Forschungsmonitor Schule
Rezensionen für die Praxis

Rebecca Schneider, Johannes Schult

Rezension zu

Fuchs, G. & Brunner, M. (2017). Wie gut können bildungsstandardbasierte Tests den schulischen Erfolg von Grundschulkindern vorhersagen? *Zeitschrift für Pädagogische Psychologie, 31*(1), 27–39.

Kommentierter Kurzbefund

Fuchs und Brunner untersuchen, wie gut bildungsstandardbasierte Tests spätere Schulleistungen (Tests, Noten) sowie die Übergangsempfehlung vorhersagen können. Diesen Forschungsfragen wird anhand längsschnittlicher Daten von 568 Grundschulkindern in Brandenburg nachgegangen.

Es zeigen sich durchgängig hohe Zusammenhänge der Tests in der 3. bzw. 4. Klasse mit den nachfolgenden Testleistungen, den Noten sowie den Gymnasialempfehlungen ($r \geq .47$). Auch wenn die Noten in der dritten Klasse, die Intelligenz und der familiäre Hintergrund berücksichtigt wurden, blieben die Testleistungen prognostisch valide (zwischen 1% und 14% zusätzliche Varianzaufklärung). Die Befunde unterstreichen die praktische Relevanz der Testergebnisse für die weitere Leistungsentwicklung und somit die Nützlichkeit von bildungsstandardbasierten Tests.

Da derartige Testaufgaben im Rahmen von Vergleichsarbeiten großflächig an Schulen eingesetzt werden, hat die vorliegende Fragestellung eine hohe Relevanz. Die Einschätzung im Manuskript, dass die Tests Lehrkräften gerade auch durch die kriteriale Verortung auf den Bildungsstandards eine wichtige Informationsgrundlage verschaffen, wird ausdrücklich geteilt. Denn so bieten sich direkte Anknüpfungspunkte für eine datengestützte Schul- und Unterrichtsentwicklung.

Hintergrund

Bundesweit werden im Rahmen der Vergleichsarbeiten (VERA) und des IQB-Ländervergleichs bildungsstandardbasierte Kompetenztests eingesetzt. In der Grundschule sind u. a. alle Lehrkräfte öffentlicher Schulen verpflichtet, jährlich Vergleichsarbeiten in Klassenstufe 3 in mindestens einem Fach durchzuführen. Die Ergebnisse der Tests sollen der Überprüfung der von der Kultusministerkonferenz gesetzten Bildungsstandards (Kultusministerkonferenz, 2006) sowie der schulischen Qualitätsentwicklung und -sicherung dienen.

Im Rahmen der Auswertung bildungsstandardbasierter Tests erhalten Lehrkräfte Rückmeldung zum aktuellen Leistungsstand ihrer Klasse und Hinweise für die weitere schulische Entwicklung ihrer Schülerinnen und Schüler. Erwünscht ist eine Auseinandersetzung der Lehrkräfte mit dieser Rückmeldung zur Weiterentwicklung ihres Unterrichts. Für Fuchs und Brunner (2017) stellt sich in diesem Zusammenhang jedoch insbesondere die Frage der prognostischen Validität bildungsstandardbasierter Tests für zentrale Schulerfolgskriterien in höheren Grundschulklassenstufen (Testleistungen, Schulnoten, Gymnasialempfehlung).

Die Autoren verweisen auf die wenigen relevanten längsschnittlichen Vorarbeiten, die substantielle Korrelationen von Kompetenztestleistungen in Mathematik und Deutsch in Klassenstufe 3 oder 4 mit

späteren Testleistungen (.63 ≤ r ≤ .67; Nachtigall, 2014) und Schulnoten (r = –.38; Hildebrandt und Watermann, 2015) im gleichen Fach fanden. Es zeigte sich zudem eine signifikante Prognosekraft bildungsstandardbasierter Kompetenztests für die Vorhersage der späteren Deutschnote über andere relevante Variablen hinaus (-.19 ≤ βadj ≤ –.17; ebenfalls Hildebrandt und Watermann, 2015). Längsschnittliche Befunde zur Vorhersage einer zukünftigen Übertritts- bzw. Gymnasialempfehlung lagen nicht vor.

Ziel der Studie von Fuchs und Brunner (2017) ist es, die in den Bildungsstandards implizierte Vorhersagekraft bildungsstandardbasierter Kompetenztests für die spätere schulische Entwicklung längsschnittlich zu untersuchen. Grundlage bilden die Daten von Schülerinnen und Schülern aus brandenburgischen Grundschulen; die Grundschule in Brandenburg umfasst die Klassenstufen 1 bis 6. Analysiert wurde die Prognosekraft bildungsstandardbasierter Kompetenztests in Deutsch und Mathematik in den Klassenstufen 3 und 4 für (1) Testleistungen und (2) Schulnoten für dieselben Fächer in höheren Klassenstufen sowie (3) die Gymnasialempfehlung in Klassenstufe 6. Die Analysen wurden jeweils ohne und mit Kontrolle weiterer wichtiger Variablen (vorherige Schulnoten, Intelligenz, familiärer Hintergrund) sowie getrennt für die Hauptfächer Mathematik und Deutsch durchgeführt.v

Design

Stichprobe: Die untersuchte Stichprobe umfasste n = 568 Grundschulkinder von 22 Grundschulen in Brandenburg (6-jährige Grundschule), die zu mindestens einem Messzeitpunkt an der Studie teilnahmen.

Durchführung: Die Datenerhebung erfolgte erstmalig im Schuljahr 2006/07 (2. Klassenstufe) und endete im Schuljahr 2010/11 (6. Klassenstufe). Daten der 2. Klassenstufe wurden nicht ausgewertet. Die Erhebung bildungsstandardbasierter Tests erfolgte durch geschulte Testleiterinnen und Testleiter.

Instrumente: Bildungsstandardbasierte Kompetenztests wurden in den Klassenstufen 3, 4, 5 und 6 in Mathematik und in den Klassenstufen 4, 5 und 6 in Deutsch erhoben. Grundlage der Tests in Mathematik bildeten Aufgaben aus dem IQB-Aufgabenpool zur Überprüfung der Bildungsstandards in der Primarstufe zu den Inhaltsbereichen Zahlen und Operationen, Muster und Strukturen, Größen und Messen sowie Raum und Form. Die Aufgaben in Deutsch sollten die Lesekompetenz der Schülerinnen und Schüler erfassen; Grundlage bildeten neben Aufgaben aus dem IQB-Pool insbesondere in Klassenstufe 5 und 6 zudem Aufgaben aus der ELEMENT-Studie (Lehmann & Lenkeit, 2008). Die Reliabilitäten der Tests waren mindestens zufriedenstellend (α ≥ .72).

Als schulische Leistungsindikatoren wurden die Halbjahresnoten in den Klassenstufen 4, 5 und 6 in den Fächern Mathematik und Deutsch erfasst. Die Empfehlung der Lehrkräfte für einen Bildungsgang ihrer Schülerinnen und Schüler wurde in Klassenstufe 6 erhoben (kodiert als Gymnasialempfehlung vs. keine Gymnasialempfehlung). Als Kontrollvariablen wurden die Intelligenz der Schülerinnen und Schüler (Untertest „Figurenanalogien“ des Kognitiven Fähigkeitstests, KFT 4–12+R, Heller & Perleth, 2000; Reliabilität α = .94) sowie deren familiärer Hintergrund (u. a. Beruf der Eltern, höchster schulischer und beruflicher Abschluss der Eltern, Buchbesitz, Einkommen) erfasst bzw. erfragt.

Analysen: Den eigentlichen Analysen zur Prognosekraft bildungsstandardbasierter Tests wurden umfangreiche Analysen der Testitems vorangestellt. U. a. wurden Items mit unzureichender Rasch-Homogenität und unbefriedigenden Trennschärfen von den Analysen ausgeschlossen. Fehlende Werte (15% – 19% bei den Kompetenztestwerten, 16% – 57% bei den übrigen Variablen) wurden mit Hilfe

eines multiplen Imputationsverfahrens ersetzt.

Alle Analysen bezüglich der zu untersuchenden Forschungsfragen erfolgten mit dem Statistikprogramm Mplus (Robust-Maximum-Likelihood-Schätzer, Berücksichtigung der hierarchischen Datenstruktur durch Spezifikation „type = complex“). Mit Hilfe von Regressionsmodellen wurden die Zusammenhänge zwischen den Ergebnissen bildungsstandardbasierter Kompetenztests in Mathematik und Deutsch (Klassenstufe 3 und 4) und späteren Testleistungen, Schulnoten sowie der Gymnasialempfehlung berechnet – ohne sowie unter Kontrolle weiterer relevanter Prädiktoren (Schulnoten, Intelligenz und familiärer Hintergrund).

Ergebnisse

(1) Prognosekraft für spätere Testleistungen: Die Testleistungen der Schülerinnen und Schüler der Klassenstufen 3 bzw. 4 können spätere Testleistungen in Mathematik substantiell vorhersagen ($.56 \leq \beta \leq .66$). Zudem erklärt die Testleistung in Mathematik 6% bis 11% der späteren Testleistungsvarianz in Mathematik über weitere Variablen hinaus auf (Schulnoten, Intelligenz und familiärer Hintergrund). Vergleichbare Ergebnisse finden sich für das Fach Deutsch (ohne Berücksichtigung von Kontrollvariablen: $.60 \leq \beta \leq .61$; 13% Varianzaufklärung über die Kontrollvariablen hinaus).

(2) Prognosekraft für spätere Schulnoten: Die Testleistungen der Schülerinnen und Schüler der Klassenstufen 3 bzw. 4 sagen spätere Schulnoten in Mathematik substantiell vorher ($-.57 \leq \beta \leq -.51$). Dabei geht ein Zugewinn von einer Kompetenzstufe im Test in Mathematik mit einer etwa um 1/3 besseren Schulnote in höheren Klassenstufen einher. Über vorherige Noten, Intelligenz und den familiären Hintergrund hinaus klärt die Testleitung 1% bis 3% der Schulnote in Mathematik in den Klassenstufen 5 und 6 auf. Für das Fach Deutsch zeigen sich vergleichbare Ergebnisse (ohne Berücksichtigung von Kontrollvariablen: $-.50 \leq \beta \leq -.47$; ein Zugewinn von einer Kompetenzstufe im Test geht einher mit einer Verbesserung um 0.21 bis 0.28 Notenpunkte; 1% bis 2% Varianzaufklärung über die Kontrollvariablen hinaus).

(3) Prognosekraft für die Gymnasialempfehlung: Die (punktbiserialen) Korrelationen zwischen den Kompetenztestwerten in Mathematik und Deutsch in Klassenstufe 4 und der Gymnasialempfehlung in Klassenstufe 6 liegen bei $r = .68$ und $r = .62$. Der Zugewinn von einer Kompetenzstufe im Test in Mathematik geht mit einer höheren Wahrscheinlichkeit von 14% bis 20% für eine Gymnasialempfehlung einher (adjustiert für die Lesekompetenz). Für den Deutschtest lag diese Wahrscheinlichkeit bei 11% bis 14% (adjustiert für die Mathematikkompetenz). Zudem zeigen Testleistungen der Schülerinnen und Schüler in Mathematik und Deutsch in der Klassenstufe 4 einen substantiellen prognostischen Mehrwert zur Vorhersage der Gymnasialempfehlung über Kontrollvariablen hinaus (8% bzw. 9%).

Diskussion und Einschätzung

Hintergrund

Aussagekräftige Testergebnisse sind eine Voraussetzung für datengestützte, kompetenzbasierte Unterrichtsentwicklung. Zu den Kompetenztestaufgaben, wie sie beispielsweise im Rahmen von VERA

eingesetzt werden, gibt es jedoch nur fragmentarische Informationen über die Testgüte. Fuchs und Brunner (2017) gehen deshalb der Frage nach, wie gut bildungsstandardbasierte Kompetenztests zukünftige Schulleistungen vorhersagen können. Die prognostische Validität ist ein wichtiges Testgütekriterium, zumal bildungsstandardbasierte Tests häufig vor dem Zeitpunkt geschrieben werden, für den die jeweiligen Bildungsstandards konzipiert wurden (z. B. VERA 3 in Hinblick auf die zu erreichenden Kompetenzen am Ende der 4. Klasse, VERA 8 in Hinblick auf den Mittleren Schulabschluss). Die Darstellung des Forschungsstandes im Artikel ist klar fokussiert auf die Vorhersagekraft der Testergebnisse und in diesem eher engen Rahmen vollständig. Die Einbettung der Fragestellung in den theoretischen ist als schlüssig zu bewerten.

Design

Das Studiendesign ist angemessen für die Forschungsfragen. Durch die Betrachtung von drei unterschiedlichen Schulerfolgsmerkmalen ergibt sich ein aussagekräftiges Gesamtbild. Im Artikel selbst liegt der Fokus auf 22 Schulen, bei denen die Durchführung und Auswertung der Arbeiten von geschultem, externem Personal durchgeführt wurde (zu den vermutlich geringen Kodiereffekten vgl. Spoden, Fleischer & Leutner, 2014). Zusätzlich werden darüber hinaus noch weitere Schulen untersucht, bei denen die Lehrkräfte die Tests durchführten (aber nicht auswerteten), wobei sich keine großen Unterschiede zeigten. Diese und weitere Limitationen werden im Artikel selbstkritisch reflektiert. Der ausschließliche Blick auf brandenburgische Schulen wird dabei erwähnt. Dieser hat aber auch den Vorteil, dass es sich um ein einheitliches Schulsystem handelt, bei dem die Schülerinnen und Schüler bis zum Ende der 6. Klasse gemeinsam die Grundschule besuchen. Dadurch können auch Beschulungsunterschiede gut kontrolliert werden. Der Einsatz von Kompetenztests mit einer Bearbeitungszeit von jeweils einer Schulstunde entspricht näherungsweise der an Grundschulen üblichen Testsituation. Der Fokus liegt dabei auf eng umrissenen Kompetenzen, so dass die Prognosekraft entsprechend geringer sein dürfte als bei umfangreicheren Testbatterien. Als einziger deutlicher Kritikpunkt muss in dieser Rezension erwähnt werden, dass Testleistung und Noten in Deutsch erst ab der 4. Klasse erhoben wurden, so dass hier nur ein kürzerer Prognosezeitraum betrachtet werden kann.

Ergebnisse

Es zeigt sich eine hohe Prognosekraft der Testleistung für alle drei Zielkriterien. Auch nach der Berücksichtigung von Kontrollvariablen bleibt eine inkrementelle Validität bestehen, die allerdings bei der Vorhersage der Noten tendenziell niedrig ausfällt. Zusätzliche Auswertungen weiterer Schulklassen, bei denen die Lehrkräfte die Tests durchführten, liefern ein sehr ähnliches Bild. Probleme wie Stichprobenausfälle wurden nach Möglichkeit z. B. durch statistische Verfahren kompensiert, wobei eher eine leichte Unterschätzung der Effektgrößen zu erwarten ist.

Obwohl die Testaufgaben nur spezifische Kompetenzen zum konkreten Testzeitpunkt erfassen, liefern sie den Lehrkräften offensichtlich deutliche Hinweise, wohin sich ihre Schülerinnen und Schüler entwickeln. Der belegte prognostische Wert bildungsstandardbasierter Tests impliziert die Notwendigkeit, die pädagogische Praxis daraufhin zu untersuchen, inwiefern die Vergleichsarbeiten tatsächlich – wie politisch intendiert – zu diagnostischen Zwecken und der Unterrichtsentwicklung genutzt werden.

Dass die Testleistung nur geringe zusätzliche Notenvarianz über die Kontrollvariablen (frühere Noten, Intelligenz, familiärer Hintergrund) hinaus erklärt, spricht dafür, dass die Testleistung in Vergleichsarbeiten nicht in die Notengebung einfließen sollte. Auch die direkte Verwendung von Testleistungen für Gymnasialempfehlungen ist kritisch zu sehen. Vielmehr sollte im Anschluss an

Vergleichsarbeiten anhand der Ergebnisse kritisch reflektiert werden, wie Schülerinnen und Schüler in der verbleibenden Zeit bis zur Übergangsentscheidung angemessen unterrichtet werden.

Da sich in der vorliegenden Studie für die Grundschule ein ähnliches Befundmuster gezeigt hat wie für die Sekundarstufe I (Graf, Harych, Wendt, Emmrich & Brunner, 2016), kann mangelnde prognostische Validität nicht als Kritikpunkt an bildungsstandardbasierten Testverfahren wie VERA 3 und VERA 8 aufgeführt werden. Vielmehr dürfte es auf Lehrkraftseite noch Informationsbedarf hinsichtlich der Testgüte und der Aufgabeneigenschaften geben, zumal sich Grundschullehrkräfte hinsichtlich ihrer aufgabenbezogenen Diagnosegenauigkeit teils deutlich unterscheiden (Schult & Lindner, 2018).

Reflexionsfragen für die Praxis

Nachfolgende Reflexionsfragen sind ein Angebot, die Befunde der rezensierten Studie auf das eigene Handeln als Lehrkraft oder Schulleitungsmitglied zu beziehen und zu überlegen, inwiefern sich Anregungen für die eigene Handlungspraxis ergeben. Die Befunde der rezensierten Studien sind nicht immer generalisierbar, was z. B. in einer begrenzten Stichprobe begründet ist. Aber auch in diesen Fällen können die Ergebnisse interessante Hinweise liefern, um über die eigene pädagogische und schulentwicklerische Praxis zu reflektieren.

Reflexionsfragen einer Lehrkraft:

- Wie haben sich die Schülerinnen und Schüler in der 4. Klasse entwickelt unter Berücksichtigung der VERA 3-Ergebnisse aus dem Vorjahr?
- Welche Schülerinnen und Schüler haben in VERA erwartungswidrige Leistungen erzielt? Können die Motivation und die Klassensituation am Testtag eine Rolle gespielt haben? Oder könnten unerwartet schwache Leistungen Förderbedarf signalisieren?
- Welche Kompetenzen sollen mit den Aufgaben in den Vergleichsarbeiten und im Bildungstrend erfasst werden? (Hinweise stehen auch in den Ergebnisdokumenten; ausführliche Beschreibungen der Kompetenzstufenmodelle finden sich unter https://www.iqb.hu-berlin.de/bista/ksm).

Reflexionsfragen einer Schulleitung:

- Welche Möglichkeiten haben die Lehrkräfte, um sich über die Ergebnisse von Vergleichsarbeiten – fachintern und fachübergreifend – auszutauschen?
- Gibt es verbindliche Abmachungen, ob/wie Lehrkräfte im Anschluss an bildungsstandardbasierte Kompetenztests Entwicklungsaktivitäten absprechen und durchführen?
- Welche Fortbildungsmöglichkeiten zum Umgang mit den Ergebnissen von Vergleichsarbeiten gibt es?
- Welche Kompetenzen sollen mit den Aufgaben in den Vergleichsarbeiten und im Bildungstrend erfasst werden? (Hinweise stehen auch in den Ergebnisdokumenten; ausführliche Beschreibungen der Kompetenzstufenmodelle finden sich unter https://www.iqb.hu-berlin.de/bista/ksm).

Literatur

Graf, T., Harych, P., Wendt, W., Emmrich, R. & Brunner, M. (2016). Wie gut können VERA-8-Testergebnisse den schulischen Erfolg am Ende der Sekundarstufe I vorhersagen? *Zeitschrift für Pädagogische Psychologie 30*, 201–204.

Heller, K. A. & Perleth, C. (2000). *KFT 4–12+R. Kognitiver Fähigkeitstest für 4. bis 12. Klassen*, Revision (3. Aufl.). Göttingen: Beltz Test.

Hildebrandt, J. & Watermann, R. (2015). *Prognostische Validität von curricular gemessenen Testleistungen am Ende der Grundschulzeit*. Vortrag auf der 3. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Bochum.

Kultusministerkonferenz (Hrsg.) (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Bonn.

Lehmann, R. & Lenkeit, J. (2008). *ELEMENT. Erhebung zum Lese- und Mathematikverständnis. Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin. Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien*. Berlin: Humboldt Universität zu Berlin.

Nachtigall, C. (2014). *Landesbericht. Thüringer Kompetenztests 2014*. Jena: Friedrich-Schiller-Universität, Institut für Psychologie.

Spoden, C., Fleischer, J. & Leutner, D. (2014). Niedrige Testmodellpassung als Resultat mangelnder Auswertungsobjektivität bei der Kodierung landesweiter Vergleichsarbeiten durch Lehrkräfte. *Journal für Mathematik-Didaktik 35*, 79–99.

Schult, J. & Lindner, M. A. (2018). Diagnosegenauigkeit von Deutschlehrkräften in der Grundschule: Eine Frage des Antwortformats? *Zeitschrift für Pädagogische Psychologie 32*, 75–87.

Rezensent/-in

Rebecca Schneider, M.Sc., wissenschaftliche Mitarbeiterin in der Fachrichtung Bildungswissenschaften an der Universität des Saarlandes, Saarbrücken. Arbeitsschwerpunkte: Schulleistungsdiagnostik, Motivation bei Grundschulkindern, Leistungsängstlichkeit.

Johannes Schult, Dr. Dipl.-Psych., PsychR am Landesinstitut für Schulentwicklung, Stuttgart. Arbeitsschwerpunkte: Lernstandserhebungen, datengestützte Schulentwicklung, Leistungsdiagnose, Psychometrie

Zitiervorschlag

Schneider, R., Schult, J. (2018). Rezension zu Fuchs, G. & Brunner, M. (2017). Wie gut können bildungsstandardbasierte Tests den schulischen Erfolg von Grundschulkindern vorhersagen? Zeitschrift für Pädagogische Psychologie, 31(1), 27–39. *Forschungsmonitor Schule, 73*. Abgerufen von https://www-

.forschungsmonitor-schule.de/print.php?id=49

Urheberrecht