415–434. <https://doi.org/10.1080/10888438.2013.764879>

- Richter, T. & Müller, B. (2017). Entwicklung hierarchieniedriger Leseprozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 51–66). Weinheim: Beltz Juventa.
- Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E. & Stanat, P. (2014). Socioeconomic and language minority classroom composition and individual reading achievement. The mediating role of instructional quality. *Learning and Instruction*, 32, 63–72. <https://doi.org/10.1016/j.learninstruc.2014.01.007>
- Rolff, H.-G. (2016). *Schulentwicklung kompakt. Modelle, Instrumente, Perspektiven* (3. vollständig überarbeitete und ergänzte Auflage). Weinheim: Beltz. Verfügbar unter http://www.content-select.com/index.php?id=bib_view&ean=9783407294838
- Rosebrock, C. & Nix, D. (2017). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung* (8. korrigierte Auflage). Baltmannsweiler: Schneider Verlag.
- Rosebrock, C., Nix, D., Rieckmann, C. & Gold, A. (2017). *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe* (5. Auflage). Seelze: Klett Kallmeyer.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (Psychologie Lehrbuch, 2. vollständig überarbeitete und erweiterte Auflage). Bern: Huber.
- Samuels, S. J. (2006). Reading fluency: Its development and assessment. In A. E. Farstrup & S. J. Samuels (Hrsg.), *What research has to say about reading instruction* (3. Edition, pp. 166–183). Newark, Del.: International Reading Association. <https://doi.org/10.1598/0872071774.8>
- Sappok, C. & Fay, J. (2018). Prosodische Aspekte von Leseflüssigkeit messen. Evaluation einer Ratingprozedur mit Audioaufnahmen von DrittklässlerInnen. *Didaktik Deutsch*, 23(44), 61–83. Verfügbar unter http://www.didaktik-deutsch.de/wp-content/uploads/2018/11/Forschungsbeitrag_Sappok_44.pdf

- Schabmann, A. & Kabicher, R. (2007). Auswirkungen früher Leseschwierigkeiten - Leseverständnis, Leseverhalten und Einstellungen zum Lesen am Ende der Grundschulzeit. *Heilpädagogische Forschung*, 33(3), 155–166.
- Scheer, D. (2014). Lernverläufe sichtbar machen – nur wie?! Eine Einführung in die Lernverlaufsdiagnostik mittels curriculumbasierter Messens. *Sonderpädagogik in Rheinland-Pfalz*, 44(2), 15–19.
- Scheerer-Neumann, G. (1981). Prozessanalyse der Leseschwäche. In R. Valtin, U. O. H. Jung & G. Scheerer-Neumann (Hrsg.), *Legasthenie in Wissenschaft und Unterricht. Leseprozessmodell, Fremdsprachenlegasthenie und Erstlesedidaktik* (S. 183–209). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Scheerer-Neumann, G. (2006). Leseschwierigkeiten. In U. Bredel, H. Günther, P. Klotz, J. Ossner & G. Siebert-Ott (Hrsg.), *Didaktik der deutschen Sprache* (2. Auflage, S. 551–567). Paderborn: Schöningh.
- Scheerer-Neumann, G. (2015). *Lese-Rechtschreib-Schwäche und Legasthenie. Grundlagen, Diagnostik und Förderung*. Stuttgart: Kohlhammer Verlag.
- Scheerer-Neumann, G., Schnitzler, C. D. & Ritter, C. (Hrsg.). (2008). *ILeA - Individuelle Lernstandsanalysen - Lesen* (3. überarbeitete Auflage). Ludwigsfelde-Struveshof: LISUM.
- Schründer-Lenzen, A. & Merckens, H. (2006). Differenzen schriftsprachlicher Kompetenzentwicklung bei Kindern mit und ohne Migrationshintergrund. In A. Schründer-Lenzen (Hrsg.), *Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang* (S. 15–44). Wiesbaden: VS Verlag für Sozialwissenschaften/ GWV Fachverlage GmbH.
- Schurig, M., Jungjohann, J. & Gebhardt, M. (2019). Minimization of a computer-based learning progress test in reading. *Manuskript zur Veröffentlichung eingereicht*.
- Schwanenflugel, P. J., Westmoreland, M. R. & Benjamin, R. G. (2015). Reading fluency skill and the prosodic marking of linguistic focus. *Reading and Writing*, 28(1), 9–30. <https://doi.org/10.1007/s11145-013-9456-1>

- Seifert, S. (2019). Chancen von Digitalisierung im inklusiven Leseunterricht. In L. Paleczek & S. Seifert (Hrsg.), *Inklusiver Leseunterricht. Leseentwicklung, Diagnostik und Konzepte* (S. 267–280). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-24221-3_12
- Shin, J., Deno, S. L. & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education, 34*(3), 164–172. <https://doi.org/10.1177/002246690003400305>
- Silberglitt, B. & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-CB; growth rates by level of performance. *Exceptional Children, 42*(1), 795–819. <https://doi.org/10.1177/001440290707400104>
- Spencer, M., Quinn, J. M. & Wagner, R. K. (2014). Specific reading comprehension disability. Major problem, myth, or misnomer? *Learning Disabilities Research & Practice, 29*(1), 3–9. <https://doi.org/10.1111/ldrp.12024>
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (Hrsg.). (2017). *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann Verlag. Verfügbar unter http://www.content-select.com/index.php?id=bib_view&ean=9783830987307
- Stecker, P. M. (2017). Reflections on teachers' data-based decision making. *Learning Disabilities Research & Practice, 32*(1), 71–72. <https://doi.org/10.1111/ldrp.12128>
- 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*(8), 795–819. <https://doi.org/10.1002/pits.20113>
- Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research, 25*(1), 78–90. <https://doi.org/10.1086/209528>
- Stemmann, J. & Lang, M. (2018). Eignet sich die logfilegenerierte Explorationsvollständigkeit als Prozessindikator für den Wissenserwerb im

- problemlösenden Umgang mit technischen Alltagsgeräten? *Journal of Technical Education (JOTED)*, 6(1), 185–199.
- Stothard, S. E. & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 36(3), 399–408.
- Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L. & Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, 20(4), 331–354.
<https://doi.org/10.1177/014662169602000403>
- Taraban, R. & McClelland, J. L. (1990). Parsing and comprehension: A multiple constraint view. In D. A. Balota, G. B. Flores d'Arcais & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 231–264). Hillsdale: Erlbaum.
- Taylor, C. R. (2012). Engaging the struggling reader: Focusing on reading and success across the content areas. *National Teacher Education Journal*, 5(2), 51–58.
- Tichá, R., Espin, C. A. & Wayman, M. M. (2009). Reading progress monitoring for secondary-school students. Reliability, validity, and sensitivity to growth of reading-aloud and maze-selection measures. *Learning Disabilities Research & Practice*, 24(3), 132–142. <https://doi.org/10.1111/j.1540-5826.2009.00287.x>
- Torgesen, J. (2005). Recent discoveries on remedial interventions for children with dyslexia. In M. J. Snowling (Eds.), *The science of reading. A handbook* (pp. 521–537). Malden, Mass.: Blackwell Publishers.
- United Nations. (2006). *Convention on the rights of persons with disabilities*. Zugriff am 13.08.2019. Verfügbar unter <https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html>

- Van den Bosch, R. M., Espin, C. A., Chung, S. & Saab, N. (2017). Data-based decision-making. Teachers' comprehension of curriculum-based measurement progress-monitoring graphs. *Learning Disabilities Research & Practice*, 32(1), 46–60. <https://doi.org/10.1111/ldrp.12122>
- Van den Bosch, R. M., Espin, C. A., Pat-El, R. J. & Saab, N. (2019). Improving teachers' comprehension of curriculum-based measurement progress-monitoring graphs. *Journal of Learning Disabilities*, 52(5), 413-427. <https://doi.org/10.1177/0022219419856013>
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J. & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 45(1), 2–40.
- Wagner, D. L., Hammerschmidt-Snidarich, S. M., Espin, C. A., Seifert, K. & McMaster, K. L. (2017). Pre-service teachers' interpretation of CBM progress monitoring data. *Learning Disabilities Research & Practice*, 32(1), 22–31. <https://doi.org/10.1111/ldrp.12125>
- Walter, J. (2009a). Eignet sich die Messtechnik "MAZE" zur Erfassung von Lesekompetenzen als lernprozessbegleitende Diagnostik? *Heilpädagogische Forschung*, 35(2), 62–75.
- Walter, J. (2009b). Theorie und Praxis Curriculumbasierten Messens (CBM) in Unterricht und Förderung. *Zeitschrift für Heilpädagogik*, 60(5), 162–170.
- Walter, J. (2010a). *LDL - Lernfortschrittsdiagnostik Lesen. Ein curriculumbasiertes Verfahren*. Göttingen: Hogrefe.
- Walter, J. (2010b). Lernfortschrittsdiagnostik am Beispiel der Lesekompetenz (LDL): Messtechnische Grundlagen sowie Befunde über zu erwartende Zuwachsraten während der Grundschule. *Heilpädagogische Forschung*, 36(4), 162–176.
- Walter, J. (2011a). Die Entwicklung eines auch computerbasiert einsetzbaren Instruments zur formativen Messung der Lesekompetenz. *Heilpädagogische Forschung*, 37(3), 106–126.

- Walter, J. (2011b). Die Messung der Entwicklung der Lesekompetenz im Dienste der systematischen formativen Evaluation von Lehr- und Lernprozessen. *Zeitschrift für Heilpädagogik*, 62(6), 204–217.
- Walter, J. (2013). *VSL. Verlaufsdiagnostik sinnerfassendes Lesen*. Göttingen: Hogrefe.
- Walter, J. (2014). Lernfortschrittsdiagnostik Lesen (LDL) und Verlaufsdiagnostik sinnerfassenden Lesens (VSL): Zwei Verfahren als Instrumente einer formativ orientierten Lesediagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 166–201). Göttingen: Hogrefe.
- Waltzman, D. & Cairns, H. (2000). Grammatical knowledge of third grade good and poor readers. *Applied Psycholinguistics*, 21(2), 263–284.
<https://doi.org/10.1017/S014271640000206X>
- Weis, M., Zehner, F., Sälzer, C., Strohmaier, A., Artelt, C. & Pfof, M. (2016). Lesekompetenz in PISA 2015: Ergebnisse, Veränderungen und Perspektiven. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 249–284). Münster/New York: Waxmann Verlag.
- Wember, F. B. (2001). Adaptiver Unterricht. *Sonderpädagogik*, 31(3), 161–181.
- Wember, F. B. & Greisbach, M. (2018). Zwischen individueller Leseförderung und inklusivem Unterricht - Didaktische Aspekte der Inklusion im Förderschwerpunkt Lernen. In R. Benkmann & U. Heimlich (Hrsg.), *Inklusion im Förderschwerpunkt Lernen* (S. 133–213). Stuttgart: Kohlhammer Verlag.
- West, R. F. & Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Development*, 49(3), 717–727.
- Wilbert, J. (2014). Instrumente zur Lernverlaufsmessung. Gütekriterien und Auswertungsherausforderungen. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 281–308). Göttingen: Hogrefe.
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik*, 3(3), 225–242.

- Wiley, H. I. & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*(4), 207–214.
<https://doi.org/10.1177/07419325050260040301>
- Wimmer, H., Hartl, M. & Moser, E. (1990). Passen "englische" Modelle des Schriftspracherwerbs auf "deutsche" Kinder? Zweifel an der Bedeutsamkeit der logographischen Stufe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 22*(2), 136–154.
- Wimmer, H., Mayringer, H. & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92*(4), 668–680. <https://doi.org/10.1037//0022-0663.92.4.668>
- Witzel, J. & Witzel, N. (2016). Incremental sentence processing in Japanese. A maze investigation into scrambled and control sentences. *Journal of Psycholinguistic Research, 45*(3), 475–505. <https://doi.org/10.1007/s10936-015-9356-4>
- Wocken, H. & Gröhlich, C. (2009). Kompetenzen von Schülerinnen und Schülern an Hamburger Förderschulen. In W. Bos, M. Bensen & C. Gröhlich (Hrsg.), *KESS 7 Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7* (S. 133–142). Münster: Waxmann Verlag.
- Wolf, M. & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5*(3), 211–239.
https://doi.org/10.1207/S1532799XSSR0503_2
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370–371. Zugriff am 28.08.2019. Verfügbar unter <https://www.rasch.org/rmt/rmt83b.htm>
- Zeuch, N., Förster, N. & Souvignier, E. (2017). Assessing teachers' competencies to read and interpret graphs from learning progress assessment. Results from tests and interviews. *Learning Disabilities Research & Practice, 32*(1), 61–70.
<https://doi.org/10.1111/ldrp.12126>

Zimmermann, T. E. (2014). *Einführung in die Semantik* (Einführung Germanistik). Darmstadt: WBG (Wissenschaftliche Buchgesellschaft). Verfügbar unter http://content-select.com/index.php?id=bib_view&ean=9783534735587

Zumeta, R. O., Compton, D. L. & Fuchs, L. S. (2012). Using word identification fluency to monitor first-grade reading development. *Exceptional Children*, 78(2), 201–220. <https://doi.org/10.1177/001440291207800204>

9. Anhang

Die veröffentlichten Fachpublikationen (1, 2, 3 & 5) werden in derselben Reihenfolge abgedruckt, wie sie in Kapitel 4 zusammengefasst wurden.

Erster Artikel

Jungjohann, J., Gegenfurtner, A. & Gebhardt, M. (2018). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit. *Empirische Sonderpädagogik*, 10(1), 100-118.

Zweiter Artikel

Jungjohann, J., DeVries, J. M., Gebhardt, M. & Mühling, A. (2018) Levumi. A web-based curriculum-based measurement to monitor learning progress in inclusive classrooms. In K. Miesenberger, G. Kouroupetroglou & P. Penaz (Eds.), *Computers helping people with special needs. ICCHP 2018. Lecture Notes in Computer Science*, (pp. 369-378). Wiesbaden: Springer. https://doi.org/10.1007/978-3-319-94277-3_58

Dritter Artikel

Jungjohann, J., DeVries, J.M., Mühling, A. & Gebhardt, M. (2018). Using theory-based test construction to develop a new curriculum-based measurement for sentence reading comprehension. *Frontiers in Education*, 3, 1. <https://doi.org/10.3389/educ.2018.00115>

Vierter Artikel

Jungjohann, J., Schurig, M. & Gebhardt, M. (2019). Konstruktion einer Lernverlaufsdagnostik im Lesen für den Längsschnitt. Messungen mit Leseflüchtigkeits- und Leseverständnistests in der 3. und 4. Jahrgangsstufe in einer inklusiven Grundschule. *Manuskript zur Veröffentlichung eingereicht.*

Fünfter Artikel

Jungjohann, J., Diehl, K., Mühling, A. & Gebhardt, M. (2018). Graphen der Lernverlaufsdagnostik interpretieren und anwenden – Leseförderung mit der Onlineverlaufsmessung Levumi. *Forschung Sprache*, 6(2), 84-91. <http://dx.doi.org/10.17877/DE290R-19806>

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 Leserwerb 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 Leserwerbsprobleme 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 lernverlaufsmessungsmethodische 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, Gastteiger-Klicpera & Schmidt, 2017). Besonders in inklusiven Klassen vergrößert sich somit die Heterogenität im Leserwerb. 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 Leseanfängerunterricht 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 (hierarchieniedrige) 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 Leistungsänderungen 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 Leseerschwierigkeiten 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 Wörtern, 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. ¹⁵ Materialien (Anpassung ans Curriculum, Fördermaterialien)	Land	Feedback (Dokumenta- tion, Übertra- gung 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 unterschied- licher 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 unterschied- licher 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 & Madeline (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 aufgefunden 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 Erken-

nung von sprachspezifischen Strategien und Stolpersteine in den Lernverlaufsdaten 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 & Leeringonderwijsvolgssysteem – 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. & Mühling, 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



Levumi: A Web-Based Curriculum-Based Measurement to Monitor Learning Progress in Inclusive Classrooms

Jana Jungjohann¹, Jeffrey M. DeVries¹(✉), Markus Gebhardt¹,
and Andreas Mühling²

¹ Technische Universität Dortmund,
Emil-Figge-Str. 50, 44227 Dortmund, Germany
jeffrey.devries@tu-dortmund.de

² Christian-Albrechts-Universität zu Kiel,
Christian-Albrechts-Platz 4, 24118 Kiel, Germany

Abstract. Our paper introduces and assesses the Levumi platform's web-based assessment of reading fluency. One challenge of inclusive education is meeting the needs of the learners with special education needs (SEN). Children with SEN and other risk factors face an increased risk of failing in schools and manifesting academic and social problems over the time. Web-based curriculum-based measurement (CBM) can provide an effective tool to track progress of learners and limit such risks. In particular, it can ease the challenges of test administration in inclusive classrooms through automation and providing multiple difficulty levels without the need of different paper-forms. Furthermore, Levumi can help educators track children and thus provide support for learners. Levumi takes advantage of the strengths of web-based CBM to assess reading fluency in primary school students. We confirmed the reading fluency test's test-retest reliability ($n = 334$), its ability to measure learning over time in individual learners with SEN ($n = 8$, across 14 MPs), and its applicability to learners with SEN ($n = 300$, including $n = 46$ with SEN). We evaluate Levumi's overall usefulness in assessing different types of learners, and discuss its contributions to CBM research.

Keywords: Curriculum-based measurement · Web-based assessment
Computer-based assessment · Reading fluency

1 Introduction

Since Germany signed the UN-Convention on the Rights of Persons with Disabilities in 2009, the German School system has been changing to meet new standards for inclusive classrooms. Before 2009, children with special educational needs (SEN) primarily visited special schools and were not included in the regular school system. However, an important goal of today's German school system is to reach academic inclusion, meaning that children with and without SEN learn together in the same classroom.

In inclusive classrooms, academic performance of children with and without SEN is very heterogeneous [1]. Children with SEN and other risk factors (e.g., minority or migration background) face an especially high risk of failing in schools and manifesting academic and social problems over the time (e.g., [2]). Issues relating to reading instructions are important to teachers because most students with SEN have difficulties learning to read. Indeed over the last several decades, numerous studies have reported that the vast majority of students with SEN have difficulties in learning to read [3, 4], that they achieve significant lower reading skills than their peers [5], and that this has consequences both within and outside the classroom [6]. These results were also applicable to students learning to read German [1, 7]. Reading problems might negatively affect skills related to early reading acquisition (e.g., reading fluency [8]) or in later skills (e.g., reading comprehension [9]). Reading fluency influences general reading development [10]. Because reading fluency already correlates to reading comprehension in primary school [11], it is a fundamental reading skill and an important goal in the early reading development for every student [12, 13]. Influencing factors, such as general cognitive ability, phonological awareness, speech perception and production, letter-name knowledge, rapid automatized naming, or consistency of the orthography, predict the early reading ability and individualize the learning processes [14, 15]. Additionally, specific reading fluency instruction can be differentially effective for students with and without SEN [16].

Furthermore, the heterogeneous conditions and needs in inclusive classrooms challenges teachers in new ways, such as designing effective instruction for the special needs students. In this regard, it is important that the teachers focus on individual learning growth instead of the social comparison between students [17]. Only in this way, can teachers provide targeted effective reading instruction [18]. This requires a new approach: first, teachers need to determine the individual needs of their students. Second, they choose a lesson and give it to the target student. Third, they monitor the individual learning growth and reflect upon the choice of the instruction. If learning growth is sufficient, the teacher can continue with the same instruction or focus on a new goal. If not, the teacher can adapt the method or can try another approach for that student.

Curriculum-based measurement (CBM) is a method for monitoring the learning growth of children and for supporting the teachers in effective decision-making [19]. CBM was designed in the United States to solve academic difficulties in special education [20]. CBM tests can be used very frequently during the lessons, take only few minutes, and show the slope of individual learning growth for a longer period graphically. The tasks are representative of end-year performance goals and integrate various subskills for competence in a domain (e.g., in reading or mathematics) [21]. Like any other test, CBM instruments need to possess quality criteria such as objectivity, reliability, validity, and be sensible of learning growth [22]. For classroom use, it should be simple and easy for educators to use the test and interpret the results. Additionally, it is important that multiple measurement points (MP) are comparable over time and the test is demonstrated to be invariant across these MPs [23].

The reliability and validity of CBM instruments has been shown in the classical test theory framework [24], and they are particularly useful for children with SEN [25]. More recently, computer based instruments have boosted the potential of CBM

instruments [26]. Computer versions can reduce the time requirements, ease the creation of parallel test versions, and provide automatic feedback to students and teachers.

Curriculum-based measurement has also been studied in the context of German schools [27]. The first German CBM tests were pen-and-paper tests tracking reading or mathematic skills [e.g., [28]]. However, in large classes pen-and-paper tests can result in a lot of additional work for teachers. Moreover, with these instruments, teachers must choose CBM tests from lower classes for children with SEN and lower ability levels (e.g., [29]), which complicates the use of the CBM instruments in inclusive classrooms [30]. Newer instruments have focused on online assessments to remedy some of these problems (e.g., [31, 32]), but such online tools often cost money for teachers or administrators [33]. Nonetheless, more research is necessary regarding the use of CBM techniques in real inclusive German classrooms [34].

The web-based platform Levumi (www.levumi.de) was founded by a multidisciplinary research team with the goal of creating a free online CBM tool to assess reading and mathematics competencies in primary education, with a focus on children with SEN, learning problems, behavior problems, or other risk factors. The three main goals of this research project are (1) to offer teachers a practical CBM tool for inclusive classrooms, (2) to improve research on CBM and the acceptance of CBM tools by teachers, and (3) to use the collected data for evaluating supporting materials for research and development in teaching and learning [33].

Levumi is currently available for teachers and researchers. Users can register on the website for free. Supporting material is also provided free of charge (e.g., [35]). Levumi tests can be used in all 16 German federal states and focus on learning goals throughout the country. Levumi runs from within any major browser. This makes it easy for teachers to use the system without having to install additional software, which usually requires administrative privileges that teachers do not have in typical school IT infrastructures. The only requirement is a permanent internet connection. In some tests, the teachers control the tests, and in other tests, the teachers activate a test for a class and the students can then take the test on their own by logging in with a personal ID code. Each such test begins with an instruction page, often with an interactive sample item, which requires multiple inputs to prevent a student from accidentally starting the test.

The design of the platform is visually simple, and the tests look similar to one another in order for learners with SEN to easily work within the platform. For each item, the actual answer and whether or not it is correct is recorded. The sum of correct responses is used as the final score of the test. Immediately after each test, our mascot, a dragon named Levumi, shows each child if he or she has improved. Teachers can see graphically how each student performs in comparison both to past performance and to other students. Furthermore, the teachers get more detailed performance information, including items a student had problems with. The multiple difficulty levels allow learners with SEN to use the same tools and tests as their peers, which makes supporting all learners in an inclusive classroom much easier for teachers. Additionally, the teachers can track student information such as age, sex, migration background, and SEN.

Levumi contains tests of important indicators of overall competence in different learning domains. These learning domains are currently reading, writing, and mathematics. For each domain, several competencies can be measured with separate test types and multiple difficulty levels. The reading test structures are outlined in Fig. 1.

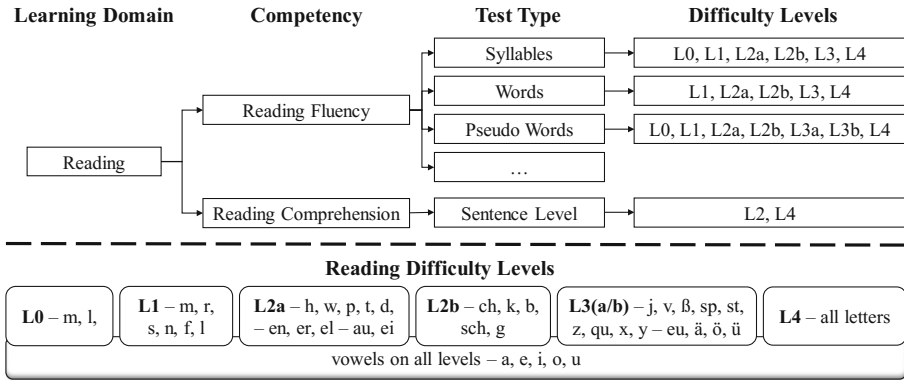


Fig. 1. Test structure of the learning domain reading in the Levumi platform

The reading domain contains two competencies: reading fluency and reading comprehension. For each competency, teachers can pick several test types and difficulty levels, which is important or practical use in class, it is very important that the teachers can easily choose a particular test and difficulty level. For each learning domain, the same difficulty structure exists across all competencies (see Fig. 1, lower part).

Each test has a unique item pool that includes approximately 40 to 200 items. All items are constructed upon theoretical models of reading acquisition. The web platform randomly orders the items for each measurement. This creates a huge number of parallel test forms for each single measurement. In some tests, additional rules are applied to produce the random item ordering. Typically, these tests involve items from multiple categories, such as item types from different dimensions. Additionally, in some tests, such as the reading fluency test, words or syllables with the same initial letter are prevented from following one another, preventing common mistakes. In these cases, items of the different categories or types were randomly in selected via round robin.

The Levumi reading fluency test measures fluency in reading aloud, a robust indicator of overall reading competence [36]. In these tests, the children read aloud items from a computer screen for one minute. Teachers rate the correct and non-correct answers by keyboard and can consider any factors specific to each learner such as a speech impairment. Reading of syllables, words, and pseudo words are assessed through separate tests. All reading fluency tests are based on the ‘Kieler Leseaufbau’ [37], which applies specifically to the German language. For each test, we incorporate multiple difficulties levels (L0 to L4) [38], based on a range of letters (see Fig. 1). Vowels are used in every difficulty level. The lower levels use stretchable consonants (e.g., /m/, /r/, /s/, /l/) in simple word structures. In the higher levels, the plosive, less common consonants, and consonant combinations are used with more complex word structures (e.g., /b/, /sch/, /qu/). In L4 all consonants and vowels are included. A student can be tested in a difficulty level after instruction on all contained letters. Because the teachers can choose a suitable Levumi test based upon the competency level of a student and not upon their age or grade, ease-of-use for teachers in inclusive class-rooms is maintained.

As a part of the completion of test development of the reading fluency assessment for Levumi, we present four research questions. First, to verify the psychometric properties of the test in respect to the item response theory, we ask: (1) how does a Rasch model fit the Levumi syllable reading test? Next, we wish to examine the applicability of the sum scores to important theoretical questions and the general reliability of the syllable reading test and its ability to measure changes over time. Thus, our second and third research questions are: (2) do the sum scores of Levumi reading fluency test possess good test-retest reliability over 2 MPs, and: (3) can the Levumi test measure learning progress of learners with SEN over multiple MPs. Lastly, we wish to examine the applicability of the assessment to learners with special education needs, so we ask: (4) how do Levumi reading fluency test takers with SEN compare to other test takers in terms of sum scores on the syllable reading test?

2 Methods

Participants include test takers of the syllable reading fluency test on the Levumi platform in five samples. The first three samples were measured twice within a period of 7 to 10 weeks, with each representing data from a single difficulty level, L2b ($n = 105$), L3 ($n = 97$), and L4 ($n = 132$). The fourth sample was a small group of learners with SEN from a single class ($n = 8$). The fourth sample included 4 MPs of difficulty level L2b, followed by 5 MPs of L3, and lastly 5 MPs of L4 over one school year. Lastly, the fifth sample was taken from inclusive and special schools ($N = 300$) and included learners with SEN ($n = 46$; 38 with learning SEN, 7 with German SEN, and 1 other).

We calculated a full Rasch analysis for the second sample. We examined item fit scores (infit and outfit) for all test items. We further report Warm's weighted likelihood estimates (WLEs) for the first three samples.

Next, we analyzed the sum scores of each sample. First, we conducted a reliability assessment over 2 MPs in the first three samples. Second, we conducted a repeated measures ANOVA on the sum scores of the 8 participants in the fourth sample to assess the test's ability to measure learning progress. Lastly, we compared the results of SEN learners to other learners in the fifth sample.

3 Results

A graphical model check confirmed that all items performed equally across both MPs for difficulty level L3, indicating good test-retest reliability. Next, we calculated a Rasch model across both measurement points for difficulty level L3. The model possessed good data fits, with the mean square (MSQ) of the outfit ranging from .726 to 1.682 and the MSQ of the infit ranging from .911 to 1.079. Good values for the MSQ of the outfit and infit measures are between 0.5 and 1.5, while only values above 2.0 are considered harmful to measurement [39]. With only 2 of 112 items with an outfit over 1.5, we concluded the Rasch model fits our data very well. Furthermore, all three difficulty levels had good reliability within the Rasch model, $WLE_{L2b} = .919$,

$WLE_{L3} = .883$, and $WLE_{L4} = .895$. Therefore, we conclude that the Rasch models fit the data well, and the test is suitably unidimensional to use sum scores.

Sum scores at difficulty level L4 indicated a very high level of test-retest reliability for difficulties L2b and L4, $r_{L2b} = .84$ and $r_{L4} = .85$. The reliability of difficulty L3 was lower, but still high, $r_{L3} = .76$. This is consistent with the other reliability analyses.

Figure 2 includes all sum scores of the tracking sample of 8 learners with SEN across one school year. Each individual line represents the performance of a single learner. The left section contains MPs of difficulty L2b, the middle has L3, and the right has L4. Separate ANOVAs for each test type confirmed significant changes over time. For the L2b test, performance significantly changed from MP1 ($M = 21.8$, $SD = 5.4$) to MP4 ($M = 30.8$, $SD = 7.9$), $F(3,21) = 13.49$, $p < .001$. For the N3 test no differences over time were found, $F(4,28) = 1.644$, $p > .10$. Lastly for the L4 test, a change from MP1 ($M = 24.0$, $SD = 7.8$) to MP5 ($M = 29.8$, $SD = 8.6$) was detected, $F(4,28) = 4.32$, $p < .01$. We concluded that individual changes over different MPs are detectable.

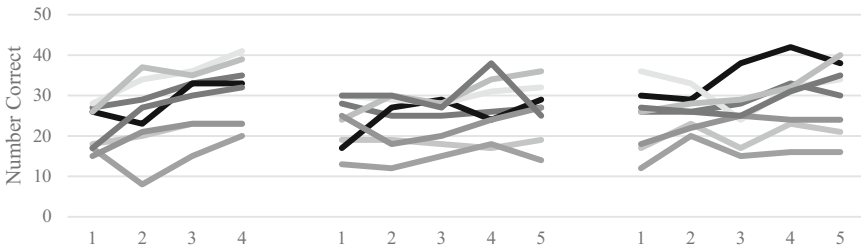


Fig. 2. Individual Tracking across Difficulty Levels L2b, L3, and L4

Lastly, the average items correctly solved for those students with SEN ($M = 28.5$, $SD = 9.7$) was no different than for those without SEN ($M = 30.4$, $SD = 14.0$), Welch's $t(83.4) = 1.08$, $p > .25$.

4 Discussion and Future Work

We assessed the quality of the syllable reading fluency test across different difficulty levels. Rasch analyses verified the test's psychometric qualities. Test-retest reliability was confirmed in a graphical model check and an analyses of sum scores. Learners with SEN performed no differently than those without SEN. Also, scores of the difficulty L2 and L4 improved significantly over the course of learning. Scores for L3 did not change, but this may represent a local plateau in the learning process.

These assessments allow for three important implications. First, good test-retest reliability indicates that any changes in student responses over the course of the school indicate changes in learner ability level not test artifacts. Second, the test is effective at measuring the learning process across multiple MPs. Lastly, equivalent test performance for learners with SEN demonstrates that the test provides a fair assessment for

those learners. However, some limitations remain. We did not assess all learners, all test types, and difficulty levels. Further work should assess other test quality criteria.

Levumi provides multiple tests to monitor learning growth in different learning domains. Development is nearly complete on other tests. An item-response theory based evaluation of test quality criteria is needed for other test types and difficulty levels. Simultaneously, new test types for use on the Levumi platform are being designed. These measure competencies in reading, writing, and mathematics, as well as behavior ratings. They include reading comprehension tests on the level of individual words and complete sentences comprehension, and mathematical assessments on early number sense and number sequencing tasks. Work is ongoing to create more difficult tests for use in secondary schools. Lastly, CBM behavior ratings are planned for primary and secondary schools. Similar test evaluations will be performed on the new tests.

We are also improving platform use on tablets. We have had the chance to collect pilot data for some tests with tablets. Some of these work well on tablets, but others need specific adaptations. Many participants were familiar with touch screens but require some time to familiarize themselves with Levumi. Further work should investigate if there are any mode effects between these two systems. Providing an app that displays the tests for students is also a possible improvement.

Our research established essential reliability and usefulness of a new web-based CBM technique. This platform will allow for rapid assessments and easy tracking of children with and without special education needs in all classroom types. Reliability of the tests for reading fluency was confirmed, and development continues on new tests of different competencies. Lastly, we found no difference in performance between students with and without SEN, and no differences were required in the test preparation and handling for learners with SEN.

References

1. Gebhardt, M., Sälzer, C., Mang, J., Müller, K., Prenzel, M.: Performance of students with special educational needs in Germany. Findings from PISA 2012. *J. Cognit. Educ. Psychol.* **14**(3), 343–356 (2015). <https://doi.org/10.1891/1945-8959.14.3.343>
2. Jones, D.E., Greenberg, M., Crowley, M.: Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *J. Inf.* **105**(11), 2283–2290 (2015). <https://doi.org/10.2105/AJPH.2015.302630>
3. Kavale, K.A., Reece, J.H.: The character of learning disabilities. *Learn. Disabil. Q.* **15**, 74–94 (1992)
4. Bental, B., Tirosh, E.: The relationship between attention, executive functions and reading domain abilities in attention deficit hyperactivity disorder and reading disorder: a comparative study. *J. Child Psychol. Psychiatry* **48**(5), 455–463 (2007). <https://doi.org/10.1111/j.1469-7610.2006.01710.x>
5. National Center for Education Statistics: The nation's report card: Reading 2011 (NCES 2012-457). Institute of Education Sciences, U.S. Department of Education, Washington, D. C. (2011)
6. Taylor, C.R.: Engaging the struggling reader: Focusing on reading and success across the content areas. *Nat. Teacher Educ. J.* **5**(2), 51–58 (2012)

7. Landerl, K., Wimmer, H.: Development of word reading fluency and spelling in a consistent orthography. An 8-year follow- up. *J. Educ. Psychol.* **100**(1), 150–161 (2008). <https://doi.org/10.1037/0022-0663.100.1.150>
8. Klicpera, C., Ehgartner, M., Gasteiger-Klicpera, B., Schabmann, A.: 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 (1993)
9. Nation, K.: Children's reading comprehension difficulties. In: Snowling, M.J., Hulme, Ch. (eds.) *The Science of Reading: A Handbook*, pp. 248–265. Blackwell Publishing Ltd. (2005)
10. National Institute of Child Health and Human Development: 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. U.S. Government Printing Office, Washington, D.C. (2000)
11. Roberts, G., Good, R., Corcoran, S.: Story retell: A fluency-based indicator of reading comprehension. *Sch. Psychol. Q.* **20**(3), 304–317 (2005). <https://doi.org/10.1521/scpq.2005.20.3.304>
12. Hudson, R.F., Pullen, P.C., Lane, H.B., Torgesen, J.K.: The complex nature of reading fluency: a multidimensional view. *Reading & Writ. Q.* **25**(1), 4–32 (2008). <https://doi.org/10.1080/10573560802491208>
13. Perfetti, C.A.: *Reading Ability*. Oxford University Press, New York (1985)
14. Bowey, J.A.: Predicting individual differences in learning to read. In: Snowling, M.J., Hulme, Ch. (eds.) *The Science of Reading: A Handbook*, pp. 155–172. Blackwell Publishing Ltd. (2005)
15. Wimmer, H., Mayringer, H., Landerl, K.: The double-deficit hypothesis and difficulties in learning to read a regular orthography. *J. Educ. Psychol.* **92**(4), 668–680 (2000). <https://doi.org/10.1037//0022-0663.92.4.668>
16. Kuhn, M.R., Stahl, S.A.: Fluency: a review of developmental and remedial practices. *J. Educ. Psychol.* **95**(1), 3–21 (2003). <https://doi.org/10.1037/0022-0663.95.1.3>
17. Gebhardt, M.: Gemeinsamer Unterricht von Schülerinnen und Schülern mit und ohne sonderpädagogischen Förderbedarf. Ein empirischer Überblick. In: Kiel, E. (ed.): *Inklusion im Sekundarbereich*, pp. 39–52. Kohlhammer, Stuttgart (2015)
18. Jungjohann, J., Gebhardt, M.: Lernverlaufsdagnostik im inklusiven Anfangsunterricht Lesen – Verschränkung von Lernverlaufsdagnostik, Förderplanung und Wochenplanarbeit. In: Hellmich, F., Görel, G., Löper, M.F. (eds.) *Inklusive Schul- und Unterrichtsentwicklung*. Kohlhammer, Stuttgart (2018, in Press)
19. Deno, S.L.: Curriculum-based measurement: the emerging alternative. *Except. Child.* **52**, 219–232 (1985). <https://doi.org/10.1177/001440298505200303>
20. Deno, S.L.: Developments in curriculum-based measurement. *J. Spec. Educ.* **37**(3), 184–192 (2003). <https://doi.org/10.1177/00224669030370030801>
21. Fuchs, L.S.: The past, present, and future of curriculum-based measurement research. *Sch. Psychol. Rev.* **33**(2), 188–192 (2004)
22. Good, R., Jefferson, G.: Contemporary perspectives on curriculum-based measurement validity. In: Shinn, M.R. (ed.) *Advanced Applications of Curriculum-Based Measurement (The Guilford School Practitioner Series)*, pp. 61–88. Guilford Press, New York (1998)
23. Gebhardt, M., Heine, J.-H., Zeuch, N., Förster, N.: Lernverlaufsdagnostik im Mathematikunterricht der zweiten Klasse. Raschanalysen zur Adaptation eines Testverfahrens für den Einsatz in inklusiven Klassen. In: *Empirische Sonderpädagogik*, **7**(3), 206–222 (2015). http://www.psychologie-aktuell.com/fileadmin/download/esp/3-2015_20150904/esp_3-2015_206-222.pdf

24. Wayman, M.M., Wallace, T., Wiley, H.L., Tichá, R., Espin, C.A.: Literature synthesis on curriculum-based measurement in reading. *J. Spec. Educ.* **41**(2), 85–120 (2007). <https://doi.org/10.1177/00224669070410020401>
25. Stecker, P.M., Fuchs, L.S., Fuchs, D.: Using curriculum-based measurement to improve student achievement: review of research. *Psychol. Sch.* **42**(8), 795–819 (2005). <https://doi.org/10.1002/pits.20113>
26. Russell, M.K.: Technology-aided formative assessment of learning: new developments and applications. In: Andrade, H.L., Cizek, G.J. (eds.) *Handbook of Formative Assessment*, pp. 125–138. Routledge, New York, London (2010)
27. Klauer, K.J.: Lernverlaufsdagnostik - Konzepte, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik* **3**(3), 207–224 (2011)
28. Diehl, K., Hartke, B.: Zur Reliabilität und Validität des formativen Bewertungssystems IEL-1. Inventar zur Erfassung der Lesekompetenz von Erstklässlern. *Empirische Sonderpädagogik* **3**(2), 121–146 (2011). <http://nbn-resolving.de/urn:nbn:de:0111-opus-93201>
29. Förster, N., Souvignier, E.: Curriculum-based measurement: Developing a computer-based assessment instrument for monitoring student reading progress on multiple indicators. *Learn. Disabil. Contemp. J.* **9**(2), 21–44 (2011)
30. Gebhardt, M., Diehl, K., Mühling, A.: Online-Lernverlaufsmessung für alle Schülerinnen und Schüler in inklusiven Klassen. *Zeitschrift für Heilpädagogik* **66**, 444–453 (2016)
31. Kuhn, J.-T., Holling, H.: Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. *Adv. Cognit. Psychol.* **10**(2), 59–67 (2014). <https://doi.org/10.5709/acp-0157-2>
32. Salaschek, M., Souvignier, E.: Web-based progress monitoring in first grade mathematics. *Front. Learn. Res.* **2**, 53–69 (2013). <https://doi.org/10.14786/flr.v1i2.5>
33. Mühling, A., Gebhardt, M., Diehl, K.: Formative Diagnostik durch die Onlineplattform Levumi. *Informatik Spectrum* **40**(6), 556–561 (2017). <https://doi.org/10.1007/s00287-017-1069-7>
34. Jungjohann, J., Gegenfurtner, A., Gebhardt, M.: Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit. *Empirische Sonderpädagogik* **10**(1), 100–118 (2018)
35. Jungjohann, J., Gebhardt, M., Diehl, K., Mühling, A.: Förderansätze im Lesen mit LEVUMI (2017). <https://doi.org/10.17877/de290r-18042>
36. Deno, S.L., Mirkin, P.K., Chiang, B.: Identifying a valid measure of reading. *Except. Child.* **49**(1), 36–45 (1982)
37. Dummer-Smoch, L., Hackethal, R.: *Kieler Leseaufbau. Handbuch*, 9th edn. Veris, Kiel (2016)
38. Gebhardt, M., Diehl, K., Mühling, A.: *Lern-Verlaufs-Monitoring LEVUMI Lehrerhandbuch. Version 1.1.* (2016). <https://doi.org/10.17877/de290r-17792>
39. Wright, B.D., Linacre, J.M.: Reasonable mean-square fit values. *Rasch Meas. Trans.* **8**(3), 370 (1994) <https://www.rasch.org/rmt/rmt83b.htm>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





Using Theory-Based Test Construction to Develop a New Curriculum-Based Measurement for Sentence Reading Comprehension

Jana Jungjohann^{1*}, Jeffrey M. DeVries¹, Andreas Mühling² and Markus Gebhardt¹

¹ Faculty of Rehabilitation Science, Research in Inclusive Education, Technische Universität Dortmund, Dortmund, Germany,

² Department of Computer Science, Computer Science Education Research, Kiel University, Kiel, Germany

OPEN ACCESS

Edited by:

Mustafa Asil,
University of Otago, New Zealand

Reviewed by:

Lisa Zimmerman,
University of South Africa, South Africa
Haci Bayram Yilmaz,
Ondokuz Mayıs University, Turkey

*Correspondence:

Jana Jungjohann
jana.jungjohann@tu-dortmund.de

Specialty section:

This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

Received: 29 August 2018

Accepted: 10 December 2018

Published: 21 December 2018

Citation:

Jungjohann J, DeVries JM, Mühling A
and Gebhardt M (2018) Using
Theory-Based Test Construction to
Develop a New Curriculum-Based
Measurement for Sentence Reading
Comprehension. *Front. Educ.* 3:115.
doi: 10.3389/feduc.2018.00115

Reading comprehension at sentence level is a core component in the students' comprehension development, but there is a lack of comprehension assessments at the sentence level, which respect the theory of reading comprehension. In this article, a new web-based sentence-comprehension assessment for German primary school students is developed and evaluated using a curriculum-based measurement (CBM) framework. The test focuses on sentence level reading comprehension as an intermediary between word and text comprehension. The construction builds upon the theory of reading comprehension using CBM-Maze techniques. It is consistent on all tasks and contains different syntactic and semantic structures within the items. This paper presents the test development, a description of the item performance, an analysis of its factor structure, and tests of measurement invariance, and group comparisons (i.e., across gender, immigration background, over two measurement points, and the presence of special educational needs; SEN). Third grade students ($n = 761$) with and without SEN finished two CBM tests over 3 weeks. Results reveal that items had good technical adequacy, the constructed test is unidimensional, and it is valid for students both with and without SEN. Similarly, it is valid for both sexes, and results are valid across both measurement points. We discuss our method for creating a unidimensional test based on multiple item difficulties and make recommendations for future test construction.

Keywords: curriculum-based measurement, reading assessment, web-based assessment, sentence reading comprehension, special educational needs, primary school age, progress monitoring

INTRODUCTION

Reading acquisition, particularly reading comprehension, is one of the most important academic skills (Nash and Snowling, 2006); however, multiple student groups [i.e., students with special educational needs (SEN), or students with immigration background] have disadvantages to reach a high reading comprehension competency level. In particular, reading comprehension is a challenge for students with SEN resulting in problems in both primary and secondary school (Gersten et al., 2001; Cain et al., 2004; Berkeley et al., 2011; Cortiella and Horowitz, 2014; Spencer et al., 2014). Many of students with SEN have language difficulties that accompany a high risk for difficulties in literacy. In the United States and the United Kingdom, the vast majority of students with SEN struggles in this area (Kavale and Reese, 1992; NAEP, 2015; Lindsay and Strand, 2016). Similar

results were reported for learners speaking in phonologically consistent languages such as German (Gebhardt et al., 2015). Additionally, in most western countries, including Germany, students with an immigration background (i.e., they were born in another country, have at least one parent were born in another country, or are a non-citizen but born in the target country; Salentin, 2014) have greater academic challenges than native students (OECD, 2016). Consequently, they also tend to have a lower reading competency (Schnepf, 2007; OECD, 2014; Lenkeit et al., 2018) which partially related to a discrepancy in language skills for the host country language (Kristen et al., 2011). Recently, Lindsay and Strand's (2016) large scale and longitudinal study emphasized the importance of identifying children with reading problems and their individual needs as early as possible. Alarming, students with reading comprehension problems during primary school demonstrate reading difficulties well into adolescence (Landerl and Wimmer, 2008; Taylor, 2012). Additionally, it is widely agreed that there is a large need for adequate assessments to achieve positive outcomes for children with SEN (OECD, 2005; Hao and Johnson, 2013; Lindsay, 2016).

One possible assessment type is curriculum-based measurement (CBM), which involves short, time-limited, and frequent tests to visualize the learning progress of low achieving students (e.g., students with SEN; Deno, 1985, 2003). For instance, the CBM-Maze task is a reading comprehension task where students receive a complete text with multiple blanks. They fill in each blank with one word from a few choices (Fuchs et al., 1992). CBM-Maze was designed to monitor the growth of intermediate and secondary students' reading comprehension. More recent studies showed that CBM-Maze measures early language skills, such as sentence level comprehension and code-related skills rather than higher language skills, such as inference-making, text comprehension, and knowledge about text structures (Wayman et al., 2007; Graney et al., 2010; Muijselaar et al., 2017). Because CBM-Maze assesses earlier reading skills, it may be adapted for younger students, including low achieving students. In this paper, we develop a new assessment for reading comprehension at the sentence level for primary school students, in the lines of reading comprehension theory and CBM framework. Our goal is to create an assessment that considers different structures of language and is highly suitable for both researchers and practitioners.

Reading Comprehension

Reading comprehension is "the ultimate goal of reading" (Nation, 2011, p. 248); it is necessary for everyone to be successful in school and society. In general, reading comprehension is a system of cognitive skills and processes (Kendeou et al., 2014). Multiple underlying skills, such as rapid naming, phonological, and orthographic processing, fluency, vocabulary, and working memory, need to interlock to allow for good comprehension performance (e.g., Cain et al., 2004; Cain and Oakhill, 2007; Kendeou et al., 2012; Keenan and Meenan, 2014). According to the simple view of reading (Gough and Tunmer, 1986; Hoover and Gough, 1990), reading comprehension is defined as a product of code-decoding skills and linguistic comprehension. Furthermore, it is divided into three related levels: word,

sentence, and text. Comprehension at the word level is the lowest-tier ability. Readers visually identify a word as a single unit and compare it with their mental representation (Coltheart et al., 2001). It includes subskills such as phonological awareness, decoding, and written word recognition. The sentence represents the middle tier. When readers connect several elements of a sentence, such as words, and phrases, they frame a local representation at this level. Sentence comprehension builds a fundamental bridge between the lower (word) and upper (text) levels using syntactic parsing and semantic integration (Frazier, 1987; Ecalte et al., 2013). At the top tier is text comprehension. In order to understand connected texts, reading learners need to establish additional complex cognitive processes, such as inference-making, coherence-making, and background knowledge (Van Dijk and Kintsch, 1983). In contrast to the simple view of reading, the hierarchical construction-integration model of text comprehension (originally by Kintsch, 1998), is divided into two mental representations: textbase and situation (Kintsch and Rawson, 2011). In textbase representations, readers combine the low level reading skills on word and sentence levels (i.e., microstructure) to build a local, coherent representation of the macrostructure of a text. The situation model relates to text content (i.e., integration of further information and knowledge) and is comparable with the text level comprehension. Both the textbase and the situation model of a text are fundamentally connected (Perfetti, 1985). Without understanding single words or sentence meaning, no reader is able to make inferences, or coherences. Consequently, in both the simple view of reading and the construction-integration model, sentence comprehension is a critical skill for advanced text comprehension.

Processes of Sentence Comprehension

General syntactic and semantic comprehension processes influence the comprehension at the word and sentence level. Additionally, word recognition is linked with the syntactic and semantic analysis (Oakhill et al., 2003; Frisson et al., 2005). All these processes can be understood as parallel, modular, or dominated by one process (Taraban and McClelland, 1990; McRae et al., 1998; Kennison, 2009). Even though research results are not consistent, relevant factors about sentence comprehension are known. For instance, word classes guide syntactic parsing. Kennison (2009) found that verb information affected syntactical parsing in undergraduate and graduate students. In contrast, Cain et al. (2005) studied primary school students' ability to choose correct conjunctions in tasks similar to CBM-Maze tasks (specifically Cloze tasks). Results showed that filling in additive and adversative conjunctions is easier than temporal and causal ones for 8–10 years old. Thus, syntactical information influences the extraction of the sentence meaning, and semantic information influences the comprehension of individual words. More precisely, their results show that the part of speech causes different difficulty levels in the children's sentence comprehension. Therefore, reading comprehension assessments need to include these possible difficulties in the test structure to be able to measure these core skills. In line with earlier claims, these results confirm that reading comprehension assessments should reveal whether students are struggling in

lower- or upper-level subskills, and where the difficulties lie (Catts et al., 2003).

Reading Difficulties and Sentence Comprehension

Comprehension problems are very heterogeneous. They can be caused by lexical processes (e.g., phonological and semantic skills or visual word recognition), by the capacity of the working memory, or by higher text processes (see Nation, 2011). Students might develop word recognition difficulties, isolated comprehension difficulties, or combined problems in both areas (Leach et al., 2003; Nation, 2011; Catts et al., 2012; Kendeou et al., 2014). Some studies suggest that younger readers' problems can be mostly ascribed to poor word recognition skills (Vellutino et al., 2007; Tilstra et al., 2009). Accordingly, less than one percent of early primary students who perform well in decoding, and vocabulary show isolated comprehension problems (Spencer et al., 2014). However, poor readers also differ from good readers in the efficiency of related cognitive processes (Perfetti, 2007). At the sentence level, poor comprehenders use sentence content for word recognition, which is especially challenging in complex semantic and syntactic structures. West and Stanovich (1978) showed that sentence content supports word recognition processes of fourth graders more than the process of advanced adult readers. Consequently, all readers can develop greater difficulties if the sentence content is not congruent to the word meaning. Martohardjono et al. (2005) investigated the relevance of the syntax of single sentences in reading comprehension on bilingual students. Results revealed that the participants could not use the syntactic structure to support their word recognition. Relatedly, poor readers in third grade struggle to extract the meaning of syntactically complex sentences when they were presented verbally (Waltzman and Cairns, 2000). Besides the semantic and syntactic deficits, poor readers take longer to read complex sentences (Graesser et al., 1980; Chung-Fat-Yim et al., 2017). The more complex the sentence, the more time is needed to process the syntax and semantics of the sentence. Thus, poor readers need more cognitive resources to read and understand single sentences.

Furthermore, sentence level comprehension tests can measure general reading comprehension. Ecalte et al. (2013) examined the role of sentence processing as a mediatory skill within reading comprehension in second through ninth graders. First, they presented the students semantically similar sentences with different complexity and vocabulary. The students had to judge whether the contents were similar or not. Second, they examined the impact of these skills on expository text comprehension. The results confirmed that sentence processing increases over age and that sentence comprehension "could constitute a good indicator of a more general level of reading comprehension irrespective of the type of text" (Ecalte et al., 2013, p. 128).

Overall, these results show that even small changes (i.e., semantical and syntactic ones) within a sentence can influence students' comprehension. Especially important is that the comprehension ability of poor readers (i.e., students with SEN or with immigration background) is sensitive to both word meaning and sentence structure. Additionally, these results affirm the need for specific reading comprehension assessments at sentence level

in contrast to word recognition or fluency tests (Cutting and Scarborough, 2006).

Curriculum-Based Measurements

Curriculum-based measurement (CBM) is a problem-solving approach for assessing the learning growth of low achieving students (e.g., students with SEN) in basic academic skills, such as reading, writing, spelling, or mathematic competencies (Deno, 1985, 2003). The main idea of CBM is to monitor the children's development with short and very frequent tests during regular lessons. This allows teachers to graphically view the slope of the individual learning growth and to evaluate the effectiveness of the instruction. Teachers can then link the results with their decision-making and lesson planning.

Curriculum-Based Measurements as an Assessment of Reading Comprehension

A lot of research relating to CBM has been conducted (Fuchs, 2017), especially in reading tasks. In primary schools, two kinds of CBM instruments are ordinarily used to measure reading competencies: CBM-R and CBM-Maze (Graney et al., 2010). CBM-R involves individual students reading aloud from a word list or a connected text, and it measures oral reading fluency and accuracy (i.e., word recognition). CBM-Maze was developed to compensate for the disadvantages of CBM-R, namely individual administration and teacher distrust of CBM-R as a reading comprehension measurement. CBM-Maze is a group administered silent reading task that measures general text reading comprehension. However, both tests provide valid measurements of students' reading comprehension skills (Ardoin et al., 2004; Marcotte and Hintze, 2009). For both types of CBM, many studies report technical adequacy, strong alternate-form reliability, moderate to strong criterion-related validity, and predictive validity (e.g., Shin et al., 2000; Graney et al., 2010; Espin et al., 2012; Ardoin et al., 2013). Furthermore, correlations between CBM-R and CBM-Maze were found (Wayman et al., 2007). CBM-R is typically used for measurements within the first three grades and CBM-Maze for fourth and higher graders.

Principles of CBM-Maze

In traditional CBM-Maze tasks, students read a timed short passage (~250 words), in which different words are deleted. For each blank, one target word and two or more distractors are presented. The students choose one word for each blank. In general, the first and the last sentence of each passage are kept intact to allow the context to guide comprehension. The number of correctly filled blanks is then used as the competence measure for reading comprehension. The CBM-Maze task is based on the Cloze (Louthan, 1965; Gellert and Elbro, 2013) test, where the students fill out the blanks without any time limit or word suggestions. Since then, many studies examined CBM-Maze test construction and administration issues.

Test administration

Fuchs and Fuchs (1992) were among the first researchers to describe the CBM-Maze task. In contrast to the CBM-R, they

highlighted the higher classroom usability because the CBM-Maze can be administered in groups and at computers. Recently, Nelson et al. (2017) found that computer adaptive tests—as a practice of CBM—are valid for progress monitoring with fourth and fifth graders. They reported that with frequent data collection, computer testing systems can examine the overall learning growth of individuals and student groups. One main feature of CBM tests is that they can be administered multiple times, which requires both alternative test forms and sensitivity to student growth (Fuchs, 2004).

In their meta-analysis, García and Cain (2014) determined general reading comprehension assessment characteristics. They observed significant differences for the linguistic material and the administration procedure (i.e., reading aloud or silently, and test time). However, this was not upheld for CBM-Maze tasks because research has not found significant differences for primary students between reading silently or aloud (Hale et al., 2011). Accordingly, the CBM-Maze assessments are mostly administered silently for higher practicability in group settings. The CBM-Maze test time is usually short, from 1 to 10 min (e.g., Fuchs and Fuchs, 1992; Brown-Chidsey et al., 2003; Wiley and Deno, 2005). While no specific limit has been agreed upon, Brown-Chidsey et al. (2003) suggested that a test time of 10 min is too long for a 250-word passage.

Item construction

Traditionally, the items of CBM-Maze tasks are connected passages. Depending on the age of the students, different kinds of passages are used, such as fables (Förster and Souvignier, 2011), newspaper articles (Tichá et al., 2009), and historical texts (Brown-Chidsey et al., 2003). Outside the CBM approach, maze tasks are commonly used to measure semantic and syntactic skills within sentence processing (Forster, 2010). In these cases, single sentences are mostly used instead of connected passages (e.g., Witzel and Witzel, 2016). January and Ardoin (2012) examined differences in the construction of the CBM-Maze probes with third, fourth, and fifth graders. They examined whether the students' performance is contingent on the content of the passages. The findings indicated that primary school students performed well on both intact (i.e., sentences in order) and scrambled (i.e., sentences out of order) CBM-Maze passages. Additionally, they concluded that the CBM-Maze task measures reading comprehension at the sentence level because the students did not need the context to perform well. Taken together with the results of Ecalle et al. (2013), these results suggest that CBM-Maze could also be administered with single sentences instead of connected passages.

Furthermore, item construction depends on the deletion pattern and on the linguistic material. Some test designers use a fixed (i.e., every seventh word) or a lexical (e.g., deletion of nouns, verbs, adjectives, or conjunctions) deletion pattern, however, Kingston and Weaver (1970) showed that the lexical deletion pattern had similar results as fixed deletion. Similarly, January and Ardoin (2012) could not find significant differences in the students' accuracy based upon different lexical deletion patterns.

While not explicitly tested, results that indicated similar accuracy for different types of lexical deletion suggest

unidimensionality. This is important for teachers and researchers because items that fall on the same dimension are easier to interpret (Gustafsson and Åberg-Bengtsson, 2010). This does not mean that the underlying construct of reading comprehension is unidimensional, but that the results are interpretable along a single dimension of reading comprehension (Reise et al., 2013). Furthermore, individual test performance differs based upon factors such as SEN and immigration background (Cortiella and Horowitz, 2014; Spencer et al., 2014; OECD, 2016). Thus, in order for a test to be fair for all test takers, the linguistic material should be similar in construction and the items need to be appropriate for diverse groups of students, such as learners with SEN, those with an immigration background, and learners of both genders (i.e., measurement invariance; Good and Jefferson, 1998; Steinmetz, 2013).

Distractors

Other studies discussed the influence of distractors on a correct answer. Early studies indicated two types of distractors: semantically plausible with incorrect syntactic structure or semantically meaningless with correct syntactic structure (Guthrie et al., 1974; Gillingham and Garner, 1992). Resulting suggestions for distractors include a similar look as the target word, a lack of contextual sense, words with a related, incompatible contextual meaning, or nonsense words (Fuchs and Fuchs, 1992). McKenna and Miller (1980) found that syntactically correct distractors are more difficult for students to exclude in comparison to similar looking words. Meanwhile Conoyer et al. (2017) found that tests using content-based and part of the speech-based distractors were similar.

The Present Study

Because reading comprehension at the sentence level is a necessary skill (Ecalte et al., 2013), we developed a new web-based test to measure this competence. Our new assessment focuses on sentence reading ability within a CBM framework for primary school students (i.e., third graders). Our study details test development and analyzes item difficulty. It also assesses dimensionality with an analysis of the factorial structure and tests measurement invariance across several relevant groups. Additionally, we track the performance of our participants across two measurement points and examine the effect of subject variables including SEN, immigration background, and gender. Accordingly, we developed two groups of research questions. The first group of questions relates to technical evaluation of the test and the second group relates to the overall performance of our participants.

The first three questions relate to the technical aspects of test construction and interpretation including item difficulty, unidimensionality, and measurement invariance:

1. *What are the item difficulties and do they relate to different item types?* We use multiple word types to create different difficulty levels and we expect that some word types will be more difficult than others (see Cain et al., 2005; Frisson et al., 2005; Kennison, 2009).

2. *Can the instrument results be interpreted unidimensionally?* A unidimensional test structure would allow for easy test interpretation for both researchers and educators, because they only need to consider overall performance on the test. We use a consistent and straightforward sentence structure with age-related words to create a test structure that is applicable to both good and poor comprehenders (e.g., students with SEN). Thus, we hypothesize that all items fit on a unidimensional test structure because the item structure is consistent and all items represent the same underlying reading competence.
3. *Does the test possess measurement invariance relating to SEN, immigration background, gender, and measurement points?* Test construction followed guidelines for CBM test construction and evaluation (Fuchs, 2004; Wilbert and Linnemann, 2011). This includes multiple alternate test forms of equal difficulty, integration of several subskills (e.g., word recognition, syntactic parsing, and semantic integration). Because we adopt these established recommendations and combine them with CBM-Maze praxis (e.g., Brown-Chidsey et al., 2003; January and Ardoin, 2012; Conoyer et al., 2017), we hypothesize that our test will be invariant over student groups and measurement points.

The last two questions focus on the performance of the participants in relation to classroom and individual factors:

4. *What is the intraclass correlation?* Although the test was given to different classrooms, we expect the results to be comparable in each classroom. Therefore, we expect a relatively low intraclass correlation, meaning that the test performed similarly across all classrooms. For comparison, Hedges and Hedberg (2007) use the guideline of 0.05–0.15 in their large-scale assessment.
5. *Did performance improve over time, and did subject variables, including SEN, immigration background, and gender influence performance?* We compare the sum scores of our participants across two measurement points. We expect that there will be an improvement in performance from measurement point one to two and learners with SEN will perform worse (see Gebhardt et al., 2015; Lindsay and Strand, 2016).

METHODS

Test Administration

The new reading assessment is administered online via a German web-based platform for CBM monitoring, called Levumi (www.levumi.de). The code for the Levumi platform is published on Github, and all tests, test materials, and teacher handbooks will be published with a creative commons license. This means that this test is free of charge for teachers and researchers (Jungjohann et al., 2018). The platform runs on all major browsers and only requires an internet connection. It records each student's response and reaction time for every item.

The test can be administered in groups or individually. Each student has his or her individual student account where the activated tests become available. Teachers or researchers activate the test for each measurement point for the participants (e.g.,

students) in the test-taker's individual account. A computer or tablet is required for each simultaneous test-taker.

At the beginning of each test, a simple interactive example is shown. This prevents an accidental test start or other interface problems. After the example, students have 8 min to answer as many items as possible. On the screen, the items appear one after another. The students see a single sentence with a blank. Underneath the sentences, the target word and the distractors are displayed in a random order. When a student chooses one of the possible words, it appears in the blank. The students can change their minds by clicking on another response, and afterwards they confirm their answer by clicking on the "next" button. When the time limit runs out, the students can finish the current item, and then the test closes automatically.

At each measurement point, the item order is different, allowing for alternate test forms for frequent measurements over a school year. The items have a fixed item order for the first measurement (see **Table 1**), with items alternating between each different word-category (as described below). This fixed item order creates a baseline for comparison for the first measurement point. Random orders are generated for the second measurement point on to allow for a large number of alternative test forms. In these orders, a category is randomly selected, and then an item from each category, but the ratio of items from each category is kept equal.

Item and Distractor Construction

The overall process of item creation followed the CBM-Maze's principles; however, some modifications were made according to reading comprehension theory in order to create a test using sentences rather than connected passages. First, all items were carefully created as individual sentences. The entire pool of 60 items can be seen in **Table 1**. To ensure that every sentence is appropriate for third graders, all important words were collated from curricula within the German primary school systems (e.g., lists of frequent words based on grade level). Every sentence is a sentoid, meaning it is semantically explicit including the distractors. All sentences are constructed in the active voice and with age-appropriate syntactic structures (i.e., avoiding sentences with multiple clauses).

Second, a lexical deletion pattern was chosen to set different item difficulties within one test. Research results showed that the use of single word types can affect a different sentence comprehension. To adopt these results for the test construction, all items were classified by the lexical deletion pattern. The essential hypothesis is that difficulty is determined by the type of word deleted (i.e., part of speech). Because all items were set with a similar sentence structure, they relate to the same competence (i.e., reading comprehension at sentence level). To build up the variation of the German language, word types were summarized in multiple categories. This new assessment considers only possibilities relevant for third graders and not all possibilities within the German language. Therefore, three categories were set. The first word-category included nouns ($n = 20$) used as both subjects and objects. The second category included verbs and adjectives ($n = 21$). The third category included conjunctions and prepositions ($n = 19$).

TABLE 1 | Overview of the items*.

Item number	Item category	Item order	Item	Item difficulty (%)
1	1	1	A face has two eyes/ <i>fingers/books/cars</i> .	86
2	1	17	My Mum sleeps in a bed/ <i>picture/cage/table</i> .	94
3	1	38	Lasse draws nice pictures/ <i>bits/air/cold</i> .	88
4	1	28	A lama has four legs/ <i>thumbs/camels/books</i> .	88
5	1	7	My friend moves into a new house/ <i>shirt/exercise book/flowerbed</i> .	84
6	1	35	I tie my shoes/ <i>keys/chest/nature</i> .	84
7	1	58	Dad locks the front door/ <i>bottle/bridge/blackboard</i> in the evening.	88
8	1	42	Jutta goes shopping with her sister/ <i>shower/cottage/pain</i> .	87
9	1	60	My Dad works in a(n) office/ <i>July/ruler/fun</i> .	90
10	1	51	The buoy floats in the water/ <i>bed/lions/cloth</i> .	90
11	1	24	The bird/ <i>dog/club/father</i> flies to the nest.	91
12	1	48	The ducks/ <i>bears/houses/hair</i> quack in the lake.	87
13	1	4	The flowers/ <i>bids/boys/chairs</i> bloom in the field.	87
14	1	22	The friends/ <i>pens/shoes/lights</i> are up to no good.	82
15	1	57	The rabbit/ <i>sand/skirt/tooth</i> runs across the field.	91
16	1	6	The scissors/ <i>pizzas/onions/foreheads</i> cut paper.	92
17	1	14	The baby/ <i>radio/packet/puzzle</i> cries for its mother.	91
18	1	32	The sun/ <i>meadow/clock/doll</i> shines every day.	92
19	1	41	The frogs/ <i>mushrooms/fruits/teeth</i> jump across the street.	89
20	1	44	The bee/ <i>tree bark/nose/flower</i> sits on a blossom.	77
21	2	40	The lemonade is sweet/ <i>quiet/wealthy/sandy</i> .	89
22	2	21	An apple is a round/ <i>long/blue/warm</i> fruit.	91
23	2	54	A needle is sharp/ <i>guilty/lovely/loud</i> .	88
24	2	13	The police arrests the bad/ <i>square/flat/warm</i> robbers.	91
25	2	23	With my glasses, I am able to see well/ <i>fresh/loud/young</i> .	92
26	2	52	My Dad buys a new/ <i>round/cold/high</i> car.	91
27	2	2	My sister always studies hard/ <i>greenly/flatly/thinly</i> .	89
28	2	43	The fat/ <i>high/round/square</i> man shouts loudly.	81
29	2	26	Lisa tells me a funny/ <i>late/wet/deep</i> joke.	91
30	2	30	At dinnertime, I am often hungry/ <i>close/big/far</i> .	90
31	2	11	The fast/ <i>full/close/fresh</i> car races down the street.	87
32	2	36	Your friend bakes/ <i>builds/asks/studies</i> a large loaf of bread.	83
33	2	47	The girl eats/ <i>runs/ drives/rotates</i> the soup.	84
34	2	16	Lukas talks/ <i>opens/packs/bakes</i> with Frida about the holidays.	70
35	2	19	The sun shines/ <i>rains/melts/shouts</i> often in the summer.	82
36	2	50	You wait/ <i>meet/love/remove</i> out the thunderstorm.	57
37	2	34	My grandma sleeps/ <i>hits/rinses/glances</i> late on Saturdays.	82
38	2	56	Paula buys/ <i>fries/chews/races</i> a present for me.	83
39	2	5	The cake tastes/ <i>sniffs/drinks/chooses</i> very good to us.	80
40	2	29	In the forest, we collect/ <i>calculate/close/place</i> leaves.	89
41	2	10	I live/ <i>want/know/let</i> with my family.	85
42	3	31	Paul has neither/ <i>that/although/because</i> of a pen nor a notebook.	61
43	3	12	I have a good grade, but/ <i>that/because/or</i> Sina unfortunately doesn't.	82
44	3	39	If/ <i>Before/Instead/Than</i> water freezes, it turns into ice.	78
45	3	45	While/ <i>So/Except/Neither</i> my Mum paints the fence, I play in the garden.	74
46	3	53	Jutta doesn't like to eat fruit, except for/ <i>from which/so/because of</i> cherries.	70
47	3	20	After/ <i>As if/But rather/Neither</i> we have moved, Mum and Dad buy new furniture.	67
48	3	9	I brush my teeth, before/ <i>after/soon/from</i> I go to bed.	74
49	3	55	The gull is able to fly, because/ <i>before/except/from</i> it is a bird.	72
50	3	25	Instead of/ <i>Because/Right after/From</i> looking, I would rather ask my Mom.	62

(Continued)

TABLE 1 | Continued

Item number	Item category	Item order	Item	Item difficulty (%)
51	3	3	As soon as/ <i>But/And/Therefore</i> , I'm ready, I'll let you know.	73
52	3	37	You climb up/ <i>to/from/amid</i> the tree.	50
53	3	49	Lots of people sit in/ <i>through/on/out</i> the plaza.	81
54	3	59	The car pulls the other car only with/ <i>between/for/to</i> difficulty.	66
55	3	27	Through/ <i>Over/Inside/To</i> the telescope, I see a tower.	66
56	3	46	The present is for/ <i>on /under /next to</i> my sister.	77
57	3	15	My notebook is in <i>between/from/inside/out</i> of the pillows.	73
58	3	8	We can't visit during/ <i>behind/over/for</i> the bad weather	78
59	3	33	We play under/ <i>through/inside/from</i> the table.	85
60	3	18	This pen is from/ <i>under/over/in</i> my grandpa.	82
	1		Category 1 (Nouns)-average difficulty	88
	2		Category 2 (Adjective/verbs)-average difficulty	85
	3		Category 3 (Conjunction/prepositions)-average difficulty	72
			All Categories-Average Difficulty	82

Italicized words denote distractors.

*CC-BY-NC-SA. This work will be licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Third, three distractors were created for every target word. Every distractor was contextually meaningless but syntactically possible. Three rules guided the distractor creation: one distractor had a similar look, another distractor was related to the contextual sense of the target word or sentence but lacked the correct meaning, and the last distractor made no contextual sense. Because of vocabulary limitations, these rules could not be implemented precisely in each item. In these cases, words from the other rules were adapted. In every case, the same number of words were presented. The following example illustrates these rules:

(Item 4—German) Ein Lama hat vier Beine/*Bücher/Daumen/Kamele*.

(Item 4—English) A llama has four legs/*books/thumbs/camels*.

Participants

Participants were third grade students attending regular elementary schools in the northwest of Germany ($n = 761$). Approximately half of the participants were female (46.5%). The participants' teachers were asked about the immigration background ($n = 344$) and SEN. SEN were listed as learning ($n = 37$), cognitive development ($n = 2$), and other (i.e., speech and language impairments, emotional disturbed, or functional disability; $n = 67$). Participants with low German language competence ($n = 40$) were also categorized as SEN.

Procedure

Trained research assistants (i.e., university students) contacted local elementary school administrators and teachers to recruit participants. All data was collected with the informed consent of participants, parents, teachers, and administrators. The research assistants tested participants in small groups. Each participant worked individually on a single school computer. After the first measurement, researchers returned 3 weeks later to collect

the data for the second measurement. Some students were not available due to illness or other reasons at the second measurement ($n = 94$). In this case, their data for the second measurement was treated as missing, but their data from the first measurement was used where appropriate.

During both measurements, research assistants followed the same scripted procedure. Children were told that the little dragon Levumi has brought many sentences with it, but that each of the sentences were missing a word. The child was asked if he or she could find the correct word in each sentence. Next, the participants were given an example item. Once they gave the correct answer for the example item, the research assistant showed the participant how to give this answer on the test with the mouse. After the example item, participants answered as many items as they could in an 8 min period.

Analysis

Item Difficulties

Average item difficulty was obtained by averaging the percentage of correct responses across both measurement points. Sufficient numbers of participants completed all test items ($n = 87$, 11.4%) to allow for a power-analysis, and a further 25% of participants completed the vast majority of the test (50 items). Therefore, missing scores were treated as not yet reached in our analyses. A repeated measures one-way ANOVA compared the average difficulties based on word-category (noun, adjective/verb, and conjunction/preposition).

Dimensionality

To assess dimensionality of the instrument, the factor structure was tested via three separate confirmatory factor analyses (CFAs). All factor analyses were conducted in Mplus 7.4 (Muthén and Muthén, 1998–2015) using a weighted least squares with mean and variance adjusted (WLSMV) estimation method. The WLSMV estimator is more appropriate for categorical data

(Muthén et al., 1997; Flora and Curran, 2004). Factor structures were based on the word type category and function in the sentence. The 3-factor model mirrored the three word-categories: nouns, verbs/adjectives, and conjunctions/prepositions. In the 2-factor model, the verbs/adjectives, and conjunctions/prepositions factors were combined. Finally, in the 1-factor structure, all items were placed on the same factor. We appraised the model fits via root mean squared error of approximation (RMSEA), CFI, and gamma-hat. We considered RMSEA < 0.08, CFI > 0.90, and gamma-hat > 0.90 acceptable fits. Meanwhile, we considered RMSEA < 0.05, CFI > 0.95, and gamma-hat > 0.95 good fits (Hu and Bentler, 1998).

Next, we compared the fits of the separate models in order of increasing complexity. We compared the 2-factor model to the 1-factor model and the 3-factor model to the 2-factor model. We examined changes in CFI (Δ CFI) and gamma-hat (Δ Gamma-hat). We set a threshold of 0.01 for Δ CFI and Δ Gamma-hat as a significantly better fit (Cheung and Rensvold, 2002; Dimitrov, 2010).

Measurement Invariance

We examined the measurement invariance of each of the three models based on presence of SEN, immigration background, gender, and measurement point. We constrained or freed thresholds and lambda together. In other words, we tested the scalar (strong) model directly against the configural (base) model, as recommended when using WLSMV analysis by Muthén and Muthén (1998–2015). As described above, differences in Δ CFI and Δ Gamma-hat greater than 0.01 were considered significant.

Intraclass Correlation

Next, the intraclass correlation was calculated using the proportion of variance explained by classroom compared to overall variance. Values were calculated based on the sum of squares in a one way ANOVA of class on average percent correct at the first measurement point.

Change Over Time and the Influence of Subject Variables

Finally, to assess the differential performance on the test by our target groups, we conducted a repeated measures ANOVA including SEN, immigration background, and gender across both measurement points on the sum score of each participant.

RESULTS

Item Difficulties

Table 1 lists all items and item difficulties across both measurement points. ANOVA results confirmed that difficulty varied across word-categories, $F_{(2,57)} = 25.215$, $p < 0.001$. Tukey's honestly significant difference test revealed that items in the easier two categories (noun and adjective/verb) were similar in difficulty, $p > 0.05$, meanwhile items in the conjunction/preposition group were significantly harder than the other two groups, $p < 0.05$.

TABLE 2 | Model Fits.

	RMSEA (90% CI)	CFI	Gamma-hat	Δ CFI	Δ Gamma-hat
1-Factor	0.013 (0.009–0.016)	0.988	0.991	–	–
2-Factor	0.012 (0.007–0.015)	0.991	0.992	0.003	0.001
3-Factor	0.011 (0.006–0.015)	0.991	0.993	0.000	0.001

Δ CFI and Δ Gamma-hat represent the change in model fit from the less complex to the more complex model.

Dimensionality

Fit metrics for all three models surpassed our criteria for good fits, as described in Table 2. Fit metrics were only slightly worse in the 2-factor model than in the 1-factor model, and virtually identical between the 2-factor and 3-factor model. None of the model comparisons exceeded the critical value of Δ CFI or Δ Gamma-hat of 0.01. Therefore, we conclude that all models fit equally well, and on the grounds of parsimony, we prefer the simpler 1-factor model. Thus, the instrument can be considered unidimensional.

Measurement Invariance

Measurement invariance test results are shown in Table 3. In each case, Δ CFI and Δ Gamma-hat are below the threshold of 0.01, meaning that the scalar model fit similar to the metric model. Therefore, we conclude that all three models possessed strong measurement invariance across presence of SEN, gender, immigration background, and measurement point. Because invariance was upheld for all models, the simpler 1-factor model is still preferable to other models. Thus, a unidimensional interpretation is equally valid for all subgroups within our data.

Intraclass Correlation

The intraclass correlation coefficient, as measured by proportion of total variance, indicated the test functioned similarly across all classrooms in our data, ICC = 0.15. This is relatively high, but still in the guidelines used in previous work (see Hedges and Hedberg, 2007).

Change Over Time and the Influence of Subject Variables

Table 4 shows the results of the sum score analysis. The repeated measures ANOVA revealed that students performed better on the test at measurement point 2, $F_{(1,658)} = 93.32$, $p < 0.001$. Additionally, learners with SEN performed worse overall than those without, $F_{(1,658)} = 89.01$, $p < 0.001$. Furthermore, a significant interaction indicated that learners with SEN did not improve from measurement point 1 to measurement point 2, $F_{(1,658)} = 5.45$, $p < 0.05$. No other interactions or main effects were found, all $ps > 0.05$.

DISCUSSION

Overview of Findings and Theoretical Implication

This study developed and evaluated a new theory-based formative assessment that measures reading comprehension at sentence level and that follows the CBM approach for

TABLE 3 | Measurement Invariance for 1-, 2-, and 3-factors.

Grouping	Model	RMSEA	CFI	Gamma-hat	Δ CFI	Δ Gamma-hat
Special education needs	1-Factor					
	Configural	0.012	0.982	0.984	–	–
	Scalar	0.012	0.982	0.984	0.000	0.000
	2-Factor					
	Configural	0.012	0.984	0.984	–	–
	Scalar	0.012	0.984	0.984	0.000	0.000
	3-Factor					
	Configural	0.011	0.985	0.987	–	–
	Scalar	0.011	0.986	0.987	–0.001	0.000
Immigration background	1-Factor					
	Configural	0.012	0.989	0.984	–	–
	Scalar	0.012	0.988	0.984	0.001	0.000
	2-Factor					
	Configural	0.011	0.990	0.987	–	–
	Scalar	0.011	0.990	0.987	0.000	0.000
	3-Factor					
	Configural	0.011	0.991	0.987	–	–
	Scalar	0.011	0.990	0.987	0.001	0.000
Gender	1-Factor					
	Configural	0.013	0.981	0.981	–	–
	Scalar	0.012	0.987	0.984	–0.006	–0.003
	2-Factor					
	Configural	0.012	0.988	0.984	–	–
	Scalar	0.011	0.989	0.987	–0.001	–0.003
	3-Factor					
	Configural	0.012	0.989	0.985	–	–
	Scalar	0.011	0.989	0.987	0.000	–0.002
Measurement point	1-Factor					
	Configural	0.010	0.990	0.989	–	–
	Scalar	0.011	0.989	0.986	0.001	0.003
	2-Factor					
	Configural	0.009	0.992	0.910	–	–
	Scalar	0.010	0.991	0.989	0.001	0.001
	3-Factor					
	Configural	0.009	0.992	0.991	–	–
	Scalar	0.010	0.991	0.989	0.001	0.002

Δ CFI and Δ Gamma-hat denote the difference between the configural and scalar models.

practical use in inclusive primary schools. The main goal was to create a unidimensional test structure with different item difficulties to allow for easy interpretation and a high usability for heterogeneous classrooms. Within our theory-based test construction, we linked common reading comprehension models at the sentence level with the principles of the CBM-Maze task. In addition, guidelines were set to assure that all finalized items from the same word-category were equivalent in construction and difficulty.

In general, the evaluation of the test construction revealed a 1-factor model with items of varying difficulty. Our results indicated significant differences between the three deletion pattern categories (e.g., word-categories) of the single items. In this study, the German third graders had fewer problems identifying correct target words for nouns (e.g.,

category 1), verbs, or adverbs (e.g., category 2) compared to conjunctions, or prepositions (e.g., category 3). These results are in line with previous results from Kennison (2009) and Cain et al. (2005), which indicated that different word types affect the syntactical parsing in different ways. Our results showed that these previous results could be generalized to the German language. Additionally, the different item difficulty between the word-categories can assist teachers to precisely screen problems in reading development. Förster and Souvignier (2011) argued that precise identification is an important feature of CBM assessments. Furthermore, the CFA demonstrated a unidimensional test structure, which allows for a simple interpretation from educators. Overall, our theory-based test construction demonstrated both adherence to reading comprehension theory and technical adequacy, making it useful

TABLE 4 | Comparison of sum scores.

	Measurement point 1		Measurement point 2		Significance
	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)	<i>p</i> -value
Special educational need					0.001***
Yes	21.4	(13.1)	21.2	(11.9)	–
No	34.4	(13.2)	39.9	(13.4)	–
Immigration background					0.083
Yes	30.0	(13.8)	34.2	(14.6)	–
No	34.0	(14.0)	38.8	(14.8)	–
Gender					0.171
Male	31.8	(14.2)	36.3	(15.4)	–
Female	32.8	(13.8)	37.3	(14.3)	–
All groups	32.2	(14.0)	36.8	(14.9)	0.001***

The significance column denotes the between subjects results of the ANOVA for the special education needs, immigration background, and gender rows, but for the all groups row, it denotes the within subjects variable of measurement point.

***Significant at $p < 0.001$.

to both teachers and researchers. Additionally, these results indicated that it is possible to set item difficulty by language related rules without creating several test types (i.e., letter, sentence, and text-based tests).

The assessment's practicability for inclusive classrooms was verified by the results of the measurement invariance tests, the intraclass correlation, and the analysis of the sum scores in three key ways. First, the intraclass correlation showed that our test performed similarly within different inclusive classroom settings. Additionally, the sum score comparison showed that changes in the students' ability were detectable without any changes in the test administration for a specific student group. This indicates that no specific class or student characteristics are required for test administration. Second, the alternative test forms were invariant across both measurement points. Meaning, that the random drawing order created good multiple alternative test forms, which can ease the test handling for both researchers and teachers. Within big classrooms, teachers do not have to track which student completed which test form or remember test dates manually, meaning that Levumi can reduce teachers' workload in multiple ways. Similarly, researchers can create a large number of alternative test forms. Thus, our web-based Levumi platform demonstrates one of the key benefits of computer-based formative assessment (see Russell, 2010). Third, especially in inclusive classrooms, the students are heterogeneous in their academic performance (Gebhardt et al., 2015). While students with SEN performed significantly lower, our Levumi reading comprehension test was invariant for different student groups (i.e., SEN, gender, and immigration background). Meaning, that teachers can use the Levumi test for all these students because the test assesses competence fairly across these groups. Furthermore, students with and without SEN can use the same test system over multiple measurement points in inclusive classrooms. Again, this reduces the teachers'

workload because teachers do not have to use other materials for special student groups within one classroom (Jungjohann et al., 2018).

Our Levumi reading comprehension test includes the test administration benefits of the CBM-Maze, such as group administration, and silent reading (see Graney et al., 2010). Additionally, it is suitable for early readers and for readers with low reading abilities (e.g., students with SEN). In particular, our test uses a sentence-based item pool, rather than a complete text-based item pool. Fuchs et al. (2004) and Good et al. (2001) argued that complete text-based tests can cause floor effects on low performance readers. Because of this, the Levumi test may be more suitable for even younger students and those who might have difficulties with complete texts. Correspondingly, these test characteristics demonstrate that it is possible to expand the established CBM assessment types in new ways. Additionally, we expand the existing techniques of evaluating CBM assessments with intraclass correlations, factor analyses, and measurement invariance tests. These three evaluation techniques are well established on other fields of test evaluation, and their use can help to rigorously evaluate existing and future CBM assessments. This demonstrated technique of theory-based test construction and evaluation provides an essential template for other researchers who may be developing a diverse range of CBM assessments.

Limitations and Future Work

Nonetheless, this study has some limitations. The findings still need to be replicated with a more varied participant pool and with a larger sample from other regions inside and outside of Germany. This actual study also focused on third graders, but reading difficulties can appear earlier in first and second grade, when students start to develop an understanding of written words and sentences (Richter et al., 2013). Therefore, further studies should also include first and second graders to expand adequate CBM assessment for this age.

Besides a broader participant pool, future longitudinal research is necessary to validate our hypotheses. One main CBM characteristic is the ability to track the students' learning growth across multiple measurement points over a long period (e.g., one school year; Deno, 2003). In our study, we confirmed that our test is invariant over two measurement points for a period of 3 weeks. For classroom use, it is necessary to analyze the test's ability to measure the students learning slope over a larger period with more than two measurement points per student.

Additionally, we have not established a concrete indicator of criterion validity yet (see Fuchs, 2004). Besides the Levumi reading comprehension test, participants could complete additional CBM assessments with a complete text-based item pool, and other established reading comprehension screenings. This would establish if the Levumi reading comprehension test relies more on to code-related skills (e.g., reading fluency) than on language related skills (e.g., reading comprehension), as suggested by Muijselaar et al. (2017). This can indicate which reading problems our test is effective at identifying. This is particularly important because established reading

comprehension tests do not agree with each other in the identification of reading problems (Keenan and Meenan, 2014).

Our last key limitation relates to the item language. Our research is limited by only using the German items. Both, the English item translation and the general theory-based test construction need to be evaluated in additional languages. At first, studies should evaluate the translated items with native speakers to test the quality of the items. Results of these studies would confirm the usability of our theory-based guidelines for CBM test construction in other languages. The original items can be translated into additional languages based upon these studies. This procedure will expand CBM offerings into new languages and regions.

Further work should also focus on instructional utility. This study followed Fuchs's (2004) recommendations for CBM test construction and examined the technical adequacy for formative learning growth monitoring (e.g., stage 1 and 2; Fuchs, 2004). Instructional utility means that teachers can include the CBM test system in their actual lessons, that they can understand, and interpret the results, and that they can link the students' learning slopes with their reading instructions. In Germany, the CBM approach is still unknown by many teachers. Therefore, the three main aspects of the instructional utility (Fuchs, 2004) are also concerns for the Levumi reading comprehension test and should be investigated in further research. First, the acceptance and the application of the Levumi platform needs to be evaluated within the school context. For the practical use of the Levumi platform, teachers need access to a computer, or a tablet and an internet connection. Some German schools already have good technical equipment, while others do not. Even assuming access to good technical equipment, teachers must be willing to use a web- and computer-based assessment. To that end, a clear user interface and teacher-focused supporting material will encourage user adoption. Second, further studies should examine how the teachers can handle the Levumi test results. Recently, studies revealed that preservice and in-service teachers can have problems in understanding CBM graphs (Van den Bosch et al., 2017; Zeuch et al., 2017). As one example, preservice teachers estimate fictitious future student achievements lower than can be expected by a linear regression model (Klapproth, 2018). In all these studies, the participants were not able to adjust the layout, visualize additional information from the tests (e.g., correct, and incorrect answers), or receive statistical help (e.g., trend line, or goal line). Consequently, future studies need to test these interpretation difficulties in order to adapt the specific Levumi output (i.e., CBM graphs, and further information). Third, especially, the web-based test system brings a high potential for automatized support in CBM graph interpretation and instruction making. For instance, the Levumi platform could highlight at risk students, automatize, or add additional information into the graph, such as statistical trend lines, instruction phases, or students' moods. Furthermore, the Levumi platform could learn typical problem patterns and suggest relevant instruction materials. Lastly, work needs to be done to identify the connection between specific

aspects of reading competency and performance on individual items and overall results. All these possibilities should be implemented carefully so that teachers are not confused by CBM data.

CONCLUSION

We created a new CBM assessment using theory-based test construction and evaluation. Evaluations indicated the test to be of high use for further research and praxis. This provides three key implications. First, researchers can adapt our approach as guidelines for further CBM assessments (i.e., further language, learning domains) to enhance the CBM research and evaluation field in new ways. Second, the web-based Levumi platform and the reading comprehension assessment are suitable for inclusive classrooms and their use can reduce the teachers' workload in multiple ways. And third, our procedure demonstrates the development of a test with multiple item difficulties which can be interpreted along a single dimension.

ETHICS STATEMENT

Permission for this study was granted through dean of the Faculty of Rehabilitation Science, Technical University of Dortmund. Following the requirements of the ministry of education of the federal state North Rhine-Westphalia (Schulgesetz für das Land Nordrhein-Westfalen, 2018), school administrators decided in co-ordination with their teachers about participation in this scientific study. An additional ethics approval was not required for this study as per Institution's guidelines and national regulations. Parents obtained written information about the study and any potential benefits. They gave their written consent for each child. Participation was supervised by school staff. Participation was voluntary and participants were free to withdraw at any time.

AUTHOR CONTRIBUTIONS

JJ developed the Levumi reading comprehension assessment and coordinated the study. As the primary author, JJ did most of the writing, research, and some of the analyses. JD did most of the data analyses. AM programmed the Levumi platform, realized the Levumi reading comprehension assessment based on the test specification of JJ, and edited the manuscript. MG gave the initial study design, supervised the entire research process, and edited the manuscript.

ACKNOWLEDGMENTS

We acknowledge financial support by Deutsche Forschungsgemeinschaft and Technische Universität Dortmund/TU Dortmund Technical University within the funding programme Open Access Publishing.

REFERENCES

- Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., and 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. *J. Sch. Psychol.* 51, 1–18. doi: 10.1016/j.jsp.2012.09.004
- Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psych. Rev.* 33, 218–233.
- Berkeley, S., Mastropieri, M. A., and Scruggs, T. E. (2011). Reading comprehension strategy instruction and attribution retraining for secondary students with learning and other mild disabilities. *J. Learn. Disabil.* 44, 18–32. doi: 10.1177/0022219410371677
- Brown-Chidsey, R., Davis, L., and Maya, C. (2003). Sources of variance in curriculum-based measures of silent reading. *Psychol. Sch.* 40, 363–377. doi: 10.1002/pits.10095
- Cain, K., and Oakhill, J. (2007). “Reading comprehension difficulties: Correlates, causes, and consequences,” in *Children’s Comprehension Problems in Oral and Written Language*, eds K. Cain and J. Oakhill (London: Guilford Press), 41–75.
- Cain, K., Oakhill, J., and Bryant, P. (2004). Children’s reading comprehension ability. Concurrent prediction by working memory, verbal ability, and component skills. *J. Educ. Psychol.* 96, 31–42. doi: 10.1037/0022-0663.96.1.31
- Cain, K., Patson, N., and Andrews, L. (2005). Age- and ability-related differences in young readers’ use of conjunctions. *J. Child. Lang.* 32, 877–892. doi: 10.1017/S0305000905007014
- Catts, H. W., Compton, D., Tomblin, J. B., and Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *J. Educ. Psychol.* 104, 166–181. doi: 10.1037/a0025323
- Catts, H. W., Hogan, T. P., and Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *J. Learn. Disabil.* 36, 151–164. doi: 10.1177/002221940303600208
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equation Model.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5
- Chung-Fat-Yim, A., Peterson, J. B., and Mar, R. A. (2017). Validating self-paced sentence-by-sentence reading. Story comprehension, recall, and narrative transportation. *Read. Writ.* 30, 857–869. doi: 10.1007/s11145-016-9704-2
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC. A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295X.108.1.204
- Conoyer, S. J., Lembke, E. S., Hosp, J. L., Espin, C. A., Hosp, M. K., and Poch, A. L. (2017). Getting more from your maze. Examining differences in distractors. *Read. Writ. Q.* 33, 141–154. doi: 10.1080/10573569.2016.1142913
- Cortiella, C., and Horowitz, S. H. (2014). *The State of Learning Disabilities: Facts, Trends and Emerging Issues*. New York, NY: National Center for Learning Disabilities.
- Cutting, L. E., and Scarborough, H. S. (2006). Prediction of reading comprehension. relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Sci. Stud. Reading* 10, 277–299. doi: 10.1207/s1532799xssr1003_5
- Deno, S. L. (1985). Curriculum-based measurement: the emerging alternative. *Exc. Child.* 52, 219–232. doi: 10.1177/001440298505200303
- Deno, S. L. (2003). Developments in curriculum-based measurement. *J. Spec. Educ.* 37, 184–192. doi: 10.1177/00224669030370030801
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Meas. Eval. Counsel. Dev.* 43, 121–149. doi: 10.1177/0748175610373459
- Ecalle, J., Bouchafa, H., Potocki, A., and Magnan, A. (2013). Comprehension of written sentences as a core component of children’s reading comprehension. *J. Res. Read.* 36, 117–131. doi: 10.1111/j.1467-9817.2011.01491.x
- Espin, C., Rose, S., McMaster, K., and Wayman, M. (2012). *A Measure of Success: The Influence of Curriculum Based Measurement on Education*. Minneapolis, MN: University of Minnesota Press.
- Flora, D. B., and Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods* 9, 466–491. doi: 10.1037/1082-989X.9.4.466
- Forster, K. (2010). Using a maze task to track lexical and sentence processing. *Mental Lexicon* 5, 347–357. doi: 10.1075/ml.5.3.05for
- Förster, N., and Souvignier, E. (2011). Curriculum-based measurement. developing a computer-based assessment instrument for monitoring student reading progress on multiple indicators. *Learn. Disabil.* 9, 65–88.
- Frazier, L. (1987). “Sentence processing: A tutorial review,” in *Attention and Performance*, ed M. Coltheart (Hove: Erlbaum), 559–586.
- Frisson, S., Rayner, K., and Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *J. Exp. Psychol. Learn. Memory Cogn.* 31, 862–877. doi: 10.1037/0278-7393.31.5.862
- Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychol. Rev.* 2, 188–192.
- Fuchs, L. S. (2017). curriculum-based measurement as the emerging alternative. Three decades later. *Learn. Disabil. Res. Pract.* 32, 5–7. doi: 10.1111/ldrp.12127
- Fuchs, L. S., and Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psych. Rev.* 21, 45–58.
- Fuchs, L. S., Fuchs, D., and Compton, D. L. (2004). Monitoring early reading development in first grade. Word identification fluency versus nonsense word fluency. *Exc. Child.* 71, 7–21. doi: 10.1177/001440290407100101
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., and Ferguson, C. (1992). Effects of expert system consultation within curriculumbased measurement using a reading maze task. *Except. Child.* 58, 436–450. doi: 10.1177/001440299205800507
- García, J. R., and Cain, K. (2014). Decoding and reading comprehension. *Rev. Educ. Res.* 84, 74–111. doi: 10.3102/0034654313499616
- Gebhardt, M., Sälzer, C., Mang, J., Müller, K., and Prenzel, M. (2015). Performance of students with special educational needs in germany. findings from programme for international student assessment 2012. *J. Cogn. Educ. Psychol.* 14, 343–356. doi: 10.1891/1945-8959.14.3.343
- Gellert, A. S., and Elbro, C. (2013). Do experimental measures of word learning predict vocabulary development over time? A study of children from grade 3 to 4. *Learn. Individ. Differ.* 26, 1–8. doi: 10.1016/j.lindif.2013.04.006
- Gersten, R., Fuchs, L. S., Williams, J. P., and Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities. A review of research. *Rev. Educ. Res.* 71, 279–320. doi: 10.3102/00346543071002279
- Gillingham, M. G., and Garner, R. (1992). Readers’ comprehension of mazes embedded in expository texts. *J. Educ. Res.* 85, 234–241. doi: 10.1080/00220671.1992.9941121
- Good, R. H., and Jefferson, G. (1998). “Contemporary perspectives on curriculum-based measurement validity,” in *Advanced Applications of Curriculum-Based Measurement*, ed M. R. Shinn (New York, NY: Guilford Press), 61–88.
- Good, R. H., Simmons, D. C., and 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. *Sci. Stud. Read.* 5, 257–288. doi: 10.1207/S1532799XSSR0503_4
- Gough, P. B., and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remed. Spec. Educ.* 7, 6–10. doi: 10.1177/074193258600700104
- Graesser, A. C., Hoffman, N. L., and Clark, L. F. (1980). Structural components of reading time. *J. Verbal Lear. Verbal Behav.* 19, 135–151. doi: 10.1016/S0022-5371(80)90132-2
- Graney, S. B., Martínez, R. S., Missall, K. N., and Aricak, O. T. (2010). Universal screening of reading in late elementary school. *Remed. Spec. Educ.* 31, 368–377. doi: 10.1177/0741932509338371
- Gustafsson, J.-E., and Åberg-Bengtsson, L. (2010). “Unidimensionality and interpretability of psychological instruments,” in *Measuring Psychological Constructs: Advances in Model-Based Approaches*, ed S. E. Embretson (Washington, DC: American Psychological Association), 97–121.
- Guthrie, J. T., Seifert, M., Burnham, N. A., and Caplan, R. I. (1974). The maze technique to assess, monitor reading comprehension. *Read. Teach.* 28, 161–168.
- Hale, A. D., Hawkins, R. O., Sheeley, W., Reynolds, J. R., Jenkins, S., Schmitt, A. J., et al. (2011). An investigation of silent versus aloud reading comprehension of elementary students using Maze assessment procedures. *Psychol. Schools* 48, 4–13. doi: 10.1002/pits.20543
- Hao, S., and Johnson, R. L. (2013). Teachers’ classroom assessment practices and fourth-graders’ reading literacy achievements. An international study. *Teach. Educ.* 29, 53–63. doi: 10.1016/j.tate.2012.08.010

- Hedges, L. V., and Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educ. Eval. Policy Anal.* 29, 60–87. doi: 10.3102/0162373707299706
- Hoover, W. A., and Gough, P. B. (1990). The simple view of reading. *Read. Writ.* 2, 127–160. doi: 10.1007/BF00401799
- Hu, L.-T., and Bentler, P. M. (1998). Fit indices in covariance structure modeling. Sensitivity to underparameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037/1082-989X.3.4.424
- January, S.-A. A., and Ardoin, S. P. (2012). The impact of context and word type on students' maze task accuracy. *School Psych. Rev.* 41, 262–271.
- Jungjohann, J., DeVries, J. M., Gebhardt, M., and Mühling, A. (2018). "Levumi: A Web-Based Curriculum-Based Measurement to Monitor Learning Progress in Inclusive Classrooms," in *Computers Helping People with Special Needs*. 16th International Conference, eds K. Miesenberger, G. Kouroupetroglou, and P. Penaz (Wiesbaden: Springer), 369–378.
- Kavale, K. A., and Reese, J. H. (1992). The character of learning disabilities. An Iowa profile. *Learn. Disabil. Q.* 15, 74–94. doi: 10.2307/1511010
- Keenan, J. M., and Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *J. Learn. Disabil.* 47, 125–135. doi: 10.1177/0022219412439326
- Kendeou, P., Papadopoulos, T. C., and Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learn. Instr.* 22, 354–367. doi: 10.1016/j.learninstruc.2012.02.001
- Kendeou, P., van den Broek, P., Helder, A., and Karlsson, J. (2014). A cognitive view of reading comprehension. Implications for reading difficulties. *Learn. Disabil. Res. Prac.* 29, 10–16. doi: 10.1111/ldrp.12025
- Kennison, S. M. (2009). The use of verb information in parsing. Different statistical analyses lead to contradictory conclusions. *J. Psychol. Res.* 38, 363–378. doi: 10.1007/s10936-008-9096-9
- Kingston, A. J., and Weaver, W. W. (1970). Feasibility of Cloze techniques for teaching and evaluating culturally disadvantaged beginning readers. *J. Soc. Psychol.* 82, 205–214. doi: 10.1080/00224545.1970.9919952
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. New York, NY: Cambridge University Press.
- Kintsch, W., and Rawson, K. A. (2011). "Comprehension," in *The Science of Reading: A Handbook*, ed M. Snowling and C. Hulme (Malden, MA: Blackwell Publ.), 209–226.
- Klapproth, F. (2018). Biased predictions of students' future achievement. An experimental study on pre-service teachers' interpretation of curriculum-based measurement graphs. *Stud. Educ. Eval.* 59, 67–75. doi: 10.1016/j.stueduc.2018.03.004
- Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., et al. (2011). The education of migrants and their children across the life course. *Zeitschrift Für Erziehungswissenschaft* 14, 121–137. doi: 10.1007/s11618-011-0194-3
- Landerl, K., and Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography. An 8-year follow-up. *J. Educ. Psychol.* 100, 150–161. doi: 10.1037/0022-0663.100.1.150
- Leach, J. M., Scarborough, H. S., and Rescorla, L. (2003). Late-emerging reading disabilities. *J. Educ. Psychol.* 95, 211–224. doi: 10.1037/0022-0663.95.2.211
- Lenkeit, J., Schwippert, K., and Knigge, M. (2018). Configurations of multiple disparities in reading performance: longitudinal observations across France, Germany, Sweden and the United Kingdom. *Assess. Educ.* 25, 52–86. doi: 10.1080/0969594X.2017.1309352
- Lindsay, G. (2016). Grand challenge: priorities for research in special educational needs. *Front. Educ.* 1:1. doi: 10.3389/feeduc.2016.00001
- Lindsay, G., and Strand, S. (2016). Children with language impairment. prevalence, associated difficulties, and ethnic disproportionality in an english population. *Front. Educ.* 1:2. doi: 10.3389/feeduc.2016.00002
- Louthan, V. (1965). Some systematic grammatical deletions and their effects on reading comprehension. *English J.* 54, 295–299. doi: 10.2307/811113
- Marcotte, A. M., and Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *J. Sch. Psychol.* 47, 315–335. doi: 10.1016/j.jsp.2009.04.003
- Martohardjono, G., Otheguy, R., Gabriele, A., DeGoeas-Malone, M., Rivero, S., Pyrzanowski, S. et al. (2005). "The role of syntax in reading comprehension: a study of bilingual readers," in *Proceedings of the 4th International Symposium on Bilingualism* (Somerville, MA: Cascadia Press), 1522–1544. Available online at: <http://www.lingref.com/isb/4/119ISB4.PDF>
- McKenna, M. C., and Miller, J. W. (1980). "The Effects of Age and Distractors Type on Maze Performance," in *Perspectives on Reading Research and Instruction: 29th Yearbook of the National Reading Conference*, ed K. L. Kamil (Washington, DC: National Reading Conference), 288–292.
- McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling thematic fit (and other constraints) within an integration competition framework. *J. Mem. Lang.* 38, 283–312. doi: 10.1006/jmla.1997.2543
- Muijselaar, M. M. L., Kendeou, P., Jong, P. F., and van den Broek, P. W. (2017). What does the CBM-maze test measure? *Sci. Stud. Read.* 21, 120–132. doi: 10.1080/10888438.2016.1263994
- Muthén, B. O., Du Toit, S., and Spisic, D. (1997). *Robust Inference using Weighted Least Squares and Quadratic Estimating Equations in Latent Variable Modeling with Categorical and Continuous Outcomes*. Available online at: https://www.statmodel.com/download/Article_075.pdf
- Muthén, L. K., and Muthén, B. O. (1998–2015). *Mplus User's Guide, 7th Edn*. Los Angeles, CA: Muthén & Muthén.
- NAEP (2015). *National Assessment of Educational Progress (NAEP; 2015). NAEP Reading Report Card*. Available online at: https://www.nationsreportcard.gov/reading_2017/#?grade=4
- Nash, H., and Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge. A controlled evaluation of the definition and context methods. *Int. J. Lang. Commun. Disord.* 41, 335–354. doi: 10.1080/13682820600602295
- Nation, K. (2011). "Children's reading comprehension difficulties," in *The Science of Reading: A Handbook*, eds M. Snowling and C. Hulme (Malden, MA: Blackwell Publ.), 248–265.
- Nelson, P. M., Van Norman, E. R., Klingbeil, D. A., and Parker, D. C. (2017). Progress monitoring with computer adaptive assessments. the impact of data collection schedule on growth estimates. *Psychol. Schools* 54, 463–471. doi: 10.1002/pits.22015
- Oakhill, J. V., Cain, K., and Bryant, P. E. (2003). The dissociation of word reading and text comprehension. Evidence from component skills. *Lang. Cogn. Proces.* 18, 443–468. doi: 10.1080/01690960344000008
- OECD (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. Paris: OECD Publishing.
- OECD (2014). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science*, Vol. 1., Paris: OECD Publishing.
- OECD (2016). *PISA 2015 Results in Focus, PISA in Focus, No. 67*. Paris: OECD Publishing.
- Perfetti, C. A. (1985). *Reading Ability*. New York, NY: Oxford University Press.
- Perfetti, C. A. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730
- Reise, S. P., Bonifay, W. E., and Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *J. Pers. Assess.* 95, 129–140. doi: 10.1080/00223891.2012.725437
- Richter, T., Isberner, M.-B., Naumann, J., and Neeb, Y. (2013). Lexical Quality and reading comprehension in primary school children. *Sci. Stud. Read.* 17, 415–434. doi: 10.1080/10888438.2013.764879
- Russell, M. K. (2010). "Technology-aided formative assessment of learning: New developments and applications," in *Handbook of Formative Assessment*, eds H. L. Andrade and G.J. Cizek (New York, NY: Routledge), 125–138.
- Salentin, K. (2014). Sampling the ethnic minority population in germany. the background to "Migration Background". *Methods Data Anal.* 8, 25–52. doi: 10.12758/mda.2014.002
- Schnepf, S. V. (2007). Immigrants' educational disadvantage: an examination across ten countries and three surveys. *J. Popul. Econ.* 20, 527–545. doi: 10.1007/s00148-006-0102-y
- Schulgesetz für das Land Nordrhein-Westfalen (2018). (*Schulgesetz NRW – SchulG vom 15. Februar 2005 (GV.NRW.S.102) Zuletzt Geändert Durch Gesetz Vom 21. Juli 2018 (SGV.NRW.223)*). [Engl. School Law of the Federal State North Rhine-Westphalia]. Available online at: <https://www.schulministerium.nrw.de/docs/Recht/Schulrecht/Schulgesetz/Schulgesetz.pdf>
- Shin, J., Deno, S. L., and Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *J. Spec. Educ.* 34, 164–172. doi: 10.1177/00224690003400305

- Spencer, M., Quinn, J. M., and Wagner, R. K. (2014). Specific reading comprehension disability: major problem, myth, or misnomer? *Learn. Disabil. Res. Pract.* 29, 3–9. doi: 10.1111/ldrp.12024
- Steinmetz, H. (2013). Analyzing observed composite differences across groups. *Methodology* 9, 1–12. doi: 10.1027/1614-2241/a000049
- Taraban, R., and McClelland, J. L. (1990). "Parsing and comprehension: A multiple constraint view," in *Comprehension Processes in Reading*, eds D. A. Balota, G. B. Flores D'Arcais, and K. Rayner (Hillsdale, NJ: Erlbaum), 231–264.
- Taylor, C. R. (2012). Engaging the struggling reader: focusing on reading and success across the content areas. *Natl. Teach. Educ. J.* 5, 51–58.
- Tichá, R., Espin, C. A., and Wayman, M. M. (2009). Reading progress monitoring for secondary-school students. Reliability, validity, and sensitivity to growth of reading-aloud and maze-selection measures. *Learn. Disabil. Res. Pract.* 24, 132–142. doi: 10.1111/j.1540-5826.2009.00287.x
- Tilstra, J., McMaster, K., van den Broek, P., Kendeou, P., and Rapp, D. (2009). Simple but complex. Components of the simple view of reading across grade levels. *J. Res. Read.* 32, 383–401. doi: 10.1111/j.1467-9817.2009.01401.x
- Van den Bosch, R. M., Espin, C. A., Chung, S., and Saab, N. (2017). Data-based decision-making. teachers' comprehension of curriculum-based measurement progress-monitoring graphs. *Learn. Disabil. Res. Pract.* 32, 46–60. doi: 10.1111/ldrp.12122
- Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York, NY: Academic Press.
- Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., and Chen, R. (2007). Components of reading ability. multivariate evidence for a convergent skills model of reading development. *Sci. Stud. Read.* 11, 3–32. doi: 10.1207/s1532799xssr1101_2
- Waltzman, D. E., and Cairns, H. S. (2000). Grammatical knowledge of third grade good and poor readers. *Appl. Psychol.* 21, 263–284. doi: 10.1017/S014271640000206X
- Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., and Espin, C. A. (2007). Literature synthesis on curriculum-based literature synthesis on curriculum-based. *J. Spec. Educ.* 41, 85–120. doi: 10.1177/00224669070410020401
- West, R. F., and Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Dev.* 49, 717–727. doi: 10.2307/1128240
- Wilbert, J., and Linnemann, M. (2011). Kriterien zur analyse eines tests zur lernverlaufsdiagnostik [engl. criteria for analyzing a test to measure Learning Progress]. *Empirische Sonderpädagogik* 3, 225–242.
- Wiley, H. I., and Deno, S. L. (2005). Oral reading and maze measures as predictors of success for english learners on a state standards assessment. *Remed. Spec. Educ.* 26, 207–214. doi: 10.1177/07419325050260040301
- Witzel, J., and Witzel, N. (2016). Incremental sentence processing in japanese. A maze investigation into scrambled and control sentences. *J. Psychol. Res.* 45, 475–505. doi: 10.1007/s10936-015-9356-4
- Zeuch, N., Förster, N., and Souvignier, E. (2017). Assessing teachers' competencies to read and interpret graphs from learning progress assessment. Results from Tests and interviews. *Learn. Disabil. Res. Pract.* 32, 61–70. doi: 10.1111/ldrp.12126

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Jungjohann, DeVries, Mühling and Gebhardt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Graphen der Lernverlaufsdiagnostik interpretieren und anwenden – Leseförderung mit der Onlineverlaufsmessung Levumi*

Interpretation and use of graphs of curriculum-based measurement – reading support with the online platform Levumi

Jana Jungjohann, Kirsten Diehl, Andreas Mühling & Markus Gebhardt

Zusammenfassung

Viele Schüler mit sonderpädagogischem Förderbedarf entwickeln Schwierigkeiten im sprachlichen Bereich. Dies hat zur Folge, dass insbesondere der Leseerwerb einer pädagogischen Aufbereitung bedarf. In dieser pädagogischen Arbeit bietet der problemlösende Ansatz der Lernverlaufsdiagnostik (engl. Curriculum-Based Measurement) Lehrkräften eine Unterstützung bei der Überprüfung der Effektivität ihres Unterrichts durch kurze und leicht handhabbare Tests. Diese Unterstützung kann durch weiterführende Materialien (z. B. Interpretationshilfen, Fördermaterialien) bestärkt werden. Für eine erfolgreiche Etablierung der Lernverlaufsdiagnostik und der damit verbundenen computergestützten Testsysteme in der Praxis müssen die aus dem Schulsystem bedingten Anforderungen berücksichtigt werden. Dieser Beitrag thematisiert literaturbasiert die Chancen und Grenzen bei der Implementation der Lernverlaufsdiagnostik in der pädagogischen Praxis. Dafür werden digitale Testsysteme, die Perspektive der Lehrkräfte sowie Schwierigkeiten bei der Interpretation von Lernverlaufsgraphen fokussiert betrachtet. Als Praxisbeispiel wird das Angebot der kostenlosen Onlineplattform Levumi (www.levumi.de), ihre Lernverlaufstests im Bereich Lesen, die Interpretation der Klassen- und Individualgraphen und Fördermaterialien vorgestellt. Damit sich der Ansatz der Lernverlaufsdiagnostik in Deutschland im inklusiven Schulsystem etablieren kann, müssen Barrieren durch Kosten oder einen fehlenden technischen Support abgebaut werden. Zusätzlich können adaptive Testsysteme die Leseförderung durch automatisierte Empfehlungen von Fördermaterialien, schülergerechte Rückmeldungen und Analysen der Kompetenzprofile der Schüler den Unterricht bereichern.

Schlüsselwörter

Lernverlaufsdiagnostik, formative Evaluation, Lesentwicklung, Lesediagnostik, Lernverlaufsgraphen, Grapheninterpretation

Abstract

Many students with special educational needs develop difficulties in language. As a result, extra preparation by teachers is needed to support the development of reading skills. Curriculum-based measurement (CBM) helps evaluate teacher lessons with short and easy-to-handle tests. Further materials, such as handbooks and other supporting materials, can mean better outcomes. School system requirements must also be considered for a successful implementation of CBM instruments and their linked digital test systems. With a basis in prior work, the following article addresses opportunities and boundaries of CBM implementation in the classroom. We focus on digital test systems, the perspective of the teachers, and problems within the interpretation of CBM graphs. As an example, we present the online platform

* Dieser Beitrag hat das Peer-Review-Verfahren durchlaufen.

Levumi (www.levumi.de) and its CBM reading tests, the interpretation of its CBM graphs (individual and group graphs), and its supporting materials. We conclude that the CBM approaches can be established in Germany, but barriers, costs must be reduced, and technical support must be more available. Additionally, adaptive test systems can enhance the school lessons by automatically providing lesson material suggestions, feedback for students, and analyses of student competency profiles.

Keywords

curriculum-based measurement, formative evaluation, reading development, reading diagnostic, graphs of curriculum-based measurement, graph interpretation

1 Herausforderungen im inklusiven Leseunterricht

Heterogene Lernvoraussetzungen und Lernbedürfnisse sind eine hohe Herausforderung für Lehrkräfte. Insbesondere im inklusiven Unterricht individualisieren Lehrkräfte stärker und passen ihre Förderungen noch mehr an die individuellen Bedürfnisse der Schüler an als im regulären Unterricht (Gebhardt, Schwab, Krammer, Gasteiger-Klicpera & Sälzer, 2014). Die schulische Umgebung ist dominiert durch Sprache, sowohl als Medium sowie als Gegenstand, sodass Schwierigkeiten im sprachlichen Bereich den Bildungserfolg negativ beeinflussen können (Lütje-Klose & Mehlem, 2015). Die Mehrheit der Schüler unterschiedlicher sonderpädagogischer Förderbedarfe (SPF) haben oder entwickeln Schwierigkeiten im sprachlichen Bereich (Lindsay & Stand, 2016, Lütje-Klose, 2012), sodass sie insbesondere im Anfangsunterricht Lesen eine erhöhte pädagogische Unterstützung benötigen. Diese Schüler starten mit geringeren Vorläuferfähigkeiten in den Leseerwerb (Klicpera, Ehgartner, Gasteiger-Klicpera & Schabmann, 1993) und ihre Lernprobleme manifestieren sich größtenteils über die Schuljahre hinweg bis ins Erwachsenenalter (Taylor, 2012, Landerl & Wimmer, 2008). Die Aufgabe der Lehrkräfte im inklusiven Leseunterricht ist es, frühzeitig potentielle Schwierigkeiten aufzudecken und den Leseerwerb fortlaufend mit effektiven Fördermaßnahmen zu unterstützen (Diehl, 2011). Ein neuer vielversprechender Ansatz ist die Lernverlaufsdagnostik. Lernverlaufsdagnostik ist eine Brücke zwischen Förderdiagnostik, Förderung und ihrer fortlaufenden Evaluierung (Jungjohann & Gebhardt, 2018).

2 Lernverlaufsdagnostik im Bereich Lesen

2.1 Lernverlaufsdagnostik von Leseleistungen

Das Konzept der Lernverlaufsdagnostik (engl. curriculum-based measurement) wurde in den USA in den 1970er-Jahren von der Forschergruppe um Stanley Deno entwickelt und ist dort ein etabliertes System zur Erfassung und Evaluation des schulischen Lernens (Deno, 2003). Instrumente, die den Leseerwerb und die weiterführende Leseentwicklung begleiten, sind besonders weit verbreitet und erforscht (Ardoin, Christ, Morena, Cormier & Klingbeil 2013; Fuchs, 2017; Reschly, Busch, Betts, Deno und Long, 2009; Wayman, Wallace, Wiley, Tichá & Espin, 2007). Lernverlaufsdagnostik stellt einen ‚problemlösenden Ansatz‘ dar, der durch unterrichtsimmanente Messungen des Lernens die Effektivität von Fördermaßnahmen überprüft (Fuchs, 2017). Für die Schulpraxis geht es dabei um die Beantwortung folgender zentraler Fragestellungen:

- Was sollen Schüler lernen?
- Wie erfahren Lehrkräfte, dass sich die Leistungen der Schüler verbessern?
- Was wird getan, wenn die Schüler etwas nicht ausreichend lernen?

Der Grundgedanke der Lernverlaufsdagnostik besteht darin, dass der Unterricht zum Kind passen muss und er als effektiv bewertet werden kann, wenn die Schüler Fortschritte im Lernen machen. Dafür werden curriculumsrelevante Konstrukte erhoben, die stellvertretend für eine übergeordnete Kompetenz sind. Zur Erfassung basaler Lesekompetenzen wird beispielsweise der robuste Indikator des flüssigen lauten Lesens stellvertretend gemessen (Deno, Mirkin & Chiang, 1982; Fuchs, Fuchs, Hosp & Jenkins, 2001). Charakteristisch für die Lernverlaufsdagnostik ist, dass sich die nur wenige Minuten dauernden Messungen leicht in den alltäglichen Unterricht einbinden lassen und sie hochfrequent (bis zu wöchentlich) durchgeführt werden. Ein weiteres Merkmal ist die visuelle Darstellung der Messergebnisse in einem Lernverlaufsgraphen, um die Interpretation für Lehrkräfte zu erleichtern. Die Interpretationen der Lernverlaufsgraphen

ermöglichen den Lehrkräften die frühzeitige Interpretation von Non-Responder (Gebhardt, Diehl, Mühling, 2016). Als Non-Responder werden Schülerinnen und Schüler bezeichnet, die nicht in gewünschter Weise von einer Förderung profitieren. Im Bereich der Leseflüssigkeit sind eine anhaltend geringe Dekodiergenauigkeit sowie eine nicht steigende Lesegeschwindigkeit Hinweise auf einen Non-Responder. Nach der Identifikation können die Lehrkräfte anhand des individuellen Kompetenzprofils ihre Leseförderungen unmittelbar an die individuellen Bedürfnisse anpassen.

2.2 Chancen und Grenzen von Lernverlaufsdagnostik

Seit der Jahrtausendwende wird ebenfalls in Deutschland zur Lernverlaufsdagnostik geforscht und die Forschungsbemühungen nehmen zu (Jungjohann, Gegenfurtner & Gebhardt, 2018; Walter, 2011). Besonders digitale und onlinebasierte Testsysteme gewinnen in der Schullandschaft an Bedeutung (Maier, 2014), da immer mehr Schulen eine technische Ausstattung erhalten. Bei der Einführung von Lernverlaufsmessung in Deutschland ergeben sich somit neue Anforderungen an die Testsysteme und die Lehrkräfte.

Anforderungen an die Testsysteme

In Deutschland sind die etablierten und überwiegend verbreiteten Testverfahren ausschließlich als Papierversion verfügbar. Allerdings verursachen analoge Tests im Schultagalltag bei Lehrkräften zusätzlichen Dokumentations- und Verwaltungsaufwand, der durch digitale Medien verringert werden kann (Gebhardt & Jungjohann, 2018). Die digitalen Testsysteme übernehmen Verwaltungsaufgaben, wie die Dokumentation über bereits gelöste Tests, und bereiten Schülerergebnisse in vielfältiger Weise auf. Gleichzeitig erfordern digitale Testsysteme von Lehrkräften neben der inhaltlichen Einarbeitung eine technische Einarbeitung. Eine technische Einarbeitung verläuft unterschiedlich, je nachdem wie vertraut Lehrkräfte bereits mit digitalen Medien sind. Ein explorativer Zugang ist besonders für Lehrkräfte geeignet, die bereits mit digitalen Medien vertraut sind. Sie probieren die Technik intuitiv aus und benötigen Hilfestellungen nur bei expliziten Fragen. Andere Lehrkräfte hingegen wünschen sich eine detaillierte Anleitung, die sie bei der Einarbeitung eng begleitet. Zusätzlich müssen onlinebasierte Testsysteme den Datenschutz gewährleisten, damit sie überhaupt in deutschen Schulen genutzt werden dürfen. Beispielsweise schließt dieser in Nordrhein-Westfalen die ausschließliche Nutzung von in Europa liegenden Servern und die sichere Verschlüsselung der Daten ein (Medienberatung NRW, 2015). Für Lehrkräfte ist die Einhaltung dieser Vorschriften unumgänglich. Folglich müssen die Testsysteme diese Anforderungen berücksichtigen und ihre Datenschutzrichtlinien für Lehrkräfte verständlich und transparent offenlegen.

Die Ausstattung in deutschen Schulen ist nicht einheitlich geregelt und hängt von vielen Faktoren ab. Die Nutzung von Lernverlaufsdagnostik wird einerseits durch die finanziellen und andererseits durch die technischen Ressourcen beeinflusst. Kosten und Lizenzen reglementieren die Einsatzmöglichkeiten von Lernverlaufsdagnostik maßgeblich. Schulen haben bisher ein Budget für Fördermaterial und Bücher, welche sie dauerhaft erwerben. Viele digitale Testsysteme sind meist nur über eine zeitlich begrenzte Lizenz verfügbar, sodass sie bei einer langfristigen Nutzung im jährlichen Rhythmus Kosten verursachen. Ein Ausweg stellen Testsysteme dar, die aufgrund staatlicher Förderung für die Schulen kostenlos angeboten werden können. In gut ausgestatteten Schulen können die Lehrkräfte zwischen Tablets und Standcomputern wählen und haben Zugang zu einer stabilen Internetverbindung. In anderen Fällen haben Lehrkräfte keine Administratorenrechte um Programme oder Apps zu installieren und müssen selbstständig die Funktionalität der Endgeräte gewährleisten. In beiden Situationen kann es zu Komplikationen bei der Benutzung kommen. Digitale Testsysteme müssen diese unterschiedlichen Ausgangslagen berücksichtigen und einen ausreichenden Support für Lehrkräfte bereitstellen. Ohne Unterstützung im Umgang mit den neuen Medien bauen diese Barrieren auf, die die Potentiale der Testsysteme für den Unterricht mindern.

Neben der Unterstützung zum Umgang mit der digitalen Lernverlaufsdagnostik, werden zusätzlich weiterführende Materialien zur Interpretation und zur Förderung benötigt (Ardoin et al., 2013). Im pädagogischen Kontext ist ein Statustest notwendig, um Problembereiche zu identifizieren, aber nicht ausreichend um einen Lernzuwachs oder eine Lernstagnation zu erfassen und effektivere Fördermaßnahmen zu prüfen. Daher sollen aus den Ergebnissen der Lernverlaufstests konkrete Hinweise auf die individuellen Schwierigkeiten der Schülerinnen und Schüler ableitbar sein. Folglich ist eine Verknüpfung des eigentlichen Lernverlaufstests mit Hinweisen zur spezi-

fischen Interpretation der Ergebnisse sinnvoll. Weiterführend können Testsysteme über potentielle Hürden im Lernbereich (z. B. Schwierigkeiten während des Leseerwerbs) und Fördermöglichkeiten (z. B. Übungen zur Erhöhung der Leseflüssigkeit) empfehlenswert sein.

Die Entwicklung neuer Lernverlaufstests für das Lesen muss die Besonderheiten der jeweiligen Sprache berücksichtigen. Die Struktur einer Sprache impliziert Regeln, die in der Testkonstruktion entweder durch Dimensionen oder schwierigkeitsgenerierende Regeln nachgebildet werden. Leseerwerbsprobleme im Deutschen können durch unterschiedliche Ursachen (z. B. Deutsch als Zweitsprache, sozial benachteiligende Lebensbedingungen; Dirim, Hauenschild, Lütje-Klose, 2008) bedingt werden. In Deutschland ist die Varianz der sprachlichen Heterogenität in (inklusi-ven Grundschul-)Klassen hoch (OECD, 2018), sodass die Auswahl des linguistischen Testmaterials sensibel erfolgt, um Benachteiligung einzelner Schülergruppen zu vermeiden. So benötigen Lehrkräfte eine breite Auswahl an Tests, um die unterschiedlichen Bedürfnisse der Schülerinnen und Schüler zu messen. Zukünftig können adaptive Testsysteme auf eine solche Heterogenität reagieren, indem sie Items oder Tests in Abhängigkeit eines Kompetenzprofils einer Schülerin oder eines Schülers automatisiert auswählen oder in einem weiteren Schritt die Lernverlaufstests an die linguistischen Schwierigkeiten anpassen. Dadurch kann eine genauere Passung zwischen Bedarfen in den Klassen und der Testentwicklung erreicht werden.

Anforderungen durch die Lehrkraftkultur

Da das Feld der Lernverlaufsdagnostik in Deutschland erst seit wenigen Jahren erforscht wird (Walter, 2011; Klauer, 2011), kennt bisher nur ein geringer Anteil der Lehrkräfte konkrete Testsysteme und ihre Einsatzzwecke. Fuchs (2004) ordnet der dritten Stufe der Erforschung von Lernverlaufsdagnostik den Nutzen und die Anwendung durch Lehrkräfte im Feld zu. Bevor Lehrkräfte Lernverlaufsdagnostik als Baustein ihres Unterrichts erwägen, prüfen sie die Systeme mit folgenden Fragen:

- Welche Einsatzmöglichkeiten bietet ein Testsystem der Lernverlaufsdagnostik?
- Inwiefern bereichert Lernverlaufsdagnostik meinen Unterricht?
- Inwiefern beeinflusst Lernverlaufsdagnostik meinen bisherigen Unterricht?

Im inklusiven Unterricht ist es für die Lehrkräfte von Vorteil, wenn die Testsysteme möglichst breit und flexibel einsetzbar sind. Manche Lehrkräfte benötigten Instrumente, die mit ganzen Klassen genutzt werden können und andere benötigten sie nur für einzelne, wenige Schülerinnen und Schüler. Zusätzlich erhöht ein zeitlich ungebundener Einsatz die Flexibilität im Unterrichtsalltag, um Schülerinnen und Schüler je nach Kompetenzprofil häufiger oder seltener zu testen. Im Fokus der Lernverlaufsdagnostik steht die individuelle Bezugsnorm (Gebhardt, Diehl, Mühling, 2016), sodass Schülerinnen und Schüler mit hohem Lerntempo mit größeren Abständen getestet werden können, da diese sogenannte Responder sind. Responder profitieren von der angebotenen Förderung und sie zeigen einen erwarteten, positiven Lernverlauf. Schülerinnen und Schüler mit Lernschwierigkeiten sollten dagegen solange häufiger getestet werden bis ein stetiger Lernzuwachs erkennbar und sichergestellt ist, dass die unterrichtliche Förderung wirkt. Testsysteme, die für möglichst viele Schülerinnen und Schüler geeignet sind, verringern den Aufwand der Einarbeitung, Einführung und Koordination im Unterricht für die Lehrkräfte.

Damit Lehrkräfte neue Innovationen in ihren Unterricht etablieren, ist die Erkenntnis über den Mehrwert für sie selbst und ihre Klassen Voraussetzung. Der Mehrwert der Lernverlaufsdagnostik liegt für die Lehrkräfte nicht alleine in den Informationen über ansteigende oder stagnierende Lernverläufe, sondern ebenfalls auf der Verknüpfung der Diagnostik mit Förderempfehlungen und Fördermaterial. Besonders automatisierte Analysen der Schwierigkeiten im Lesen und der verwendeten Strategien bei der Bearbeitung von Aufgaben liefern Lehrkräften Anregungen für Alternativen in der Förderplanung. Zudem wünschen sich Lehrkräfte, dass datenbasierte Übungsformate (z. B. Arbeitsblätter, Aufgabenstellungen) und weiterführendes Übungsmaterial (z. B. individueller Übungswortschatz, Lesetexte) bereitgestellt werden. Des Weiteren erhöht ein schülergerechtes, fachbezogenes Feedback den Mehrwert für den Unterricht. Viele Schülerinnen und Schüler möchten wissen, ob sie ihre Leistungen verbessern konnten. Automatisierte und altersadäquate Rückmeldungen für Selbstkontrollen des Lernprozesses können den Workload der Lehrkräfte ebenfalls reduzieren und können Lernverlaufsdagnostik im Unterrichtsalltag profitabler machen.

Eine Etablierung von Lernverlaufsdagnostik im Unterricht hängt ebenfalls davon ab, wie sehr sie in das alltägliche Unterrichtsgeschehen eingreift. Lehrkräfte mit viel Berufserfahrung greifen

auf ein großes Repertoire an didaktischen sowie fachlichem Wissen zurück und haben eigene Unterrichtskonzepte mit bevorzugten Methoden und Materialien entwickelt. Für diese Lehrkräfte ist es wichtig, dass Lernverlaufsdagnostik nicht als Zusatz zum bisherigen Unterricht verstanden wird, sondern dass sie sich in bestehende Strukturen integrieren lässt. Hingegen ist für Berufsanfängerinnen und Berufsanfänger ein Testsystem attraktiv, welches sie in der Unterrichtsplanung und -gestaltung unterstützt und leitet. Erst wenn diese Voraussetzungen geschaffen sind, können alle Lehrkräfte die Effektivität ihres Unterrichts anhand der Lernverlaufsgraphen evaluieren.

2.3 Lernverlaufsgraphen

In den Lernverlaufsgraphen werden die Ergebnisse der Tests abgetragen, damit eine Interpretation dieser gut gelingen kann. Viele Graphen haben einen einheitlichen Aufbau. Auf der Abszissenachse (x-Achse) wird meist der zeitliche Verlauf abgebildet (z. B. Anzahl der Schulwochen). Hier können Lehrkräfte die Zeitpunkte sowie die Zeitabstände der durchgeführten Lernverlaufstests ablesen. Auf der Ordinatenachse (y-Achse) wird hingegen das Kompetenzmaß des jeweiligen Lernverlaufstests abgetragen (z. B. Anzahl der richtig gelesenen Wörter pro Minute). Wenn mehrere Testergebnisse vorliegen, werden diese zu einer Lernverlaufskurve verbunden. Die Steigung der Kurve zeigt an, ob der Lernverlauf einer Schülerin oder eines Schülers über die Zeit steigt, stagniert oder negativ abfällt. Traditionell wird zwischen Klassengraphen und Individualgraphen differenziert (Klauer, 2014). Die Klassengraphen zeigen die Lernverläufe aller Schülerinnen und Schüler einer sozialen Bezugsgruppe und die Individualgraphen eine Verlaufskurve im Sinne der individuellen Bezugsnorm an. Mit einem Individualgraphen werden häufig noch zusätzliche Informationen angeboten, z. B. Angaben zur Anzahl falsch gelöster Antworten, eine Abbildung der sozialen Bezugsnorm, eine Trendlinie (durchschnittliche Steigung des Lernverlaufs) und eine Ziellinie (erwartete Steigung des Lernverlaufs vor Förderbeginn; van den Bosch, Espin, Chung & Saab, 2017). Neben Erweiterung der schülerbezogenen Daten können auch unterrichtliche Maßnahmen in einem Lernverlaufsgraphen markiert werden. Hierzu nennen van den Bosch et al. (2017) vertikale Linien, die Änderungen in der unterrichtlichen Förderung zueinander abgrenzen. Zusätzlich können qualitative Informationen über einzelne Antworten bereitgestellt werden. Alle zusätzlichen Informationen haben das Ziel, die Interpretation der Daten zu vereinfachen.

Lehrkräfte benötigen diagnostische und förderpädagogische Kompetenzen, um Graphen richtig zu interpretieren und mit einer Förderung in Bezug zu setzen. Grundsätzlich beschreiben Wagner, Hammerschmidt-Snidarich, Espin, Seifert und McMaster (2017) die Grafeninterpretation als drei aufeinander aufbauende Hauptprozesse. Der erste Prozess ist die visuelle Mustererkennung, bei dem der Verlauf des Graphens betrachtet wird. Im zweiten interpretativen Prozess werden allgemeines Graphen- sowie Inhaltswissen miteinander verknüpft. Im letzten integrativen Prozess werden zusätzliche fachliche Theorien (z. B. Wissen über die Leseentwicklung) mit dem Verlauf des Graphens verbunden. Zeuch, Förster und Souvignier (2017) spezifizieren diese grundlegenden Kompetenzen und beschreiben drei Kompetenzstufen von Lehrkräften. In der ersten Stufe *Lesen der Daten* ermitteln Lehrkräfte wichtige Datenpunkte in einem Graphen und beschreiben ihn. Dazu gehört, dass die Lehrkräfte den Trend der Kurve erkennen und verstehen. In der zweiten Stufe *Lesen zwischen den Daten* verbinden die Lehrkräfte konkrete schülerbezogene Fähigkeiten mit dem individuellen Lernverlauf. In der höchsten Kompetenzstufe *Lesen über die Daten hinaus* ziehen Lehrkräfte fachlich begründete Schlüsse über zukünftige Entwicklungen, kindliche Schwierigkeiten und gestalten begründet Fördermaterialien und Übungsphasen.

Manche Lehrkräfte erreichen auch ohne spezifisches Training die erste Kompetenzstufe. Trotzdem benötigen die Lehrkräfte besondere Unterstützung sowie Begleitmaterialien für die Interpretation, damit es nicht zu Fehlinterpretation kommt (Ardoin et al., 2013). Lehrkräfte können Schwierigkeiten haben, Daten systematisch zu beschreiben oder neigen dazu, sich auf einzelne, irrelevante Datenpunkte zu konzentrieren (Zeuch et al., 2017). Außerdem wird ihre Vorhersage des zukünftigen Lernverlaufs durch Extremwerte, eine niedrige Variabilität der richtigen Antworten, die Steigung der Kurve sowie die Ausprägung des letzten Datenpunkts beeinflusst (Klapproth, 2018). Optische Hilfsmittel (z. B. eine Trendlinie) helfen den Lehrkräften bei der Interpretation. Allerdings fokussieren sie sich so stark auf diese Hilfsmittel, dass andere Aspekte vernachlässigt werden (Newell & Christ, 2017). Diese aktuellen Forschungsergebnisse müssen allerdings zunächst noch in die Unterstützung der Lehrkräfte übertragen werden. In Deutschland ist der Einsatz der Lernverlaufsdagnostik noch selten, sodass Unterstützung vor allem an der Schnittstelle Grapheninterpretation und lesedidaktische Förderplanung benötigt wird (Jungjohann, Gegenfurtner & Gebhardt, 2018).

3 Die Onlineplattform Levumi

Die Onlineplattform Levumi (www.levumi.de) bietet kompetenzorientierte Lernverlaufstests an und verbindet diese mit Fördermaßnahmen. Dieses ineinandergreifende Angebot verbindet Förderdiagnostik, Förderplanung und deren Evaluierung für den inklusiven Unterricht in einem Testsystem. Die Plattform ist 2015 in einem multidisziplinären Forscherteam entwickelt worden und wird kontinuierlich erweitert (Mühling, Gebhardt & Diehl, 2017). Lehrkräfte können sich kostenlos einen Account anlegen und alle Funktionen explorativ in einer Testklasse kennenlernen. Zusätzlich stehen Handbücher sowie Video-Tutorials für die Benutzung, Interpretation und Förderung zum Download bereit. Alle Lernverlaufstests und Fördermaterialien sind theoriegeleitet konstruiert und werden in Schulstudien evaluiert. Neben den Tests im Lernbereich Lesen stehen Lehrkräften Tests zur Erfassung mathematischer Kompetenzen sowie dem Verhalten zur Verfügung.

3.1 Lernverlaufstests im Bereich Lesen

Im Lernbereich Lesen sind alle Tests am Kieler Leseaufbau orientiert (Dummer-Smoch & Hacketal, 2016). Die Grapheme sind nach Schwierigkeitsgrad gestaffelt und für die Levumi Plattform bis maximal zu fünf Niveaustufen zusammengefasst. Damit Lehrkräfte die Levumi Tests unabhängig zu ihrem sonstigen Unterrichtsmaterial nutzen können, wurde bei der Testkonstruktion die Reihenfolge der Buchstabeneinführung verschiedener Leselehrwerke berücksichtigt.

3.1.1 Leseflüssigkeitstests

Alle Leseflüssigkeitstests sind lehrerzentriert und werden digital durchgeführt. Nach einer mündlichen Einführung startet die Lehrkraft den Test, der dann eine Minute lang zufällig gezogene Items anzeigt. Bei der zufälligen Ziehung werden schwierigkeitsgenerierende Regeln berücksichtigt, sodass beispielsweise zwei aufeinander folgende Items nicht das gleiche Anfangsgraphem haben. Das Kind liest jedes Item laut vor und die Lehrkraft bewertet über die Tastatur, ob das Kind korrekt gelesen hat. Das Kompetenzmaß ist die Anzahl der richtig gelesenen Items. Nach jeder Testung erhält das Kind leistungsbezogene Rückmeldung über den Drachen Levumi. Levumi bietet mehrere Testtypen an, die unterschiedliche Teilfähigkeiten des basalen Lesens erfassen (siehe Jungjohann & Gebhardt, 2018).

3.1.2 Tests zum sinnentnehmenden Lesen

Das sinnentnehmende Lesen wird in mehreren Niveaustufen auf Satzbasis erfasst. Das Satzlesen ist das Bindeglied zwischen basalen Lesefähigkeiten und dem Textverstehen (Ecalte, Bouchafa, Potocki & Magnan, 2013) und erhebt diese Teilfähigkeiten. Neben den Buchstaben der Niveaustufen werden zusätzlich unterschiedliche Argument-Prädikat-Strukturen innerhalb der Sätze berücksichtigt. Die Kinder führen diese Tests eigenständig in einem individuellen Schüleraccount durch. Alle schülerzentrierten Tests können als Gruppentest angeleitet werden. Analog zu den Leseflüssigkeitstests wird die Anzahl richtig gelöster Items als Kompetenzmaß herangezogen und anschließend gibt der Drache Levumi Feedback.

3.2 Interpretation der Levumi Graphen

Die Plattform wertet alle Ergebnisse unmittelbar für die Lehrkräfte aus und stellt drei unterschiedliche Ansichten zur Verfügung. Dazu zählt ein Klassengraph, ein Individualgraph sowie eine qualitative Auswertung, in der alle Antworten sowie die Lösungswahrscheinlichkeiten für jede Messung gelistet sind (siehe Gebhardt, Diehl & Mühling, 2016). Hilfestellungen zur Interpretation werden im Lehrerhandbuch kontinuierlich aktualisiert. Die Klassengraphen werden dazu genutzt um potentielle Förderkinder innerhalb der Klasse zu identifizieren. Meist haben die Lehrkräfte zuvor subjektive Theorien, die mithilfe der sozialen Bezugsnorm bestätigt werden können. Zusätzlich können Kinder entdeckt werden, die ihre Schwierigkeiten im Unterrichtsalltag kompensieren. Anhand des Individualgraphs wird das kindspezifische Leseschwierigkeitsprofil ermittelt. Die qualitativen Auswertungen ergänzen die Individualgraphen, indem sie systematisch linguistische Fehlerquellen aufdecken. Analysen zur Dekodiergenauigkeit oder der Lesegeschwindigkeit können ebenfalls leicht mithilfe der Lehrerhandbücher erfolgen. Alle gewonnenen Informationen dienen als Basis für die Förderplanung. Durch eine automatisierte Auswertung der Plattform können zukünftig gezielte Hilfen bei der Interpretation für Lehrkräfte angeboten werden. So könnten beispielsweise auffällige Lernverläufe erkannt und visuell hervorgehoben oder spezifische Hinweise auf mögliches Fördermaterial gegeben werden.

3.3 Fördermaterialien

Für die Verschränkung von Diagnostik und Förderung bietet Levumi das Förderhandbuch Lesen an, in dem zentrale Entwicklungsbereiche und Schwierigkeiten des Leseerwerbs sowie Angebote zur Förderung dargestellt sind (siehe Jungjohann, Gebhardt, Diehl & Mühling, 2017). Zudem enthält es exemplarische kopier- und bearbeitbare Formatvorlagen. Das Förderhandbuch stellt stets eine Verbindung zwischen Aufgabentypen, deren Intentionen und den Niveaustufen der Tests her. Die fachliche Begründung der Förderplanung geht damit einher und erfolgt nach linguistischen Kriterien. Lehrkräfte können sich durch dieses Material informieren und bei Bedarf alternative Förderansätze entwickeln.

4 Fazit

Lehrkräfte sind die Experten für das schulische Lernen ihrer Schülerinnen und Schüler. Sie entscheiden frei über den Einsatz von Fördermaterialien und gestalten Förderungen auf der Basis ihrer fachlichen und pädagogischen Kompetenzen. Der Ansatz der Lernverlaufsdagnostik stellt kein neues Unterrichtskonzept dar, sondern versteht sich als Erweiterungsangebot bekannter Strukturen. Die Lernverlaufstests ermöglichen die Evaluation des eigenen Unterrichts und geben Anlass zur Reflexion. Für die Etablierung von Lernverlaufsdagnostik in Deutschland müssen die Besonderheiten des inklusiven Schulsystems berücksichtigt werden. Das angebotene Fördermaterial sowie die Aufbereitung des lesedidaktischen Wissens dienen als Anregungen zur Förderplanung. Die Intention der Onlineplattform Levumi ist es, Lehrkräfte in der Erweiterung ihres bestehenden Repertoires zu unterstützen um den Herausforderungen im inklusiven Leseunterricht effektiv begegnen zu können.

Literatur

- 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. DOI: 10.1016/j.jsp.2012.09.004.
- Deno, S. L. (2003). *Curriculum-Based Measures: Development and Perspectives. Curriculum-Based Measures: Development and Perspectives: 28* (3-4), 3-12. <https://doi.org/10.1177/073724770302800302>.
- 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 Lese-Rechtschreibschwierigkeiten. *Zeitschrift für Heilpädagogik, 5*, 164-172.
- Dirim, I., Hauenschild, K. & Lütje-Klose, B. (2008). Ethnische Vielfalt und Mehrsprachigkeit an Schulen. Einführung. In Dirim, I., Hauenschild, K. & Lütje-Klose, B., Löser, J. & Sievers, I. (Hrsg.) *Ethnische Vielfalt und Mehrsprachigkeit an Schulen. Beispiel aus verschiedenen nationalen Kontexten*. (9-12) Frankfurt am Main: Brandes & Apsel.
- Dummer-Smoch, L., & Hackethal, R. (2016). *Kieler Leseaufbau. Handbuch*, 9. Auflage. Kiel: Veris.
- Ecalte, J., Bouchafa, H., Potocki, A., & Magnan, A. (2013). Comprehension of Written Sentences as a Core Component of Children's Reading Comprehension. *Journal of Research in Reading, 36*(2), 117-131.
- Fuchs, L. (2017). Curriculum-Based Measurement as the Emerging Alternative: Three Decades Later. *Learning Disabilities Research & Practice, 32*(1), 5-7. DOI: 10.1111/ldrp.12127.
- Fuchs, L. S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research. *School Psychology Review, 33*(2), 188-192.
- Fuchs, L. S., Fuchs, D., 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*(3), 239-256. doi:10.1207/S1532799XSSR0503_3.
- Gebhardt, M. & Jungjohann, J. (2018, im Druck). Digitale Unterstützung bei der Dokumentation von Verhaltens- und Leistungsbeurteilungen. In Meyer, B., Tretter, T., Englisch, U. (Hrsg.), *Praxisleitfaden auffällige Schüler und Schülerinnen*. Weinheim, Basel: Beltz.
- Gebhardt, M., Diehl, K., & Mühling, A. (2016). *Lern-Verlaufs-Monitoring LEVUMI Lehrerhandbuch. Version 1.1*. DOI: 10.17877/DE290R-17792.
- Gebhardt, M., Schwab, S., Krammer, M., Gasteiger-Klicpera, B., & Sälzer, C. (2014). Erfassung von individualisiertem Unterricht in der Sekundarstufe I. Eine quantitative Überprüfung der Skala „Individualisierter Unterricht“ in zwei Schuluntersuchungen in der Steiermark. *Zeitschrift für Bildungsforschung, 4*(3), S. 303-316. DOI: 10.1007/s35834-014-0095-7.
- Jungjohann, J., & Gebhardt, M. (2018). Lernverlaufsdagnostik im inklusiven Anfangsunterricht Lesen – Verschränkung von Lernverlaufsdagnostik, Förderplanung und Wochenplanarbeit. In F. Hellmich, G. Görel & M.F. Löper (Hrsg.), *Inklusive Schul- und Unterrichtsentwicklung* (160-173). Stuttgart: Kohlhammer.
- Jungjohann, J., Gebhardt, M., Diehl, K., & Mühling, A. (2017). *Förderansätze im Lesen mit LEVUMI*. DOI: 10.17877/DE290R-18042.
- Jungjohann, J., Gegenfurtner, A., & Gebhardt, M. (2018). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Lesefähigkeit. *Empirische Sonderpädagogik, 10*(1), 100-118.
- Klapproth, F. (2018). Biased Predictions of Students' Future Achievement: An Experimental Study on Pre-Service Teachers' Interpretation of Curriculum-Based Measurement Graphs. *Studies in Educational Evaluations, 59*, S. 67-75. DOI: 10.1016/j.stueduc.2018.03.004.
- Klauer, K.J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdagnostik, Tests und Trends N.F. Band 12* (1-17). Göttingen: Hogrefe.
- Klauer, K.J. (2011). Lernverlaufsdagnostik – Konzepte, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik, 3*(3), 207-224.
- 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.
- 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*(1), 150-161. DOI: 10.1037/0022-0663.100.1.150.
- Lindsay, G., & Strand, S. (2016). Children with Language Impairment. Prevalence, Associated Difficulties, and Ethnic Disproportionality in an English Population. *Frontiers in Education, 1*, 1102. DOI: 10.3389/educ.2016.00002.

- Lütje-Klose, B. & Mehlem, U. (2015). Die Inklusion mehrsprachiger Kinder in der Grundschule. In Huf, C. Katzenbach, D. & Schnell, I. (Hrsg.) *Inklusive Bildung – Herausforderungen für Erziehungswissenschaft, Fachdidaktik und pädagogische Psychologie*. Bad Heilbrunn: Klinkhardt.
- Lütje-Klose, B. (2012). Sprachlich-kommunikative Beeinträchtigung bei Schülerinnen und Schülern mit dem Förderschwerpunkt Lernen. In Braun, O. & Lütke, U. (Hrsg.) *Sprache und Kommunikation. Behinderung, Bildung und Partizipation. Enzyklopädisches Handbuch der Behindertenpädagogik: Band 8*. Stuttgart: Kohlhammer, 646-652.
- Maier, U. (2014). Computergestützte, formative Leistungsdiagnostik in Primar- und Sekundarschulen. Ein Forschungsüberblick zu Entwicklung, Implementation und Effekten. *Unterrichtswissenschaft, 42(1)*, 69-86.
- Medienberatung NRW (2015): Datenschutz an Schulen in NRW. Handreichung für Schulleitungen. Abrufbar unter: http://www.medienberatung.schulministerium.nrw.de/Medienberatung-NRW/Publikationen/Broschuere_Datenschutz_Schulen_NRW_Final.pdf_download_web.pdf (Abruf: 19.10.18).
- Mühling, A., Gebhardt, M., & Diehl, K. (2017). Formative Diagnostik durch die Onlineplattform Levumi. *Informatik Spectrum, 40(6)*, S. 556-561. DOI: 10.1007/s00287-017-1069-7.
- Newell, K.W., & Christ, T.J. (2017). Novice Interpretations of Progress Monitoring Graphs: Extreme Values and Graphical Aids. *Assessment for Effective Intervention, 42(4)*, 224-236. DOI: 10.1177/1534508417694855.
- OECD (2018), The Resilience of Students with an Immigrant Background: Factors that Shape Well-being, OECD Reviews of Migrant Education, OECD Publishing, Paris, <https://doi.org/10.1787/9789264292093-en>.
- 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(6)*, 427-469. <https://doi.org/10.1016/j.jsp.2009.07.001>.
- Taylor, C.R. (2012). Engaging the struggling reader: Focusing on reading and success across the content areas. *National Teacher Education Journal, 5(2)*, 51-58.
- van den Bosch, R.M., Espin, C.A., Chung, S., & Saab, N. (2017). Data-Based Decision-Making: Teachers' Comprehension of Curriculum-Based Measurement Progress-Monitoring Graphs. *Learning Disabilities Research & Practice, 32(1)*, 46-60. DOI: 10.1111/ldrp.12122.
- Wagner, D., Hammerschmidt-Snidarich, S.M., Espin, C.A., Seifert K. & McMaster K.L. (2017). Pre-service Teachers' Interpretation of CBM Progress Monitoring Data. *Learning Disabilities Research & Practice, 32(1)*, 22-31.
- 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.
- Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R. & Espin, C. A. (2007). Literature Synthesis on Curriculum-Based Measurement in Reading. *The Journal of Special Education, 41(2)*, 85-120. <https://doi.org/10.1177/00224669070410020401>.
- Zeuch, N., Förster, N., & Souvignier, E. (2017). Assessing Teachers' Competencies to Read and Interpret Graphs from Learning Progress Assessment: Results from Tests and Interviews. *Learning Disabilities Research & Practice, 32(1)*, S. 61-70. DOI: 10.1111/ldrp.12126.

Zu den Autoren

Jana Jungjohann hat 2015 das erste Staatsexamen für das Lehramt Sonderpädagogik mit der Fachrichtung Sprache und Kommunikation sowie Lernen abgeschlossen. Seitdem arbeitet sie an der Technischen Universität Dortmund als wissenschaftliche Mitarbeiterin mit den Forschungsschwerpunkten Lernverlaufsdagnostik im Anfangsunterricht Lesen in inklusiven Schulen und Verknüpfung von Lernverlaufsdagnostik, Dateninterpretation und Fördermaterialien.

Prof. Dr. Kirsten Diehl leitet die Abteilung Inklusion und pädagogische Entwicklungsförderung an der Europa-Universität Flensburg. Ihre Arbeitsschwerpunkte sind u. a. Heterogenität und Bildungsungleichheiten im Kontext Sonderpädagogik, Diagnostik und Förderung im Schriftspracherwerb und Lernfortschrittsdiagnostik.

Prof. Dr. Andreas Mühling ist Professor der Arbeitsgruppe Didaktik der Informatik an der Christian-Albrecht-Universität zu Kiel. Seine Arbeitsschwerpunkte sind u. a. Lehren und Lernen des Programmierens, Konzeptionen der Informatik, Educational Data Mining sowie die selbst programmierte Onlineplattform www.levumi.de.

Prof. Dr. Markus Gebhardt ist Professor für die Entwicklung und Erforschung inklusiver Bildungsprozesse an der Technischen Universität Dortmund. Seine Forschungsschwerpunkte liegen in der Lernverlaufsdagnostik (Schwerpunkt: die Onlineplattform www.levumi.de), Fragebogen und Testkonstruktion für die inklusive Schule sowie in der Zusammenarbeit unterschiedlicher Akteure im Feld schulischer Bildung.

Korrespondenzadresse

Jana Jungjohann, M.Ed.
Technische Universität Dortmund
Fakultät Rehabilitationswissenschaften
Entwicklung und Erforschung inklusiver Bildungsprozesse
Emil-Figge-Straße 50
44227 Dortmund
jana.jungjohann@tu-dortmund.de