

Empirische Sonderpädagogik, 2019, Nr. 4, S. 279-293
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Unter der Norm – Kompetenz und Diagnostik in IGLU 2016

Anke Hußmann & Michael Schurig

Technische Universität Dortmund

Zusammenfassung

Large-Scale Assessments (LSA) sind fester Bestandteil der ‚Gesamtstrategie zum Bildungsmonitoring‘ des deutschen Bildungssystems. Hinsichtlich der Berücksichtigung von Schülerinnen und Schülern mit und ohne sonderpädagogischen Förderbedarf (SPF) stellt sich jedoch die Frage, inwieweit ihre Leistungen in LSA erfasst und interpretiert werden können. Der vorliegende Beitrag diskutiert am Beispiel der *Internationalen Grundschul-Lese-Untersuchung* (IGLU 2016; Hußmann, Wendt, Bos et al., 2017) und damit für den Kompetenzbereich Lesen diesbezügliche Grenzen und Möglichkeiten für Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe. Im Fokus stehen schwache Leseleistungen, d.h. Leistungen, die im IGLU-Kompetenzstufenmodell als \leq Kompetenzstufe III zu bezeichnen sind und damit ‚unter der Norm‘ liegen. ‚Unter der Norm‘, weil seitens des Sekretariats der Kultusministerkonferenz (KMK) Leseleistungen ab Kompetenzstufe III als Regelstandard gelten (KMK, 2011a), niedrigere Leistungen entsprechend jenseits dieses Regelstandards bzw. dieser Norm liegen.

Deskriptive Befunde zeigen, dass die Leseleistungen von Schülerinnen und Schülern mit SPF im Mittel am Ende der vierten Jahrgangsstufe orientiert am IGLU-Kompetenzstufenmodell meist nur die Kompetenzstufe III oder weniger erreichen. In Rückgriff auf Daten des *Leseverständnistests für Erst- bis Siebtklässler, Version II* (ELFE II; Lenhard, Lenhard & Schneider, 2017), der mit den Untertests auf Wort- und Satzebene ergänzend in IGLU 2016 eingesetzt wurde, zeigt sich, dass eine Orientierung der Leseleistungen von Schülerinnen und Schülern mit SPF am bestehenden IGLU-Kompetenzstufenmodell möglich scheint. Die Zuordnung zu den Kompetenzstufen ist dann aber inhaltlich stärker zu differenzieren.

Schlüsselwörter: Large Scale Assessments, Lesekompetenzen, Inklusion, Diagnostik, Lese-Rechtschreibschwäche

Below the norm – competence and diagnosis in PIRLS Germany 2016

Abstract

Large Scale Assessments are an integral part of the “Strategy on educational Monitoring” of the German educational system. However, particularly considering students with and without special education needs (SEN), the question arises whether or rather to which degree student achievements can be assessed and reasonably interpreted at all. The paper at hand discusses the respective possibilities and limits for students at the end of 4th grade in reference to reading competence, based on PIRLS data for Germany (IGLU 2016; Hußmann, Wendt, Bos et al., 2017). The analysis focuses on low reading performance, i.e. those students which are classified as below Proficiency Level III. These proficiencies are regarded as ‘below standard’. The determination ‘below standard’ results from the definition of Competence Level III by the Secretary

of the German Kultusministerkonferenz (KMK). The German KMK has defined Competence Level III as the overall reference level; lower proficiencies accordingly lie beyond this rule or norm. Descriptive findings indicate that average students with SEN rarely reach the PIRLS Competence Level III or above at the end of 4th grade. A diagnostic test in reading comprehension (ELFE II; Lenhard, Lenhard & Schneider, 2017) that was employed supplementing PIRLS 2016 in Germany for the first time, including subtests on the level of wording and sentence structure, revealed that it is at least possible to use reading performance as a reference. However, the critical analysis of the data in question reveals the need for further differentiation of the Competence Levels.

Keywords: large scale assessment, reading literacy, inclusive education, diagnostics, dyslexia

Die Ergebnisse internationaler Large-Scale Assessments (LSA) dienen als Basis einer langfristigen Form des Bildungsmonitorings. Im Rahmen der 2006 vom Sekretariat der Kultusministerkonferenz (KMK) verabschiedeten ‚Gesamtstrategie zum Bildungsmonitoring‘ sollen sie „Stärken und Schwächen des Bildungssystems“ (KMK, 2016, S. 3) aufdecken und dabei unterstützen, „geeignete Maßnahmen umzusetzen“ (ebd.). Damit dies umfassend gelingen kann, sind bildungspolitische Entscheidungen wie die Umsetzung von Inklusion, insbesondere bei einem weiten Verständnis des Begriffs (z.B. Göransson & Nilholm, 2014), bei der Konzeption von LSA-Studien zu berücksichtigen. Seit Inkrafttreten der UN-Konvention über die Rechte von Menschen mit Behinderung 2009 in Deutschland ist das inklusive Bildungssystem zum Leitbild der Bildungspolitik geworden, taucht aber bis heute nicht als differenzierte Analysekategorie auf. Gründe dafür finden sich unter anderem darin, dass die Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf (SPF) zwar zunehmend an allgemeinbildenden Schulen unterrichtet werden (KMK, 2011b), die amtlichen Feststellungsverfahren in den Bundesländern aber unterschiedlich gestaltet sind (z.B. Sälzer, Gebhardt, Müller & Pauly, 2015). So besteht keine Einigkeit darüber, wann Schülerinnen und Schüler das Label SPF erhalten (Gresch, Kölm & Kocaj, 2017). Auch wird der Leistungsvergleich durch ungleiche Verteilungen von Schülerinnen und Schülern mit und

ohne SPF und anteilig geringe Häufigkeiten sowie hochvariable Leistungsfähigkeiten innerhalb der Gruppen (Gebhardt, Heine & Sälzer, 2015) erschwert. Messmodelle müssen ein breites Spektrum abbilden und jede Gruppe muss in der Stichprobe repräsentiert werden. Skalenwerte sind für populationsbeschreibende oder gruppenspezifische Zwecke optimiert (Carstensen, Knoll, Rost & Prenzel, 2004) und Schätzungen an den Rändern des Leistungsspektrums können unschärfer werden (Rost, 1996). Sind also Gruppen, deren Fähigkeiten sich systematisch von denen der Population unterscheiden, in der Stichprobenziehung nicht expliziert worden, kann dies für die Schätzungen ihrer Leistungen zu Problemen führen. Somit wäre auch die Repräsentativität dieser Gruppen nicht gewährleistet. Neben diesen praktischen und methodischen Herausforderungen ist hinsichtlich der inhaltlichen Belastung der Messwerte unklar, ob sich die Fähigkeiten von Schülerinnen und Schülern mit SPF überhaupt in Kompetenzstufenmodellen von LSA trennscharf abbilden lassen und ob es geeignete Vergleichskategorien gibt.

Im Rahmen dieses Beitrags reflektieren und diskutieren wir einige dieser Herausforderungen am Beispiel der *Internationalen Grundschul-Lese-Untersuchung* (IGLU 2016; Hußmann, Wendt, Bos et al., 2017), also am Beispiel von Lesekompetenzen der Viertklässlerinnen und Viertklässler in Deutschland. Dazu stehen in IGLU 2016 zweierlei Daten zur Verfügung: Zum einen

die mittels des internationalen Lesekompetenztests ermittelten Leseleistungen von Schülerinnen und Schülern. Zum anderen wurde in Deutschland ergänzend der *Leseverständnistest für Erst- bis Siebtklässler, Version II* (ELFE II; Untertests zum Wort- und Satzverständnis; Lenhard, Lenhard & Schneider, 2017) eingesetzt. Bei dem ELFE II handelt es sich um ein Testinstrument zur differenzierten Diagnostik des Leseverstehens. Er wurde eingesetzt, um niedrige Lesekompetenzen (Leistungen verortet auf den Kompetenzstufen I und II, s. Abschnitt 2) auch auf individueller Ebene angemessen, also beispielsweise mit Verweis auf ein Risiko zum Vorliegen einer Lese-Recht Schreibstörung, beschreiben zu können. Die Kombination von IGLU-Lesekompetenztest und ELFE II hat das Potential, Leseleistungen, die ‚unter der Norm‘ liegen, genauer in den Blick zu nehmen. ‚Unter der Norm‘ deshalb, weil die Bildungsstandards im Fach Deutsch für den Primarbereich (KMK, 2011b) Leseleistungen ab Kompetenzstufe III als Regelstandard definieren (ebd., S. 12), demnach niedrige Leistungen jenseits dieses Regelstandards bzw. dieser Norm liegen. Die Bezugnahme des IGLU-Kompetenzstufenmodells auf das KMK-Kompetenzstufenmodell begründet sich durch ihre konzeptionelle Verknüpfung mit dem Beschluss der Bildungsstandards (KMK, 2004).

Der Fokus dieses Beitrags liegt auf der Beschreibung von Leseleistungen der Schülerinnen und Schüler mit SPF und Schülerinnen und Schülern ohne SPF mit niedrigen Leseleistungen. Dazu skizzieren wir zunächst das in IGLU zugrunde liegende Verständnis von Lesekompetenz, den aktuellen Forschungsstand zu Leseleistungen von Schülerinnen und Schülern mit SPF und jenen ohne SPF mit niedrigen Leseleistungen am Ende der Grundschulzeit und leiten die aus unserer Sicht notwendig erscheinenden Prüfschritte zur Klärung der Forschungsfragen her (Abschnitt 2). Nach der Darstellung der Stichprobe und des methodischen Vorgehens (Abschnitt 3) werden

mittlere Leistungswerte für Schülerinnen und Schüler mit SPF und mit schwachen Leseleistungen dargestellt (Abschnitt 4). Es folgt die Klärung der inhaltlichen Angemessenheit des IGLU-Kompetenzmodells für Schülerinnen und Schüler mit SPF und schwachen Leseleistungen auf der Basis eines Abgleichs der IGLU-Ergebnisse mit den Ergebnissen des ELFE II (Abschnitt 4). Die Diskussion dient weiteren Untersuchungen als Vergleichsbasis und adressiert Analysemöglichkeiten von LSA-Daten für Fragestellungen im Kontext von schulischer Inklusion (Abschnitt 5).

Lesekompetenzen von Schülerinnen und Schülern mit SPF

Lesekompetenz gilt als die Schlüsselkompetenz für gesellschaftliche Teilhabe, den Erwerb von Wissen und lebenslanges Lernen (z.B. Hurrelmann, 2007). Lesekompetenz, wie sie in IGLU erhoben wird, basiert auf theoretischen Annahmen zum Lesen als einem konstruktiven, interaktiven Lernprozess (Bremerich-Vos, Wendt & Bos, 2017). Die Kompetenz zu lesen umfasst nicht nur die Fähigkeit, schriftsprachliche Formen verstehen und nutzen zu können, sondern auch, dem Geschriebenen Bedeutung beizumessen und aus ihm heraus Bedeutung zu konstruieren, also, mit konkreten Intentionen zu lesen. Angesprochen ist mit dieser Dimension das sich zwischen Text und Leserin bzw. Leser als konstruktiv-interaktiver Prozess vollziehende Verstehen beim Lesen (z.B. Kintsch, 2012). LSA-Studien berücksichtigen in ihrem Studiendesign einzelne Aspekte dieses Kompetenzbereichs. In IGLU sind es die folgenden drei: (1) Leseverstehensprozesse, (2) Lesentionen sowie (3) das Leseselbstkonzept, die Lesemotivation und das Leseverhalten. Letztere (2+3) werden im vorliegenden Beitrag nicht betrachtet. Wie in Abschnitt 3 beschrieben, handelt es sich bei den berichteten Leseverstehensprozessen

(1) um als Skalenwerte quantifizierte Leseleistungen. Diese werden hinsichtlich ihres Kompetenzniveaus in einem Kompetenzstufenmodell beschrieben (Tabelle 1). Bei der Herleitung solcher Modelle werden inhaltlich sinnvoll voneinander abgrenzbare Abschnitte im Leistungskontinuum gebildet, um eine inhaltliche Beschreibung von beobachteten Schülerkompetenzen vorzunehmen.

Eine ausführliche Beschreibung der Kompetenzstufen findet sich in Bremerich-Vos et al. (2017). Während Leistungen auf Kompetenzstufe II von der KMK noch als „Mindeststandard“ (KMK, 2011a, S. 11) gelten, liegen Leistungen auf Kompetenzstufe I „deutlich unterhalb der Erwartungen der KMK-Bildungsstandards“ (ebd., S. 10). Leseleistungen in diesem Bereich bewegen sich zwischen rudimentärem Leseverständnis (Kompetenzstufe I) und dem Identifizieren explizit im Text angegebener Informationen sowie der Herstellung von Kohärenz auf lokaler Ebene, d.h. dem Erkennen nicht explizit angegebener Beziehungen unter Berücksichtigung von genannten Gründen oder Ursachen in benachbarten Sätzen innerhalb eines Textes (Kompetenzstufe II).

Der Anteil an Schülerinnen und Schülern, die Kompetenzstufe III am Ende der vierten Jahrgangsstufe in Deutschland im

Lesen nicht erreichen, lag im Jahr 2016 bei rund 19 % (Hußmann, Wendt, Kasper et al., 2017). Nur rudimentär lesen konnten fast 6 % (Kompetenzstufe I), knapp 13 % erreichten Leseleistungen auf Kompetenzstufe II. Seit Jahren ist bekannt, dass – neben Aspekten wie Geschlecht, Lesesozialisation im Elternhaus, Lesemotivation, Leseselbstkonzept (z.B. Klicpera & Gasteiger-Klicpera, 1995; Schneider, 2016), die nicht im Fokus des Beitrags stehen – soziale und migrationsbedingte Aspekte (z.B. armutsgefährdete Elternhäuser, Bildungsniveau der Eltern, Affinität zu schriftsprachlichen Gütern im Elternhaus (Bücher), im Elternhaus gesprochene Sprache) der Schülerfamilien eng mit der Leistung von Schülerinnen und Schülern gekoppelt sind (z.B. Hußmann, Stubbe & Kasper, 2017; Wendt & Schwippert, 2017).

Tabelle 1: Kompetenzstufen und Skalenwerte IGLU 2016

Kompetenzstufe		Skalenbereich der Fähigkeit
V	Unter Bezug auf Textpassagen bzw. den Gesamttext Informationen ordnen und Aussagen selbstständig interpretierend und kombinierend begründen	> 625
IV	Für die Herstellung von Kohärenz auf der Ebene des Textes relevante Aspekte des Inhalts und der Darstellung erfassen und komplexe Schlüsse ziehen	551 - 625
III	„Verstreute“ Informationen verständlich miteinander verknüpfen	476 - 550
II	Explizit angegebene Informationen identifizieren und auf lokaler Ebene Kohärenz herstellen	400 - 475
I	Rudimentäres Leseverständnis	< 400

In differenzierter Betrachtung der Schülerinnen und Schüler mit SPF gibt es keine vergleichbaren, an einem Kompetenzmodell orientierten und für Deutschland repräsentativen Befunde. Lediglich der *IQB-Bildungstrend 2016* (Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017) berichtet Lesekompetenzen für Schülerinnen und Schüler mit den SPF Lernen, Sprache und Emotionale und soziale Entwicklung differenziert danach, ob sie an einer Förderschule oder einer allgemeinen Schule unterrichtet werden. Die Mittelwerte liegen für alle drei SPF auf Kompetenzstufe I, wenn die Schülerinnen und Schüler an einer Förderschule unterrichtet werden. Höhere Leistungen (Kompetenzstufe II) erreichen im Mittel Schülerinnen und Schüler mit den SPF Sprache sowie Emotionale und soziale Entwicklung an allgemeinen Schulen (Kocaj, Kuhl, Haag, Kohrt & Stanat, 2017). Leseleistungen von Schülerinnen und Schülern mit weiteren SPF, die an Kompetenzstufen orientiert beschrieben werden, liegen für Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe in Deutschland bislang nicht vor.

Durch Aufgreifen dieses Desiderats sowie der einleitend formulierten Herausforderung, die Leistungen von SuS mit SPF inhaltlich im Kompetenzstufenmodell interpretieren zu können, werden folgende Forschungsfragen am Beispiel von IGLU 2016 näher betrachtet und diskutiert:

- 1) Zu den Lesekompetenzen der Schülerinnen und Schüler mit SPF: Welche Lesekompetenzen erreichen Schülerinnen und Schüler mit SPF am Ende der vierten Jahrgangsstufe in Deutschland? Wie groß sind die Anteile von Schülerinnen und Schülern mit SPF, die unter Kompetenzstufe III liegen?
- 2) Zur Verortung der Lesekompetenzen der Schülerinnen und Schüler mit SPF im IGLU-Kompetenzstufenmodell: Wie können geringe Leseleistungen und Leseleistungen von Schülerinnen und Schülern mit SPF im Kompetenzmodell verortet werden? Wie können diese inhaltlich interpretiert werden?

Stichprobe und methodisches Vorgehen

Stichprobe

Die Stichprobe unserer Analysen entspricht der Stichprobe von IGLU 2016. Es wurden deutschlandweit 208 Schulen mit je einer vierten Klasse gezogen, Details sind in Martin, Mullis und Foy (2015) nachzulesen. Insgesamt nahmen 3959 Schülerinnen und Schüler am Lesetest teil. Die hier verwendeten Hintergrundmerkmale entstammen bis auf das Haushaltseinkommen dem Fragebogen der Schülerinnen und Schüler. Das Haushaltseinkommen wurde im Elternfragebogen erfasst. Das Durchschnittsalter lag bei 10.3 Jahren, 49% waren Mädchen. 19% der Schülerinnen und Schüler wiesen einen Migrationshintergrund auf (beide Elternteile im Ausland geboren) und für 7% der Schülerinnen und Schüler lag nach Angaben der Lehrkräfte ein diagnostizierter SPF vor. Von der internationalen Studienleitung sind Ausschlusskriterien definiert, die zum Tragen kommen, wenn beispielsweise Schülerinnen und Schüler aus körperlichen, emotionalen oder geistigen Gründen nicht dazu fähig sind, den Test selbstständig zu bearbeiten (in IGLU 2016: 28 Schülerinnen und Schüler; <1%), oder aber, wenn sie weniger als ein Jahr in der Testsprache unterrichtet wurden und ihre Muttersprache nicht die Testsprache ist (in IGLU 2016: 74 Schülerinnen und Schüler; 2%). Details zur deutschen Stichprobe und weiteren Ausschlüssen sind im Berichtsband gegeben (Hulsmann, Wendt, Kasper, Bos & Goy, 2017).

Zu allen Schülerinnen und Schülern der Stichprobe wurden Hintergrunddaten erhoben. Bezüglich des SPF wurden die Klassenlehrkräfte angehalten anzugeben, ob die Schülerinnen und Schüler einen diagnostizierten SPF aufweisen und, falls ja, welchen. Die Antworten wurden im offenen Format gegeben und wie folgt kategorisiert: Schülerinnen und Schüler ohne SPF, Schülerinnen und Schüler mit SPF. Vertiefend können Ein-

zelgruppen betrachtet werden, die $\geq 0.5\%$ (mind. ~20 Schülerinnen und Schüler) umfassen. Dies betrifft die SPF Lernen ($n = 94$), Sprache ($n = 38$) und Emotionale und soziale Entwicklung ($n = 44$). Gering besetzte Kombinationen von SPF und weitere Gruppen werden nicht detailliert dargestellt. Dies umfasst die SPF Sehen (S ; $n = 8$), Körperlich und motorische Entwicklung (KME; $n = 5$), Lernen und Sprache (LS; $n=13$), Lernen und Emotionale und soziale Entwicklung (LESE; $n = 16$), Lernen, Sprache und Emotionale und soziale Entwicklung (LSESE; $n = 7$). Alle weiteren Kategorien und Kombinationen umfassen weniger als vier Kinder und werden gar nicht berücksichtigt. Bei 65 Schülerinnen und Schülern lagen mangelhafte oder unpräzise Angaben zum SPF vor.

Bei den 208 Schulen wurden 13 Förderschulen mit insgesamt 82 Schülerinnen und Schülern berücksichtigt, von denen 20 (24%) nicht am Lesekompetenztest teilnahmen.

Methodisches Vorgehen

Die Analysen entsprechen dem Vorgehen in IGLU 2016 (Hußmann, Wendt, Kasper et al., 2017). Die Leseleistungen werden mittels Leistungstests erhoben, die aus insgesamt sechs Erzähl- und sechs Sachtexten inklusive Verständnisaufgaben bestehen (pro Text 12 bis 17 Aufgaben in offenen oder geschlossenen Formaten). Jede Schülerin bzw. jeder Schüler bearbeitete einen Erzähl- und einen Sachtext. Die Leistungswerte wurden in Form von fünf *Plausible Values* (PV; z.B. Lüdtke & Robitzsch, 2017) auf Basis eines multiplen Matrixdesigns geschätzt, wobei die finalen Werte unter Berücksichtigung der Imputationsvarianz abgeleitet wurden.

Als statistisches Modell wurde ein mehrdimensionales dreiparametrisches logistisches Modell (3-PL-Modell) gewählt, als Schätzer der *marginal Maximum-Likelihood* (MML) Ansatz. In diesem Beitrag werden durchgängig die PV des allgemeinen Leseverständnisses verwendet.

Mit PV zu arbeiten erlaubt die Korrektur von Messfehlern auf der Ebene von Populationsschätzungen bei einer getrennten Schätzung des Messmodells und möglicher Strukturmodelle. Ergänzend wurden alle Standardfehler (SE) mit dem *Jackknife*-Verfahren geschätzt, um ein Unterschätzen der SE bei mehrfach geschichteten Stichproben zu vermeiden.

Als Normwerte für den ELFE II wurden die Tabellen für Schülerinnen und Schüler mit sieben bis acht Schulmonaten in der vierten Klassenstufe verwendet, da dies den überwiegenden Teil der Schülerinnen und Schüler zutraf. Die *split-half odds even* Reliabilität befand sich im Bereich der im Manual angegebenen Werte ($\alpha_{\text{Gesamt}} = .89$; $\alpha_{\text{Wortverständnis}} = .99$; $\alpha_{\text{Satzverständnis}} = .98$). Um auf diagnostische Schwellenwerte zu schließen, wurden die Normwerte der Untertests summiert und anhand der Normtabelle für die ELFE II-Kurzform ausgewertet. Dadurch ist es möglich, durch Standardabweichungen definierte Ergebnisbereiche zu schätzen und eine datengestützte Risikoeinschätzung zum Vorhandensein einer möglichen Lese- störung abzuleiten (‚starker Verdacht‘, ‚Verdacht‘, ‚kein Verdacht‘).

Ergebnisse

Die Ergebnisdarstellung erfolgt analog zu den Forschungsfragen (s. Abschnitt 2). Der Forschungsfragenblock 1 zielt auf die Beschreibung der Lesekompetenzen von Schülerinnen und Schülern mit SPF am Ende der vierten Jahrgangsstufe in Deutschland. Antworten auf diese Frage sind in den Tabellen 2 und 3 zusammengetragen. Tabelle 2 zeigt die mittleren Lesekompetenzen und entsprechende Verteilungswerte differenziert nach der Kategorisierung von SPF, wie in Abschnitt 3 vorgestellt. Zur besseren Interpretierbarkeit wurden standardisierte Effektmaße für ungleiche Gruppen- größen (Hedges g) ergänzt.

Tabelle 2: Mittlere Kompetenzwerte pro Analysekatgorie

		<i>n</i>	<i>M</i>	<i>SE</i>	<i>SD</i>	<i>g</i> (95% <i>KI</i>)
SPF	Kein SPF (Referenz)	3669	546.46	2.40	145.37	-
	SPF	205	422.94	18.48	264.59	0.80 (0.66, 0.94)
	SPF Lernen	66	388.54	25.10	203.91	1.08 (0.83, 1.32)
	SPF Sprache	31	440.97	13.43	74.78	0.73 (0.37, 1.08)
	SPF Emotionale und soziale Entwicklung	37	509.63	15.55	94.59	0.25 (-0.07, 0.58)
	Andere SPF	156	451.74	14.76	184.35	0.64 (0.48, 0.81)
Kompetenzstufe	Auf oder über Kompetenzstufe III (Referenz)	3274	563.96	1.47	84.11	-
	Unter Kompetenzstufe III	685	419.81	6.32	165.41	1.40 (1.31, 1.49)

Anmerkungen: Hervorgehobene Werte unterscheiden sich signifikant auf einem Niveau von $p \leq .05$ gegenüber der markierten Referenzkategorie. *n* = Stichprobenumfang, *M* = Mittelwert, *SE* = Standardfehler, *SD* = Standardabweichung; *g* = Hedges *g*

Schülerinnen und Schüler ohne SPF erreichten im Jahr 2016 in Deutschland im Mittel Lesekompetenzen von 546 Punkten ($SE = 2.4$) (zum Vergleich: Int. Mittelwert: 521 ($SE = 0.4$), Vergleichsgruppe EU: 540 ($SE = 0.5$); Bremerich-Vos et al., 2017, S. 114), Schülerinnen und Schüler mit SPF 423 Punkte ($SE = 18.5$). Betrachtet nach SPF-Kategorisierungen, erreicht die Gruppe mit dem SPF Emotionale und soziale Entwicklung die höchsten Lesekompetenzen ($M = 510, SE = 15.6$), die niedrigsten die Gruppe mit dem SPF Lernen ($M = 389, SE = 25.1$). Übertragen auf das Kompetenzstufenmodell (Tabelle 1) erreichte die Gruppe mit dem SPF Emotionale und soziale Entwicklung im Mittel Leseleistungen auf Kompetenzstufe III, die Gruppe mit dem SPF Sprache im Mittel Kompetenzstufe II und die Gruppe mit dem SPF Lernen im Mittel Kompetenzstufe I an der Grenze zu Kompetenzstufe II. Der Mittelwert der Leseleistungen unter Kompetenzstufe III betrug in Deutschland 420 Punkte (Schülerinnen und Schüler mit und ohne SPF). Bezüglich der Kompetenzniveaus erreichten 685 Schülerinnen und Schüler (17%) nicht die Kompetenzstufe III. Dies betraf 14% ($n = 522$) der

Schülerinnen und Schüler ohne SPF und 65% ($n = 133$) mit SPF. Weiterhin erreichten Schülerinnen und Schüler mit den SPF Sehen 430 Punkte ($SE = 76.0$), KME 415 Punkte ($SE = 52.4$), LS 375 Punkte ($SE = 39.5$), LESE 387 Punkte ($SE = 35.7$) und LSESE 408 Punkte ($SE = 12.4$), das entspricht Lesekompetenzen im Bereich von Kompetenzstufe I und II. Zwischen jeder der Analysekatgorien in der Tabelle 2 und der Referenzkategorie ‚kein SPF‘ zeigen sich statistisch bedeutsame Unterschiede in der Höhe der Kompetenzwerte im Lesen. Zudem unterscheiden sich alle SPF-spezifischen Gruppen in Paarvergleichen statistisch bedeutsam voneinander ($p \leq .05$).

In Tabelle 3 sind Befunde weiterer Hintergrundvariablen der Schülerinnen und Schüler aufgeführt, die als Prädiktoren für Minderleistungen gelten (s. Abschnitt 2) und die Stichprobe näher beschreiben. Auch hier wurden standardisierte Effektmaße (Hedges *g*) ergänzt, um die Effektgrößen zu quantifizieren.

Tabelle 3: Hintergrundvariablen

	SPF								g (95% KI)
	sonderpädagogischer Förderbedarf				kein sonderpädagogischer Förderbedarf				
	n	M/%	SE	SD	n	M/%	SE	SD	
Gesamtanteil	205	7%	0.76	10.88	3669	94%	0.76	46.03	–
Geschlecht (Anteil der Mädchen)	162	35%	0.64	85.40	3264	49%	0.79	45.13	0.291 (0.15, 0.43)
Sprache im Haushalt	159	1.58	0.09	1.13	3241	1.50	0.02	1.14	-0.07 (-0.21, 0.07)
Anzahl der Bücher zu Hause	156	2.58	0.14	1.75	3214	3.15	0.03	1.70	0.335 (0.19, 0.48)
Monatliches Nettohaushaltseinkommen	96	10.93	0.64	6.27	2287	15.49	0.21	10.04	0.462 (0.32, 0.60)

	Kompetenzstufe								g (95% KI)
	unter Kompetenzstufe III				auf oder über Kompetenzstufe III				
	n	M/%	SE	SD	n	M/%	SE	SD	
Gesamtanteil	685	18%	685	33.24	3274	82%	1.27	72.67	–
Geschlecht (Anteil der Mädchen)	546	45%	546	46.97	2960	51%	0.83	45.16	0.13 (0.05, 0.21)
Sprache im Haushalt	538	1.85	538	1.39	2941	1.44	0.02	1.08	0.36 (0.28, 0.44)
Anzahl der Bücher zu Hause	512	2.42	512	1.81	2907	3.24	0.03	1.62	0.496 (0.41, 0.58)
Monatliches Nettohaushaltseinkommen	309	10.92	309	6.15	2118	15.84	0.22	10.12	0.529 (0.45, 0.61)

Anmerkungen: Hervorgehobene Werte unterscheiden sich innerhalb der Kategorien SPF oder Kompetenzstufe signifikant auf einem Niveau von $p \leq .05$, n = Stichprobenumfang, M = Mittelwert, SE = Standardfehler, SD = Standardabweichung, g = Hedges g ; Sprache im Haushalt: 1-4, 1 = ‚Ich spreche immer Deutsch zu Hause‘, 4 = ‚Ich spreche nie Deutsch zu Hause‘; Anzahl der Bücher zu Hause: 1-5, 1 = ‚keine oder wenige (0-10 Bücher)‘, 5 = ‚genug, um drei oder mehr Regale zu füllen (200 oder mehr Bücher)‘; Monatliches Nettohaushaltseinkommen: 1-25, 1 = ‚< 400 €‘, 25 = ‚> 8000 €‘

Der Vergleich der beiden Gruppen ‚SPF‘ und ‚unter Kompetenzstufe III‘ zeigt in Bezug auf die Hintergrundvariablen, dass ein höherer Anteil an Jungen einen SPF aufweist. Die mittlere Anzahl von Büchern im elterlichen Haushalt ebenso wie das mittlere Nettohaushaltseinkommen ist bei der Gruppe mit SPF geringer als in der Gruppe ohne SPF. Auch ist der Anteil von Jungen unter der Kompetenzstufe III erhöht. Bei einer Zuordnung der erreichten Werte unter-

halb der Kompetenzstufe III wird im Elternhaus Deutsch gesprochen, die Anzahl der Bücher und das monatliche Einkommen sind geringer. Für alle Hintergrundvariablen außer der für den SPF konditionalen Häufigkeit, mit der im Elternhaus Deutsch gesprochen wird, sind statistisch bedeutsame Unterschiede zwischen den konditionalen Mittelwerten zu identifizieren.

Mit dem Forschungsfragenblock 2 wird die inhaltliche Verortung der Lesekompetenzen von Schülerinnen und Schülern mit SPF im IGLU-Kompetenzstufenmodell in den Blick genommen. In der Analyse der Werte des ELFE II wird geprüft, wie stark die

mittels Normwerten erstellten *t*-Werte mit den generellen *PV* in der Leseleistung zusammenhängen. Ist der Zusammenhang stark, deutet dies darauf hin, dass der ELFE II gute Möglichkeiten bietet, Leseleistungen im unteren Bereich des IGLU-Kompetenz-

Tabelle 4: Korrelationen der *PV* im Lesen und der *t*-Werte des ELFE II (*n* = 3948)

	Lesen (PV)		ELFE II		ELFE II – Satzverständnis	
	<i>r</i>	<i>SE</i>	<i>r</i>	<i>SE</i>	<i>r</i>	<i>SE</i>
ELFE II	0.53	0.02	1.00			
ELFE II – Satzverständnis	0.60	0.02	0.92	0.01	1.00	
ELFE II – Wortverständnis	0.38	0.03	0.92	0.01	0.69	0.02

Anmerkungen: Alle dargestellten Zusammenhangsmaße erreichen eine statistische Bedeutsamkeit ($p \leq .001$). *r* = Bravais-Pearson Korrelationskoeffizient; *SE* = Standardfehler der Korrelation

Tabelle 5: Konditionale Korrelationen der *PV* im Lesen und der *t*-Werte des ELFE II

	Variable	Lesen (PV)		ELFE II		ELFE II – Satzverständnis	
		<i>r</i>	<i>SE</i>	<i>r</i>	<i>SE</i>	<i>r</i>	<i>SE</i>
Kein SPF	ELFE II	0.46	0.02	1.00			
	ELFE II – Satzverständnis	0.53	0.02	0.91	0.01	1.00	
	ELFE II – Wortverständnis	0.30	0.02	0.91	0.01	0.66	0.02
SPF	ELFE II	0.64	0.02	1.00			
	ELFE II – Satzverständnis	0.67	0.02	0.93	0.01	1.00	
	ELFE II – Wortverständnis	0.55	0.02	0.95	0.01	0.76	0.03
SPF Lernen	ELFE II	0.64	0.11	1.00			
	ELFE II – Satzverständnis	0.65	0.08	0.95	0.01	1.00	
	ELFE II – Wortverständnis	0.59	0.15	0.96	0.01	0.82	0.06
SPF Sprache	ELFE II	0.50	0.14	1.00			
	ELFE II – Satzverständnis	0.57	0.16	0.94	0.02	1.00	
	ELFE II – Wortverständnis	0.40	0.15	0.96	0.01	0.82	0.06
SPF Emotionale und soziale Entwicklung	ELFE II	0.69	0.12	1.00			
	ELFE II – Satzverständnis	0.60	0.16	0.89	0.04	1.00	
	ELFE II – Wortverständnis	0.65	0.13	0.92	0.03	0.65	0.11
Unter Kompetenzstufe III	ELFE II	0.41	0.07	1.00			
	ELFE II – Satzverständnis	0.42	0.07	0.91	0.01	1.00	
	ELFE II – Wortverständnis	0.35	0.07	0.94	0.01	0.71	0.02
Auf oder über Kompetenzstufe III	ELFE II	0.33	0.02	1.00			
	ELFE II – Satzverständnis	0.41	0.02	0.90	0.01	1.00	
	ELFE II – Wortverständnis	0.19	0.02	0.90	0.01	0.63	0.02

Anmerkungen: Alle dargestellten Zusammenhangsmaße erreichen eine statistische Bedeutsamkeit ($p \leq .001$). *r* = Bravais-Pearson Korrelationskoeffizient; *SE* = Standardfehler der Korrelation

stufenmodells differenzierter zu deuten, als es das Kompetenzmodell gegenwärtig anbietet. Dies käme einem kriterialen Abgleich eines validen Instruments zur Kompetenzmessung mit einem validen Instrument der Individualdiagnostik gleich. Dafür wurden Pearson Korrelationen zwischen den ELFE II-Testteilen und den *PV* herangezogen (Tabelle 4).

Wie dargestellt, hängen das Satzverständnis und das Wortverständnis hoch (Cohen, 1988) zusammen ($r = .69$; $SE = 0.02$) und beide Testteile erwartungsgemäß sehr hoch mit der gesamten Kurzform des ELFE II ($r_w = .92$; $SE_w = 0.01$; $r_s = .92$; $SE_s = 0.01$). Zwischen der Kurzform des ELFE II und den *PV* im Lesen liegt bei Berücksichtigung des Konfidenzintervalls ein mittlerer bis großer Effekt ($r = .53$; $SE = 0.02$). Ein deutlicher Unterschied besteht in den Zusammenhängen des Satzverständnisses ($r = .60$; $SE = 0.02$) und des Wortverständnisses ($r = .38$; $SE = 0.03$) mit den *PV*. Während der Zusammenhang für das Satzverständnis groß ist, ist der für das Wortverständnis nur von mittlerer Größe.

Tabelle 5 zeigt die Zusammenhänge konditional für Leseleistungen der kategorisierten SPF-Gruppen und orientiert an der Zuordnung im Kompetenzmodell.

Wie dargestellt variieren die Zusammenhangsmaße des ELFE II mit den Kompetenzwerten aus IGLU zwischen Schülerinnen und Schülern mit und ohne SPF bedeutsam ($\Delta r = -.18$; $SE_{\Delta r} = .06$). Insbesondere gilt dies für das Wortverständnis zwischen Schülerinnen und Schülern mit und ohne SPF ($\Delta r = -.25$; $SE_{\Delta r} = .05$). Ein vergleichbares Muster ergibt sich für das Wortverständnis im Vergleich zur Schwelle um Kompetenzstufe III ($\Delta r = -.16$; $SE_{\Delta r} = .02$). Auch innerhalb der kategorisierten SPF-Gruppen liegen Differenzen vor, die aber ob der vergrößerten *SE* in den verkleinerten Gruppen nicht inhaltlich belastet werden. Die Varianz der Zusammenhänge deutet auf eine Nichtlinearität in der Zusammenhangsstruktur hin. Diese wird visualisiert, indem die Lesekompetenz (basierend auf den mitt-

leren *PV*) als für Intervalle des ELFE II konditionale Gauß'sche Kerndichteschätzungen ($k(t)$) abgetragen werden (Abbildung 1). Die abgetragenen Kurven zeigen also die Verteilungen der Leseleistungen innerhalb von Gruppen, welche auf der Basis der ELFE II-Ergebnisse gebildet wurden. In der Abbildung 1 werden Gruppen auf Basis der Standardabweichungen der ELFE-II-Ergebnisse gebildet.

Anhand der hohen Überlappung der Perzentile zwischen $\leq 2 SD$ und $\geq 2 SD$ ist nachzuvollziehen, dass zwar eine klare Rangreihung, aber keine klare Differenzierung zwischen den Perzentilen in den höheren Leistungslagen vorliegt. In den niedrigen Leistungslagen ist diese Differenzierung jedoch deutlicher (Abbildung 2). Hier wird, orientiert am Manual (Lenhard et al., 2017), eine Kategorisierung der ELFE II-Ergebnisse auf der Basis von Schwellenwerten für den Verdacht einer Lese-Rechtschreibschwäche (LRS) vorgenommen ($\leq 1.5 SD =$ Starker Verdacht auf LRS; $\leq 1 SD =$ Verdacht auf LRS).

Abbildung 2 zeigt, dass diese inhaltliche Bewertung der ELFE II-Ergebnisse sich bedingt auf die Verteilung der IGLU-Ergebnisse übertragen lassen: Die Risikoeinschätzungen für einen starken Verdacht auf LRS streuen, um den Schwellenwert des Erreichens der Kompetenzstufe II, Risikoeinschätzungen für den Verdacht auf LRS streuen um den Schwellenwert des Erreichens der Kompetenzstufe III.

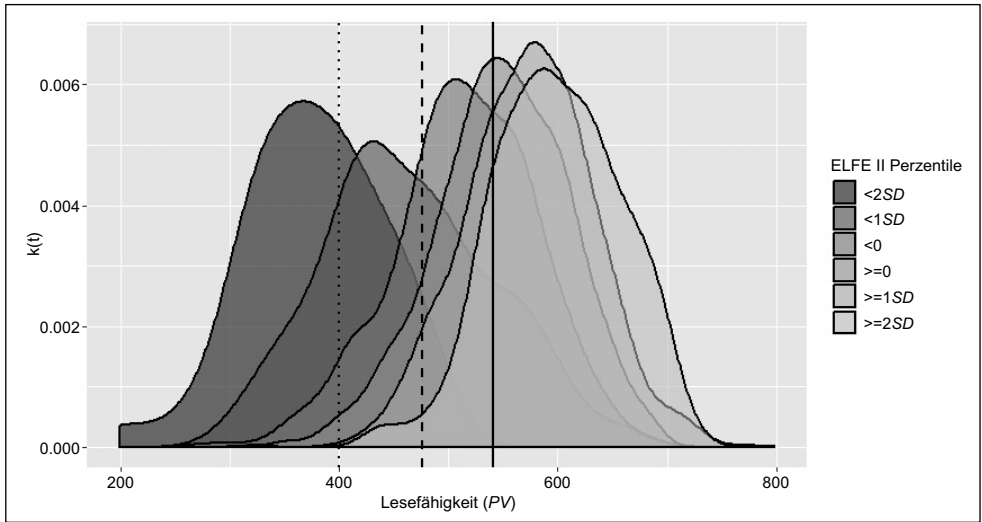


Abbildung 1: Dichteverteilung der Lesefähigkeit (PV) konditional für Perzentile des ELFE II.

Anmerkungen: Die durchgezogene Linie zeigt den Gesamtmittelwert der Lesekompetenz. Die gestrichelte Linie zeigt den Schwellenwert, ab dem die Kompetenzstufe III erreicht wird. Die gepunktete Linie zeigt den Wert, ab dem die Kompetenzstufe II erreicht wird.

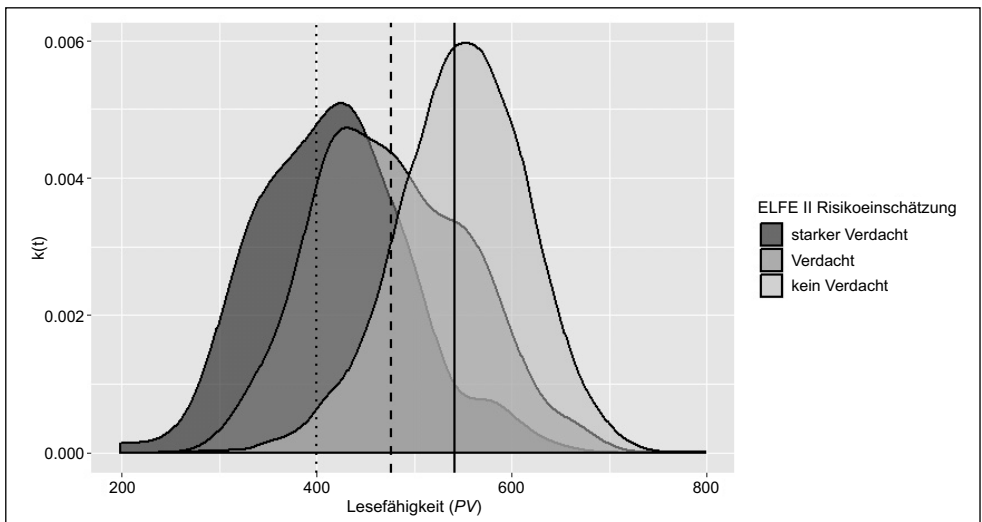


Abbildung 2: Dichteverteilung der Lesefähigkeit (PV) konditional für die Risikoeinschätzung des ELFE II in Bezugnahme auf LRS.

Anmerkungen: Die durchgezogene Linie zeigt den Gesamtmittelwert der Lesekompetenz. Die gestrichelte Linie zeigt den Schwellenwert, ab dem die Kompetenzstufe III erreicht wird. Die gepunktete Linie zeigt den Wert, ab dem die Kompetenzstufe II erreicht wird.

Diskussion

Das Leistungskontinuum, das in IGLU angelegt wird, erscheint mit den in Abschnitt 4 berichteten Befunden zwar angemessen um auch Leistungen ‚unter der Norm‘ abzubilden, aber die inhaltliche Belastung der Bereiche ober- und unterhalb der Schwellen im Vergleich zu einem Instrument, welches Prävalenzen bestimmen kann, ist problematisch.

Zur Verortung der Lesekompetenzen von Schülerinnen und Schülern mit SPF im IGLU-Kompetenzmodell ist demnach festzuhalten, dass die Ergänzung der ELFE-II-Testteile zur differenzierten Beschreibung niedriger Leseleistungen im Kompetenzstufenmodell möglich und hilfreich ist, um die bislang wenig praxisorientierten Stufen von Lesekompetenz ‚unter der Norm‘ mit inhaltlicher Bedeutung zu füllen. So zeigt sich der Mehrwert der ELFE II-Testteile vor allem in der zusätzlichen Beschreibbarkeit der unteren Kompetenzstufen durch den Verdacht auf LRS, da Leseleistungen unterhalb von Kompetenzstufe III starke Zusammenhänge mit einem Verdacht auf LRS aufweisen.

Kompetenzmodelle in LSA, wie hier für IGLU 2016 aufgegriffen, sind weder dafür konzipiert noch dazu geeignet individuelle diagnostische Empfehlungen zu formulieren. Vielmehr sollen sie robuste Vergleichswerte auf a priori definierten Strukturebenen zur Verfügung stellen, wie in IGLU zum internationalen Vergleich. Ungeachtet vieler Bemühungen zur Umsetzung einer gelingenden Inklusion werden Schülerinnen und Schüler mit SPF oder mit geringen Leistungen noch immer als Sonderfälle begriffen, welche, abseits einer definierten Norm, für die Ableitung populationsbeschreibender Kennwerte von geringem Interesse sind, obwohl sie einen relevanten Anteil der Schülerschaft ausmachen. Dabei könnte gerade die Adressierung *aller* Schülerinnen und Schüler in internationalen LSA im Sinne eines vollständigen Monitorings des deutschen Bildungssystems ein Gradmesser für

das Erreichen von Inklusion sein. Gelingt es, alle Schülerinnen und Schüler in das Bildungsmonitoring zu integrieren, dann sind (spätestens!) auch die Kompetenzstufenmodelle, mit denen die ermittelten Leistungen der Schülerinnen und Schüler inhaltlich interpretiert werden können, entsprechend anzupassen. Für das IGLU-Kompetenzstufenmodell zeigen sich insbesondere für die niedrigen Stufen Lücken, die mit Inhalten zu füllen wären, um zu beschreiben, was Schülerinnen und Schüler mit Leseleistungen auf diesen Stufen können. Mit der Ergänzung der ELFE II-Testteile in IGLU 2016 ist der Versuch unternommen worden, entsprechendes zusätzliches Wissen zu generieren. Der vorliegende Beitrag hat sich mit dem Fokus auf die Leseleistungen von Schülerinnen und Schülern mit SPF und niedrige Leseleistungen von Schülerinnen und Schülern ohne SPF diesbezüglichen Möglichkeiten gewidmet. Als relevant erachtet haben wir dabei auch die Frage, ob die Verfahrensweisen über die Kompetenzmodellierung und die Kompetenzstufen die SPF-Gruppen inhaltlich bewerten können.

Bevor die Ergebnisse zusammenfassend diskutiert werden, möchten wir auf Grenzen des Beitrags hinweisen: Kritisch ist 1.) zu bewerten, dass ein kleiner Anteil an Schülerinnen und Schülern mit SPF nicht an der Untersuchung teilnehmen konnte, weil Voraussetzungen nicht gegeben waren. Jedweder Test kann nur eine praktisch begrenzte Spannweite von vergleichbaren Befähigungen abbilden. So die Anteile der betroffenen Schülerinnen und Schüler klein genug sind, dass diese keinen relevanten Einfluss auf die adressierten Analysen haben, so muss akzeptabel sein, dass diese nicht berücksichtigt werden. Zudem wurden 2.) Hintergrundmerkmale zwar beschrieben, nicht aber für die Analysen kontrolliert. Hier ist anzumerken, dass in vorangegangener Forschung die Bedeutsamkeit dieser Merkmale hinreichend belegt wurde (s. Abschnitt 2). In unseren Analysen wurden die Hintergrundmerkmale allein zur Beschreibung und Bestätigung der Belastbarkeit der

Stichprobe verwendet. Dies kann zu Konfundierungen führen. Ein zentrales Hintergrundmerkmal konnte zudem nicht kontrolliert werden: das Bundesland. Regelmäßige Testungen erreichter Lesekompetenzen am Ende der vierten Jahrgangsstufe im *IQB-Bildungstrend* zeigen, dass die Leseleistungen zwischen den Bundesländern zum Teil stark variieren (Stanat et al., 2017), und es ist anzunehmen, dass die variable Umsetzung inklusiver Beschulung große Varianzen zwischen den Bundesländern bedingt. Da IGLU 2016 für internationale Vergleiche konzipiert ist, ist ein Vergleich der Bundesländer nicht möglich, aber auch nicht nötig, weil es andere Studien wie den *IQB-Bildungstrend* gibt, die hierzu Auskunft erteilen.

Zur Datengrundlage ist festzuhalten, dass die Stichprobenziehung in IGLU 2016 zwar populationsbeschreibenden Charakter hat, das Stratum des SPF und der inklusiven Beschulung insbesondere aber das Stratum der Förderschulen in der Stichprobenziehung bislang nur bedingt berücksichtigt wurden. Hier ist anzuregen, dass in zukünftigen Studienzyklen wenigstens Förderschulen jedes Bundeslandes als Ergänzungstichprobe und Schulen mit inklusivem Anteil in angemessener Form gezogen werden sollten. Zudem ist zu klären, ob das internationale Modell zur Kompetenzschätzung, wie es in IGLU basierend auf Modellen der *Item-Response*-Theorie erfolgt, trennscharf für Schülerinnen und Schülern mit niedrigen Leistungen und SPF ist. Dies kann beispielsweise über Analysen auf Messinvarianz und die Evaluation der Personen-Fit Werte nachvollzogen werden. Für den vorliegenden Beitrag wurde davon abgesehen, da eine Orientierung an der internationalen Skalierung erfolgte.

Unsere Analysen zeigen, dass die mittleren Leseleistungen von Schülerinnen und Schülern mit SPF am Ende der vierten Jahrgangsstufe orientiert am Kompetenzstufenmodell zum größten Teil höchstens Kompetenzstufe III erreichen. Inwiefern diese Orientierung sinnvoll ist, zeigen die vertie-

fenden Analysen des IGLU-Lesekompetenztests mit dem ELFE II: die Ergebnisse der Leistungstests aus IGLU und des ELFE II weisen hohe Übereinstimmungen auf, aber auch gegenseitige Unschärfen. Der ELFE II zeigt für den unteren Leistungsbereich sowohl eine trennschärfere Unterscheidungsfähigkeit als auch eine wertvolle inhaltliche Deutungsmöglichkeit auf, die das Kompetenzstufenmodell in IGLU nicht leistet. Es konnte eine überzufällige Übereinstimmung zwischen den durch den ELFE II ausgewiesenen Prävalenzen und den Kompetenzschwellenwerten ausgewiesen werden, welche ein Hinterfragen der inhaltlichen Belastbarkeit des Kompetenzstufenmodells unterhalb der Kompetenzstufe III und gegebenenfalls auch eine Diskussion zur Notwendigkeit des Stufenmodells nötig macht. Da das Kompetenzstufenmodell den Bildungsstandards entlehnt ist, muss herausgestellt werden, dass eine praxisorientierte, also an der konkreten Lesefähigkeit orientierte Definition der Standards im unteren Leistungsbereich erstrebenswert sein sollte. Es ist nachfolgenden Forschungsarbeiten in Zusammenarbeit mit Expertinnen und Experten fachlicher Expertise vorbehalten, im Anschluss an diese Erkenntnisse das Kompetenzstufenmodell im unteren Leistungsbereich inhaltlich weiter auszudifferenzieren und den Transfer in die Schulpraxis beispielsweise mit der Entwicklung und Erprobung entsprechender Materialien zu forcieren. So kann der besonders starke Zusammenhang der Dimension des Wortverständnisses (ELFE II) mit den IGLU-Kompetenzwerten im unteren Leistungsbereich als Beleg dafür verstanden werden, dass a) eine hierarchische Struktur von Wort- und Satzverständnis valide ist (Lenhard et al., 2017), b) die IGLU-Werte auch in diesem Bereich tendenziell trennscharf sind und c) die Kompetenzstufen I und II für zukünftige Studienzyklen stärker differenziert werden müssen. Alle drei Punkte bieten konkrete Ansatzpunkte für die Ausgestaltung des Kompetenzstufenmodells.

Literaturverzeichnis

- Bremerich-Vos, A., Wendt, H. & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In A. Hußmann et al. (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79-142). Münster: Waxmann.
- Carstensen, C. H., Knoll, S., Rost, J. & Prenzel, M. (2004). Technische Grundlagen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 371-388). Münster: Waxmann.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Auflage). Hillsdale, NJ: Erlbaum.
- Gebhardt, M., Heine, J.-H. & Sälzer, C. (2015). Schulische Kompetenzen von Schülerinnen und Schülern ohne sonderpädagogischen Förderbedarf im gemeinsamen Unterricht. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete VHN*, 84(3), 246-258.
- Göransson, K. & Nilholm, C. (2014) Conceptual diversities and empirical shortcomings – a critical analysis of research on inclusive education. *European Journal of Special Needs Education*, 29(3), 265-280, DOI: 10.1080/08856257.2014.933545
- Gresch, C., Kölm, J. & Kocaj, A. (2017). Amtlich festgestellter sonderpädagogischer Förderbedarf und sonderpädagogische Förderung. 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. 282-290). Münster: Waxmann.
- Hußmann, A. et al. (Hrsg.). (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Hußmann, A., Wendt, H., Kasper, D., Bos, W. & Goy, M. (2017). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU 2016). In A. Hußmann et al. (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 29-78). Münster: Waxmann.
- Hußmann, A., Stubbe, T. C. & Kasper, D. (2017). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann et al. (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195-218). Münster: Waxmann.
- Hurrelmann, B. (2007). Modelle und Merkmale der Lesekompetenz. In A. Bertschi-Kaufmann (Hrsg.), *Lesekompetenz – Leseleistung – Leseförderung. Grundlagen, Modelle und Materialien* (S. 18-29). Seelze-Velber & Zug: Friedrich Verlag & Klett & Balmer.
- Kintsch, W. (2012). Psychological models of reading comprehension and their implications for assessments. In J. Sabatini, E. Albro & T. O'Reilly (Hrsg.), *Measuring up: Advances in how to assess reading ability* (S. 21-37). Plymouth: Rowman & Littlefield Publishers.
- Klicpera, C. & Gasteiger-Klicpera, B. (1995). *Psychologie der Lese- und Rechtschreibschwierigkeiten: Entwicklung, Ursachen, Förderung*. Weinheim: Beltz.
- KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2016). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Zugriff am 15.04.2019 unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf
- KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2011a). *Kompetenzstufenmodell zu den Bildungsstandards für das Fach Deutsch im Kompetenzbereich „Lesen – mit Texten und Medien umgehen“ für den Primarbereich*. Zugriff am 19.04.2019 unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2011/Kompetenzstufenmodell_zu_den_Bildungsstandards_fuer_das_Fach_Deutsch_im_Kompetenzbereich_Lesen_mit_Texten_und_Medien_umgehen_fuer_den_Primarbereich.pdf

- iqb.hu-berlin.de/bista/ksm/KSM_GS_Deutsch_L.pdf
- KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2011b). *Inklusive Bildung von Kindern und Jugendlichen mit Behinderungen in Schulen* (Beschluss vom 20.10.2011). Zugriff am 23.11.2017 unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2011/2011_10_20-Inklusive-Bildung.pdf
- KMK – Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2004). *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung*. Zugriff am 30.06.2019 unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Konzeption-Entwicklung.pdf
- 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.
- Lenhard, W., Lenhard, A. & Schneider, W. (2017). *ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler – Version II*. Göttingen: Hogrefe.
- Lüdtke, O. & Robitzsch, A. (2017). Eine Einführung in die Plausible-Values-Technik für die psychologische Forschung. *Diagnostica*, 63, 193-205. DOI: 10.1026/0012-1924/a000175
- Martin, M. O., Mullis, I. V. S. & Foy, P. (2015). Assessment design for PIRLS, PIRLS Literacy, and ePIRLS in 2016. In I. V. S. Mullis & M. O. Martin (Hrsg.), *PIRLS 2016 assessment framework* (2. Auflage, S. 55–70). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Huber.
- Sälzer, C., Gebhardt, M., Müller, K. & Pauly, E. (2015). Der Prozess der Feststellung sonderpädagogischen Förderbedarfs in Deutschland. In P. Kuhl, P. Stanat, B. Lütje-Klose, C. Gresch, H.A. Pant & M. Prenzel (Hrsg.), *Inklusion von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in Schulleistungserhebungen* (S.129-152). Wiesbaden: Springer.
- Schneider, W. (2016). *Lesen und Schreiben lernen: Wie Kinder die Schriftsprache erobern*. Heidelberg: Springer-Spektrum.
- Stanat, P. Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (2017). *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.
- Wendt, H. & Schwippert, K. (2017). Lesekompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund. In A. Hußmann et al. (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 219-234). Münster: Waxmann.

Dr. Anke Hußmann

TU Dortmund

Fakultät Rehabilitationswissenschaften

Unterrichtsentwicklungsforschung

mit dem Schwerpunkt Inklusion

Emil-Figge-Straße 50

44227 Dortmund

Tel.: 0231 – 755 5831

Erstmalig eingereicht: 24.04.2019

Überarbeitung eingereicht: 21.08.2019

Angenommen: 29.08.2019