

PRECIRE JobFit

LOTHAR SCHMIDT-ATZERT, Philipps-Universität Marburg, Fachbereich Psychologie
 JANINA KÜNECKE, Psychologische Hochschule Berlin
 JOHANNES ZIMMERMANN, Universität Kassel

Beschreibung des Tests und der diagnostischen Zielsetzung

Das Verfahren soll anhand der Sprachprobe eines 15-minütigen Telefoninterviews berufsrelevante Merkmale automatisiert erfassen. Dazu gehören konkrete sprachliche Merkmale – wie unter anderem Komplexität und positive Wirkung –, Persönlichkeitseigenschaften – wie unter anderem emotionale Stabilität, Verantwortungsübernahme und Zielorientierung – und verschiedene Skills, also weniger stabile Fähigkeiten und Fertigkeiten – wie unter anderem Durchsetzungsvermögen und Belastbarkeit. Zudem soll das Verfahren die Passung einer Bewerberin bzw. eines Bewerbers für eine Stelle auf Grundlage einer Anforderungsanalyse feststellen.

Bewertung des Informationsgehalts der Verfahrenshinweise

Die Anforderungen an die allgemeinen Informationen sind teilweise erfüllt. Die Dokumentation der empirischen Studien in den Verfahrenshinweisen weist große Lücken auf. Dies gilt insbesondere für die Überprüfung der Konstrukt- und Kriteriumsvalidität in neuen (das heißt von der Testkonstruktion unabhängigen) Stichproben. In der Psychologie ist die Erkenntnis fest verwurzelt, dass Ergebnisse stets methodenabhängig sind; daher ist das Vertrauen in die Ergebnisse der unzureichend dokumentierten Studien gering.

Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion

Das Verfahren orientiert sich nicht an einer psychologischen oder linguistischen Theorie, sondern am »Linguistic Inquiry and Word Count« von Pennebaker et al. »PRECIRE« wertet nach eigenen Angaben 564.367 (!) sogenannte »Features« aus, die sich aus lexikalischen, syntaktischen, prosodischen Informationen und deren Kombinationen ergeben.

Insgesamt 5.201 Personen lieferten eine Sprachprobe ab und wurden hinsichtlich 61 Außenkriterien untersucht. Zu deren Messung wurden Selbsteinschätzungsfragebögen (zum Beispiel »B5T«, »AVEM«, »PHQ«) sowie Fremdeinschätzungen auf Grundlage von Beobachtungen in »klassischen Assessmentsituationen« (S. 12) verwendet. Zur Skalenkonstruktion wurden Verfahren aus dem Bereich des maschinellen Lernens angewendet, wobei der Datensatz hierzu mehrmals in eine Trainings- und eine kleinere Validierungsstichprobe aufgeteilt wurde. Dabei wurden für jede Skala jeweils diejenigen Features ausgewählt, die sich für die Vorhersage eines Außenkriteriums als passend erwiesen (pro Skala bis zu 10.000 Features).

Objektivität

In einem telefonischen Vorgespräch wird die Probandin bzw. der Proband individuell über den Testablauf informiert. Die Sprachaufnahme erfolgt dann in einem automatisierten Telefoninterview (maximal 25 Minuten) mit einfachen Fragen zum eigenen Leben, wobei einzelne Fragen ausgelassen werden können. Die Ausführungen sprechen für eine leicht eingeschränkte Durchführungsobjektivität.

Die Sprachprobe wird teilautomatisiert transkribiert. Dieser Text und die dazugehörige Audiodatei werden maschinell analysiert. Die Auswertungsobjektivität ist damit gegeben.

Die Interpretation erfolgt durch von »PRECIRE« geschulte Beraterinnen und Berater gemäß den Erläuterungen im Manual. Dabei fehlen Hinweise dazu, bei welchen Normwerten zum Beispiel von einer durchschnittlichen Ausprägung gesprochen werden soll. Lediglich für die Aussagen zur Passung werden präzise Angaben gemacht. Zudem gibt das Manual keine Auskunft zur Präzision der Skalenwerte (zum Beispiel Angaben zum Standardmessfehler). Die Interpretationsobjektivität ist also eingeschränkt.

Insgesamt sind die Anforderungen an die Objektivität weitgehend erfüllt.

Normierung

Die 5.201 Personen, an denen das Verfahren entwickelt wurde, fungieren als Eichstichprobe. Sie sind alle mindestens 18 Jahre alt und beherrschen die deutsche Sprache mindestens auf dem Niveau C1. Das Durchschnittsalter wird mit 37,6 Jahren ($SD = 18,3$) angegeben; Männer und Frauen sind etwa gleich stark vertreten.

Laut Autorinnen und Autoren haben Alter, Geschlecht und Bildungsgrad keinen Einfluss auf die Skalenwerte (S. 18), weshalb einheitliche Normwerte präsentiert werden. Dies ist merkwürdig, da viele Persönlichkeitsmerkmale sehr wohl von solchen Faktoren abhängen (siehe zum Beispiel Roberts, Walton & Viechtbauer, 2006).

Zuverlässigkeit

Die interne Konsistenz (Cronbachs Alpha) der 22 Skalen wird mit .68 bis .92 angegeben. »Zur Messung [...] wurden die Ergebnisse der für ein Merkmal relevanten Featurekombinationen pro Bereich miteinander korreliert« (S. 36). Da die Features anhand der gleichen Daten zu Skalen zusammengefasst und hinsichtlich interner Konsistenz beurteilt wurden, ist von einer Überschätzung der Reliabilität auszugehen.

PRECIRE JobFit	Die TBS-DTK-Anforderungen sind erfüllt			
	voll	weitge- hend	teilwei- se	nicht
Informationsgehalt der Verfahrensweise			●	
Objektivität		●		
Zuverlässigkeit			●	
Gültigkeit			●	
			ja	nein
In den Verfahrenshinweisen ist verzeichnet, wo die nach dem DTK-Testinformationsstandard notwendigen Informationen zu finden sind.				●

Zur Schätzung der Retest-Reliabilität wurden 500 Probandinnen und Probanden nach drei Monaten erneut getestet. Nähere Angaben zur Stichprobe fehlen. Die Analyse beschränkt sich auf 11 Skalen zu Merkmalen, von denen eine hohe Stabilität angenommen wurde. Die Koeffizienten variieren zwischen $r = .67$ (Neugier) und $r = .81$ (Kooperationsbereitschaft).

Insgesamt sind die Anforderungen an die Reliabilität teilweise erfüllt.

Gültigkeit

Für eine Teilstichprobe der Normierungsgruppe ($n = 895$) werden Korrelationen mit »Fragebogenskalen und Beobachtungsverfahren« berichtet. Diese variieren zwischen $r = .59$ (Belastbarkeit) und $r = .81$ (Kooperationsbereitschaft). Zugleich fallen die Korrelationen mit konstruktdivergenten Skalen überwiegend sehr niedrig aus. Dieses Ergebnis kommt vermutlich dadurch zustande, dass das Auswertungsprogramm mehrfach an dem Datensatz trainiert und validiert wurde (das heißt, dass eine Überanpassung vorliegt). Die berichteten Korrelationen zeigen nur, dass das Prozedere für diesen Datensatz sehr erfolgreich war; über die Validität im eigentlichen Sinne sagen die Kennwerte nichts aus.

In einer unabhängigen Stichprobe ($N = 303$; Alter: $M = 28.8$, $SD = 10.6$; keine weiteren Angaben) korrelieren die »PRECIRE«-Skalenwerte mit den korrespondierenden Skalen des »HEXACO«-Fragebogens relativ hoch: $r = .49$ bis $r = -.66$. Belastbare Belege zur Kriteriumsvalidität (hinsichtlich beruflicher Eignung) liegen nicht vor.

Die Anforderungen an die Validität sind insgesamt teilweise erfüllt.

Weitere Gütekriterien

Im Manual finden sich nur sehr knappe Ausführungen zur Ökonomie (kurze Durchführungszeit, von zu Hause per Telefon), Akzeptanz (hohe Akzeptanz auf der Seite der Anwenderinnen und Anwender) und Ver-

fälschungstendenzen (auszuschließen). Die Annahme der Unverfälschbarkeit erscheint weitgehend plausibel, wobei bisher keine empirischen Ergebnisse zum Einfluss von Faking vorliegen. Bei ausreichenden Deutschkenntnissen ist die Fairness des Tests gewährleistet, wobei bisher auch diesbezüglich keine empirischen Ergebnisse vorliegen. So wäre zum Beispiel wünschenswert, die Messinvarianz der Skalen für verschiedene Personengruppen sicherzustellen (zum Beispiel Personen mit unterschiedlichem Alter oder Geschlecht, Personen mit Stimm- und Sprachproblemen, Personen mit Deutsch als Zweitsprache).

Abschlussbewertung

Bei »PRECIRE JobFit« handelt es sich um ein ungewöhnliches Testverfahren, das aus dem Potenzial von Methoden des maschinellen Lernens schöpft. Unverständlich ist, dass keine Testautorinnen und -autoren genannt werden, die mit ihrem Namen für die Qualität bürgen; ungewöhnlich ist die schlechte Dokumentation des methodischen Vorgehens. Das Verhältnis von untersuchten Personen ($N = 5.201$) zu analysierten Variablen (564.367 Features) bei der Erstellung eines Vorhersagemodells für psychologische Merkmale ist sehr ungünstig. Es ist nahezu unvermeidbar, dass dabei zufällig im Datensatz vorhandene Konstellationen von Features ausgebeutet werden. Die interne Konsistenz und weitgehend auch die Konstruktvalidität werden anhand der Daten aus der Stichprobe zur Testkonstruktion bestimmt und so vermutlich überschätzt. Die Überprüfung der Konstruktvalidität an einer neuen Stichprobe ist schlecht dokumentiert, die Angaben zur Kriteriumsvalidität sind ungenügend.

Vor allem fehlt ein Nachweis, dass das Verfahren eine inkrementelle Validität gegenüber den Persönlichkeitsfragebogen aufweist, anhand derer es entwickelt wurde. Mit anderen Worten: »PRECIRE« erfasst bestimmte Persönlichkeitsmerkmale, die ansonsten mit Fragebogen gemessen werden. Es bleibt aber offen, ob damit die berufliche Eignung besser erkannt werden kann. Es fehlt sogar der Nachweis, dass »PRECIRE« diesbezüglich den Fragebogen ebenbürtig ist. Insgesamt ist die psychometrische Qualität des Verfahrens nicht hinreichend belegt.

Das Verfahren wirft ethische und rechtliche Fragen auf. Der Datenschutz wird im Manual nur bezüglich der Sicherheit der persönlichen Daten diskutiert. Zum Datenschutz gehört aber auch die informationelle Selbstbestimmung. Wer eine Sprachprobe abliefern, weiß nicht, was er damit über sich verrät. Probandinnen und Probanden haben keine Möglichkeit, Angaben zu einzelnen analysierten Merkmalen zu verweigern. Die Freiwilligkeit der Teilnahme darf bei einer Bewerberinnen- und Bewerberauswahl in vielen Fällen bezweifelt werden. Potenzielle Nutzerinnen und Nutzer sollten vorab mit den zuständigen Datenschutzbeauftragten klären, welche rechtlichen Konsequenzen die Anwendung für sie haben kann.

TESTINFORMATIONEN

Diese Testrezension wurde im Auftrag des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen (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 Psychologinnenvereinigungen. 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 Psychologinnenvereinigungen. Revidierte Fassung vom 3. Januar 2018. *Psychologische Rundschau*, 69 (2), 109–116.

PRECIRE Technologies GmbH (2016). PRECIRE JobFit.

Bezugsquelle:

Precire Technologies GmbH, Charlottenburger Allee 40, 52068 Aachen. Für Einzeltestungen im Top Executive Management bis 695 Euro; in skalierenden Projekten je Testung meist zwischen 295 Euro und 495 Euro, je nach Fragestellung. Für die Bereitstellung der Testergebnisse sind meist noch die Erstellung einer Soll-Profilierung, die Bereitstellung einer digitalen Plattform inklusive Benutzungsrechte in einem Backend sowie eine PDF-Erstellung integriert.

Literatur:

Roberts, B. W., Walton, K. E. & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132 (1), 1–25.

Bitte zitieren Sie diesen Artikel wie folgt:

Schmidt-Atzert, L., Künecke, J. & Zimmermann, J. (2019). TBS-DTK Rezension: »PRECIRE JobFit«. *Report Psychologie*, 44 (7/8), 19–21.

Die Gemeinsamkeiten emotionaler Störungen

Die bahnbrechende, radikale Abkehr von der krankheitsspezifischen Behandlung emotionaler Störungen – nun endlich in deutscher Übersetzung!

Transdiagnostische Behandlung emotionaler Störungen

David H. Barlow
Todd J. Farchione
Shannon Sauer-Zavala
Heather Murray-Latin
Kristen K. Ellard
Jacqueline R. Bullis
Kate H. Bentley
Hannah T. Boettcher
Clair Cassiello-Robbins

Therapeutenmanual

Deutschsprachige Ausgabe herausgegeben von Franz Caspar

 hogrefe

David H. Barlow et al.
Transdiagnostische Behandlung emotionaler Störungen
Therapeutenmanual

Deutsche Ausgabe herausgegeben und übersetzt von Franz Caspar.
2019. 208 S., 4 Tab., Kt
€ 39,95 / CHF 48.50
ISBN 978-3-456-85240-9

Parallel dazu das Arbeitsbuch für Betroffene:
ISBN 978-3-456-85241-6
€ 29,95 / CHF 39.90



 hogrefe