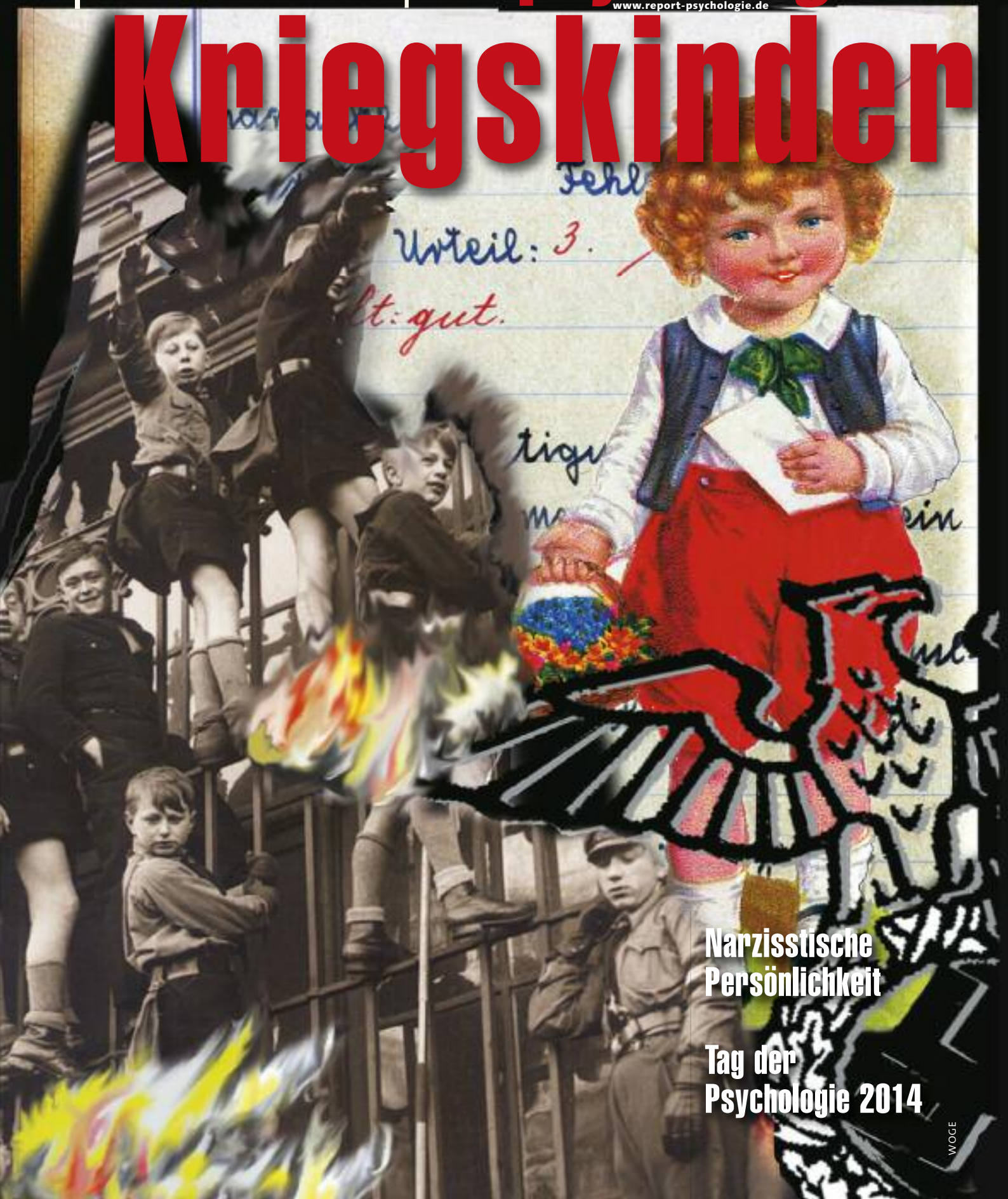
G 3777
FACHZEITSCHRIFT DES BDP
ZEITSCHRIFT DES BERUFSVERBANDES DEUTSCHER PSYCHOLOGINNEN UND PSYCHOLOGEN E.V.
39. JAHRGANG
OKTOBER 2014

10|2014

report psychologie

www.report-psychologie.de

Kriegskinder



Narzisstische Persönlichkeit

Tag der Psychologie 2014

WOGE

Testbeurteilungssystem – Testkuratorium der Föderation deutscher Psychologenvereinigungen

TBS-TK Rezension

TBS-TK
Rezension

Intelligenz-Struktur-Batterie (INSBAT)

Andreas Frey, Friedrich-Schiller-Universität Jena
Morten Moshagen, Universität Kassel

Allgemeine Information über den Test

Die Intelligenz-Struktur-Batterie (INSBAT, Version 27, 2012) ist ein computerbasierter Test zur Messung von Intelligenz für Jugendliche ab 16 Jahren. Der Test umfasst 14 Subtests, die separat mit dem eindimensionalen Rasch-Modell skaliert wurden. Zwölf der Subtests werden als adaptive Tests vorgegeben. Der Itempool enthält 1396 Items. Das Verfahren liegt in einer variablen sowie in einer festen Form (Globalform) vor. Es können Testwerte für »allgemeine Intelligenz« (G-Faktor), fünf Sekundärfaktoren (»fluide« und »kristalline Intelligenz«, »visuelle Verarbeitung«, »quantitatives Denken«, »Kurzzeitgedächtnis«) und 14 Subtests ermittelt werden. Die INSBAT kann als Einzel- oder Gruppentest vorgegeben werden. In den Standardeinstellungen erfordert die Durchführung der variablen Form 255 Minuten. Die feste Globalform benötigt 172 Minuten.

Theoretische Grundlagen

Der theoretische Hintergrund der INSBAT wird in ausreichender Form im Testmanual beschrieben. Das Verfahren basiert auf dem Cattell-Horn-Carroll-Modell (CHC-Modell) und geht von einer hierarchischen Struktur der Intelligenz mit drei Ebenen aus. Die Autoren referieren mehrere Studien, bei denen sich diese Struktur als weitgehend stabil erwiesen hat. Seitens der INSBAT werden sechs der acht Sekundärfaktoren des CHC-Modells berücksichtigt. Es wurden jene Sekundärfaktoren gewählt, die sich empirisch zuverlässig nachweisen ließen und in hohem Zusammenhang mit beruflichem und schulischem Erfolg stehen. Jeder Sekundärfaktor wird durch ein oder mehrere Subtests operationalisiert. Die Itemkonstruktion erfolgte mit subtestspezifischen, theoriebasierten Konstruktionsrationalen, sodass die resultierenden Items einen expliziten Bezug zu den kognitiven Operationen aufweisen, die dem zu messenden Fähigkeitskonstrukt zugrunde liegen.

Objektivität

Die Auswertungs- und Interpretationsobjektivität des Verfahrens sind aufgrund des hohen Standardisierungsgrades gegeben. Die inhaltliche Interpretation der gemessenen Fähigkeiten wird klar beschrieben. Die Testergebnisse werden in Bezug zu einer Normstichprobe gesetzt, und es knappe Vorschläge zur verbalen Vermittlung der Testergebnisse gemacht. Einschränkungen zeigen sich hinsichtlich der Durchführungsobjektivität. Hier ist unklar, inwieweit die vom Testleiter wählbare Reihenfolge und Zusammenstellung der Subtests einen Einfluss auf die Testergebnisse haben. Insgesamt gesehen wünscht man sich zuweilen deutlichere Angaben im Manual. So wird bei einigen Subtests nicht angegeben und auch den Testpersonen nicht durch die Instruktion mitgeteilt, dass übersprungene Aufgaben als falsch bewertet werden (was erst bei Kontrolle der Parameterschätzungen ersichtlich wird). Auch legt zwar die Natur einiger Aufgabengruppen (etwa zum Arbeitsgedächtnis) nahe, dass keine Verwendung von Hilfsmitteln gestattet ist, jedoch wird dies weder in den Instruktionen noch im Manual erwähnt.

Normierung

Für die INSBAT liegen Normen für Erwachsene (laut Auskunft Testautoren: 15–73 Jahre; laut Manual: 16–73 Jahre) und Jugendliche (12–15 Jahre) vor. Die standardisierte Ergebnisdarstellung umfasst Werte auf der Logit-Metrik, Prozentränge, T-Werte, IQ-Werte (jeweils mit 95 Prozent Konfidenzintervall) und Profildarstellungen. Die Normierungsstichprobe der Erwachsenen (N = 904) wurde in Österreich gemäß einem geschichteten Quotenstichprobenplan untersucht. Die Interpretation der Testergebnisse erfolgt bezüglich geschlechts-, bildungs- oder altersspezifischer Normen. Die Normierungsstichprobe der Jugendlichen (N = 1595) ist als Zusatz anzusehen, da einerseits die INSBAT für einen Altersbereich ab 16 Jahren vorgesehen ist und andererseits nicht alle Testvariablen erhoben wurden. Ohne Einschränkungen ist die normorientierte Interpretation somit erst ab einem Alter von 16 Jahren möglich.

Zuverlässigkeit

Adaptives Testen ist hocheffizient, sodass mit der INSBAT eine große Anzahl an Intelligenzfacetten in akzeptabler Testzeit präzise gemessen werden kann. Bei den adaptiven Subtests kann die Zielreliabilität zwischen .50 und .90 gewählt werden. Das Manual enthält indes jenseits von Simulationsstudien keine Angaben, ob die Zielreliabilitäten bei empirischen Anwendungen auch erreicht werden. Dies ist insbesondere bei Stichproben mit extremen Intelligenzausprägungen fraglich, da nur wenige Items mit entsprechenden Schwierigkeiten vorliegen. Für die Globalform wurden in der Normstichprobe interne Konsistenzen zwischen .73 und .89 realisiert. Für acht der 14 Subtests werden Test-Retest-Reliabilitäten (Zeitspanne: 15 Monate; N = 60) angegeben (.52-.87). Die sechs Sekundärfaktoren werden durch Aggregation von Subtests gebildet. Bei Zielreliabilitäten von .70 für die adaptiven Subskalen ergeben sich rechnerisch interne Konsistenzen von .80 und .83 für die Sekundärfaktoren und von .96 für den G-Faktor.

Literatur

Asseburg, R. (2011). Leistungsbereitschaft in Testsituationen. Motivation zur Bearbeitung adaptiver und nicht-adaptiver Leistungstests. Marburg: Tectum.

Frey, A., Hartig, J., & Moosbrugger, H. (2009). Effekte des adaptiven Testens auf die Motivation zur Testbearbeitung. Diagnostica, 55, 20-28.

Ortner, T. M., Weißkopf, E. & Koch, T. (2014). I will probably fail – higher ability students' motivational experiences during adaptive achievement testing. European Journal of Psychological Assessment, 30, 48-56.

INSBAT	INSBAT	Die TBS-TK-Anforderungen sind erfüllt			
		voll	weitgehend	teilweise	nicht
	Allgemeine Informationen, Beschreibung und diagnostische Zielsetzung		●		
	Objektivität		●		
	Zuverlässigkeit	●			
	Validität		●		

Gültigkeit

Die theoretische Fundierung der Itemkonstruktion wird dadurch gestützt, dass beachtliche 50 bis 94 Prozent der Itemparametervarianz durch die Konstruktionsrationale erklärt werden konnten. Auf der Ebene der Subtests sind somit valide Testwertinterpretationen gemäß den theoretisch spezifizierten Konstrukten möglich. Verschiedene Subtests eignen sich zudem zur Vorhersage verschiedener leistungsbezogener Kennwerte, wie Leistung im Flugsimulator, Ausbildungserfolg und Studienerfolg. Mehrere konfirmatorische Faktorenanalysen erbrachten übereinstimmend Evidenz für die angenommene hierarchische Struktur. Aus einer dieser Studien wurden Gewichte entnommen, die seitens der INSBAT für die Berechnung der Testwerte für die fünf Sekundärfaktoren und den G-Faktor genutzt werden. Bislang offen ist, ob andere Subtestzusammenstellungen zu substanziell anderen Gewichten und damit anderen Testergebnissen für Sekundärfaktoren und G-Faktor führen oder ob die von den Autoren getroffene Invarianzannahme trägt. Aufgrund des aktuellen Erkenntnisstands sind konstruktvalide Testwertinterpretationen für Sekundär- und G-Faktor jedenfalls nur für die bei der Testentwicklung genutzte spezifische Subtestzusammenstellung gesichert. Zudem basieren die Strukturprüfungen auf einer früheren Version der INSBAT mit weiteren Subtests, was die Übertragbarkeit auf die aktuelle Version zusätzlich einschränkt.

Weitere Gütekriterien

Die Verfälschbarkeit von Leistungstests ist als gering zu beurteilen und beschränkt sich im Wesentlichen auf Abschreiben in Gruppenuntersuchungen sowie vorheriges Vertrautmachen mit den Aufgaben. Beides ist aufgrund der adaptiven Itemvorgabe und des individuellen computergestützten Testens sehr unwahrscheinlich. Eine gewisse Störanfälligkeit besteht als Folge der computergestützten Administration, die entsprechende Anforderungen an die Systemstabilität stellt. Ein ungeplanter Testabbruch führt zum Verlust der aktuell angefallenen Daten. Für die Subtests ist, teilweise nach Itemeliminationen, Rasch-Homogenität gegeben. Ein im Manual nicht ausreichend aufgegriffener Aspekt ist die Zumutbarkeit der INSBAT, welche angesichts der langen Testzeiten und insbesondere aufgrund der adaptiven Itemvorgabe zentral erscheint. Beim adaptiven Testen werden Aufgaben mit individueller Lösungswahrscheinlichkeit von .50 vorgelegt, was zwangsläufig zu hoher Unsicherheit beim Antworten führt. Entsprechend zeigen aktuellere Studien (Asseburg, 2011; Frey, Hartig & Moosbrugger, 2009; Ortner, Weißkopf & Koch, 2014), dass adaptives Testen nicht immer die im Manual postulierte förderliche Wirkung auf die Motivation zur Testbearbeitung hat, sondern dass durchaus auch mit nachteiligen motivationalen und emotionalen Effekten und vor allem einer hohen Testbelastung zu rechnen ist.

Abschlussbewertung/Empfehlung

Die INSBAT ist ein Instrument zur Messung von Intelligenzstrukturen mit überdurchschnittlicher psychometrischer Untermauerung und hohem Gestaltungsspielraum für den Testleiter. Die elaborierte psychometrische Vorgehensweise erlaubt aber auch genaue Einschätzungen der Güte des Verfahrens und bringt im Falle des adaptiven Testens neue Aspekte mit sich, die beachtet werden müssen. Dank der adaptiven Itemvorgabe lässt sich ein guter Kompromiss zwischen Erhebungsökonomie und Messgenauigkeit erzielen. Allerdings sind Testergebnisse für die fünf Sekundärfaktoren und den G-Faktor bislang nur dann valide interpretierbar, wenn die gleichen Subtests gewählt werden wie bei der Studie, bei der Gewicht geschätzt wurden. Für andere Subtestauswahlen fehlen bislang entsprechende Validitätsstudien. Künftig sollte die Vergleichbarkeit der Testergebnisse bei variierenden Testzeiten und Testabfolgen ebenso empirisch untersucht werden wie die Frage, inwiefern die vom Anwender angegebenen Zielreliabilitäten tatsächlich erreicht werden. Schließlich sind die adaptiven Subtests der INSBAT als sehr anstrengend einzustufen, sodass negative motivationale und emotionale Auswirkungen nicht auszuschließen sind. Jenseits dieser Problematik sind der Einsatz und die Interpretation der Subtests ohne Einschränkungen zu empfehlen. Technisch wären Verbesserungen beim Testsystem hinsichtlich der Datensicherung bei Ausfällen nützlich. Trotz der genannten aktuell noch gegebenen Einschränkungen handelt es sich bei der INSBAT um ein hochwertiges Intelligenztestverfahren mit belastbarer theoretischer Fundierung, zeitgemäßer Konstruktion und hoher Anwenderfreundlichkeit für die Testadministratoren.

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

Testkuratorium (2009). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 09. September 2009. Report Psychologie, 34, 470-478.

Testkuratorium (2010). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 09. September 2009. Psychologische Rundschau, 61, 52-56.

TESTINFORMATIONEN

Arendasy, M., Hornke, L. F., Sommer, M., Wagner-Menghin, M., Gittler, G., Häusler, J., Bognar, B. & Wenzl, M. (2012). Intelligenz-Struktur-Batterie – INSBAT. Mödling, Österreich: Schuhfried GmbH. Bezugsquelle: Schuhfried GmbH, Hyrtlstraße 45, 2340 Mödling, Österreich. Kosten: eine Testvorgabe: 9,13 Euro; 50 Testvorgaben: 456,50 Euro; Lizenz für ein Jahr: 684,75 Euro; Lizenz für sieben Jahre: 2739 Euro

Bitte zitieren Sie diesen Artikel wie folgt:
Frey, A. & Moshagen, M. (2014). TBS-TK Rezension: »Intelligenz-Struktur-Batterie (INSBAT)«. Report Psychologie, 10, 407-408.