

Jungjohann, Jana; Ebenbeck, Nikola; Liebers, Katrin; Diehl, Kirsten; Gebhardt, Markus  
**Das Lesescreening LES-IN für inklusive Grundschulklassen. Entwicklung  
und psychometrische Prüfung einer Paper-Pencil-Version als Basis für  
computerbasiertes adaptives Testen (CAT)**

*Empirische Sonderpädagogik 15 (2023) 2, S. 141-156*



Quellenangabe/ Reference:

Jungjohann, Jana; Ebenbeck, Nikola; Liebers, Katrin; Diehl, Kirsten; Gebhardt, Markus: Das Lesescreening LES-IN für inklusive Grundschulklassen. Entwicklung und psychometrische Prüfung einer Paper-Pencil-Version als Basis für computerbasiertes adaptives Testen (CAT) - In: Empirische Sonderpädagogik 15 (2023) 2, S. 141-156 - URN: urn:nbn:de:0111-pedocs-290236 - DOI: 10.25656/01:29023; 10.2440/003-0003

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-290236>

<https://doi.org/10.25656/01:29023>

in Kooperation mit / in cooperation with:

Pabst Science Publishers <https://www.psychologie-aktuell.com/journale/empirische-sonderpaedagogik.html>

#### Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nc/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und das Werk bzw. den Inhalt nicht für kommerzielle Zwecke verwenden.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by-nc/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work, provided that the work or its contents are not used for commercial purposes.

By using this particular document, you accept the above-stated conditions of use.



#### Kontakt / Contact:

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

**Empirische Sonderpädagogik**, 2023.15:141-156

DOI <https://doi.org/10.2440/003-0003>

ISSN 1869-4845 (Print) · ISSN 1869-4934 (ebook)

## Das Lesescreening LES-IN für inklusive Grundschulklassen: Entwicklung und psychometrische Prüfung einer Paper-Pencil-Version als Basis für computerbasiertes adaptives Testen (CAT)

*Jana Jungjohann<sup>a</sup>, Nikola Ebenbeck<sup>b</sup>, Katrin Liebers<sup>c</sup>, Kirsten Diehl<sup>d</sup> und Markus Gebhardt<sup>b</sup>*

<sup>a</sup> TU Dortmund

<sup>b</sup> LMU München

<sup>c</sup> Universität Leipzig

<sup>d</sup> Europa-Universität Flensburg

### Zusammenfassung

Lesescreenings für den inklusiven Unterricht ermöglichen eine ökonomische und faire Testung des Leistungsstandes aller Schüler:innen, um eine datengestützte Hilfe für individuelle Förderentscheidungen zu schaffen. Das hier präsentierte Paper-Pencil-Verfahren wurde als Basis für eine spätere computerbasierte adaptive Testung (CAT) entwickelt. Das Lesescreening LES-IN umfasst die vier Dimensionen *Phonologische Bewusstheit*, *Sicherheit im lexikalischen Abruf*, *Geschwindigkeit im lexikalischen Abruf* und *Satzlesen* mit Aufgaben, die in Gruppensettings durchgeführt werden können. Es soll Schüler:innen mit gering ausgebildeten Voraussetzungen für das Textverständnis identifizieren. Die psychometrische Güte wurde bei  $N = 709$  Schüler:innen der dritten Klassenstufe mit und ohne sonderpädagogischem Unterstützungsbedarf geprüft. Die Dimensionen *Phonologische Bewusstheit* und *Satzlesen* bilden die Personenfähigkeiten inklusiver Lerngruppen ausreichend gut ab, sodass die Ergebnisse für pädagogische Förderentscheidungen nutzbar sind. Beide Dimensionen *Sicherheit* und *Geschwindigkeit im lexikalischen Abruf* benötigen noch weitere mittelschwere Items, um die gewünschte Differenzierung für eine inklusive Lerngruppe zu erreichen. Die Verbesserung und Weiterentwicklung des Screenings, dessen Implementation als frei verfügbarer computerbasierter adaptiver Test und der Einsatz des Screenings in der Schulpraxis werden diskutiert.

*Schlagwörter:* Assessment, Grundschule, Inklusion, Lesen, Screening

## The reading screening LES-IN for inclusive primary classrooms: Development and psychometric evaluation of a paper-pencil version as the basis for computerized adaptive testing (CAT)

### Abstract

Reading screenings for inclusive education allow for economical and fair testing of the achievement levels of all students to create a data-driven resource for individualized decisions. The paper-pencil method that we present in this paper was developed as a pilot version for later computerized adaptive testing (CAT). The reading screening LES-IN includes the four dimensions *phonological awareness* in the narrower sense, *fluency in lexical route*, *speed in lexical route*, and *sentence reading* with tasks that can be administered in group settings. It is designed to identify students with low-level reading skills necessary for text comprehension. Psychometric quality was tested in  $N = 709$  students with and without need for special education in third grades. The dimensions *phonological awareness* and *sentence reading* represent the person abilities of inclusive learning groups sufficiently well so that the results can be used for educational support decisions. Both the dimensions *fluency* and *speed in lexical route* need more moderately difficult items to achieve the desired differentiation for an inclusive learning group. The improvement and further development of the screening, its implementation as a computerized adaptive test, and the use of the screening in school practice are discussed.

*Keywords:* assessment, elementary school, inclusion, reading, screening

Inklusive Grundschulen sollen Kindern den Erwerb einer grundlegenden Bildung ermöglichen. Der Schlüsselkompetenz Lesen, welche eine Basis für alle Bildungsbereiche und die Teilhabe an der Gesellschaft darstellt, kommt eine besondere Bedeutung zu. Allerdings hat sich in den letzten Jahren der Anteil der Schüler:innen, die eine unzureichende Lesekompetenz erwerben, stetig erhöht. 2021 erreichten durchschnittlich 19 Prozent der Viertklässler:innen die Mindeststandards basaler Lesekompetenz nicht, in einzelnen Ländern wie Bremen oder Berlin trifft dies sogar auf fast ein Drittel zu (Stanat et al., 2022). Als Ursachen werden demografische Veränderungen sowie zuletzt der Corona-Lockdown diskutiert. Besonders betroffen sind Schüler:innen mit Migrationshintergrund, niedrigem sozio-ökonomischen Status und sonderpädagogischem Unterstützungsbedarf (SPU), wobei Unterschiede von bis zu mehreren Schuljahren zwischen den Leistungen Gleichaltriger entstehen (Gebhardt et al., 2015, Stanat et al., 2022).

Für eine frühzeitige Prävention im Sinne datenbasierter Förderentscheidungen werden Lesescreenings meist querschnittlich zur Identifizierung von Schüler:innen mit einem Risiko für Leseschwierigkeiten eingesetzt (Arias-Gundín & Llamazares, 2021). Basierend auf den Screeningergebnissen werden Entscheidungen über weitere diagnostische Strategien, die inhaltliche und methodische Ausrichtung und die Intensität zukünftiger Förderungen für einzelne Schüler:innen und Kleingruppen getroffen. Screenings gelten daher als ein wesentliches Instrumentarium im Prozess datenbasierter Förderentscheidungen (Liebers et al., 2019). Je nach Einsatzzweck und Zielgruppe unterscheiden sie sich hinsichtlich des Praxiseinsatzes und der dimensional Struktur (Glover & Alberts, 2007).

Der Schulpraxis fehlt es aktuell an Lesescreenings, welche spezifisch für inklusive Lerngruppen mit einem breiten Entwicklungsspektrum entwickelt und geprüft wurden. Des Weiteren gibt es bislang keine

deutschsprachigen adaptiven Lesescreenings, die durch eine computerbasierte Itemziehung eine besonders ökonomische Testung ermöglichen. Im vorliegenden Beitrag wird das Lesescreening LES-IN für inklusive Klassen als Paper-Pencil-Test mit einer Parallelversion vorgestellt, welche die Basis für die Weiterentwicklung zum computerbasierten adaptiven Test darstellt. Das Screening soll Grundschüler:innen mit noch fehlenden Voraussetzungen für das zu entwickelnde Textverständnis in ökonomischer Weise identifizieren. Es umfasst vier Dimensionen (*Phonologische Bewusstheit, Sicherheit und Geschwindigkeit im lexikalischen Abruf und Satzlesen*), um trotz einer kurzweiligen und gruppenbasierten Durchführung der schulischen Förderpraxis differenzierte Informationen bereitstellen zu können. Das Ziel der Studie ist die psychometrische Prüfung auf Itemebene, um eine Grundlage für einen computerbasierten adaptiven Test mit mehreren Versionen für eine jahrgangsübergreifende Zielgruppe zu schaffen.

## **Anforderungen an die Testkonstruktion eines inklusiven Lesescreenings**

### **Praxiseinsatz**

Inklusive Lesescreenings sollen in heterogenen Leistungsgruppen einschließlich der Leistungen von Schüler:innen mit gravierenden Leseschwierigkeiten möglichst reliabel und fair messen und Ergebnisse für pädagogische Zwecke zurückmelden (Watkins, 2007). Andernfalls werden Schüler:innen entgegen dem Grundgedanken der schulischen Inklusion von der Teilnahme ausgeschlossen. Screenings mit ausschließlich altersbezogenen oder klassenstufenbasierten Normen differenzieren meist nicht ausreichend für Schüler:innen, deren Leistungen von den curricularen oder sozialen Normwerten stark abweichen. Dies betrifft besonders Schüler:innen, die den Schriftspracherwerb mit ungünstigen lese-spez-

fischen Vorläuferfähigkeiten und geringen kognitiven Grundlagen beginnen, da sie in allen Schritten des basalen Lesens Schwierigkeiten entwickeln können (Nagler et al., 2018, Torgesen, 2002, Valtin et al., 2015). Dadurch besteht die Gefahr, dass deren Leistungen nicht ausreichend reliabel erfasst werden. Um sicherzustellen, dass Subtests fair gegenüber allen Schüler:innen sind, wird die Prüfung der Testgüte mittels der Item-Response-Theorie (IRT) empfohlen (Klauer, 1978). Hierfür eignen sich Raschmodelle, um Aussagen auf Itemebene sowie zur Dimensionalität der Skalen treffen zu können (Rost, 2004).

Papierbasierte Lesescreenings sind in der Grundschule etablierter als computergestützte Tests (Meile & Liebers, 2022). Sie bieten meist wenige Parallelversionen an und werden einmalig oder mit großen zeitlichen Abständen im Sinne einer wiederholten Statusdiagnostik eingesetzt (Jungjohann & Gebhardt, 2023). Computergestützte Screenings stellen ein neueres Medium dar und ermöglichen eine zufällige Itemziehung (Klauer, 2006) oder einen adaptiven Ziehalgorithmus von Items, der auf den Antworten der Schüler:innen basiert (Ebenbeck & Gebhardt, 2022). Dadurch werden diskriminierungsfreie Testsituationen möglich (Liebers et al., 2019). Sie generieren in ökonomischer Weise faire Paralleltestversionen für einen hochfrequenten Einsatz in der Praxis. Zusätzlich reduzieren adaptive Screenings die Anzahl der Testaufgaben und somit die Testzeit, wodurch die Kapazitäten von Schüler:innen mit geringer Konzentrationsdauer nicht überlastet werden (Schurig et al., 2021). Allerdings setzen computerbasierte Tests voraus, dass die Lernenden neben den Aufgaben auch die Endgeräte eigenständig handhaben können (Liebers et al., 2019).

### **Lesebezogene Dimensionen**

Die Auswahl der Dimensionen und Aufgabenstellungen eines Screenings bestimmt, welche Teilprozesse des Lesens abgebildet

werden und wie hoch der informative Gehalt für die Förderplanung ist. Lesescreenings mit mehreren Dimensionen liefern differenziertere Informationen als Screenings mit nur einem stellvertretenden Aufgabentyp, da sie mehrere Teilprozesse der Kompetenz durch unterschiedliche Aufgabenstellungen abbilden können (Seifert et al., 2022). Für ein ökonomisches Lesescreening können nur wenige Dimensionen ausgewählt werden, die gleichzeitig möglichst viele Schüler:innen mit Schwierigkeiten in unterschiedlichen lesebezogenen Lernschritten identifizieren sollen.

Um Schüler:innen mit gering ausgebildeten Voraussetzungen für das Textverständnis zu identifizieren, bietet sich die Messung unterschiedlicher proximaler Vorläuferfähigkeiten und basaler Lesefähigkeiten mit einem starken Einfluss auf die allgemeine Lesekompetenz an. Denn sicher erworbene basale Lesekompetenzen auf Wort- und Satzebene schaffen freie Kapazitäten für die Sinnkonstruktion auf Textebene (Perfetti, 2007). Nach Lenhard (2019) zählen dazu die phonologische Bewusstheit, der lexikalische Abruf von Wörtern sowie die Bildung und das Verständnis der lokalen Kohärenz zwischen Sätzen.

### *Phonologische Bewusstheit*

Bei der Auseinandersetzung mit gesprochener und geschriebener Sprache müssen Schüler:innen Informationen über die Lautstruktur nutzen. Diese phonologische Informationsverarbeitung (Wagner & Torgesen, 1987) wird in vier Teilbereiche unterschieden: die phonologische Bewusstheit im engeren und weiteren Sinn, das phonologische Rekodieren im Arbeitsgedächtnis und das phonologische Rekodieren beim Zugriff auf das semantische Lexikon. Die phonologische Bewusstheit im engeren Sinn (Phonemebene) gilt als signifikanter Prädiktor für den Schriftspracherwerb in der Schuleingangsphase und darüber hinaus für die Lesegeschwindigkeit, die Lesegenauigkeit und das Leseverständnis (Pfost, 2015). Ob-

wohl bei deutschsprachigen Drittklässler:innen Leseschwierigkeiten vermehrt durch Funktionen des phonologischen Kurz- und Langzeitgedächtnisses bedingt werden (Landerl & Wimmer, 2000), zeigen einzelne Schüler:innen nach der zweiten Klasse immer noch unzureichende phonologische Fähigkeiten und damit ein erhöhtes Risiko für Lese-Rechtschreib-Schwierigkeiten (Schnitzler, 2015).

Für die Diagnose der phonologischen Bewusstheit weisen etliche Verfahren standardisierte Vergleichswerte für die Schuleingangsphase auf (z. B. ILEA plus, Gottlieb et al., 2022; TEPHOBE, Mayer, 2020). Ab Klassenstufe drei existieren nur noch wenige normierte Verfahren (BAKO 1-4; Stock et al., 2003). Um die phonologische Bewusstheit zu messen, werden beispielsweise Aufgaben zur An-, Mit- und Endlauterkennung, Phonemsynthese, -vertauschung oder -ersetzung, zu Reimen und zur Lautkategorisierung sowie Aufgaben zum phonologischen Wortgedächtnis genutzt (Schnitzler, 2015).

### *Sicherheit im lexikalischen Abruf*

Der Wortschatz ist im mentalen Lexikon gespeichert und wird in den Umfang des rezeptiven und des expressiven Wortschatzes sowie in qualitative Aspekte des semantischen, relationalen und formalen Wortwissens differenziert (Juska-Bacher & Jakob, 2014). Im Allgemeinen gilt ein unzureichender Wortschatz als limitierender Faktor für den Leseerwerb und das Leseverständnis (Röthlisberger et al., 2021). Für deutschsprachige Schüler:innen wurden unterschiedlich starke Einflüsse nachgewiesen, die über die verschiedenen Messmethoden begründet werden (Juska-Bacher et al., 2022). Für die dritte Jahrgangsstufe berichteten beispielsweise Berendes et al. (2010) von einer mittelstarken Korrelation zwischen rezeptivem Wortschatzumfang und Leseleistungen.

Deutschsprachige Wortschatztests stützen sich auf unterschiedliche Dimensionen.

Aufgaben zum rezeptiven Wortschatz für Gruppen unterscheiden sich dahingehend, ob die Versuchsleitung das Zielwort vorspricht und die Schüler:innen ein passendes Bild ankreuzen (z. B. GraWo, Seifert et al., 2017) oder lexikalische Entscheidungsaufgaben genutzt werden (ProDi-L; Richter et al., 2012, WOR-TE; Trautwein & Schroeder, 2018). Bei Entscheidungsaufgaben lesen die Schüler:innen leise eine Liste an (Pseudo-)Wörtern. Diese Aufgaben sind für die praktische Umsetzung im Unterricht gut geeignet, da viele Items in kurzer Zeit mit geringem kognitiven Aufwand im Gruppen-setting dargeboten werden können.

### *Geschwindigkeit im lexikalischen Abruf*

Die Geschwindigkeit des Wortabrufs aus dem mentalen Lexikon bildet eine Voraussetzung für die Ausbildung einer angemessenen Leseflüssigkeit (Ennemoser et al., 2012). Nach dem Dual-Route-Modell (Coltheart et al., 2001) erfolgt das Wortlesen entweder über den direkten Weg des Ganzwortabrufs aus dem sogenannten Sichtwortschatz oder über den indirekten Weg des Rekodierens. Beide Wege unterscheiden sich hinsichtlich Geschwindigkeit und Effizienz. Eine hohe Geschwindigkeit des Wortabrufs aus dem Sichtwortschatz über die direkte Route gilt in allen Klassenstufen als Voraussetzung für eine angemessene Leseflüssigkeit, da sie die für das Lesen benötigte Kapazität des Arbeitsgedächtnisses entlastet (Perfetti, 2007). Silverman et al. (2013) berichten, dass Viertklässler:innen mit guter Dekodierfähigkeit und einer geringen Leseflüssigkeit niedrigere Leistungen im Leseverständnis zeigen als Schüler:innen mit einer hohen Leseflüssigkeit.

In der deutschsprachigen Schulpraxis werden Vorleseaufgaben im Einzelsetting in Screenings (z. B. Eine-Minute-Wortlesetest (IEL-1), Diehl & Hartke, 2021) oder als Lernverlaufdiagnostik (Levumi.de, Jungjohann et al., 2018) als robuster Indikator für die Leseflüssigkeit eingesetzt. Leise-Leseaufgaben für Gruppen, bei denen die Geschwin-

digkeit der Worterkennung gemessen wird, werden in den Individuellen Lernstandsanalysen ILEA plus (Gottlebe et al., 2022) oder in der Förderpraxis (z. B. Blitzlesen; Spilles & Leidig 2020, Mayer, 2018) eingesetzt.

### *Leseverständnis auf Satzebene*

Zum Verständnis von Sätzen müssen sowohl die semantischen Bedeutungen der einzelnen Wörter als auch die syntaktischen Strukturen gemeinsam für die Sinnkonstruktion interpretiert werden (Lenhard, 2019). Das Leseverständnis auf Satzebene hat eine vermittelnde Funktion zwischen Wort- und Textverständnis und liefert als robuster Indikator Hinweise auf das generelle Leseverständnis (Ecalte et al., 2013). Schüler:innen mit Leseschwierigkeiten lesen im Vergleich zu ihren Mitschüler:innen auf Satzebene signifikant langsamer und fehlerhafter (Stanat et al., 2022). Für Schüler:innen mit SPU trifft dies umso mehr zu (Lucas & Norbury, 2015).

Um das Leseverständnis auf Satzebene zu screenen, werden neben Satzverifikationsaufgaben Maze-Aufgaben eingesetzt (Graney et al., 2010), bei denen kurze Textpassagen oder Sätze mit Wortlücken präsentiert werden. Aus einer Auswahl von bis zu vier Wörtern wählen die Schüler:innen ein Wort aus, um den Satz zu vervollständigen. Dieser Aufgabentyp eignet sich nach Seifert et al. (2022) insbesondere für gruppenbasierte und digitale Screenings, da sie eine schnelle Durchführung und Auswertung ermöglichen. Die Schwierigkeit der Maze-Aufgaben wird von weniger Determinanten beeinflusst als bei anderen Aufgabentypen. Für den deutschsprachigen Raum liegen papierbasierte Tests mit Normwerten (ELFE II, Lenhard et al., 2017; VSL, Walter, 2013) sowie nicht kommerzielle Tests in webbasierten Plattformen als Computertestungen vor (Blumenthal et al., 2022; ILEA plus; Gottlebe et al., 2022).

## Fragestellung

Das übergeordnete Ziel der Arbeitsgruppe ist die Konstruktion des computerbasierten adaptiven Lesescreenings LES-IN, welches in inklusiven Lerngruppen Schüler:innen mit unzureichenden basalen Lesekompetenzen für das Textverständnis identifiziert. Das Screening soll in Gruppensettings mehrere lesebezogene Dimensionen mit einer geringen Testzeit möglichst ökonomisch messen, um zusätzlich zur Identifikation differenzierte Informationen für die Förderpraxis bereitstellen zu können. Dafür werden ein Itempool mit Ziehungsregeln, welcher den Anforderungen des Raschmodells für einzelne Messungen entspricht, und eine aufwendige Programmierung mit adaptivem Ziehalgorithmus benötigt. Für die Entwicklung des LES-IN wurde als erster Schritt eine papierbasierte Parallelversion entworfen, welche vier lesebezogene Dimensionen als eigenständigen Gruppentest umfasst. In der vorliegenden Studie wird diese Version pilotiert. Die gewählten Aufgabenstellungen sind aus der Test- und Förderpraxis abgeleitet und sowohl papierbasiert als auch digital anwendbar.

Die zentralen Fragestellungen zur psychometrischen Prüfung auf Itemebene im Sinne der IRT lauten:

1. Liegt für jede einzelne Dimension eine ausreichende Passung des Raschmodells vor?
2. Kann das vollständige Lesescreening innerhalb eines Raschmodells interpretiert werden?

## Methode

### Stichprobe und Durchführung

Nach einer Pilotierung in zwei inklusiven Klassen bearbeiteten 709 Schüler:innen aus dritten oder jahrgangsgemischten (3/4) Klassen (45,7% weiblich,  $M_{\text{Alter}} = 10,06$  Jahre,  $SD_{\text{Alter}} = 0,82$ ) in einer Welle (Juni/Juli 2022)

das Screening als Paper-Pencil-Version im Klassenverband (siehe für Beispielaufgaben Abbildung 1). 564 Schüler:innen besuchten inklusive Grundschulen aus drei Bundesländern (ca. 5% Schüler:innen mit SPU). Um ausreichend Schüler:innen mit SPU für die Prüfung der Testfairness in die Stichprobe einzubeziehen, bearbeiteten zusätzlich 145 Schüler:innen (71% Lernen, 14% Sprache, 15% emotionale und soziale Entwicklung) aus Förderschulen das Lesescreening. Basierend auf der Pilotierung wurde die maximale Bearbeitungszeit auf 20 Minuten (5 Minuten pro Dimension) als Abbruchkriterium festgelegt und uneindeutige Items wurden entfernt. Geschulte wissenschaftliche Mitarbeitende leiteten die Testungen und alle Schüler:innen nahmen freiwillig und mit dem Einverständnis der Erziehungsberechtigten teil.

### Vorstellung des inklusiven Lesescreenings LES-IN

Die Testbeschreibung inklusive der regelgeleiteten Itemkonstruktion ist unter Ebenbeck et al. (2022) einsehbar.

Die Dimension *Phonologische Bewusstheit* (51 Items) wurde in Anlehnung an Pfost (2015) konstruiert. Die Schüler:innen sehen ein Bild und einen Ziellaut, den sie innerhalb des Wortes lokalisieren müssen. Die Schüler:innen kreuzen an, ob der Ziellaut am Anfang, in der Mitte oder am Ende des Wortes oder gar nicht verlautet wird. Alle Bilder stellen Nomen dar. Die angegebenen Phoneme sind im Wort lautgetreu und durch eine Lautanalyse erschließbar.

Die Aufgaben zur Dimension *Sicherheit im lexikalischen Abruf* (42 Items) greifen die Ja/Nein-Methode ohne Zeitbegrenzung auf (Richter et al., 2012, Trautwein & Schroeder, 2018). Die Schüler:innen bekommen ein Schriftbild präsentiert und entscheiden durch Ankreuzen, ob dieses Wort existiert. Die Wörter sind Nomen, Verben, Adverbien, Adjektive, Konjunktionen und Partikel. Die verwendeten Pseudowörter wurden mit Hilfe des Pseudowortgenerators Wuggy

(Keuleers & Brysbaert, 2010) auf Grundlage des Grundwortschatzes generiert.

Die Leise-Leseaufgaben der Dimension *Geschwindigkeit im lexikalischen Abruf* (30 Items) lehnen sich an die Übungs- und Förderpraxis (Spilles & Leidig, 2020) an, um eine möglichst hohe Praktikabilität für das Feld zu wahren. In der Durchführung werden den Schüler:innen Nomen, Verben, Adverbien, Adjektive und Konjunktionen einzeln für zwei Sekunden computergestützt über eine Präsentation mittels Beamer visuell durch die Testleitung präsentiert. Die Schüler:innen wählen aus vier ähnlichen Antworten eine aus.

Der Itempool zur Dimension *Satzlesen* setzt sich aus einer Auswahl von 75, bereits in früheren Schulstudien geprüften, Items mit unterschiedlichen linguistischen Satzstrukturen (Sinn-L-Levumi Niveaustufen N2, N4 und N6; Jungjohann et al., 2021, Jung-

johann, 2022) zusammen. Alle Items sind in Anlehnung an Maze-Aufgaben (Graney et al., 2010) konstruiert. Den Lesenden werden ein bis zwei Sätze mit einer Wortlücke präsentiert. Die Schüler:innen wählen aus vier Auswahlmöglichkeiten ein Wort aus.

**Item- und Modellanalysen**

Alle Analysen werden mit GNU R (R Core Team, 2013) durchgeführt. Zuerst wird die Dimensionalität der *einzelnen Dimensionen des Lesescreenings* als unabhängige Tests mittels Raschanalysen (R-Paket pairwise, Heine, 2021) geprüft. Gute Werte für das mittlere Quadrat (MSQ) der Outfit- und Infit-Maße für eine heterogene Stichprobe liegen zwischen 0,5 und 1,5 (Wright & Linacre, 1994). Itemwerte unter 0,5 sowie über 1,5 sind für die Messung wenig ergiebig, während Werte über 2,0 als schädlich

**Phonologische Bewusstheit**

	A	<input type="checkbox"/> Anfang	<input type="checkbox"/> Mitte	<input type="checkbox"/> Ende	<input type="checkbox"/> Nirgends
------------------------------------------------------------------------------------	---	------------------------------------	-----------------------------------	----------------------------------	--------------------------------------

**Sicherheit im lexikalischen Abruf**

Nohnseher	Gibt es <input type="checkbox"/>	Gibt es nicht <input type="checkbox"/>
-----------	-------------------------------------	-------------------------------------------

**Geschwindigkeit im lexikalischen Abruf**

<input type="checkbox"/> Montag	<input type="checkbox"/> Mieter	<input type="checkbox"/> Mittel	<input type="checkbox"/> Mittag
------------------------------------	------------------------------------	------------------------------------	------------------------------------

**Sinnkonstruierendes Satzlesen**

Die _____ blühen auf der Wiese.			
<input type="checkbox"/> Vögel	<input type="checkbox"/> Jungen	<input type="checkbox"/> Blumen	<input type="checkbox"/> Stühle

**Abbildung 1**  
Aufgabentypen der vier Dimensionen



für die Messung angesehen werden. Zur Prüfung der Testfairness werden graphische Modelltests mit dem Splitkriterium SPU durchgeführt. Um die Reliabilität der Dimensionen zu bestimmen, wird Cronbachs Alpha pro Dimension berechnet.

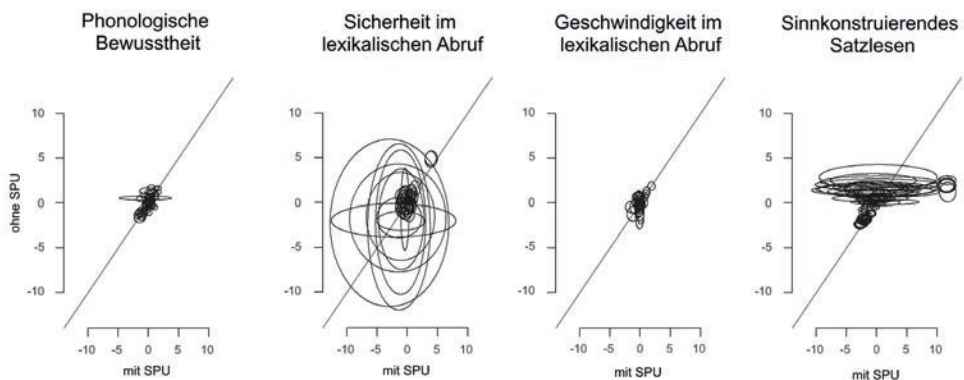
Zur Prüfung der Dimensionalität des gesamten Screenings werden ein- und multidimensionale Raschmodelle (R-Paket TAM, Robitzsch et al., 2022) gegenübergestellt. Es werden vier Modelle angenommen, die aus den Testformaten der Subtests abgeleitet wurden. Modell 1 ist ein eindimensionales Raschmodell. Modell 2 ist ein zweidimensionales Raschmodell, bei dem die Dimension Geschwindigkeit im lexikalischen Abruf aufgrund des Antwortformates ohne Zeitbegrenzung einzeln betrachtet wird. Modell 3 ist ein vierdimensionales Raschmodell, bei dem für jeden Aufgabentyp eine eigene Dimension angenommen wird. Modell 4 ist sechsdimensional. Es umfasst die Aufgaben der drei ersten Dimensionen sowie die Itempools der drei Niveaustufen N2, N4 und N6 des sinnkonstruierenden Satzlesens als jeweils eigene Dimensionen. Alle Modelle werden anhand des Akaike Informationskriteriums (AIC) und des Bayes'schen Informationskriteriums (BIC) verglichen. Niedrigere AIC- und BIC-Werte weisen auf eine bessere Modellpassung hin (Aho et al., 2014).

## Ergebnisse

### Item- und Modellanalysen der einzelnen Tests

In der Dimension *Phonologische Bewusstheit* waren 15 Items im Hinblick auf die Stichprobeninvarianz auffällig (Abbildung 2) und wurden aus dem Itempool entfernt. Letztendlich wurden 36 Items in das Raschmodell einbezogen. Die Items haben eine Schwierigkeit zwischen -1,49 und 1,80. Die Infit-Werte liegen zwischen 0,67 und 1,21 und die Outfit-Werte zwischen 0,40 und 1,35. In Bezug auf die Personenfähigkeiten der Stichprobe sind die Items eher leicht zu bearbeiten und erfassen dementsprechend Schüler:innen mit geringen Fertigkeiten differenziert (Abbildung 3).

In der Dimension *Sicherheit im lexikalischen Abruf* wurden acht nicht stichprobeninvariante Items (Abbildung 2) entfernt. 34 Items wurde in das Raschmodell einbezogen, deren Itemschwierigkeit zwischen -2,59 und 4,46 liegt. Die breite Spannweite in der Schwierigkeit resultiert maßgeblich aus den im Test vorhandenen Pseudowörtern ( $r_i = 0,77-4,46$ ; Abbildung 3). Die Infit-Werte (0,43-0,85) und Outfit-Werte (0,07-1,29) liegen alle im ergiebigen Bereich. Die Pseudowörter werden als wesentlicher Bestandteil und als Kontrollmöglichkeit der



**Abbildung 2**

Graphischer Modelltest zur Gegenüberstellung der Itemschwierigkeit für Schüler:innen mit und ohne SPU

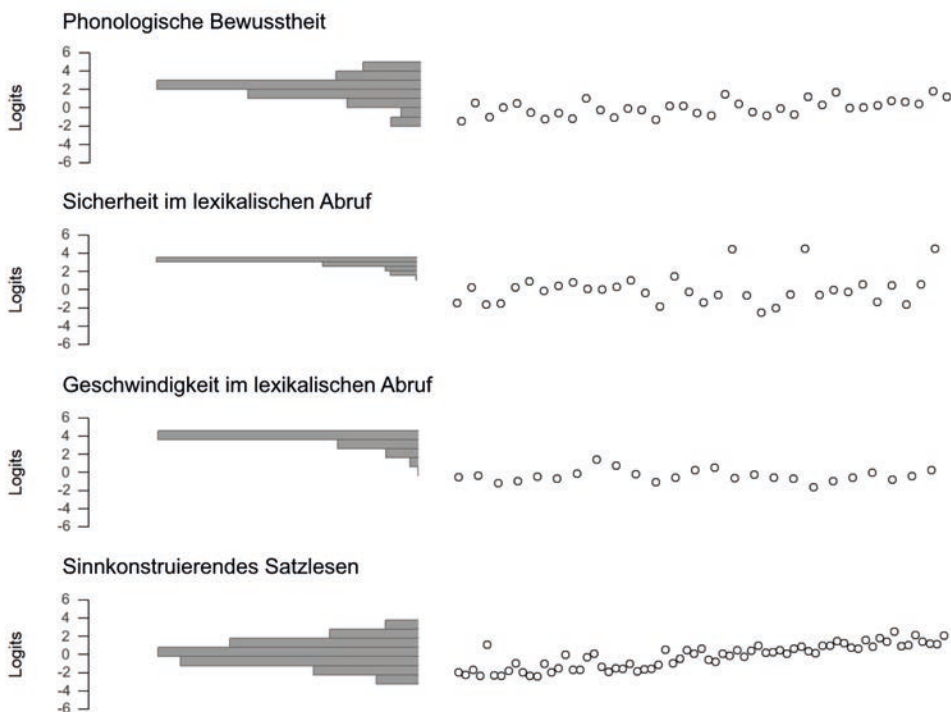
Aufgabe angesehen, sodass alle 34 Items im Pool verbleiben.

In der Dimension *Geschwindigkeit im lexikalischen Abruf* waren fünf Items nicht invariant (Abbildung 2). Diese wurden entfernt. Der reduzierte Itempool umfasst 25 Items mit einer Schwierigkeit von -1,26 bis 1,82. Die Fitwerte zeigen, dass alle Items für die Messung ergebnisfähig sind (Infit: 0,45–1,11, Outfit: 0,15–0,72). Die Person-Item-Map (Abbildung 3) verdeutlicht allerdings, dass die Items das Kompetenzniveau der Stichprobe nur unzureichend abdecken. Die Aufgaben sind für die Schüler:innen zu leicht und differenzieren dadurch vorwiegend im niedrigen Kompetenzbereich.

Beim Stichprobeninvarianztest der Dimension *Satzlesen* wurden sechs auffällige Items entfernt (Abbildung 2). Die verbleibenden 69 Items zeigen eine ansteigende Schwierigkeit von -2,18 bis 2,74 (Abbil-

dung 3). Diese ansteigende Schwierigkeit wird durch frühere Evaluationen gestützt (Jungjohann, 2022) und war durch das Matching aus unterschiedlichen Niveaustufen erwartbar. Der vorliegende Test ist demnach geeignet, sowohl leseschwächere als auch lesestärkere Schüler:innen reliabel zu erfassen. Die Items haben Infit-Werte zwischen 0,57 und 1,43 sowie Outfit-Werte zwischen 0,36 und 1,93.

Die interne Konsistenz der Dimensionen nach Entfernung der Items ist hoch (Phonologische Bewusstheit:  $\alpha = .93$ , Sicherheit im lexikalischen Abruf:  $\alpha = .75$ , Geschwindigkeit im lexikalischen Abruf:  $\alpha = .95$ , Sinnkonstruierendes Satzlesen:  $\alpha = .97$ ).



**Abbildung 3**

*Person-Item-Map der Dimensionen des Lesescreenings*

### Item- und Modellanalysen des Lesescreenings LES-IN

Das angenommene vierdimensionale Raschmodell (Modell 3) hat die niedrigsten AIC- und BIC-Werte (Tabelle 1), woraus die beste Modellpassung abgeleitet wird. Die bessere Passung des Modells 3 gegenüber Modell 1 ( $X^2 [9] = 3721,65$ ),  $p < 0,001$ ) und Modell 2 ( $X^2 [7] = 2982,44$ ),  $p < 0,001$ ) sowie die geringere Passung des Modells 4 gegenüber Modell 3 ( $X^2 [11] = 0,71$ ),  $p > 0,9$ ) werden durch die Likelihood-Quotienten-Tests bestätigt.

Innerhalb des vierdimensionalen Modells 3 decken die Items den gesamten Fähigkeitsbereich der Stichprobe ab, da ihre Schwierigkeit erwartungsgemäß breit streut (Abbildung 4). Der unterste Fähigkeitsbereich, der Schüler:innen mit gravierenden Leseschwierigkeiten zugeschrieben werden kann, wird durch die leichten Items der Dimension *Geschwindigkeit im lexikalischen Abruf* am besten abgedeckt. Der hohe Fähigkeitsbereich wird ausreichend durch die Items des *Satzlesens* bedient. Die Items der Dimensionen *Sicherheit im lexikalischen*

*Abruf* und *Phonologische Bewusstheit* decken einen annähernd ähnlichen Fähigkeitsbereich ab, allerdings schließen die Items der *Phonologischen Bewusstheit* die Lücke im mittleren Fähigkeitsbereich zwischen den Items der *Geschwindigkeit im lexikalischen Abruf* und des *Satzlesens* mit größerer Überlappung.

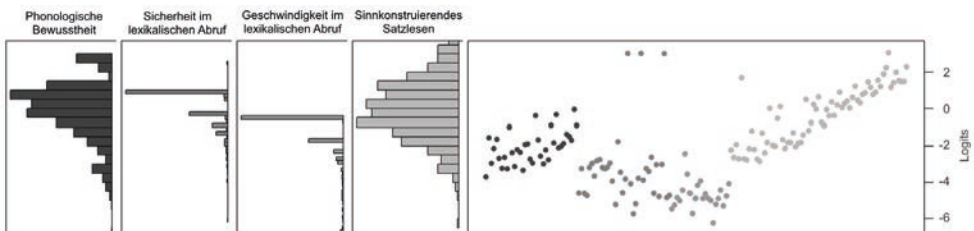
### Diskussion

Für datenbasierte Förderentscheidungen in der inklusiven Grundschule werden reliable Diagnoseinstrumente benötigt, die sowohl das breite Leistungsspektrum der gesamten Lerngruppe erfassen als auch zeitökonomisch sind (Liebers et al., 2019). Dafür wurde in der vorliegenden Studie ein neu konzipiertes Lesescreening LES-IN mit vier Dimensionen (*Phonologische Bewusstheit*, *Sicherheit im lexikalischen Abruf* und *Geschwindigkeit im lexikalischen Abruf* sowie *Satzlesen*) für inklusive Klassen entwickelt und als Paper-Pencil-Test psychometrisch geprüft. Für die Testkonstruktion galt es einerseits, analoge Aufgabenformate für Gruppensettings zu gestalten

**Tabelle 1**

*Graphischer Modelltest zur Gegenüberstellung der Itemschwierigkeit für Schüler:innen mit und ohne SPU*

Modell	AIC	BIC	Expected-A-Posteriori-Reliabilität
1	52768.78	53521.82	0.93
2	52033.57	52795.74	0.93 / 0.56
3	49065.13	49859.24	0.85 / 0.56 / 0.75 / 0.89
4	49086.42	49930.74	0.85 / 0.60 / 0.74 / 0.85 / 0.88 / 0.88



**Abbildung 4**

*Person-Item-Map des Lesescreenings als vierdimensionales Raschmodell*

ten, die nachfolgend auch für einen noch ökonomischeren, computerbasierten adaptiven Test eingesetzt werden können. Andererseits sollte eine Auswahl weniger Dimensionen erfolgen, um Schwierigkeiten zu erkennen und Testzeiten für Schüler:innen mit Konzentrationsschwierigkeiten möglichst gering zu halten. Die Ergebnisse der multidimensionalen Raschanalysen bestätigen ein vierdimensionales Screening. Für die Schulpraxis macht dies eine getrennte Nutzung der vier Subtests möglich.

Eine Weiterentwicklung des Screenings zu einem zusammenhängenden Test kann nach einer testtheoretischen Anpassung der Schwierigkeit einzelner Dimensionen oder mittels einer förderpraktisch begründeten reduzierten Auswahl der Dimensionen erfolgen. In der aktuellen Version messen die beiden Dimensionen *Sicherheit* und *Geschwindigkeit im lexikalischen Abruf* ausschließlich Leistungen von Schüler:innen mit sehr geringen Fähigkeiten differenziert. Diese Einzeltests können daher als Erweiterung des Testrepertoires in der Schulpraxis genutzt werden.

Aus testtheoretischer Sicht kann der Anteil schwierigerer Items erhöht werden, um den Fähigkeitsbereich der Stichprobe besser abbilden zu können. In der Dimension *Sicherheit im lexikalischen Abruf* wurde zur Messung des lexikalischen Abrufs eine Entscheidungsaufgabe in Anlehnung an Richter et al. (2012) sowie Trautwein und Schroeder (2018) konstruiert, die Rückschlüsse auf die Qualität der orthographischen Vergleichsprozesse ermöglicht. Die Schwierigkeit der Items unterscheidet sich stark zwischen den leichteren Wörtern ( $r_i = -2,60-1,39$ ) und den schwierigeren Pseudowörtern ( $r_i = 0,77-4,46$ ). Dadurch entsteht auch eine niedrigere interne Konsistenz der Items dieser Dimension ( $r = .75$ ) im Vergleich zu den übrigen Dimensionen ( $r > .93$ ). Durch eine hochfrequenter Präsentation von schwierigen Pseudowörtern reduziert sich der Anteil leichter Items und die Schwierigkeit des Tests erhöht sich. Zur Messung der Geschwindigkeit des Wortabrufs wurde in

der Dimension *Geschwindigkeit im lexikalischen Abruf* die Fördermethode Blitzlesen (Mayer, 2018, Spilles & Leidig, 2020) in ein Testformat transferiert. Diese Transformation wurde im Lesescreening LES-IN erstmalig erprobt und basiert auf der Idee, den Schüler:innen möglichst bekannte Aufgabenformate anzubieten. In Abgrenzung zu den Leise-Leseaufgaben zur Worterkennung von ILEA plus (Gottlebe et al., 2022) bestand zudem die Hoffnung, den Einfluss des Wortschatzumfangs in dieser Dimension stärker zu minimieren. Die Anzeigedauer der Zielwörter wurde in Anlehnung an das Verfahren ILEA plus (Gottlebe et al., 2022) auf zwei Sekunden festgelegt. Um die Schwierigkeit dieses Tests besser an die Personfähigkeiten anzunähern, ist es denkbar, die Präsentationsdauer zu verkürzen, wie es sich in computergestützten Trainingsprogrammen bereits als reliabel erwiesen hat (Konerding et al., 2021). Durch eine geringere Lesezeit können langsamere Leser:innen das Wort über die indirekte Route (Coltheart et al., 2001) nicht ausreichend erfassen und die Itemschwierigkeit steigt.

Aus Sicht der Förderpraxis kann ein praktikables Ineinandergreifen von Diagnostik und Förderung (Jungjohann et al., 2022) durch die Streichung der Dimension *Sicherheit im lexikalischen Abruf* ermöglicht werden. Die psychometrische Prüfung bestätigt die Eindimensionalität der Dimensionen *Sicherheit* und *Geschwindigkeit im lexikalischen Abruf*, allerdings testen ihre Aufgaben Schüler:innen mit einem ähnlichen Fähigkeitsniveau. Eine Kürzung des Screenings auf drei Dimensionen ermöglicht Klassenlehrkräften inklusiver Lerngruppen eine praktikable Einteilung in drei Fördergruppen mit Schwerpunktsetzung mit Übungen zur Phonologischen Bewusstheit, zum lexikalischen Abruf und zum Satzverständnis. In der schulischen Förderpraxis sind Treppenstufenmodelle mit dreigeteilter Förderstruktur verbreitet (Arias-Gundín & García Llamazares, 2021). Ohne weitere testtheoretische Adaption bildet aktuell die Dimension *Geschwindigkeit im lexi-*

*kalischen Abruf* die niedrigste, die *Phonologische Bewusstheit* die mittlere und das *Satzlesen* die höchste Treppenstufe des Fähigkeitsbereichs der Schüler:innen ab. Die Aufgaben zur *Phonologischen Bewusstheit* wurden als eigenständige Dimension in Anlehnung an Lenhard (2019) entwickelt, da Schwierigkeiten in der phonologischen Verarbeitung zu den wichtigsten proximalen Vorläuferfähigkeiten von Leseschwierigkeiten zählen und das Leseverständnis am Ende der Grundschule beeinflussen (Schnitzler, 2015, Pfost, 2015). Daher wurde für diese Dimension die geringste Schwierigkeit erwartet. Im dimensionalen Vergleich (Abb. 4) wurde allerdings eine mittlere Aufgabenschwierigkeit beobachtet, welche durch die fehlende Schwierigkeit in den Dimensionen zum *lexikalischen Abruf* erklärt werden kann. Für eine datenbasierte Förderentscheidung (Blumenthal et al., 2022) sollte eine dimensionenbasierte Gruppeneinteilung kontinuierlich durch Instrumente der Lernverlaufsdiagnostik (Jungjohann et al., 2018) begleitet und geprüft werden, um die Passung zwischen den aktuellen Lernschritten und dem Förderangebot sicherzustellen.

Die vorliegende Studie präsentiert die erste Evaluierung des papierbasierten Lesescreenings LES-IN, die mit Limitationen behaftet ist. Die nichtrepräsentative Gelegenheitsstichprobe wurde über die Schulnetzwerke der beteiligten Projektpartner:innen akquiriert und sollte durch ein breites Altersspektrum der Schüler:innen erweitert werden. Die Raschanalysen deuten darauf hin, dass die Dimensionen *Sicherheit* und *Geschwindigkeit im lexikalischen Abruf* in der aktuellen Fassung eher für Schüler:innen mit geringeren Fähigkeiten geeignet sind als für diejenigen in der vorliegenden Stichprobe.

Für die im Projektverlauf geplante Folge-studie zum Screening können die Dimensionen passgenau für die Zielgruppe überarbeitet werden. Ziel ist es, die einzelnen Tests sowie das gesamte Screening LES-IN als computerbasierten adaptiven Test in Levumi.de (Jungjohann et al., 2018) zu

realisieren und der Schulpraxis kostenfrei bereitzustellen. In der Folge-studie werden das digitalisierte Screening an einer neuen, breiteren Stichprobe geprüft und die Aufgaben mit etablierten Lesetests validiert. Bereits jetzt kann die Schulpraxis von den vorhandenen Entwicklungsschritten profitieren, weil die einzelnen Subtests vorab über Levumi.de zur Verfügung gestellt werden. So können ab sofort die Fähigkeiten von Drittklässler:innen mit einem Risiko für manifeste Leseschwierigkeiten erfasst werden, um unzureichender Lesekompetenz eher zu begegnen.

## Literatur


- Aho, K., DeWayne, D. & Peterson, T. (2014). Model selection for ecologists. The Worldviews of AIC and BIC. *Ecology*, 95(3), 631–636.
- Arias-Gundín, O. & García Llamazares, A. (2021). Efficacy of the RtI Model in the Treatment of Reading Learning Disabilities. *Education Sciences*, 11(5), 209. <https://doi.org/10.3390/educsci11050209>
- Berendes, K., Schnitzler, C. D., Willmes, K. & Huber, W. (2010). Die Bedeutung von Phonembewusstheit und semantisch-lexikalischen Fähigkeiten für Schriftsprachleistungen in der Grundschule. *Sprache · Stimme · Gehör*, 34(3), 33–41. <https://doi.org/10.1055/s-0029-1246203>
- Blumenthal, S., Gebhardt, M., Förster, N. & Souvignier, E. (2022). Internetplattformen zur Diagnostik von Lernverläufen von Schülerinnen und Schülern in Deutschland. Ein Vergleich der Plattformen Lernlinie, Levumi und quop. *Zeitschrift für Heilpädagogik* 73(4), 153–167. <https://doi.org/10.5283/epub.52069>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. <https://doi.org/10.1037/0033-295X.108.1.204>
- Diehl, K. & Hartke, B. (2012). *IEL-1 - Inventar zur Erfassung der Lesekompetenzen von Erstklässlern*. Hogrefe.


- Ebenbeck, N., & Gebhardt, M. (2022). Simulating computerized adaptive testing in special education based on inclusive progress monitoring data. *Frontiers in Education*, 7, Article 945733. <https://doi.org/10.3389/educ.2022.945733>
- Ebenbeck, N., Jungjohann, J. & Gebhardt, M. (2022). *Testbeschreibung des Lesescreenings LES-IN für dritte inklusive Klassen*. Universität Regensburg. <https://doi.org/10.5283/epub.53204>
- 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>
- Ennemoser, M., Marx, P., Weber, J. & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Leseschwindigkeit, des Leseverständnisses und des Rechtschreibens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(2), 53–67. <https://doi.org/10.1026/0049-8637/a000057>
- 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>
- Glover, T. A. & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, 45(2), 117–135. <https://doi.org/10.1016/j.jsp.2006.05.005>
- Gottlebe, K., Latzko, B., Dietrich, S. & Liebers, K. (2022). ILEA plus Deutsch A. Psychometrische Güte eines digitalen Instruments zur Erfassung früher Literalität in der Jahrgangsstufe 1. *Frühe Bildung*, 11, 186–193. <https://doi.org/10.1026/2191-9186/a000586>
- 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>
- Heine, J.-H. (2021). *pairwise: Rasch Model Parameters by Pairwise Algorithm. R package version 0.4.4-5.1*. <https://CRAN.R-project.org/package=pairwise>
- Jungjohann, J. (2022). Komplexe Nebensätze, Kohärenz- oder Inferenzbildung: Unterschiede im satzübergreifenden Leseverständnis von Jugendlichen mit sonderpädagogischem Unterstützungsbedarf im Bereich Sprache. *Forschung Sprache*, 10(2). <https://doi.org/10.5283/epub.53198>
- Jungjohann, J. & Gebhardt, M. (2023). Dimensions of Classroom-Based Assessments in Inclusive Education: A Teachers' Questionnaire for Instructional Decision-Making, Educational Assessments, Identification of Special Educational Needs, and Progress Monitoring. *International Journal of Special Education (IJSE)*, 38(1), 131–144. <https://doi.org/10.52291/ijse.2023.38.12>
- Jungjohann, J., DeVries, J. M., Gebhardt, M. & Mühlhling, A. (2018). Levumi: A web-based curriculum-based measurement to monitor learning progress in inclusive classrooms. In K. Miesenberger & G. Kouroupetroglou (Hrsg.), *Computers helping people with special needs. ICCHP 2018*. (S. 369–378). Springer. [https://doi.org/10.1007/978-3-319-94277-3\\_58](https://doi.org/10.1007/978-3-319-94277-3_58)
- Jungjohann, J., Diehl, K. & Gebhardt, M. (2022). Datenbasierte Leseförderung im (inkluisiven) Grundschulunterricht. Eine Umsetzung adaptiver Förderung mit Unterrichtsmaterialien und Lernverlaufstests. In Y. Blumenthal, S. Blumenthal & K. Mahlau (Hrsg.), *Kinder mit Lern- und emotional-sozialen Entwicklungsauffälligkeiten in der Schule: Diagnostik – Prävention – Förderung* (S. 177–189). Kohlhammer.
- Jungjohann, J., Schurig, M. & Gebhardt, M. (2021). Pilotierung von Leseflüchtigkeits- und Leseverständnistests zur Entwicklung von Instrumenten der Lernverlaufdiagnostik. Ergebnisse einer Längsschnittstudie in der 3ten und 4ten Jahrgangsstufe. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete Plus*, 90, 1–19. <https://doi.org/10.2378/vhn2021.art12d>
- Juska-Bacher, B. & Jakob, S. (2014). Wortschatzumfang und Wortschatzqualität und ihre Bedeutung im fortgesetzten Spracherwerb. *Zeitschrift für Angewandte Linguistik*, 61, 49–75. <https://doi.org/10.1515/zfal-2014-0016>


- Juska-Bacher, B., Zangger, C. & Röthlisberger. (2022). Methoden zur differenzierten Auswertung von Wortschatzkompetenzen und ihre Bedeutung für das frühe Lesen. In M. Philipp & S. Jambor-Fahlen (Hrsg.), *Lesen: Prozess- und Produktperspektiven von der Wortebene bis zu multiplen Texten* (S. 34–55). Beltz Juventa.
- Klauer, K. J. (2006). Erfassung des Lernfortschritts durch curriculumbasierte Messungen. *Heilpädagogische Forschung*, 32(1), 16–26.
- Klauer, K. J. (1978). Handbuch der Pädagogischen Diagnostik. Pädagogischer Verlag Schwann.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, 42(3), 627–633.
- Konerding, M., Bergström, K., Lachmann, T., & Klatte, M. (2021). Wirksamkeit des computergestützten grapho-phonologischen Trainingsprogramms Lautarium bei Kindern mit Lese-Rechtschreibstörung. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 70(4), 333–355.
- Landerl, K. & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics*, 21(2), 243–262. <https://doi.org/10.1017/S0142716400002058>
- Lenhard, W., Lenhard, A. & Schneider, W. (2017). *ELFE II – ein Leseverständnistest für Erst- bis Siebtklässler. Version II* (1. Auflage). Hogrefe Schultests.
- Lenhard, W. (2019). *Leseverständnis und Lesekompetenz: Grundlagen – Diagnostik – Förderung* (2. aktualisierte Auflage). Kohlhammer Verlag.
- Liebers, K., Kanold, E., & Junger, R. (2019). Digitale Lernstandsanalysen in der inklusiven Grundschule? In Bartusch (Ed.), *Lernprozesse begleiten* (pp. 209–221). Springer VS. [https://doi.org/10.1007/978-3-658-21924-6\\_16](https://doi.org/10.1007/978-3-658-21924-6_16)
- Lucas, R. & Norbury, C. F. (2015). Making Inferences From Text: It's Vocabulary That Matters. *Journal of Speech Language and Hearing Research*, 58(4), 1224–1232. [https://doi.org/10.1044/2015\\_JSLHR-L-14-0330](https://doi.org/10.1044/2015_JSLHR-L-14-0330)
- Mayer, A. (2018). Benennungsgeschwindigkeit und Lesen. *Forschung Sprache*, 6(1), 20–42.
- Mayer, A. (2020). *Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit* (4. Aufl.). Reinhardt.
- Meile, D., & Liebers, K. (2022). Digitale Lernstandsanalysen in der Grundschule. *Die Deutsche Schule*, (3), 280–283. <https://doi.org/10.31244/dds.2022.03.04>
- Nagler, T., Lindberg, S. & Hasselhorn, M. (2018). Leseentwicklung im Grundschulalter. Kognitive Grundlagen und Risikofaktoren. *Lernen und Lernstörungen*, 7(1), 33–44. <https://doi.org/10.1024/2235-0977/a000185>
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Pfost, M. (2015). Children's Phonological Awareness as a Predictor of Reading and Spelling. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(3), 123–138. <https://doi.org/10.1026/0049-8637/a000141>
- R Core Team (2013). R: A language and environment for statistical computing. <https://cran.r-project.org/web/packages/dplR/vignettes/xdate-dplR.pdf>
- Richter, T., Isberner, M. B., Naumann, J., & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern. *Zeitschrift für Pädagogische Psychologie*, 26(4), 313–331.
- Robitzsch, A., Kiefer, T., Wu, M. & Robitzsch, M. A. (2022). Package 'TAM'. Test Analysis Modules–Version, 3–5.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Psychologie Lehrbuch. Huber.
- Röthlisberger, M., Schneider, H. & Juska-Bacher, B. (2021). Lesen von Kindern mit Deutsch als Erst- und Zweitsprache – Wortschatz als limitierender Faktor. *Zeitschrift für Grundschulforschung*, 14(2), 359–374. <https://doi.org/10.1007/s42278-021-00115-w>
- Schnitzler, C. (2015). Schriftsprache und phonologische Verarbeitung bei Grundschulkindern mit im Vorschulalter überwundenen phonologischen Aussprachestörungen. *Sprache · Stimme · Gehör*, 39(1), 24–30. <https://doi.org/10.1055/s-0035-1545270>
- Schurig, M., Jungjohann, J. & Gebhardt, M. (2021). Minimization of a short computer-based test in reading. *Frontiers in Education*, 6, Artikel 684595. <https://doi.org/10.3389/feduc.2021.684595>


- Seifert, S., Maitz, K., Riedl, S., Neumaier, A., Stoff, I., & Gasteiger-Klicpera, u. B. (2022). Screening zur Erfassung des Leseverständnisses in den Klassen 6 – 7 (SELV 6 – 7). *Diagnostica*, 68(2), 101–112. <https://doi.org/10.1026/0012-1924/a000289>
- Seifert, S. Palaczek, L., Schwab, S. & Gasteiger-Klicpera, B. (2017). *Grazer Wortschatztest*. Hogrefe.
- Silverman, R. D., Speece, D. L., Haring, J. R. & Ritchey, K. D. (2013). Fluency Has a Role in the Simple View of Reading. *Scientific Studies of Reading*, 17(2), 108–133. <https://doi.org/10.1080/10888438.2011.618153>
- Spilles, M., & Leidig, T. (2020). Förderung von Lesekompetenz und Soziabilität durch tutorielle Lernverfahren–Hinweise zur Unterrichtsgestaltung. *Zeitschrift für Heilpädagogik*, 71(4), 185–199.
- Stanat, P., Schipolowski, S., Schneider, R., Sächse, K. A., Weirich, S. & Henschel, S. (2022). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Waxmann Verlag GmbH. <https://doi.org/10.31244/9783830996064>
- Stock C., Marx, P. & Schneider, W. (2003). *Basiskompetenzen für Lese-Rechtschreibleistungen (BAKO 1–4). Ein Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Grundschuljahr*. Hogrefe.
- Trautwein, J. & Schroeder, S. (2019). WOR-TE: Ein Ja / Nein-Wortschatztest für Kinder verschiedener Altersgruppen. *Diagnostica*, 65(1), 37–48. <https://doi.org/10.1026/0012-1924/a000212>
- Wagner, R. K. & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192–212. <https://doi.org/10.1037/0033-2909.101.2.192>
- Walter, J. (2013). *VSL: Verlaufsdiagnostik sinnerfassendes Lesen*. Hogrefe.
- Watkins, A. (2007). *Assessment in Inclusive Settings: Key Issues for Policy and Practice*. European Agency for Development in Special Needs Education.
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370–371. <https://www.rasch.org/rmt/rmt83b.htm>


## Autorinnen- und Autorenhinweis

 Jana Jungjohann  
<https://orcid.org/0000-0002-9985-4780>

 Nikola Ebenbeck  
<https://orcid.org/0000-0002-4167-981X>

 Katrin Liebers  
<https://orcid.org/0000-0001-9094-1629>

 Kirsten Diehl  
<https://orcid.org/0000-0002-5865-9305>

 Markus Gebhardt  
<https://orcid.org/0000-0002-9122-0556>

### Korrespondenzadresse:

Jana Jungjohann  
Fachgebiet Digitale Förderung und  
inklusive Bildung  
Technische Universität Dortmund  
Emil-Figge-Str. 50  
D-44227 Dortmund  
[jana.jungjohann@tu-dortmund.de](mailto:jana.jungjohann@tu-dortmund.de)

Erstmals eingereicht: 18.11.2022  
Überarbeitung eingereicht: 24.03.2023  
Angenommen: 02.04.2023



Offene Daten	Der Datensatz "Daten.csv" ist verfügbar unter: <a href="https://osf.io/b476z/">https://osf.io/b476z/</a>
Offener Code	Die Syntax "Syntax.R" für GNU R ist verfügbar unter: <a href="https://osf.io/b476z/">https://osf.io/b476z/</a>
Offene Materialien	Die Testbeschreibung und das Testheft des verwendeten Instruments sind verfügbar unter: <a href="https://doi.org/10.5283/epub.53204">https://doi.org/10.5283/epub.53204</a>
Präregistrierung	Nein
Votum Ethikkommission	Die Ethikkommissionen der Universitäten Leipzig (30.11.2021) und Regensburg (10.12.2021) haben das forschungsmethodische Vorgehen sowie den Datenschutz der vorliegenden Studie positiv begutachtet.
Finanzielle und weitere sachliche Unterstützung	Die Studie wurde vom BMBF im Verbundprojekt „Digitale alltagsintegrierte Förderdiagnostik - Lesen in der inklusiven Bildung (DaF-L)“ gefördert unter den Förderkennzeichen 01NV2116A, 01NV2116B, 01NV2116C, 01NV2116D
Autorenschaft	JJ, NE und MG haben die Studie, das Lesescreening sowie die statistische Auswertung konzipiert und angeleitet. Alle Autor:innen waren an der Datenerhebung beteiligt. JJ und NE schrieben das Manuskript. NE schrieb die Syntax und analysierte die Daten unter Rücksprache mit MG und JJ. KD, KL und MG überarbeiteten und editierten.