

Kompetenzzentrum Schulpsychologie Hessen

Diagnostik schulischer Fertigkeiten bei Kindern mit Deutsch als Zweitsprache (DaZ)

Janin Brandenburg, Deutsches Institut für Internationale Pädagogische Forschung | Leibniz-Institut für
Bildungsforschung und Bildungsinformation
brandenburg@dipf.de

Nach neusten Schätzungen ist Deutschland mittlerweile das zweitstärkste Einwanderungsland innerhalb der OECD (OECD, 2018). Diese Entwicklung spiegelt sich auch in der Zusammensetzung der Schülerschaft wider: Kinder mit Migrationshintergrund¹ bilden eine zunehmend wachsende Subpopulation an deutschen Grundschulen und machen gegenwärtig einen Anteil von 34-37% unter den 6- bis 15-jährigen Schülerinnen und Schülern aus (Autorengruppe Bildungsberichterstattung, 2018). Im Zuge der Debatte um *Bildungsgerechtigkeit* ist der Bildungserfolg dieser Schülergruppe immer wieder Gegenstand der schulpolitischen und wissenschaftlichen Diskussion in Deutschland, denn Schülerinnen und Schüler mit Migrationshintergrund zeigen im Vergleich zu ihren Klassenkameraden ohne Migrationshintergrund oftmals Kompetenznachteile im Leseverständnis (Schwippert, Wendt & Tarelli, 2012) sowie in Mathematik und den Naturwissenschaften (Gebhardt, Rauch, Mang, Sälzer & Stanat, 2013; Tarelli, Schwippert & Stubbe, 2012).

Daher haben sich mittlerweile viele deutsche Forschungsarbeiten der besonderen Bildungssituation dieser Schülergruppe gewidmet und mehr Erkenntnisse darüber gewonnen, wie sich diese herkunftsbedingten Disparitäten erklären lassen. Eine entscheidende Determinante für den Bildungserfolg von Kindern mit Migrationshintergrund stellt dabei die Kompetenz in der Unterrichtssprache Deutsch dar. So lässt sich ein beachtlicher Anteil der Leistungsunterschiede zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund auf den familiären Sprachgebrauch zurückführen (Tiedemann &

Billmann-Mahecha, 2007). Beispielsweise waren in PISA 2015 die Bildungsdisparitäten zwischen Jugendlichen mit und ohne Migrationshintergrund um 43 Leistungspunkte reduziert, wenn die zugewanderten Familien zu Hause vornehmlich Deutsch sprachen (Rauch, Mang, Härtig & Haag, 2016). Auch in der internationalen Grundschulvergleichsstudie IGLU 2016 war ein deutlicher Effekt des familiären Sprachgebrauches zu verzeichnen: Hier betrug der Leistungsrückstand in der Lesekompetenz von Viertklässlerinnen und Viertklässlern, die zu Hause nie oder nur manchmal Deutsch sprachen, 40 Leistungspunkte und entsprach somit der Leistungsentwicklung eines ganzen Schuljahres im Vergleich zu den Kindern, die zu Hause (fast) immer Deutsch sprachen (Wendt & Schwippert, 2017). Der familiäre Sprachgebrauch nimmt dabei eine medierende Rolle zwischen dem Zuwanderungshintergrund und der Leistungskompetenz ein (z. B. McElvany, Becker & Lüdtkke, 2009): Deutsch als Familiensprache ist prädiktiv für einen höheren Wortschatz in der deutschen Sprache, was sich beispielsweise positiv auf das Leseverständnis auswirkt. Umgekehrt können Schülerinnen und Schüler, die die Unterrichtssprache nur unzureichend beherrschen, auch den Lerninhalten nur schwer folgen und zeigen daher geringere Lernfortschritte (vgl. Butler & Stevens, 2001). Sie können die Lernmöglichkeiten, die ihnen im Unterricht geboten werden, also weniger gut nutzen. Gute Kompetenzen in der Unterrichtssprache gelten daher als wesentliche Schlüsselkompetenz für den schulischen Wissens- und Kompetenzerwerb. Daneben trägt auch das (sozio-)ökonomische und kulturelle Kapital der Familie sowie

¹ Ein Migrationshintergrund ist im Folgenden über eine familiäre Migrationserfahrung in der ersten, zweiten oder dritten Generation definiert. Diese Definition deckt sich mit der Operationalisierung eines Migrationshintergrundes im deutschen Bildungsbericht (Autorengruppe

Bildungsberichterstattung, 2018) und entspricht auch der Operationalisierung, die in der internationalen Vergleichsstudie PISA verwendet wird.

das Bildungsniveau der Eltern wesentlich zur Erklärung der Leistungsunterschiede zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund bei. Unterschiede gibt es zusätzlich in Abhängigkeit der Herkunftsstaaten und der Zuwanderungsgeneration (Rauch, Mang, Härtig & Haag, 2016).

Neben ihrer Bedeutung für den Erwerb domänenspezifischen Wissens wird die Relevanz sprachlicher Kompetenzen auch in anderen Bereichen zunehmend diskutiert, beispielsweise im Kontext des standardisierten Testens oder im Rahmen der pädagogisch-psychologischen Diagnostik. So wird bei (inter-)nationalen Vergleichsstudien im Rahmen des *Large-Scale Assessments* aber auch bei der Konstruktion von standardisierten Leistungstests für die Individualdiagnostik zunehmend kritisch eruiert, inwiefern es zu einem sog. *Test Bias* für Schülerinnen und Schüler mit nicht-deutscher Muttersprache kommt (z. B. Schwabe, McElvany, Trendtel, Gebauer & Bos, 2014; García, McKoon, & August, 2008). Auch wird empirisch geprüft, durch welche Maßnahmen und/oder Veränderungen in der Testprozedur ein solcher Bias behoben werden kann (z. B. Kieffer, Lesaux, Rivera & Francis, 2009). Der vorliegende Beitrag gibt einen Überblick über das Konzept des Test Bias und leitet Folgerungen für die Anwendung diagnostischer Verfahren bei Kindern mit Deutsch als Zweitsprache im pädagogisch-psychologischen Kontext ab.

Was bedeutet Test Bias?

Ein Test Bias liegt vor, wenn ein Testverfahren (unabsichtlich) so konstruiert ist, dass es Personen entsprechend ihrer Zugehörigkeit zu bestimmten ethnischen oder (sozio-)kulturellen Gruppen systematisch benachteiligt und andere bevorzugt (vgl. Hartig, Frey & Jude, 2012) und zwar in der Form, dass Personen trotz gleicher zugrundeliegender Fähigkeiten unterschiedlich gute Testergebnisse erzielen. Das sei an einem Beispiel veranschaulicht: Mit den beiden Drittklässlern Jakob und Tayyeb wurde ein standardisierter Leseverständnistest durchgeführt. Jakobs Muttersprache ist Deutsch, Tayyeb hingegen hat erst seit dem 6. Lebensjahr regelmäßig Kontakt mit der deutschen Sprache und erwirbt Deutsch daher als Zweitsprache. Nehmen wir nun an, die Schüler erzielen in dem Leseverständnistest unterschiedliche Leistungen (gemessen an den beobachtbaren Testrohwerten). So konnte Jakob acht von insgesamt 24 Aufgaben mehr lösen als Tayyeb. Bei der Interpretation von Testergebnissen gehen wir davon aus, dass der *beobachtbare Rohwert* – also die gezeigte *Leistung* – als Schätzer für die zugrundeliegende „wahre“ *Kompetenz* im Lesen herangezogen werden kann. Wir würden anhand der Testergebnisse also zu dem Schluss kommen, dass die wahre (aber nicht direkt beobachtbare) Lesekompetenz von Jakob entsprechend höher ist als die von Tayyeb. Sofern Jakobs wahre Lesekompetenz tatsächlich entsprechend höher ist,

bildet der verwendete Leseverständnistest diese Kompetenzunterschiede gut ab. Ein Test Bias hingegen läge vor, wenn Jakob und Tayyeb in Wirklichkeit über eine vergleichbar hohe wahre Lesekompetenz verfügen, aber Tayyeb (und andere Kinder mit DaZ) in dem Test systematisch schlechter abschneiden – also systematisch geringere Rohwerte erzielen – als Kinder mit Deutsch als Muttersprache (DaM). Der Test führt in diesem Fall also zu Verzerrungen bei Kindern mit DaZ. Eine Unterschätzung der zugrundeliegenden Lesekompetenz von Kindern mit DaZ ist die Folge.

Ein Test Bias liegt also vor, wenn Personen unterschiedlicher (ethnischer oder soziokultureller) Gruppen, die über die gleiche wahre aber nicht direkt beobachtbare Kompetenz verfügen, in einem standardisierten Leistungstests systematisch unterschiedliche beobachtete Rohwerte erhalten. Wichtig dabei ist, dass ein Test Bias nicht zwangsläufig vorliegt, sobald Schülerinnen und Schüler mit und ohne DaZ im Mittel unterschiedlich gute Testergebnisse erzielen. Vielmehr liegt ein Test Bias nur dann vor, wenn Schülerinnen und Schüler mit und ohne DaZ, die nach zugrundeliegender Fähigkeit parallelisiert sind, unterschiedlich gut abschneiden (siehe auch Lenhard & Lenhard, 2017). Ein Test, der über einen Test Bias verfügt, verletzt das Gütekriterium der *Testfairness*.

Was sind mögliche Gründe für einen Test Bias bei Kindern mit DaZ?

Damit ein Testverfahren als fair angesehen werden kann, ist es zentral, dass mögliche Verzerrungen – sofern sie denn vorliegen – nicht unterschiedlich stark ausfallen je nach Zugehörigkeit zu einer bestimmten Personengruppe. Bei der (standardisierten) Leistungsdiagnostik von Schülerinnen und Schülern, deren Muttersprache nicht die Unterrichtssprache ist, kann nach García und Pearson (1994) ein *Test Bias* im Vergleich zu Kindern ohne Zweitspracherwerb vor allem auf drei Arten entstehen:

1. Ein *Linguistic Bias* kann beispielsweise vorliegen, wenn die Aufgabenstellungen aufgrund sprachlicher Barrieren falsch oder nur unvollständig verstanden werden. Gerade bei standardisierten Schulleistungstests, deren Instruktionen zur Gewährleistung der Durchführungsobjektivität wörtlich aus dem Manual vorzulesen sind, sollte man sich daher vergewissern, dass die Aufgabenstellung wirklich verstanden wurde. Von einem Linguistic Bias wird ebenso gesprochen, wenn bei der Bearbeitung eines standardisierten Leistungstests der Wortschatz oder die allgemeinen Sprachkompetenzen des Kindes nicht ausreichen, um sein domänenspezifisches Wissen auch entsprechend zu demonstrieren. Ein Linguistic Bias läge beispielsweise vor, wenn ein standardisierter Mathematiktest so hohe sprachliche

Anforderungen an die Schülerinnen und Schüler stellt, dass die gezeigte Testleistung nicht nur maßgeblich durch die zugrundeliegende mathematische Kompetenz hervorgerufen wird, sondern auch stark durch die sprachlichen Kompetenzen der Schülerinnen und Schüler beeinflusst ist. In diesem Fall kann beim Vorliegen eines schwachen Testergebnisses bei Kindern mit DaZ nicht einwandfrei abgeleitet werden, dass das Ergebnis tatsächlich durch eine entsprechend niedrige mathematische Kompetenz hervorgerufen wurde (vgl. Schwabe & Gebauer, 2013). Eine Unterschätzung der mathematischen Fähigkeiten bei Kindern mit DaZ ist die Folge.

Nach García und Pearson (1994) bergen insbesondere sprachliche Aufgaben im Multiple-Choice-Format die Gefahr eines Linguistic Bias: Da sich die jeweiligen Antwortalternativen oftmals nur leicht voneinander unterscheiden, kommt dem korrekten Verständnis von Syntax und Wortbedeutung eine größere Bedeutung zu als dies bei anderen Aufgabenformaten typischerweise der Fall ist. Sofern das Testverfahren sprachliche Kompetenzen gezielt erfassen möchte (etwa weil es sich um ein Sprachverständnistest handelt), ist dies unproblematisch. Problematisch wird es hingegen dann, wenn der Test auf die Erfassung gänzlich anderer Kompetenzen abzielt, aber das Sprachverständnis durch die Formulierung der Items unbeabsichtigt mit in das Testergebnis reinspielt.

2. Ein *Content Bias* entsteht, wenn die Lösungsgüte in einem Leistungstest auf unbeabsichtigte Weise durch das kulturelle Wissen über das Aufnahmeland beeinflusst wird bzw. davon abhängt, wie stark eine kulturelle Integration an das Aufnahmeland stattgefunden hat. Dies kann beispielsweise dann der Fall sein, wenn die in einem Intelligenz- oder Leseverständnistest vorgegebenen Szenarien Lebenssituationen widerspiegeln, die zwar typisch für die Kultur des Aufnahmelandes sind, aber nicht für die des Herkunftslandes oder wenn ein Wortschatztest kulturelle Besonderheiten des Aufnahmelandes abfragt. Im weiteren Sinne fassen Schwabe und Gebauer (2013) hierunter auch, wenn Testinhalte für unterschiedliche Personengruppen unterschiedlich interessant sind und daher je nach Gruppenzugehörigkeit unterschiedliche Motivation und Anstrengungsbereitschaft hervorrufen.
3. Ein *Norming Bias* liegt schließlich vor, wenn die Testnormen den Zweitspracherwerb der Schülerinnen und Schüler nicht ausreichend berücksichtigen. Besonders augenscheinlich ist dies dann, wenn ein standardisierter Leistungstest (fast) ausschließlich an Schülerinnen und Schüler mit deutscher Muttersprache normiert wurde, aber in der diagnostischen Praxis dennoch zur

Leistungseinschätzung von Kindern mit anderer Muttersprache herangezogen wird.

Während sich der Linguistic Bias und Content Bias auf die sog. *prozedurale Testfairness* beziehen, wird der Norming Bias der sog. *interpretationsbezogenen Testfairness* zugeordnet. Die prozedurale Testfairness beschäftigt sich also mit der Frage, ob die Gestaltung der Testinstrumente oder die Testdurchführung bestimmte Personengruppen systematisch und unbeabsichtigt benachteiligt – oder anders ausgedrückt: inwiefern die Testaufgaben über verschiedene Gruppen hinweg gleich gut funktionieren. Die interpretationsbezogene Testfairness hingegen stellt sicher, dass die aus den Testergebnissen resultierenden Entscheidungen über verschiedene Subgruppen hinweg gleichermaßen valide sind (siehe Schwabe & Gebauer, 2013, für einen Überblick zu beiden Aspekten von Testfairness).

Zusammenfassend ergibt sich also aus dem Gütekriterium der Testfairness, dass diagnostische Instrumente prozedural fair gestaltet sein sollten und dass die aus ihnen abgeleiteten Schlussfolgerungen und Diagnosen nicht interpretativ verzerrt sein sollten gegenüber solchen Kindern, für die die Testsprache nicht die Muttersprache ist. Wissenschaftliche Diskussionen, die sich um das Phänomen Test Bias bzw. Testfairness drehen, sind jedoch nicht neu, sondern greifen auf eine lange Forschungstradition zurück, die in den frühen 60-er Jahren zur Entwicklung der sog. *kultur-fairen Intelligenztests* führte. In der aktuellen Diskussion wird die Frage des Test Bias jedoch in einem breiteren Kontext eingebettet, indem sie verstärkt auch auf die Erfassung anderer Kompetenzbereiche (wie etwa die Schriftsprache und Mathematik) bezogen wird.

Wie lässt sich Testfairness sicherstellen bzw. erhöhen?

Für die pädagogisch-psychologische Leistungsdiagnostik ergibt sich schließlich die Frage, wie sich bei Kindern mit DaZ eine faire Leistungsdiagnostik sicherstellen lässt. In der Forschung wird die Frage, welche Kriterien und Faktoren beim Vorliegen eines Zweitspracherwerbes eine hohe Validität der diagnostischen Entscheidung sicherstellen, aktuell stark untersucht. Im Folgenden werden erste Empfehlungen für die Diagnostik abgeleitet.

1. *Abklärung des Sprachstandniveaus im Deutschen*
Bei der standardisierten (Schul-)leistungsdiagnostik von Schülerinnen und Schülern mit DaZ ist zunächst eine Überprüfung des sprachlichen Entwicklungsstandes im Deutschen sinnvoll, um zu eruieren, wie weit der Zweitspracherwerb bereits fortgeschritten ist und über welches Sprachstandsniveau das Kind verfügt (vgl. auch Durkin, 2000). Neben einer ersten Abklärung des hinreichenden (Instruktions-)

verständnis kann dieses Wissen hilfreich sein, um ein geeignetes Testverfahren für die anschließende Leistungsdiagnostik auszuwählen. Sollen beispielsweise bei einem Kind, das noch etwas Schwierigkeiten mit der deutschen Sprache hat, die Rechenfertigkeiten überprüft werden, bietet es sich zunächst an einen Test zu verwenden, der basale Rechenoperationen relativ sprachfrei testet (z. B. der DIRG von Grube, Weberschock, Blum & Hasselhorn, 2010 oder der KR 3-4 von Roick, Gölitz & Hasselhorn, 2011) anstatt einen Mathematiktest zu verwenden, dessen Rechenaufgaben sprachlich sehr komplex sind (z. B. durch einen hohen Anteil an Sach- und Textaufgaben).

Schließlich kann die Berücksichtigung des Sprachstandsniveaus auch bei der abschließenden Ergebnisinterpretation der Diagnostik von Relevanz sein: Erzielt ein Schüler oder eine Schülerin mit DaZ ein niedriges Ergebnis in einem standardisierten Leistungstest bei gleichzeitig gutem Abschneiden in einem Sprachtest, lassen sich unzureichende Deutschkenntnisse als alternative Erklärung der schwachen Leistungsergebnisse ausschließen. Es lässt sich also abschätzen, inwiefern ein *Linguistic Bias* die Testergebnisse verzerrt haben könnte und bei der Interpretation daher entsprechend berücksichtigt werden müsste.

Insbesondere bei Schülerinnen und Schülern, die erst seit kurzem in Deutschland leben, ist zusätzlich auch die Kontaktdauer mit der deutschen Sprache zu berücksichtigen, um zu entscheiden, wann sinnvollerweise eine erste Sprachtestung erfolgen kann. So dauert es typischerweise etwa sechs bis zehn Monate bis ein Basisniveau an sprachlichen Fähigkeiten erreicht ist (vgl. Grimm, 2012).

2. Finden sich im Testmanual Hinweise auf das Vorliegen prozeduraler Testfairness?

Zur Überprüfung prozeduraler Testfairness existieren verschiedene statistische Analysemethoden, die im Rahmen von Itemanalysen zunehmend an Bedeutung gewinnen und hoffentlich bald standardmäßig bei der Konstruktion neuer Leistungstests durchgeführt werden. So kann beispielsweise mit Analysen zum *Differential Item Functioning* (kurz: DIF) die Fairness der einzelnen Testitems empirisch überprüft werden. Bei einem Item liegt DIF vor, wenn Personen aus verschiedenen Gruppen, aber mit denselben zugrundeliegenden Fähigkeiten, systematisch unterschiedliche Lösungswahrscheinlichkeiten für dieses Item zeigen (vgl. Schwabe & Gebauer, 2013). Ist die Lösungswahrscheinlichkeit hingegen gleich, wird von *Messinvarianz* gesprochen. DIF-Analysen erlauben somit eine Aussage darüber, ob die Testitems über verschiedene Gruppen hinweg (z.

B. bei Kindern mit DaM oder mit DaZ) gleichermaßen gut funktionieren und das gleiche messen. Zur Leistungsdiagnostik von Kindern mit DaZ sollte also bestenfalls ein Testverfahren herangezogen werden, dessen Messinvarianz für Kinder mit DaM und mit DaZ empirisch gewährleistet ist.

Einige neue standardisierte Testverfahren geben in ihren Manualen bereits Hinweise auf DIF-Analysen und das Vorliegen von Messinvarianz zwischen Kindern mit und ohne DaZ. Als vorbildhaft ist hier der neue ELFE-II (*Ein Leseverständnistest für Erst- bis Siebtklässler – Version II*) von Lenhard und Kolleg(inn)en (2018) herauszustellen, da die Autor(inn)en die Itemformulierungen und -inhalte des ELFE II ausführlich in Bezug auf Test Bias getestet haben und Items, bei denen sich DIF gezeigt hat, aus der Endversion des Tests ausgeschlossen haben. So erlauben diese Analysen den Schluss, dass der ELFE II das Leseverständnis von Kindern mit und ohne Migrationshintergrund gleichermaßen gut misst.

Aber was, wenn für ein Testverfahren keine direkte Überprüfung der Itemfairness vorliegt? Diese Frage ist nicht trivial, denn aktuell dürfte das noch für die meisten standardisierten Testverfahren gelten. Nach Schwabe und Gebauer (2013) empfiehlt es sich in diesem Fall die im Testmanual angegebene Reliabilität zu prüfen: Liefert das Testmanual beispielsweise Hinweise darauf, dass die interne Konsistenz (also die Homogenität) der Items als vergleichbar gelten kann für Kinder mit DaM bzw. DaZ? Auch die im Testmanual angegebenen Hinweise zur Validität können ggf. indirekt einen Aufschluss geben. Werden beispielsweise zur Überprüfung der divergenten Validität eines Leistungstests die korrelativen Zusammenhänge mit einem Sprachtest angegeben, lässt sich daraus eruieren, inwieweit die Testergebnisse mit der Sprachleistung eines Kindes zusammenhängen. Auch geben manche Testverfahren in ihren Manualen Hinweise darauf, wie Kinder mit DaZ bei der Normierung abgeschnitten haben. So untersuchten beispielsweise Petermann und Daseking (2012) für ihren Lesetest ZLT-II, ob Kinder mit DaZ signifikant andere Leistungen erzielten als Kinder mit DaM, die in Bezug auf Alter, Klassenstufe, Schulform und Geschlecht mit der DaZ-Gruppe gemacht waren. Dabei zeigte sich, dass die beiden Gruppen im Wort- und Textlesen vergleichbare Ergebnisse erzielten.

Im Minimalfall sollte die Testanwenderin und der Testanwender zumindest die Iteminhalte und -formulierungen selbst einmal aufmerksam betrachten und versuchen die kulturelle und sprachliche Tauglichkeit der Items kritisch einzuschätzen.

3. Interpretationsbezogene Testfairness

Um das Kriterium der interpretationsbezogenen Testfairness zu berücksichtigen, sollte bei der Diagnostik von schulischen Fertigkeiten bei Kindern mit DaZ sichergestellt werden, dass das ausgewählte Testverfahren an einer (annähernd) repräsentativen Stichprobe von Kindern mit Deutsch als Erst- bzw. Zweitsprache normiert wurde. Dies mag trivial erscheinen, dennoch gibt es nach wie vor einige Testverfahren, die in ihren Manualen keine Hinweise darauf geben, ob und wie viele Kinder mit nicht-deutscher Muttersprache in der Eichstichprobe enthalten waren. Daneben gibt es auch aktuelle standardisierte Testverfahren, bei denen der Anteil an Kindern mit DaZ in der Normierungsstichprobe verschwindend gering ist.

Manche standardisierten Schulleistungstests verfügen neben einer (repräsentativen) Gesamtnorm zusätzlich über eine Extra-Norm für Kinder mit DaZ. Für den Bereich des Lesens sind es beispielsweise für das Dekodieren der DiLe-D (Paleczek, Seifert, Obendrauf, Schwab & Gasteiger-Klicpera, 2017) und für das Leseverständnis der ELFE-II (Lenhard et al., 2018). Für das Rechtschreiben liegen gesonderte Normen für DaZ beispielsweise in der WRT-Reihe von Birkel (2007) vor. Zur Einschätzung der Sprachleistungen bieten das Verfahren LiSe-DaZ (Schulz & Tracy, 2011) sowie der Wortschatztest GraWo (Seifert, Paleczek, Schwab & Gasteiger-Klicpera, 2017) einen getrennten Vergleich nach Spracherwerbstyp an. Die DaZ-Normen können indirekt eine Möglichkeit bieten abzuschätzen, inwieweit sich schwache Testergebnisse, die sich unter Anwendung der Gesamtnorm gezeigt haben, durch den Zweitspracherwerb erklären lassen. Ist das Testergebnis des Kindes unter Anwendung der DaZ-Norm nicht mehr im unterdurchschnittlichen Bereich, so lassen sich die schwachen Testergebnisse indirekt zu einem Stück weit durch die Herausforderung des Zweitspracherwerbs erklären. Dennoch sei an dieser Stelle betont, dass die Verwendung von DaZ-Normen lediglich ergänzend bei der Diagnostik eingesetzt werden sollte und eine Leistungseinschätzung anhand der Gesamtnorm nicht ersetzen darf. Geht es etwa um die Klärung der Frage, ob ein Kind gemessen an seiner Klassenstufe im Lernstand zurückliegt und Förderung benötigt, sind Referenzwerte am aussagekräftigsten, die an der gesamten Schülerschaft normiert sind. Die Verwendung der Gesamtnormen hilft in diesem Fall abzuschätzen, ob bei einem Kind (und zwar unabhängig von seinem Spracherwerbstyp) ein spezifischer Förderbedarf besteht, um einen Leistungsrückstand aufzuholen. Die DaZ-Norm sollte lediglich im Anschluss dazu genutzt werden, die möglichen Gründe der gezeigten Lernschwierigkeiten näher zu spezifizieren.

Fazit

Die Diagnostik schulischer Fertigkeiten von Kindern mit DaZ stellt die Schulpsychologie vor besondere Herausforderungen: Um ein schwaches Testergebnis tatsächlich auch als Ausdruck eines zugrundeliegenden Defizites im domänenspezifischen Wissen interpretieren zu können, sollte sichergestellt werden, dass das verwendete Testverfahren keinen Bias gegenüber DaZ hat. Es sollte im Sinne der Testfairness also sichergestellt werden, dass Kinder mit DaZ nicht systematisch durch das verwendete Testverfahren benachteiligt werden.

Literatur

- Autorengruppe Bildungsberichterstattung. (2018). *Bildung in Deutschland 2018 – Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung*. Bielefeld: W. Bertelsmann Verlag.
- Birkel, P. (2007). *Weingartener Grundwortschatz Rechtschreib-Test für zweite und dritte Klassen*. Göttingen: Hogrefe.
- Birkel, P. (2007). *Weingartener Grundwortschatz Rechtschreib-Test für dritte und vierte Klassen*. Göttingen: Hogrefe.
- Birkel, P. (2007). *Weingartener Grundwortschatz Rechtschreib-Test für vierte und fünfte Klassen*. Göttingen: Hogrefe.
- Butler, F. A. & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K-12: Current trends and old dilemmas. *Language Testing*, 18, 409–427. doi:10.1177/026553220101800406
- Durkin, C. (2000). Dyslexia and bilingual children – Does recent research assist identification? *Dyslexia*, 6, 248–267. doi:10.1002/1099-0909(200010/12)6:4<248::AID-DYS173>3.0.CO;2-O
- García, G. E., McKoon, G. & August, D. (2008). Language and literacy assessment. In D. August & T. Shanahan (Eds.), *Developing reading and writing in second-language learners: Lessons from the report of the National Literacy Panel on language-minority children and youth* (pp. 251–274). New York, NY: Routledge.
- García, G. E. & Pearson, P. D. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 20, pp. 337–392). Washington, DC: American Educational Research Association.
- Gebhardt, M., Rauch, D., Mang, J., Sälzer, C. & Stanat, P. (2013). Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012 – Fortschritte und Herausforderungen in Deutschland* (S. 275–308). Münster: Waxmann.
- Grimm, H. (2012). *Handreichung für Erzieherinnen und Erzieher zum SETK 3–5. Was wird warum untersucht?*. Göttingen: Hogrefe.
- Grube, D., Weberschock, U., Blum, M. & Hasselhorn, M. (2010). *Diagnostisches Inventar zu Rechenfertigkeiten im Grundschulalter* (DIRG). Göttingen: Hogrefe.

- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 144–171). Berlin: Springer.
- Kieffer, M. J., Lesaux, N. K., Rivera, M. & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79, 1168–1201. doi:10.3102/0034654309332490
- Lenhard, W. & Lenhard, A. (2017). *Diagnostik von Lesestörungen mit ELFE II bei Kindern mit Migrationshintergrund*. Dettelbach: Psychometrica.
- Lenhard, W., Lenhard, A. & Schneider, W. (2017). *Ein Leseverständnistest für Erst- bis Siebtklässler – Version II (ELFE II)*. Göttingen: Hogrefe.
- McElvany, N., Becker, M. & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 41, 121–131. doi:10.1026/0049-8637.41.3.121
- OECD (2018), *International Migration Outlook 2018*, OECD Publishing, Paris. http://dx.doi.org/10.1787/migr_outlook-2018-en
- Paleczek, L., Seifert, S., Obendrauf, T., Schwab, S. & Gasteiger-Klicpera, B. (2017). *Differenzierter Lesetest – Dekodieren (DiLe-D)*. Göttingen: Hogrefe.
- Petermann, F. & Daseking, M. (2012). *Züricher Lesetest – II (ZLZ-II)*. Bern: Huber.
- Rauch, D., Mang, J., Härtig, H. & Haag, N. (2016). Naturwissenschaftliche Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015 – Eine Studie zwischen Kontinuität und Innovation* (S. 317–347). Münster: Waxmann.
- Roick, T., Göllitz, D. & Hasselhorn, M. (2011). *Kettenrechner für dritte und vierte Klassen (KR 3-4)*. Göttingen: Hogrefe.
- Schulz, P. & Tracy, R. (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe.
- Schwabe, F. & Gebauer, M. M. (2013). (Test-)Fairness – eine Herausforderung an standardisierte Leistungsdiagnostik. In N. McElvany, M. M. Gebauer, W. Bos & H. G. Holtappels (Hrsg.), *Jahrbuch der Schulentwicklung: Daten, Beispiele und Perspektiven* (Bd. 17, S. 217–235). Weinheim: Beltz Juventa.
- Schwabe, F., McElvany, N., Trendtel, N., Gebauer, M. M. & Bos, W. (2014). Vertiefende Analysen zu migrationsbedingten Leistungsdifferenzen in Leseaufgaben: Differentielles Itemfunktionieren für Kinder mit und ohne Migrationshintergrund auf Basis der Internationalen Grundschul-Lese-Untersuchung (IGLU). *Zeitschrift für Pädagogische Psychologie*, 28, 93–103. doi:10.1024/1010-0652/a000126
- Seifert, S., Paleczek, L., Schwab, S. & Gasteiger-Klicpera, B. (2017). *Grazer Wortschatztest (GraWo)*. Göttingen: Hogrefe.
- Schwippert, K., Wendt, H. & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 191–207). Münster: Waxmann.
- Tarelli, I., Schwippert, K. & Stubbe, T. C. (2012). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, H. Wendt, O. Köller & H. Selter (Hrsg.), *TIMMS 2011 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 247–267). Münster: Waxmann.
- Wendt, H. & Schwippert, K. (2017). Lesekompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, et al. (Hrsg.), *IGLU 2016 Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 219–234). Münster: Waxmann.