

Qualitätssicherung und -optimierung im Aufbruch: Die ersten Testrezensionen nach dem neuen TBS-TK-System

Helfried Moosbrugger, Gerhard Stemmler und Martin Kersting

In diesem Heft werden erstmalig Testrezensionen veröffentlicht, die nach dem Testbeurteilungssystem des Testkuratoriums (TBS-TK) der Föderation Deutscher Psychologinnenvereinigungen erstellt wurden. Unter den zahlreichen Methoden, die von psychologischen Diagnostikern geschaffen und genutzt werden, kommt dem »Test« im weitesten Sinne eine herausragende Rolle zu. Die Entwicklung und Anwendung von Tests gilt nach Swenson (1997) als die höchste technische Leistung der Psychologie und stellt nach Ansicht der American Educational Research Association (AERA), der American Psychological Association (APA) sowie des National Council on Measurement in Education (NCME) einen der wichtigsten Beiträge der Verhaltenswissenschaften zur Gesellschaft dar (AERA, 1999, S. 1). Das TBS-TK leistet einen Beitrag zur Qualitätssicherung und -optimierung dieser Verfahren. Es stellt eindeutige Anforderungen an den Informationsgehalt von Verfahrenshinweisen (Testhandbücher/-manuale) und formuliert in Form von Beurteilungsrichtlinien Qualitätsstandards für alle Instrumente, die zur Erfassung menschlichen Erlebens und Verhaltens eingesetzt werden. Damit trägt das TBS-TK maßgeblich zur Verbreitung von zuverlässigen Informationen sowie nachvollziehbaren Expertenurteilen über derartige Instrumente bei. Zugleich dient das System als Leitfaden für die fachgerechte Entwicklung und Weiterentwicklung von Instrumenten.

Das TBS-TK-System wurde 2006 erstmals publiziert (Testkuratorium, 2006). Der vorliegende Artikel begleitet die Veröffentlichung der ersten TBS-TK-Testrezension, indem er kurz erläutert, warum ein neues System nach Ansicht des Testkuratoriums notwendig wurde. Anschließend wird der dreistufige Ablauf der Testbeurteilung skizziert, bevor der Artikel mit einem Ausblick auf zukünftige Entwicklungen schließt.

Zur Notwendigkeit eines neuen Systems

Kersting (2006) zieht ein kritisches Resümee bezüglich der bisherigen deutschen Praxis der Information über Tests sowie Beurteilungen von Tests. Demzufolge gibt es zwar eine Vielzahl an Informationen und Beurteilungen in Büchern, Fachzeitschriften und Datenbanken, diese Bewertungen sind aber sowohl quantitativ als auch qualitativ unbefriedigend:

Sie sind häufig unsystematisch aufbereitet, in ihrem Urteil beliebig, überzogen kritisch oder aber auch uneindeutig, sie sind schließlich verstreut publiziert, sodass ein zentraler Zugriff erschwert wird.

Außerdem repräsentieren die vorhandenen Testrezensionen nicht die vorhandenen Tests: Trotz der absolut

gesehen beachtlichen Anzahl an Testrezensionen liegen zu einigen in der Praxis häufig genutzten Tests keine Informationen und Rezensionen vor, was u. a. darauf zurückzuführen ist, dass es bislang an einer steuernden Instanz mangelte.

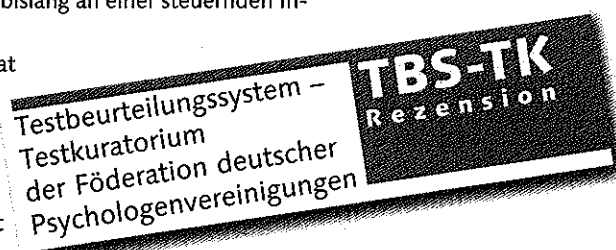
Das Testkuratorium hat sich im Zuge der Entwicklung eines neuen Systems zur Information über Tests und deren Beurteilung zunächst ausführlich mit den in an-

deren Ländern etablierten Testinformations- und Testbeurteilungssystemen (z.B. dem niederländischen COTAN-System, Evers, 2001b) sowie vor allem mit der Initiative der European Federation of Psychologists Associations (EFPA) zur Entwicklung eines »common set of European criteria for test reviews« auseinandergesetzt und zahlreiche bewährte Ideen dieser Systeme übernommen. Gegen eine vollständige Übernahme eines etablierten Systems wie COTAN oder EFPA sprach vor allem die dort vorgesehene Bewertung der numerischen Ausprägung von Kennwerten nach vorgegebenen Regeln. So verlangt das COTAN-System beispielsweise, dass eine Reliabilität kleiner als 0,80 für Tests³ als »unzureichend« zu bewerten sei usw. Derartig starre Beurteilungsvorschriften übersehen, dass die Ausprägung der Kennwerte nicht nur von der Qualität des Tests, sondern auch von Merkmalen der jeweils herangezogenen Untersuchungsgruppe abhängt (siehe Kersting, 2006).

Grundprinzipien des dreistufigen TBS-TK-Systems

Das TBS-TK begnügt sich nicht mit der schematischen Klassifikation von Koeffizienten, sondern setzt für die Testbewertung auf die Urteilskraft ausgewiesener und einschlägig qualifizierter Experten, die über ein tiefes Verständnis der jeweiligen Konstrukte und Operationalisierungen, der diagnostischen Zielsetzungen und Rahmenbedingungen sowie über ein fundiertes psychometrisches Wissen verfügen. Die Experten werden vom Testkuratorium angesprochen und gebeten, eine Rezension zu einem vom Testkuratorium ausgewählten Verfahren zu erstellen. Die Subjektivität der Urteile wird durch die Vorgabe einer klaren Beurteilungsrichtlinie eingeschränkt. Darüber hinaus wird jede Testbeurteilung von zwei Experten vorgenommen, die zunächst unabhängig voneinander arbeiten.

Konkret sieht das TBS-TK drei Schritte vor (für eine ausführliche Darstellung siehe Testkuratorium, 2006, 2007).



1 Hier wie im Folgenden ist die weibliche Form stets mit gemeint.

2 »Obwohl ein psychologischer Test im engeren Sinne nur eine besondere Untergruppe (...) psychologisch-diagnostischer Verfahren darstellt, soll die Bezeichnung »Test« im vorliegenden Zusammenhang als Oberbegriff gelten: Es sind also neben Intelligenz- und allgemeinen Leistungstests insbesondere Persönlichkeitsfragebogen, Objektive Persönlichkeitstests sowie Projektive Verfahren, aber auch standardisierte Interviews sowie Erhebungsverfahren zur Arbeitsplatzanalyse gemeint.« (Testkuratorium, 2006, S. 493)

3 Gemeint sind Tests, die für wichtige Entscheidungen auf individueller Ebene eingesetzt werden.

In einem ersten Schritt prüfen die Rezensenten, ob und in welchem Ausmaß die Anforderungen der DIN 33430 (DIN, 2002) an Verfahrenshinweise (Testmanuale) erfüllt sind. Die Operationalisierung dieser Anforderungen erfolgt mit der »Checkliste 1« der Publikation »DIN Screen« von Kersting (2008). Diese Checkliste dient als »Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen« und steht auch als Online-Version zur Verfügung (www.kersting-internet.de/DIN-Screen.html). Die Checkliste sollte bereits durch die Testautoren oder den Verlag ausgefüllt werden, sodass die Rezensenten sich auf eine Kontrolle der Angaben konzentrieren können. Aktuelle Testverfahren wie z.B. der WIT-2 (Kersting, Althoff und Jäger, 2008) veröffentlichen bereits im Testhandbuch eine Übersicht, an welcher Stelle des Handbuchs die 140 geforderten Informationen zu finden sind. Auf Basis der vorliegenden Informationen stellen die Rezensenten fest, ob der Test »prüffähig« ist. Ein Test, der in diesem Sinne nicht prüffähig ist, weil wesentliche Angaben gemäß DIN 33430 fehlen, erhält ohne weitere Prüfung die Beurteilung: »Der Test erfüllt die durch DIN 33430 festgelegten Anforderungen bezüglich Information und Dokumentation nicht.« Während sich die DIN 33430 insgesamt auf die Eignungsbeurteilung bezieht, wird die hier in Frage stehende Teilmenge der DIN 33430 im Umfang von 140 ausgewählten Informationsanforderungen auf Tests aus allen Anwendungsbereichen angewendet.

Im zweiten Schritt erfolgt eine Testkategorisierung, und es werden formale Merkmale des Tests für Datenbanken angegeben. Hierzu werden das ZPID-System (www.zpid.de) und Teile des EPPA-Systems benutzt.

Im dritten und letzten Schritt erfolgt die eigentliche Beurteilung des Tests durch die Rezensenten anhand der vorgegebenen Besprechungs- und Beurteilungskategorien. Die abgedruckten Testrezensionen umfassen nur diesen dritten Schritt des umfassenden Gesamtbeurteilungsprozesses, auch die Arbeiten des ersten und zweiten Schritts liefern wertvolle Ergebnisse, die hier nicht dokumentiert sind.

Die drei genannten Schritte werden von den beiden Rezensenten unabhängig voneinander vorgenommen. Erst nachdem die beiden Rezensionen eingegangen sind, hebt das Testkuratorium die gegenseitige Anonymität der beiden Rezensenten auf und bittet die beiden Experten um die Erstellung einer gemeinsamen Fassung. Diese Fassung wird den Testautoren zur Kenntnis gegeben, die somit Gelegenheit zur Stellungnahme erhalten. In dem gesamten Prozess behält sich das Testkuratorium eine aktive Rolle vor.

Neben der Veröffentlichung in »report psychologie« sowie in der »Psychologischen Rundschau« werden die Rezensionen auch in einem eigens eingerichteten, online verfügbaren Datenbanksegment von PSYNDEX veröffentlicht (www.zpid.de/Testkuratorium). Andere Medien können die Rezensionen auf Anfrage als Nachdruck veröffentlichen. Der Nachdruck kann vollständig, aber auch gekürzt erfolgen, sofern durch die Kürzung kein irreführender Eindruck vom Gesamtbild entsteht.

Ausblick

Das TBS-TK-System ist ein Meilenstein für die Qualitätssicherung und -optimierung von Tests. Die hier publizierten ersten TBS-TK-Rezensionen zum BDI-II, IST-2000-R und OLMT demonstrierten, dass die anspruchsvollen Ziele des Systems erfüllbar sind und die Vorgaben zu sehr nützlichen Resultaten führen. Selbstverständlich wertet das Testkuratorium die ersten Erfahrungen auch mit dem Ziel aus, das System weiter zu verbessern.

Weitere TBS-TK-Rezensionen stehen kurz vor der Vollendung und können voraussichtlich in Kürze unter www.zpid.de/Testkuratorium sowie in »report psychologie« publiziert werden. Andere Rezensionen sind in Vorbereitung; bei der Auswahl wird darauf geachtet, dass Verfahren aus allen Anwendungsgebieten sowie Konstruktbereichen Berücksichtigung finden. Sobald eine größere Anzahl von Testrezensionen vorliegt, wird das Testkuratorium eine testübergreifende Analyse vornehmen, um z.B. zu ermitteln, welche grundsätzlichen Stärken und Schwächen Testverfahren aufweisen (vgl. Evers, 2001a). Darüber hinaus sind für die Zukunft Sammelpublikationen der TBS-TK-Testrezensionen geplant. Von zentraler Bedeutung ist es, aktive Unterstützung für das neue System zu mobilisieren. Hierzu zählen beispielsweise das Engagement der TBS-TK-Rezensenten, die Bereitschaft der Verlage und Autoren, die Vorarbeiten (Schritte 1 und 2) vorzunehmen, sowie die Entscheidung einschlägiger Organe, die Beurteilungskriterien des TBS-TK zu übernehmen (wie es beispielsweise die Fachzeitschrift »Diagnostica« angekündigt hat).

Zu hoffen ist vor allem, dass sich die Testautoren proaktiv an dem neuen Qualitätsstandard orientieren und somit bessere Produkte herstellen. Die Qualitätssicherung eröffnet die Chance, einen Gegenpol zur Beliebigkeit zu prägen, seriöse Anbieter und informierte Anwender zu stärken und die testbasierte Diagnostik und Intervention noch leistungsstärker werden zu lassen.

Literatur

American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME). (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

DIN (2002). DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. Berlin: Beuth.

Evers, A. (2001a). Improving Test Quality in the Netherlands: Results of 18 years of Test Ratings. *International Journal of Testing*, 1, 137-153.

Evers, A. (2001b). The Revised Dutch Rating System for Test Quality. *International Journal of Testing*, 1, 155-182.

Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243-253.

Kersting, M. (2008). DIN Screen, Version 2. Leitfaden zur Kontrolle und Optimierung der Qualität von Verfahren und deren Einsatz bei beruflichen Eignungsbeurteilungen. In: M. Kersting, Qualität in der Diagnostik und Personalauswahl – der DIN Ansatz. Göttingen: Hogrefe.

Kersting, M., Althoff, K. & Jäger, A. O. (2008). WIT-2. Der Wilde-Intelligenztest. Verfahrenshinweise. Göttingen: Hogrefe.

Swenson, L. C. (1997). Psychology and law for the helping professions (2nd ed.). Pacific Grove, CA.: Brooks/Cole.

Testkuratorium (2006). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. *Report Psychologie*, 31, 492-499, sowie *Psychologische Rundschau*, 58, 25-30.

Testkuratorium (2007). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. *Psychologische Rundschau*, 58, 25-30.