

Psychologische Rundschau

2/14

www.hogrefe.de/zeitschriften/pru

Schwerpunktthema
Ein national gefördertes Onlinelabor
als Infrastruktur für die
psychologische Forschung

Herausgegeben im Namen
der Deutschen Gesellschaft für Psychologie

HOGREFE



Originalia	Ulrich, F. & Petermann, F.: Zur Bedeutung von sozialer Unterstützung während der Schwangerschaft <i>Impact of social support during pregnancy</i>	57
	Bruder, M., Göritz, A. S., Reips, U.-D. & Gebhard, R. K.: Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung <i>A nationally funded online laboratory as infrastructure for psychological research</i>	75

Kommentare	Kommentare zu Bruder, M., Göritz, A. S., Reips, U.-D. & Gebhard, R. K. (2014). Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung	
	Abele-Brehm, A.: Stellungnahme zum Beitrag "Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung"	86
	Funke, J.: Kommentar zum Beitrag "Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung"	87
	Hesse, F. W.: Kommentar zum Manuskript "Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung"	88
	Krampen, G.: Konzept für eine Onlinelabor-Infrastruktur der Psychologie. Inhaltlich und fachpolitisch nützlich und sinnvoll	89
	Strack, F.: Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung	91
	Bruder, M., Göritz, A. S., Reips, U.-D. & Gebhard, R. K.: Aufbruch zum Onlinelabor Begrüßung der Kommentare	92

Psychologie für die Gesellschaft	Rothermund, K.: Die Gestaltung des Alters. Ein Plädoyer für mehr Psychologie	95
---	---	----

Nachrichten	Deutsche Gesellschaft für Psychologie e.V.	101
	Berufsverband Deutscher Psychologinnen und Psychologen e.V.	109
	Nachrichten aus Instituten und Institutionen	110
	Veranstaltungen und Ankündigungen	115
	TBS-TK Rezension Vock, M. & Zimmermann, C.: KLASSE 4. Kombiniertes Leistungsinventar zur allgemeinen Schulleistung und für Schullaufbahneempfehlungen in der vierten Klasse	118

TBS-TK Rezension

KLASSE 4. Kombiniertes Leistungsinventar zur allgemeinen Schulleistung und für Schullaufbahneempfehlungen in der vierten Klasse

M. Vock, Profilbereich Bildungswissenschaften, Universität Potsdam

C. Zimmermann, Regionale Schulberatungsstelle, Stadt Leverkusen

Allgemeine Informationen über den Test, Beschreibung des Tests und seiner diagnostischen Zielsetzung

Mit Klasse 4 lässt sich der allgemeine Leistungsstand von Schülern zur Schuljahresmitte der vierten Grundschulklasse in Form eines besonders ökonomischen Gruppentests (Durchführungszeit: 45 min) erfassen. Als Leistungsbereiche werden das Sachrechnen, Geometrie, Lesen und Schreiben überprüft. Zusätzlich wird das akademische

Selbstkonzept in den Bereichen Deutsch und Mathematik ermittelt. Die Leistungsuntertests wurden aus bewährten Verfahren übernommen: Sachrechnen/Geometrie wird mit Untertests aus dem DEMAT 4 (Göltz, Roick & Hasselhorn, 2006), das Textverstehen mit einem Teil aus dem ELFE 1-6 (Lenhard & Schneider, 2006) erfasst, das Diktat ist (leicht verändert) dem DERET 3-4+ (Stock & Schneider, 2008) entnommen. Ziel ist es, Grundschullehrkräfte bei den individuellen Grundschulübergangsempfehlungen zu unterstützen. Außerdem soll KLASSE 4 als förderdiagnostisches

Instrument ermöglichen, Über- und Unterforderungssituationen bei einzelnen Schülern zu bestimmen.

Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion

Die Untertests wurden ausgewählt, um das pragmatische Ziel, „einen ökonomischen Überblick über die Leistungen in zentralen schulischen Fächern zu geben“ (S. 20) zu erreichen und eine Lücke bei den kombinierten Schulleistungstests zu schließen. Eine theoretische Begründung für die Auswahl der Skalen erfolgt im Manual nicht. Bei der Testkonstruktion wurden aus den Verfahren DEMAT 4, ELFE 1-6 und DERET 3-4+ (s.o.) auf Basis einer Vorstudie ($N = 219$) diejenigen Untertests ausgewählt, welche zur Vorhersage der Schullaufbahmpfehlungen der Lehrkräfte am meisten beitrugen. Aus diesem Grund wurde zwar ein Geometrie-, nicht aber ein Arithmetik-Untertest aufgenommen. Es werden somit solche Leistungen erfasst, die Lehrkräften bei ihrer Entscheidungsfindung für eine Schulempfehlung besonders wichtig sind. Die Lehrkräfteentscheidungen sind daher hier der entscheidende Maßstab. Bei den Selbstkonzeptskalen orientierten sich die Autoren an den in der PISA-Studie verwendeten Skalen.

Objektivität

Klasse 4 ist durch detaillierte Angaben zum Vorgehen, feste Bearbeitungszeiten und wörtlich ausformulierte Instruktionen als standardisiert zu bezeichnen. Die Schüler/innen blättern im Testheft gleichzeitig um, so dass alle am gleichen Untertest arbeiten. Das Manual enthält Hinweise zum Umgang mit Schülerfragen. Die Auswertungsobjektivität ist durch eindeutige Angaben zur Ermittlung der Rohwerte und Bestimmung der Normwerte gegeben; im Untertest Schreiben weichen die Auswertungsregeln jedoch vom DERET 3-4+ ab. T-Wertbänder und Informationen zur Berechnung des Konfidenzintervalls erleichtern die Interpretation; für das Gesamtprofil kann abgelesen werden, ab wann T-Wert-Diskrepanzen zwischen Untertests signifikant sind. Anhand gewichteter T-Werte kann die Wahrscheinlichkeit für eine Empfehlung für eine bestimmte Schulform ermittelt werden. Da in der Praxis Lehrkräfte den Test einsetzen, wären Angaben zu erforderlichen Kenntnissen in Testdiagnostik hilfreich gewesen.

Normierung (Eichung)

Die Normierung mit $N = 396$ Kindern aus 5 Bundesländern fand zwischen Januar und März (vermutlich 2010, Angabe fehlt) statt. Dieser Zeitraum erscheint relativ spät, da in einigen Bundesländern die Empfehlung für die weiterführende Schule mit dem Halbjahreszeugnis ausgesprochen wird. Die Normwerte sollen für den Zeitraum zwei Monate vor bis zwei Monate nach dem Zwischenzeugnis gültig sein. Es fehlen Angaben zum Alter und zu Klassenwiederholungen; der Anteil von Kindern mit Migrationshintergrund (11%) erscheint relativ gering. Ob die Normen auch in anderen Bundesländern gelten, insbesondere in Ländern mit 6-jähriger Grundschule, wird nicht thematisiert. Zudem sind die Quoten von Kindern, die auf eine bestimmte Schulform übergehen, regional unterschiedlich, die Klassifikation erfolgte über verschiedene Länder hinweg, worauf die Verfasser auch hinweisen (S.

22). Kinder mit einer nicht erkannten Lese-Rechtschreib- oder Rechenschwäche könnten in KLASSE 4 benachteiligt werden.

Zuverlässigkeit (Reliabilität, Messgenauigkeit)

Die internen Konsistenzen der Skalen liegen zwischen Cronbachs $\alpha = .72$ (Sachrechnen) und $.93$ (Schreiben) und sind damit akzeptabel; lediglich die Skala Geometrie hat mit 7 Items keine befriedigende interne Konsistenz ($.57$). Es ist fraglich, ob die beiden Untertests der Skala, die einerseits mentale Rotation („Lagebeziehungen“) und andererseits Spiegelungen („Spiegelbeziehungen“) erfordern, nicht sinnvoller separat betrachtet werden sollten. Die übergeordneten Skalen Deutsch ($.88$) und Mathematik ($.75$) sowie der Gesamttest ($.80$) sind hinreichend konsistent, wobei der Wert für Deutsch anhand der Angaben aus den Originalhandbüchern (ELFE / DERET) für die beiden Untertests Lesen und Schreiben durch Mittelung nach Fisher-Z-Transformation geschätzt wurde. Es werden Retestreliabilitäten über einen Zeitraum von 4 Monaten berichtet (2 Klassen, $N = 35$); die Leistungen und das Selbstkonzept sind insgesamt stabil (Gesamttest: $r_{tt} = .83$), lediglich das Sachrechnen ist eher instabil ($r_{tt} = .60$).

Gültigkeit (Validität)

Hinsichtlich der konvergenten Validität verweisen die Testautoren auf die Manuale der Originaltests, aus denen die Untertests entnommen sind. Es werden Korrelationen zwischen den Skalenwerten (Deutsch, Mathematik), dem Gesamtergebnis und den Lehrerratings hinsichtlich Schulleistung und Lernmotivation ($r = .42$ bis $.74$) sowie zu den Schulnoten in Deutsch und Mathematik ($r = -.47$ bis $-.64$) berichtet. Die Korrelationen für die Kinder mit Migrationshintergrund ($n = 44$) fallen höher aus, was die Autoren als Hinweis auf ein noch präziseres Lehrerurteil werten. Es fehlen Angaben zur Konstrukt- und zur Lehrplanvalidität. Kritisch ist die Berechnung einer Empfehlung für eine Schulform zu sehen. Die Autoren stellen zwar einige Warnhinweise voran und machen klar, dass die prognostische Validität der Aussagen zur Wahl einer weiterführenden Schulform nicht geprüft ist (S. 17). Der Nutzen dieser Berechnung ist damit eher gering, eine Fehlinterpretation durch Anwender jedoch nicht auszuschließen.

Weitere Gütekriterien (Störanfälligkeit, Unverfälschbarkeit und Skalierung)

Der Test kann mit der ganzen Klasse gemeinsam innerhalb einer Schulstunde durchgeführt werden. Dabei sind Störungen durch die Mitschüler/innen nicht auszuschließen, auch das Abschreiben wird nicht ganz zu verhindern sein, zumal keine Paralleltestversion vorhanden ist. Im Manual werden jedoch Hinweise gegeben, wie die Gefahr des Abschreibens verringert werden kann. Desweiteren können die Testergebnisse durch die Gestaltung des Testhefts von Kindern verfälscht werden: Da sich Beispiele und Testaufgaben wiederholt auf einer Doppelseite befinden, kann nicht ausgeschlossen werden, dass einzelne Kinder vorzeitig mit der Aufgabenbearbeitung beginnen. Eine empirische Prüfung der Skalenbildung durch IRT Modelle erfolgte nicht. Die Zusammenfassung von Untertests zu

Skalen wird inhaltlich über die Zuordnung zu Fächern begründet, nicht aber empirisch geprüft.

Abschlussbewertung/Empfehlung

KLASSE 4 ist ein ökonomisches Verfahren, das schnell einen Überblick über Aspekte der schulischen Leistung in Mathematik und Deutsch gibt. Hiermit schließt der Test eine Lücke, denn alternative Verfahren sind entweder veraltet oder aufwändiger.

Die diagnostische Zielstellung bleibt jedoch etwas unbefriedigend: Über die reine Deskription hinaus wird eine Empfehlung für die Wahl der weiterführenden Schule angestrebt, es fehlt bisher jedoch eine Validierung im Hinblick auf den Schulerfolg in den verschiedenen Schulformen (eine solche Studie ist in Vorbereitung). Kritisch erscheint die Interpretation einer Diskrepanz zwischen einem niedrigen Selbstkonzept und einer höheren Leistung als Hinweis auf ein ungünstiges Attributionsmuster, zu hohe Erwartungen der Eltern oder Konflikte mit Lehrkräften. Hier bleibt unberücksichtigt, dass für die Ausbildung des akademischen Selbstkonzepts Vergleichsprozesse innerhalb der Klasse entscheidend sind (insbes. Big-Fish-Little-Pond-Effekt). Das Manual ist relativ kurz, wünschenswert wären ausführlichere Erläuterungen zu den Konstrukten (z.B. Selbstkonzept) und den nötigen Qualifikationen für Testleiter (i.d.R. Lehrkräfte). Durch die fehlenden Aussagen zur Konstrukt- und Lehrplanvalidität bleibt offen, wie gut die Skalen relevante schulische Leistungen in der Grundschule in verschiedenen Bundesländern abbilden und welche Bezüge etwa zu Bildungsstandards bestehen. Die Skalenauswahl erfolgte rein auf Basis maximaler Vorhersagegüte für das Kriterium „Lehrkräftempfehlung“. Die Empfehlung, Teilergebnisse für eine Auswertung des Ursprungstest zu verwenden, muss kritisch betrachtet werden, da sich Durchführungszeiten, Instruktionen, Aufgabenblätter, Auswertungsregeln und Normierungszeiträume z. T. unterscheiden.

Insgesamt ist KLASSE 4 ein empfehlenswertes Testverfahren, das objektive Daten über den Leistungsstand einer Schulklasse liefert; die Ergebnisse sollten jedoch insbes. bei Übergangsentscheidungen zurückhaltend interpretiert werden.

Klasse 4	Die TBS-TK-Anforderungen sind erfüllt	
	voll	teilweise nicht
Allgemeine Informationen, Beschreibung und diagnostische Zielsetzung	x	
Objektivität		x
Zuverlässigkeit		x
Validität		x

Diese Testrezension wurde im Auftrag des Testkuratoriums der Föderation deutscher Psychologinnenvereinigungen (DGPs und BDP) gemäß den TBS-TK-Richtlinien (Testkuratorium, 2009, 2010) erstellt.

Literatur

- Gölitz, D., Roick, T. & Hasselhorn, M. (2006). *Deutscher Mathematiktest für vierte Klassen (DEMAT 4)*. Göttingen: Hogrefe.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.
- Stock, C. & Schneider, W. (2008). *Deutscher Rechtschreibtest für das dritte und vierte Schuljahr (DERET 3-4+)*. Göttingen: Hogrefe.
- Testkuratorium (2009). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 09. September 2009. *Report Psychologie*, 34, 470–478.
- Testkuratorium (2010). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 09. September 2009. *Psychologische Rundschau*, 61, 52–56.

Testinformationen

Lenhard, W., Hasselhorn, M. & Schneider, W. (2011). *KLASSE 4. Kombiniertes Leistungsinventar zur allgemeinen Schulleistung und für Schullaufbahneempfehlungen in der vierten Klasse*. Göttingen: Hogrefe.

Bezugsquelle: Testzentrale Göttingen,
Herbert-Quandt-Straße 4,
37079 Göttingen.

E-Mail: testzentrale@hogrefe.de,
<http://www.testzentrale.de>

Preise: Test komplett 98,00 €. Manual 49,00 €. 25 Testhefte 48,75 €. Schablonensatz 29,00 €. 50 Auswertungsbogen 30,00 €.

Bitte zitieren Sie diesen Artikel wie folgt: Vock, M. & Zimmermann, C. (2014). TBS-TK Rezension: „KLASSE 4. Kombiniertes Leistungsinventar zur allgemeinen Schulleistung und für Schullaufbahneempfehlungen in der vierten Klasse“. *Psychologische Rundschau*, 65, 118–120.

DOI: 10.1026/0033-3042/a000201