Arbeitsbezogene Belastbarkeits- und Gewissenhaftigkeitsskalen (ABGS)

DOROTHEA KISSEL, FOM Hochschule Frankfurt a. M.
CHRISTIAN WARNEKE, Hochschule für Angewandte Wissenschaften Hamburg

Beschreibung des Tests und seiner diagnostischen Zielsetzung

Die »Arbeitsbezogenen Belastbarkeits- und Gewissenhaftigkeitsskalen« (ABGS) sind ein nach der klassischen Testtheorie entwickeltes, eignungsdiagnostisches Verfahren zur Messung von Belastbarkeit (Facetten »soziale Belastbarkeit«/»Dauerbelastbarkeit«) und Gewissenhaftigkeit (Facetten »Fleiß«/»Ordnung«) mit 35 Items für die Personalauswahl und -entwicklung. Die häufig genutzten Erfolgsindikatoren »biografische Daten« und »Leistungsindikatoren« sollen durch testbasierte Messungen von Persönlichkeitsfacetten ergänzt werden.

Die Items weisen einen Berufsbezug auf. Dabei verwundert bei der Facette »soziale Belastbarkeit«, dass drei der fünf Items eine Vortragssituation im beruflichen Kontext beinhalten, obwohl z. B. bei Untergruppen der Stichprobe (wie »Führungskräfte Produktion«) diese Situation vermutlich nicht zum beruflichen Alltag gehört.

Das Auswertungsergebnis ist ein Kurzprofil. In der computergestützten Version sind zusätzlich kurze Interpretationshinweise enthalten. Testdurchführung und -auswertung werden mit jeweils weniger als zehn Minuten angegeben.

Bewertung des Informationsgehalts der Verfahrenshinweise

Die Autorinnen und Autoren berichten strukturiert und umfangreich zur Test- und Item-Entwicklung sowie Evaluierung. Eine detaillierte Tabelle ordnet die Informationen den DTK-Testinformationsstandards zu.

Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion

Die Autorinnen und Autoren unterstreichen die Bedeutung von Gewissenhaftigkeit und Belastbarkeit für die berufsbezogene Eignungsdiagnostik und diskutieren beide Dimensionen fundiert. Die für die ABGS gewählte Zwei-Facetten-Struktur der beiden Hauptdimensionen, die von der üblichen Sechs-Facetten-Struktur innerhalb des »Fünf-Faktoren-Modells« der Persönlichkeit abweicht, wird theoretisch gut nachvollziehbar begründet.

Objektivität

Durchführungsobjektivität wird bei den ABGS durch Standardisierung (Instruktionen für Testleitung/Testteilnehmende, konkrete Hinweise zur Testsituation) gesichert. Auswertungsobjektivität ist durch das geschlossene Antwortformat, Auswertungsschablonen und Hinweise zum Umgang mit fehlenden Werten gegeben (offen bleibt dabei nur, wie mit zu vielen fehlenden Werten umzugehen ist).

Leider werden nur bei der computerbasierten Auswertung Konfidenzintervalle angegeben, was die Interpretationsobjektivität der Papierversion der ABGS einschränkt.

Die Auswertung erfordert qualifiziertes Personal. Diese Empfehlung fördert die Objektivität in allen Punkten.

Die Interpretationsobjektivität wird durch diverse Normtabellen und Fallbeispiele unterstützt. Bei den Fallbeispielen werden neben den ABGS-Resultaten weitere Verhaltensindikatoren diagnostisch berücksichtigt. Es besteht die Gefahr, dass dies bei der Interpretation zu falschen Schlüssen verleitet. Ein deutlicher Hinweis auf eine vorsichtige Interpretation von hohen Extremwerten wäre wünschenswert, da extrem hohe Ausprägungen in den AGBS auch weniger positive Aspekte beinhalten. Obwohl die letztendliche Wahl einer Normgruppe natürlich den Testanwenderinnen und -anwendern obliegt, wäre es wünschenswert, wenn es klare Empfehlungen zur Wahl der Normgruppe geben würde.

Normierung

In 19 Tabellen werden aktuelle Normen für die Gesamtbevölkerung und verschiedene Geschlechts-, Alters- und Berufsgruppen für »Soziale Belastbarkeit« und »Dauerbelastbarkeit« (N = 2.580) bzw. »Fleiß« und »Ordnung« (N = 5.729) berichtet. Die Normstichproben wurden vor allem in Personalauswahlsituationen (zum Teil Personalentwicklung) erhoben, was für den Einsatz der ABGS in der Personalauswahl grundsätzlich von Vorteil ist. Angaben zur Repräsentativität der Normgruppenverteilung sowie Angaben zur Erhebungssituation (Ausnahme: Erhebungszeitraum) sind vorhanden.

Große Teilstichproben finden sich vor allem für Gruppen von Auszubildenden. Einzelne Normkategorien sind zu klein, z. B. Belastbarkeitsfacetten bei »Technischen Auszubildenden ab 20 Jahren« (n = 69) und »hochrangigen Führungskräften« (n = 64).

Testbeurteilungssystem – Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen

Arbeitsbezogene Belastbarkeits- und Gewissenhaftigkeitsskalen (ABGS)	Die TBS-DTK-Anforderungen sind erfüllt			
	voll	weitgehend	teilweise	nicht
Informationsgehalt der Verfahrensweise	●			
Objektivität	●			
Zuverlässigkeit			●	
Validität		●		
			ja	nein
In den Verfahrenshinweisen ist verzeichnet, wo die nach dem DTK-Testinformationsstandard notwendigen Informationen zu finden sind.			●	

Gesamt- und Teilstichproben sind unausgewogen hinsichtlich der Geschlechterverteilung. Dies wird nicht transparent dargelegt, und damit gegebenenfalls verbundene Problematiken werden kaum thematisiert. In der Normierung für »hochrangige Führungskräfte« (Tabelle 37) finden sich z. B. lediglich für »Fleiß« und »Ordnung« Angaben zum Geschlechterverhältnis (*n* = 105 Personen, 29 Frauen und 76 Männer). Für die anderen Dimensionen »Soziale Belastbarkeit« und »Dauerbelastbarkeit« (*n* = 64 Personen) fehlt eine diesbezügliche Angabe.

Als Normwerte stehen Stanine-Werte zur Verfügung, welche die Vergleichbarkeit mit anderen Testergebnissen erleichtern, in der Interpretation aber eher grobe Normwerte sind.

Zuverlässigkeit

Die interne Konsistenz für verschiedene Gruppen fällt für die vier Subskalen mit $\alpha = .71$ (»Dauerbelastbarkeit«) bis $\alpha = .87$ (»Fleiß«) befriedigend bis sehr gut aus, vor allem mit Blick auf die Skalenlänge (fünf bis 15 Items). Einen Ausreißer stellt $\alpha = .61$ für »soziale Belastbarkeit« (»Führungskräfte Produktion«, Gruppen-AC, *N* = 313, davon 16 oder 17 Frauen) dar. Für die Gesamtstichprobe sind die Werte durchgehend gut.

Im Manual nicht adäquat adressiert wird das starke geschlechtsspezifische Ungleichgewicht der zur Reliabilitätsbestimmung herangezogenen Stichproben (z. B. »Gewissenhaftigkeit«: 1.677 Frauen und 4.052 Männer). Zudem sind die Werte zumindest von zwei Gruppen ausschließlich bei den Mitarbeitenden (vermutlich je) einer Firma (Produktion eines Haushaltsgeräteherstellers bzw. > 90 % »Ingenieurdienstleister«) gewonnen worden. Mögliche Aussageeinschränkungen dieser hoch selektiven Stichproben insbesondere für die Eignungsdiagnostik werden nicht thematisiert.

Die Retest-Reliabilitäten zweier Auszubildendengruppen von 14 bis 20 Jahren (Alter $M_{1/2}$ = 17.18/17,27, $SD_{1/2}$ = 1.49/1.90) nach 2,5 Jahren fallen mit $r_{tt} = .34$ bis $r_{tt} = .71$ insgesamt niedrig aus, was den Einsatz der ABGS in der Altersgruppe der 14- bis 20-Jährigen für prognostische Fragestellungen als weniger geeignet erscheinen lässt. Auch Forschungsbefunde diskutieren Persönlichkeitsdimensionen als relativ stabil erst nach der Jugend (z. B. Specht, Egloff & Schmukle, 2011).

Für die älteren Teilnehmenden zweier Gruppen-Assessments (Alter M_1 = 34.78, *SD* = 6.93; M_2 = 33.36, *SD* = 6.21) fällt die Retest-Reliabilität nach einem Jahr mit $r_{tt} = .64$ bis $r_{tt} = .76$ befriedigend bis gut aus.

Gültigkeit

Zur itemmetrischen Analyse wurden Trennschärfen berechnet (überwiegend im guten bis sehr guten, teils im befriedigenden Bereich); trennscharfe Items ab .30 wurden im Test belassen. Itemschwierigkeiten werden nicht berichtet; die Mittelwerte lassen insbesondere für die Gewissenhaftigkeitsskalen eine Zustimmungstendenz erkennen, was durch die Erhebungssituation erklärbar scheint.

Eine explorative Faktorenanalyse lässt eine klare Faktorladungsstruktur erkennen. Doppelladungen werden nicht mit aufgeführt. Das postulierte Vier-Faktoren-Modell konnte mit guten Fit-Indizes bestätigt werden (CFI: .92; TLI: .91; RMSEA: .03; SRMR: .05). Die Skaleninterkorrelationen fallen überwiegend erwartungsgemäß aus; überraschenderweise korreliert »Fleiß« aus der Skala »Gewissenhaftigkeit« deutlich höher mit »sozialer Belastbarkeit« und »Dauerbelastbarkeit« als mit der zweiten Gewissenhaftigkeitsfacette »Ordnung«.

Die Konstruktvalidität wurde konvergent und diskriminant mit Leistungsmerkmalen und Persönlichkeit weitgehend überzeugend geprüft. Die Korrelation von »Fleiß« mit der Skala »Belastbarkeit« im »Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung« (BIP) von .61 verwundert und verstärkt die Befunde der oben genannten Skalen-Interkorrelationen.

Die Autorinnen und Autoren gehen zu wenig auf die hohen Korrelationswerte der ABGS-Facetten zu »Extraversion« (gemessen mit dem NEO-Fünf-Faktoren-Inventar [NEO-FFI], *r* = .41 bis .47 bei »Fleiß« und *r* = .43 bei »soziale Belastbarkeit«) ein und attestieren dem ABGS hier ohne weitere Begründung divergente Validität. Die Kriteriumsvalidität wurde durch bivariate Korrelationen in zwei Gruppen an objektiven (Schulnote) und subjektiven (Vorgesetzteneinschätzung) Kriterien geprüft, mit größtenteils erwartungsgemäßen Ergebnissen, was für die Validität der ABGS spricht.

Zur inkrementellen Validität: Für jüngere Auszubildendengruppen fällt die zusätzliche Varianzaufklärung zu NEO-FFI bzw. BIP mit 1 bis 2 % eher klein (kaum Zusatznutzen), für eher ältere Führungskräfte im Gruppen-Assessment mit bis zu 8 % beachtlich aus.

Insgesamt kann die Validität als weitgehend gesichert gelten. Deutlichere A-priori-Hypothesen wären wünschenswert gewesen.

Weitere Gütekriterien
Die Ökonomie ist durch die kurze Testdauer (fünf bis zehn Minuten) und geringe Testkosten erfüllt. Zumutbarkeit und Nützlichkeit sind gegeben, Akzeptanz durch die Verwendung arbeitsbezogener Items ist zu erwarten. Die Problematik der Verfälschbarkeit bei Persönlichkeitsverfahren wird lösungsorientiert diskutiert; die Autorinnen und Autoren räumen hinsichtlich der Unverfälschbarkeit Schwächen des Verfahrens ein.

Abschlussbewertung
Mit den ABGS ist ein ökonomisches, weitgehend valides und praxisorientiertes Instrument zur Unterstützung der Personalauswahl und -entwicklung gelungen. Die deduktive Vorgehensweise führt in der Testentwicklung über von Expertinnen und Experten entwickelte Items zu arbeitsplatzbezogenen Skalen. Inwieweit dies den ausgerufenen »Brückenschlag zwischen Wirtschaft und Wissenschaft« (S. 5) darstellt, ist unklar.

Mit Blick auf den praktischen Einsatz sind die Verständlichkeit und gute Lesbarkeit des Manuals positiv hervorzuheben, welches mit wenigen Ausnahmen übersichtlich und transparent gestaltet ist. Wünschenswert wären weitere Studien zur Skala »Fleiß« und zur prognostischen Stärke der ABGS, geschlechterausgewogene Normgruppen sowie altersbezogene Hinweise der Ergebnisinterpretation.

Literatur
Specht, J., Egloff, B. & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology, 101*(4), 862–882.

Testinformation
Moldzio, T., Peiffer, H., Dreier, K., Gergovska, T., Reiner, A. & Felfe, J. (2019). Arbeitsbezogene Belastbarkeits- und Gewissenhaftigkeitsskalen (ABGS). Göttingen: Hogrefe.

Bezugsmöglichkeit
Testzentrale Göttingen, Herbert-Quandt-Str. 4, 37081 Göttingen. Verfahren komplett 132,00 Euro, 25 Fragebogen 35,00 Euro (Nettopreise)

Diese Testrezension wurde im Auftrag des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen (DGPs und BDP) gemäß TBS-DTK (Diagnostik- und Testkuratorium, 2018) erstellt.

Diagnostik- und Testkuratorium (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 3. Januar 2018. *Report Psychologie, 43*(3), 106–113.

Diagnostik- und Testkuratorium (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 3. Januar 2018. *Psychologische Rundschau, 69*(2), 109–116.

Bitte zitieren Sie diesen Artikel wie folgt:
Kissel, D. & Warneke, C. (2024). TBS-DTK-Rezension: »Arbeitsbezogene Belastbarkeits- und Gewissenhaftigkeitsskalen (ABGS)«. Report Psychologie, 49(1), 29–31