Sarges & Wottawa

Handbuch wirtschaftspsychologischer Testverfahren

2., überarbeitete und erweiterte Auflage

Band I:
Personalpsychologische Instrumente

PABST

Handbuch wirtschaftspsychologischer Testverfahren

Band I: Personalpsychologische Instrumente

2., überarbeitete und erweiterte Auflage

herausgegeben von

Prof. Dr. Werner Sarges
Universität der Bundeswehr Hamburg

und

Prof. Dr. Heinrich Wottawa
Ruhr-Universität Bochum

PABST SCIENCE PUBLISHERS
Lengerich, Berlin, Bremen, Miami,
Riga, Viernheim, Wien, Zagreb

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://dnb.ddb.de> abrufbar.

Das Werk, einschließlich aller seiner Teile, ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

© 2004 Pabst Science Publishers, D-49525 Lengerich

Konvertierung: Dirk Brackmeyer

Druck: Drukarnia Dimograf

ISBN 3-935357-55-9

Danksagungen

Zur 1. Auflage

Aus der Vielzahl derer, die zum Gelingen dieses Handbuches beigetragen haben, möchten wir etlichen Beteiligten danken, die durch Ihr Mitwirken das Erscheinen dieses Bandes ermöglicht haben.

Zu vorderst sind dazu zu nennen die Autoren der einzelnen Beiträge, die entweder jeweils ihr eigenes oder ein von ihnen vertriebenes Verfahren beschrieben bzw. in späteren Abschnitten der gemeinsamen Zusammenarbeit auch gewünschte Änderungen eingearbeitet haben.

Bei der Recherche nach zu beschreibenden Testverfahren gab uns Dr. Martin Scherm (Wissenschaftlicher Assistent an der Professur für Quantitative Methoden, Universität der Bundeswehr Hamburg) interessante Hinweise. Darüber hinaus entlastete uns Dr. Scherm in der frühen Phase des Entstehens dieses Bandes durch die Übernahme etlicher organisatorischer Arbeiten.

Martin Elfert (Studentische Hilfskraft an der Professur für Quantitative Methoden, Universität der Bundeswehr Hamburg) hat die meisten der zuerst eingegangenen Test-Beschreibungen in ein für jedes Textverarbeitungsprogramm lesbares Datei-Format konvertiert und somit weitere Bearbeitungsschritte am PC ermöglicht. Stefanie Roos (Freie Mitarbeiterin am Institut für Management-Diagnostik) hat diese Arbeit fortgesetzt, bei einem Großteil der eingegangenen Beiträge nötige Umkonvertierungen vorgenommen und schließlich sämtliche vorliegende Beschreibungen in den von uns gewünschten Satzspiegel überführt. Das einheitliche Layout der Beiträge trägt somit maßgeblich ihre Handschrift.

Die Einarbeitung inhaltlicher Verbesserungen, um die wir die einzelnen Autoren ersuchten, organisierte schließlich Christian Roos (Mitarbeiter am Institut für Management-Diagnostik), dem über die Steuerung inhaltlicher Korrekturen hinaus auch wesentliche Anteile der redaktionellen Abschlussarbeit nebst der Vereinheitlichung mancher Formulierung oder mittlerweile wieder geänderter Formatvorlage oblagen.

Nicht zuletzt möchten wir dem Verleger, Herrn Pabst, für die offene Art der Zusammenarbeit danken.

Hamburg und Bochum, im März 2001

Die Herausgeber

Zur 2. Auflage

Auch zur 2. Auflage sei den Autoren der Beiträge an erster Stelle gedankt. Weiterhin gebührt unser Dank Philipp Dabelstein (cand. psych. an der Universität Heidelberg) für seine umfangreichen und gründlichen editorischen Detailarbeiten sowie Christian Roos (Institut für Management-Diagnostik) für seine tatkräftige Unterstützung in der Endphase der Arbeit zu diesem Handbuch (Organisatorisches, Korrekturlesen und Registererstellen).

Schließlich konnte das Ganze nur gelingen durch die Geduld und Kooperation unseres Verlegers Wolfgang Pabst; vielen Dank auch dafür. Insbesondere hat uns seine spontane Bereitschaft gefreut, das *Handbuch wirtschaftspsychologischer Testverfahren* nunmehr in zwei Bänden herausgegeben zu können: *Band I: Personalpsychologische Instrumente* und *Band II: Organisationspsychologische Instrumente*.

Der hier vorliegende *Band I: Personalpsychologische Instrumente* enthält die meisten Tests der 1. Auflage sowie ca. 40% neu hinzugekommene; dies sind alles Tests, die sich auf die einzelne Person beziehen und deren Anwendungszwecke Personalauswahl, –beratung und –entwicklung sind. Der in ca. 15 Monaten folgende *Band II: Organisationspsychologische Instrumente* wird die (nur wenigen) organisationspsychologischen Tests der 1. Auflage enthalten, dann aber viele weitere Verfahren, die sich auf andere organisationale Entitäten als Einzelpersonen beziehen, d.h. auf Teams, Abteilungen, Personengruppen, Arbeitsbedingungen etc.

Hamburg und Bochum, im Oktober 2004

Die Herausgeber

Zur Einführung

Testverfahren im HR-Management

In der Personalarbeit werden zunehmend leistungsfähige und kosteneffiziente Verfahren zur Einschätzung neuer oder schon vorhandener Mitarbeiter benötigt. Nur auf dieser Grundlage kann die Personalauswahl, die Personalplatzierung (Wer passt am besten auf welche Position?) und die Personalentwicklung (Bei wem ist was entwickelbar? Wie können knappe Personalentwicklungsressourcen am sinnvollsten genutzt werden?) optimal für das Unternehmen gestaltet werden. Treffsichere Einschätzung von Personen ist aber nicht nur für das Unternehmen, sondern auch für die Mitarbeiter und Bewerber wichtig. Das Scheitern von beruflichen Ambitionen, sei es durch den Erhalt einer überfordernden beruflichen Position oder die Teilnahme an einer zu anspruchsvollen Weiterbildungsmaßnahme, ist für Betroffene nicht nur belastend, sondern kann auch deren weiteren Lebenslauf stark negativ beeinflussen.

Es kann kein Zweifel daran bestehen, dass hinsichtlich des Verhältnisses von Aufwand und Ertrag gute Testverfahren im Vergleich zu anderen Techniken der Eignungseinschätzung hervorragend abschneiden. Immerhin sind solche Tests, die von Fachleuten nach psychometrischen Kriterien sorgfältig konstruiert wurden, das messtechnisch Anspruchsvollste, was die psychologische Eignungsdiagnostik zu bieten hat. Die Methodologie der Testkonstruktion ist weit entwickelt, und man hat in einer langen Tradition eine Fülle wertvoller Erfahrungen sammeln können. Schließlich haben Theorie und Praxis der Testkonstruktion sogar die methodischen Standards für die anderen diagnostischen Verfahren, wie Interview, Tätigkeitssimulationen oder Assessment Center, geliefert.

In Abgrenzung zu den anderen Verfahren der Eignungsdiagnostik verstehen wir unter Tests nur Instrumente, mit Hilfe derer wir psychometrisch vergleichbare und gültige Informationen über Verhalten und Erleben einzelner Personen erhalten können – was häufig bei Zeugnisnoten oder Interviews so nicht der Fall ist. Gut konstruierte Tests müssen den Gütekriterien Standardisierung, Reliabilität und Validierung genügen. In der Regel werden psychometrische Tests von an Wissenschaftsinstitutionen arbeitenden Psychologen entwickelt und von Testverlagen oder Beratungsfirmen vertrieben. Doch entwickeln und validieren gelegentlich auch Psychologen in größeren Unternehmen oder Organisationen eigene Tests, um ihren ganz spezifischen Zielgruppen und Zwecken zu genügen, beispielsweise Tests für Anlageberater, Offiziere oder Piloten.

Nur am Rande sei erwähnt, dass die Validität vieler Tests im Hinblick auf Berufserfolg (d.h. als substantieller Prädiktor dafür) in den verschiedensten beruflichen Bereichen auch durch Forschungszusammenfassungen der jüngsten Zeit eindrucksvoll belegt wird. Dabei haben sich vor allem zwei testmäßige Prädiktoren als durchgängig bedeutsam in allen möglichen Berufsfeldern etabliert: Intelligenz (aus dem Leistungsbereich) und Integrität (aus dem Persönlichkeitsbereich).

Trotz dieser – in Fachkreisen völlig unumstrittenen – Vorteile von Testverfahren ist der Einsatz solcher Verfahren in der deutschen Wirtschaft, insbesondere auch im Vergleich zum westlichen Ausland, noch immer etwas zurückhaltend. Falls Tests bisher herangezogen wurden, dann vor allem im Rahmen der Auswahl von Auszubildenden oder für Berufe mit mittlerem Anforderungsniveau, weniger auch für qualifizierte Experten oder Führungsnachwuchskräfte bzw. Führungskräfte selbst. In etlichen anderen europäischen Ländern und in den USA ist man diesbe-

züglich deutlich aufgeschlossener, bis hin zur Besetzung oberster Management-Positionen: Carly Fiorina, Chefin von Hewlett-Packard, musste sich vor ihrer Nominierung einer zweistündigen Prüfung mit 900 Ja/Nein-Fragen unterziehen. In dieser Hinsicht ist Deutschland also noch Entwicklungsland. Organisationen und Bewerber sind allerdings oft gar nicht einmal so schwer davon zu überzeugen, dass auch Tests zum selbstverständlichen Bestandteil einer guten Eignungsdiagnostik gehören.

Im Übrigen werden alle Verfahren der Eignungsdiagnostik, und damit auch und insbesondere Tests, umso wichtiger, je mehr der Lebenslauf und die Zeugnisse generell oder in spezifischen Kandidatenstichproben speziell an Aussagekraft verlieren – was seit geraumer Zeit der Fall ist.

Testverfahren können nun auf viele verschiedene Arten klassifiziert werden. Inzwischen hat sich zumindest die Grobklassifikation in Leistungstests und Persönlichkeitstests durchgesetzt. Zwar gehört im alltäglichen Sprachgebrauch auch der Bereich der Leistung zur Persönlichkeit, die von der Psychologie getroffene Unterscheidung von Leistungs- vs. Persönlichkeitstests wird aber dadurch sinnvoll, dass jeweils unterschiedliche Reaktionsweisen gefordert sind: Leistungstests sollen maximales Verhalten (z.B. im Merkmalsbereich der Intelligenz), Persönlichkeitstests typisches Verhalten (z.B. im Merkmalsbereich der Extraversion) erfassen.

Bei der noch immer zu geringen Anwendungshäufigkeit von Tests im Personalmanagement in Deutschland werden außer reinen Leistungstests inzwischen vermehrt auch Persönlichkeitstests eingesetzt. Dadurch möchte man nichtleistungsmäßige Eigenschaften im fachlichen und außerfachlichen Bereich, die aber die Arbeitsleistung im Einzelfall sehr stark mitbedingen können, ebenfalls erfassen (z.B. Emotionale Stabilität, Kooperationsbereitschaft oder Leistungsmotivation).

Dies ist besonders zu begrüßen, sind doch Persönlichkeitsmerkmale für die Prognose der beruflichen Leistung kaum weniger wichtig als Leistungsmerkmale. Dies haben jüngst erst wieder Forschungsbefunde gezeigt, die einen beachtlichen Zusammenhang gefunden haben zwischen Berufserfolg und den sog. "big five", den fünf großen Persönlichkeitsfaktoren, nach denen Menschen andere Menschen auf einer generellen Ebene einschätzen, nämlich Extraversion, Emotionale Stabilität, Offenheit für neue Erfahrungen, Gewissenhaftigkeit und Verträglichkeit. Manche Facetten beruflicher Leistung, wie Führung, Engagement oder Disziplin, lassen sich durch Persönlichkeitsmerkmale sogar besser prognostizieren als durch kognitive oder andere Leistungsmerkmale. Experten schätzen sogar, dass über 90 % der Beschäftigungsverhältnisse, die scheitern, nicht aufgrund von fachlichen oder sonstigen Könnens-Kompetenzen des betreffenden Mitarbeiters beendet werden (die man offenbar leichter feststellen kann), sondern wegen Unstimmigkeiten von Merkmalen seiner Persönlichkeit mit den Anforderungsbedingungen der Position (was man bisher wohl nur unzureichend in Erfahrung bringen konnte). Zum Beispiel: Eine Führungskraft, die das Führen anderer Personen nicht als reizvoll und motivierend erlebt, wird unter Umständen über lange Jahre hinweg genau diejenigen Situationen meiden, in denen aktiv geführt werden müsste. Welche Konsequenzen hieraus für ein Unternehmen erwachsen können, lässt sich leicht ausmalen. Die tatsächlichen Effekte solcher und vieler anderer zeitlich relativ stabiler Verhaltensdispositionen, die im Bereich der Persönlichkeit eines Mitarbeiters angesiedelt sind, werden noch vielfach unterschätzt.

Trotz der Leistungs*bezüglichkeit* vieler Persönlichkeitsmerkmale lassen sich diese aber leider kaum leistungs*mäßig* (also über maximale Leistung in irgendeinem Bereich) erfassen, sondern praktisch nur über Angaben des Probanden mit Bezug auf sein normalerweise übliches (d.h. ty-

pisches) Verhalten in dieser oder jener Situation. Von daher wird häufig beklagt, dass Testanden Persönlichkeitstests (im Gegensatz zu Leistungstests) leicht in eine als positiv angenommene Richtung verfälschen könnten, so dass die erhaltenen Informationen eigentlich invalide und damit unbrauchbar seien. Dem aber ist dreierlei entgegen zuhalten:

- Erstens kann eine beschönigende Selbstdarstellung, wenn man sie unkorrigiert zulässt, für Kontaktberufe sogar validitätserhöhend sein.
- Zweitens ist die Verfälschungswahrscheinlichkeit dann geringer, wenn Testanden mit einer grundsätzlichen Überprüfbarkeit ihrer Angaben in nachfolgenden Interviews, Assessment Centern o.ä. rechnen müssen.
- Drittens lassen sich, wenn man Normwerte aus echten Bewerberstichproben gewinnen kann, die Trennwerte für Ablehnungsentscheidungen bei Negativauswahl entsprechend adjustieren.

Weiterhin ist zu bedenken, dass die Forschung zur Verfälschbarkeit von Tests nach wie vor gut im Gange ist und weitere praxisdienliche Erkenntnisse liefert: Z.B. kann man bei durch Verfälschung gefährdeten Items (Frageninhalte und -formate) durch besondere Antworttechniken (z.B. durch Zwangswahl gleich attraktiv erscheinender Antwortalternativen – "forced choice" – als Antwortformat) derartigen Tendenzen entgegenwirken.

Aber nicht nur wegen der Unterschätzung der Bedeutung von mit Tests besonders gut messbaren Merkmalen für den Berufserfolg werden solche Verfahren noch zu selten herangezogen, sondern auch wegen einer mit Besorgtheit für betroffene Mitarbeiter bemäntelten ideologischen Ablehnung. Die wesentlichen Argumente dieser Meinungsmacher gegen Tests lauten:

- unberechtigtes Eindringen in die Privatsphäre des Kandidaten oder Mitarbeiters,
- Möglichkeit von Missbrauch bzw. Weitergabe erhobener Daten,
- erhaltene Persönlichkeitsbilder bleiben unvollständig,
- häufig wenig augenfälliger Bezug zum Arbeitsinhalt,
- Nichtberücksichtigung von situativen Besonderheiten der späteren Arbeit,
- Vortäuschen von Sicherheit, obwohl man im Einzelfall auch falsch liegen kann.

Doch sind derartige Argumente gegen Tests gleichermaßen zutreffend für die meisten anderen Verfahren der praktizierten Eignungsdiagnostik, man denke nur an die oft wenig professionelle Handhabung von Interviews. Allerdings ist ein weiteres Argument in besonderer Weise testspezifisch, nämlich das gegen Tests als quantitative Methode: In seiner allgemeinsten Form lautet es, der Mensch sei nicht messbar. Dies ist in der Tat zutreffend. Indes: Auch ein Stein ist nicht messbar, aber man kann Eigenschaften an ihm messen, z.B. sein Gewicht, seine Größe, seine Farbe. Entsprechend kann man berufsrelevante Eigenschaften auch bei einem Menschen messen, wenn die betreffenden Merkmale nur die notwendige Bedingung der Stabilität über die Zeit erfüllen, z.B. seine Integrität, seine Dominanz, seine Toleranz für Unbestimmtheit.

Die Grenzen, die jede dieser Messungen hinsichtlich der Genauigkeit und Wiederholbarkeit hat, werden aufgrund der anspruchsvollen teststatistischen Auswertungen bei Testverfahren deutlicher als z.B. bei der Durchsicht von Bewerbungsunterlagen, bei Interviews oder bei Verhaltensbeobachtungen (in Assessment Centern etwa). Dies erweckt dann allzu leicht den Eindruck, als hätten nur oder vor allem Testverfahren in diesem Bereichen Mängel, obwohl gerade sie den genannten anderen Verfahren hinsichtlich Objektivität, Messgenauigkeit und Validität in den meisten Einsatzbereichen überlegen sind. Die sorgfältige wissenschaftliche Analyse, die bei

Testverfahren selbstverständlich ist (oder, wenn man an manche nicht ganz so hervorragende Angebote denkt, doch selbstverständlich sein sollte), macht die von irrealen Idealvorstellungen abweichenden Messeigenschaften nur sichtbar, bei nicht standardisierten Instrumenten bleiben sie in den meisten Fällen verborgen und werden durch subjektive „Sicherheiten" ersetzt.

Selbst ein früher sehr wichtiges Argument gegen den praktischen Testeinsatz verliert zunehmend seine Gültigkeit: Um aussagekräftige Ergebnisse zu erzielen, muss bei Testverfahren in besonderem Maße die richtige und verständliche Instruktion, die Sicherstellung des Verständnisses der Testaufgaben, die genaue Erhebung der Antworten der getesteten Personen sowie deren Auswertung und Verrechnung gewährleistet sein. Während dies früher den Einsatz von psychologischem Fachpersonal erforderlich machte, wird heute die Objektivität und Fehlerfreiheit der Erhebung durch die Vorgabe und Auswertung von Testverfahren in computergestützter Form gesichert.

Allerdings haben Testverfahren naturgemäß auch ihre Begrenzungen. Das Interview ist den Testverfahren hinsichtlich seiner großen Bandbreite bei den erhobenen Informationen überlegen, Verhaltensübungen zeigen wesentlich besser das aktuelle Sprachvermögen, wichtige Facetten der Sozialen Kompetenz und auch die Sympathie bzw. „Ausstrahlung" eines Bewerbers, und oft sind die Fachkenntnisse aus dem Lebenslauf besser abzuschätzen als durch „Prüfungen". Nicht das missionarische Beharren auf *einer* Methode, sondern der für die jeweilige Fragestellung passend zusammengestellte Methodenmix („Multimethodalität") stellt den optimalen Ansatz in der Personalarbeit dar.

Insofern sollte man skeptisch sein gegenüber jedweder Einseitigkeit, wie beispielsweise für die Auswahl von Bewerbern nur Interviews zu empfehlen (was noch sehr häufig in der Praxis geschieht) oder bei Assessment Centern nur Tätigkeitssimulationen zuzulassen. Tests sollten in eignungsdiagnostischen Arrangements nie die einzige, vielmehr nur *eine* Verfahrensklasse neben anderen sein.

Zielsetzung des vorliegenden Handbuchs

Für die sinnvolle Gestaltung des Methodenmix leistungsfähiger Eignungsdiagnostik steht heute eine Vielzahl von Testverfahren zur Verfügung, die sich hinsichtlich ihrer speziellen Zielsetzung und auch hinsichtlich ihrer Qualität und Kosten sehr stark unterscheiden. Für den HR-Manager ist es daher notwendig, sich im Bedarfsfall schnell einen breiten Überblick über die verfügbaren Methoden zu verschaffen und die Passung eines bestimmten Testverfahrens für seine Fragestellung näher beurteilen zu können. Dazu benötigt er übersichtlich zusammengestellte Informationen. Genau dabei will das vorliegende Handbuch behilflich sein.

Die Vielzahl der allein im deutschen Sprachraum verfügbaren Testverfahren (weit über 1000) macht es notwendig, eine Auswahl für die Aufnahme in die hier vorliegende Übersicht vorzunehmen. Es wurden zunächst nur Verfahren ausgewählt, die einen eindeutigen Bezug zu Fragestellungen der Personalarbeit in der Wirtschaft haben, also keine „klinischen" oder nur im schulisch-pädagogischen Bereich einsetzbaren Instrumente. Darüber hinaus wurden bevorzugt jene aufgenommen, die einen gewissen Bekanntheits- und Verbreitungsgrad im deutschen Sprachraum haben oder erwarten lassen. Und schließlich konnten nur jene Verfahren berück-

sichtigt werden, deren Autoren bzw. Vertreiber sich bereit erklärten, Angaben zu allen in unserem Beschreibungsraster vorgesehenen Aspekten zu machen.

Trotz allem Bemühen um eine größtmögliche Vollständigkeit dieser Zusammenstellung mussten über die genannten Aspekte hinaus weitere Restriktionen vorgenommen werden, um die Übersichtlichkeit einer Darstellung in Buchform nicht zu gefährden. So wurde gerade von Anbietern mit einer größeren Anzahl von für die Wirtschaft brauchbaren Instrumenten nur eine kleinere, möglichst typische Auswahl aufgenommen. Auch war es manchmal notwendig, Testverfahren von der Aufnahme auszuschließen, wenn die Darstellung durch die Autoren zu „werbend" oder zu wenig seriös und plausibel hinsichtlich der Leistungsversprechen der Verfahren waren. Und manche der von uns angesprochenen Testvertreiber haben erst gar nicht geantwortet, meist wahrscheinlich aus gutem Grund.

Allerdings bedeutet die Nichtaufnahme eines Verfahrens nicht zwangsläufig, dass es sich um ein „schlechtes" Instrument handeln muss. In Anbetracht der Vielzahl der angebotenen Verfahren wurden von uns vermutlich auch manche gute Instrumente übersehen, vor allem dann, wenn sie erst seit kurzer Zeit verfügbar sind und damit noch nicht den Bekanntheitsgrad erreichen konnten wie länger etablierte Tests. Hinweise auf solche Lücken sind dringend erbeten, sie werden gerne bei der Erstellung der dritten Auflage dieses Handbuches berücksichtigt. Immerhin haben wir in der hier vorliegenden zweiten Auflage schon eine ganze Reihe weiterer Tests aufgenommen: mehrheitlich neue, in kleinerer Zahl aber auch schon länger auf dem Markt befindliche. Zu dieser letzteren Gruppe von Instrumenten zählen vor allem sog. Persönlichkeitstypentests, die einen gewissen Verbreitungsgrad erreicht haben. Auch wenn etliche dieser Verfahren psychometrisch nicht immer voll den Anforderungen der DIN 33430 entsprechen, haben wir sie trotzdem aufgenommen, wenn sie u.E. für Zwecke der Personalentwicklung nützlich sein können.

Eine weitere Neuerung dieser 2. Auflage besteht darin, dass wir das *Handbuch wirtschaftspsychologischer Testverfahren* nunmehr in zwei Bänden herausgeben. Der hier vorliegende *Band I: Personalpsychologische Instrumente* enthält die meisten Tests der 1. Auflage sowie die eben erwähnten neu hinzugekommenen; dies sind alles Tests, die sich auf die einzelne Person beziehen und deren Anwendungszwecke Personalauswahl, –beratung und –entwicklung sind. Der in ca. 15 Monaten folgende *Band II: Organisationspsychologische Instrumente* wird die (nur wenigen) organisationspsychologischen Tests der 1. Auflage enthalten, dann aber viele weitere Verfahren, die sich auf andere organisationale Entitäten als Einzelpersonen beziehen, d.h. auf Teams, Abteilungen, Personengruppen, Arbeitsbedingungen etc.

Um dem Ziel einer Entscheidungshilfe für HR-Manager gerecht zu werden, wurde in unserem Testbeschreibungsraster versucht, die für eine praxisdienliche Auswahl von Verfahren besonders relevanten Aspekte zu berücksichtigen. Die Angaben dazu stammen von den Verfassern der einzelnen Beiträge (meist Testautoren bzw. Testverleger/ -vertreiber), die allerdings durch Hin- und Rückfragen, manchmal mehrfach, optimiert wurden. Dennoch kann unsererseits keine Garantie für die Richtigkeit der Angaben übernommen werden.

Anordnung der Testverfahren und Suchmöglichkeiten

Um ein einfaches und schnelles Suchen von entscheidungsrelevanter Information eines dem Namen nach bekannten Testverfahrens zu ermöglichen, wurden alle Tests in alphabetischer Reihenfolge, und zwar gemäß den gebräuchlichen Testnamen in abgekürzter Form (z.B. IST, NEO-

FFI, MBTI), aufgeführt. Die Abkürzungen statt der vollen Testnamen wurden verwendet, weil sich gerade in informativen Darstellungen über Testverfahren in Zeitschriften, aber auch bei Anzeigen und anderen Informationsmaterialien die jeweilige Kurzform besonders eingebürgert hat.

Um dem Leser das Finden von eventuell geeigneten Testverfahren zu erleichtern, wurden darüber hinaus drei verschiedene Indexregister erstellt:

I. Im Register *„Getestete Merkmale“* sind jene Begriffe, die sich als durch den Test erfasste Merkmale in den Testbeschreibungen wiederfinden, alphabetisch gereiht und mit Hinweisen auf die jeweiligen Testverfahren verbunden.
II. Im Register *„Vorgesehene Einsatzbereiche“* sind die in den Testbeschreibungen enthaltenen Anwendungsgebiete aufgenommen (jeweils geordnet nach 1. Zweck, 2. Zielgruppen und 3. Inhalt) und ebenfalls mit den Hinweisen auf die Testverfahren versehen.
III. Im Register *„Inhaltsgruppen von Tests“* sind die Testkategorien nach der oben schon erwähnten groben Zweiteilung in Leistungs- vs. Persönlichkeitstests inhaltlich weiter untergliedert, aber darüber hinaus sind noch weitere gebräuchliche Inhaltskategorien hinzugekommen (z.B. spezielle Berufstests, 360-Grad-Beurteilungen) – beides wiederum mit Hinweisen auf die entsprechenden Verfahren.

Leider ist festzustellen, dass die Vielzahl der eingebürgerten Begriffe sowohl für Anwendungsbereiche als auch für zu erfassende Merkmale eine befriedigende Suchstrategie erschwert. Auch sind gerade bei Testverfahren mit einem breiten Anwendungsspektrum keineswegs alle möglichen Einsatzbereiche in der Beschreibung genannt, dies trifft insbesondere für spezielle Berufsbezeichnungen zu. Man sollte daher zum Finden der geeigneten Verfahren auch „Oberbegriffe“ berücksichtigen.

Psychologische Testverfahren sind normalerweise nicht für einzelne Berufe speziell, sondern durchaus genereller für ein breites Spektrum von Berufen aussagefähig. Die entscheidenden Unterschiede liegen weniger in der Anwendbarkeit des betreffenden Verfahrens als in den für den jeweiligen Beruf erforderlichen spezifischen Ausprägungsgraden der erfassten Merkmale. So sind etwa Merkmale wie „räumliches Vorstellungsvermögen“, „verbale Denkfähigkeit“ oder „Konfliktbereitschaft“ in vielen beruflichen Positionen interessant, das optimale Profil eines guten Hüttenfacharbeiters auf diesen Dimensionen sieht aber sicher anders aus als z.B. das eines erfolgreichen Mitarbeiters im Vertrieb beratungsintensiver technischer Produkte.

Wir hoffen nun, dass die hier zusammengestellten Informationen geeignet sind, den für das Recruiting und die Personalentwicklung Verantwortlichen nützliche Entscheidungshilfen zu bieten und damit den sachgerechten Einsatz personalpsychologischer Testverfahren innerhalb des breiten Spektrums eignungsdiagnostischer Methoden der Personalarbeit zu fördern.

Hinweise, Anregungen und Verbesserungsvorschläge für die nächste Auflage sind uns selbstverständlich immer willkommen.

Werner Sarges und Heinrich Wottawa

Inhalt

KAPITEL I: Tests

16PA
16-Persönlichkeits-Adjektiv-Skalen
Hermann Brandstätter 3

16PF-R
Der 16 Persönlichkeits-Faktoren-Test − Revidierte Fassung
Klaus A. Schneewind, Johanna Graf 9

A3DW
Adaptiver Dreidimensionaler Würfeltest
Georg Gittler 17

ABAT-R
Revidierter Allgemeiner Büroarbeitstest
Gustav A. Lienert, Heinz Schuler 23

ABLE – BDA
Business Decision Analysis
Psychometric Research & Development (PRD) 29

ABLE – FA
Financial Appraisal
Psychometric Research & Development (PRD) 35

ACIL
Adaptive Computergestützte Intelligenz-Lerntestbatterie
Jürgen Guthke, Jens F. Beckmann, Henrik Stein, Silke Rittner, Hans Vahle 33

AHA
Arbeitshaltung
Klaus D. Kubinger, Johannes Ebenhöh 47

AIST-R
Allgemeiner Interessen-Struktur-Test
Christian Bergmann, Ferdinand Eder 53

ALS
Arbeitsleistungsserie
Gernot Schuhfried 59

AMT
Adaptiver Matrizen Test
Lutz F. Hornke, Stefan Etzel, Klaus Rettig 65

APM
Advanced Progressive Matrices
Kurt A. Heller, Heinrich Kratzmeier, Angelika Lengfelder 71

ASK
Analyse Schlussfolgernden und Kreativen Denkens
Heinz Schuler, Benedikt Hell 77

ASO
Ability Screening Online
SHL Group 81

ATLANTIS
René Hirsig und Astrid E. de With 89

AVEM
Arbeitsbezogenes Verhaltens- und Erlebensmuster
Uwe Schaarschmidt, Andreas W. Fischer 93

AZUBI-BK
Arbeitsprobe zur Berufsbezogenen Intelligenz –
Büro- und kaufmännische Tätigkeiten
Heinz Schuler, Yvonne Klingner 99

B5PO
Big Five Plus One – Persönlichkeitsinventar
Stefana Holocher-Ertl, Klaus D. Kubinger, Stefan Menghin 105

Benchmarks®
Michael Lombardo und Cynthia McCauley, Stefan Oppitz und Thomas Lorenz 109

BET
Berufseignungstest
Hugo Schmale, Heinz Schmidtke 115

BFI
Big Five Inventory
Oliver P. John, Beatrice Rammstedt 119

BIM
Besorgnis-Inventar Management
Joachim Stöber, Beate Seidenstücker 123

BIMO
Bochumer Inventar zu Mitarbeiterzufriedenheit und Organisationsklima
Rüdiger Hossiep, Karin Godovsky 127

BIP
Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung
Rüdiger Hossiep, Michael Paschen 131

BIS-4
Berliner Intelligenzstruktur-Test − Form 4
Adolf Otto Jäger, Heinz-Martin Süß, André Beauducel 139

BIS-r-DGP
(sowie die Kurzformen A-1 und K-1)
Berliner Intelligenzstruktur-Test der Deutschen Gesellschaft für Personalwesen e.V.
Martin Kersting, André Beauducel 149

Bits&Bytes
Stefan Etzel, Anja Etzel 159

BOMAT–A
Bochumer Matrizentest – advanced – short version
Rüdiger Hossiep, Daniela Turck, Michele Hasella 167

BOMAT–S
Bochumer Matrizentest – standard version–
Rüdiger Hossiep, Michele Hasella 173

BOT
Berufsorientierungs-Test
Robert T. Hogan, Joyce C. Hogan, Ansfried B. Weinert 179

BPI
Bambeck-Persönlichkeits-Instrument
Jörn J. Bambeck 185

BPM
Bonner Postkorb-Module
Jochen Musch, Brigitte Rahn, Walter Lieberei 191

CALL ME
Christian Montel, Veronika von Rüden 197

CAPTain
Computer Aided Personnel Test Answers Inevitable
Ulla Petäjävaara, Christoph Nagler 203

CAS
CAE-Aircrew-Selection-System (CAE-Piloten-Selektionssystem)
Peter Braun, GAPFmbH.....209

CASA
Computer Aided Single Assessment
Christoph Schneider, Wolfgang Müskens.....217

C-A-T
CallCenter Aptitude Test (Version 2.0)
Udo Konradt, Guido Hertel, Karin Joder.....225

CAT 4
Computer-Assistiertes Testen – 4. Generation
Psychologischer Dienst der Bundeswehr.....229

CCSQ
Customer Contact Styles Questionnaire
SHL Group plc.....237

Choices Architect®
Talent Management Tool
Michael M. Lombardo, Robert W. Eichinger.....245

COG
Cognitrone
Gernot Schuhfried.....251

CPI
California Psychological Inventory („Revidierter Deutscher CPI")
Ansfried B. Weinert, Harrison G. Gough.....257

d2
Aufmerksamkeits-Belastungs-Test
Rolf Brickenkamp.....265

DAC
Dynamischer Air Traffic Control Test
Hinnerk Eißfeldt.....271

DAT
Daueraufmerksamkeitstest
Hanspeter Dvorak.....275

DAUF
Daueraufmerksamkeit
Gernot Schuhfried.....281

DCT
Dyadic Co-operation Test
Dirk Stelling 285

DEKRA Planspiel
Der Manager im Industriebetrieb
DEKRA Akademie 289

DISG-Persönlichkeitsprofil®
John Geier 293

DISKO
Management Planspiel
Uwe Funke 299

EPPD
Eysenck Personality Profiler – Deutsche Form
Hans J. Eysenck 305

Explorix®
deutschsprachiger Self-Directed Search
John Holland, Simone Jörin, François Stoll, Christian Bergmann, Ferdinand Eder 309

FAF2002
Fragebogen zur Auswahl von Führungskräften
Matthias Schubert, Klaus Podufal 317

FAIR
Frankfurter Aufmerksamkeits-Inventar
Helfried Moosbrugger, Jens Oehlschlägel 325

FAKT-II
Frankfurter Adaptiver Konzentrationsleistungs-Test II
Helfried Moosbrugger, Frank Goldhammer 331

F-DUP
Fragebogen zur Diagnose unternehmerischer Potentiale
A.S. King, Günter F. Müller 337

F-I-T
Foto-Interessen-Test
François Stoll, Daniel Jungo 343

FIT
Fragebogen zur Erfassung individueller Einstellungen zur Teamarbeit
Changiz Mohiyeddini 349

Founders Check
Thomas Lang-von Wins 353

FSF
Führungsstilfragebogen
Niclas Schaper, Walter Lieberei 359

FSK
Fragebogenn zur sozialen Kompetenz
Heinz Schuler, Uwe Funke 369

FSYS 2.0
Dietrich Wagener,Wolfgang Conrad 375

FVVB
Fragebogen zur Vorgesetzten-Verhaltens-Beschreibung
Heide Fittkau-Garthe, Bernd Fittkau 379

GPI
Global Personality Inventory
Mark J. Schmit, Jenifer A. Kihm et al. 383

GPOP
Golden Profiler of Personality
John Golden, Richard Bents und Reiner Blank 387

H.D.I.®
Herrmann Dominanz Instrument bzw. Herrmann Brain Dominance Instrument (HBDI)
Ned Herrmann 393

HPI
Hamburger Persönlichkeits-Inventar
Burghard Andresen 397

ILICA
Doris Mösender, Johannes Ebenhöh 403

INKA
Inventar Komplexer Aufmerksamkeit
Gerd Heyde 409

INSIGHTS
Insights Discovery Profil
Andi Lothian, Andrew Lothian 413

IPS
Inventar zur Persönlichkeitsdiagnostik in Situationen
Uwe Schaarschmidt, Andreas W. Fischer 421

ISA
Intelligenz-Struktur-Analyse
Institut für Test- und Begabungsforschung und Georg Gittler 427

ISIS
Interaktives System zur Identifikation Sozialer Kompetenzen
Bernd Runde, Frauke Bastians, Sandra Kluge, Klaus Wübbelmann 433

I-S-T 2000 R
Intelligenz-Struktur-Test 2000 R
Rudolf Amthauer, Burkhard Brocke, Detlev Liepmann, André Beauducel 439

Jobfidence®
Norbert Gantner, Georg Sieber 447

KLT-R
Revidierter Konzentrations-Leistungs-Test
Heinrich Düker, Gustav A. Lienert, Helmut Lukesch, Susanne Mayrhofer 453

KMF
Kienbaum Management Fragebogen
Kienbaum Management Consultants 459

Leadership Mirror 3.0
DDI Development Dimensions International 463

LIFO®-Methode
Stuart Atkins und Allan Katcher, René Bergermaier 471

Linguaskill
Manpower Inc. Europe 477

LMI
Leistungsmotivationsinventar
Heinz Schuler, Michael Prochaska 481

LVT
Linienverfolgungstest
Bernhard Biehl 487

MANAGE!
Heinz W. Kreuzig, Peter Enzenberger 493

MarP
Die Managementarbeitsprobe
Stefan Etzel, Anja Etzel 497

MAST
Management-Systemtest
Theo Wehner, Katharina Lehmann, Silvia Deplazes 505

Mathematiktest
Grundkenntnisse für Lehre und Beruf
Nedina Ibrahimovic, Stephan Bulheller, Hartmut O. Häcker 513

MBO
Münchner Erfassung von Berufsorientierungen
Lutz von Rosenstiel, Herbert Einsiedler, Thomas Lang-von Wins, Jürgen Kaschube, Günter Maier, Angela Wittmann 517

MBTI
Myers-Briggs-Typen-Indikator
Katherine C. Briggs und Isabel Briggs Myers, Stefan Oppitz und Thomas Lorenz 525

MDF-360°
Multidirektionales Feedback
Georg P. Fennekels 531

META
Erkennen von Meta-Regeln
Jutta Gatternig, Klaus D. Kubinger 537

MFA
Management-Fallstudien
Georg P. Fennekels, Simone D´Souza 541

MGIB
Management Graduate Item Bank
SHL Group 547

MMG
Multi-Motiv-Gitter
Heinz-Dieter Schmalt, Kurt Sokolowski, Thomas Langens 553

MSM
Motive Superiority Measure
Jan Eichstaedt, David Scheffer 559

MTA
Mechanisch-Technisches Auffassungsvermögen
Konrad Liedl 565

NEO-FFI
NEO-Fünf-Faktoren-Inventar
Paul T. Costa und Robert R. McCrae, Peter Borkenau und Fritz Ostendorf 569

NEO-PI-R
NEO-Persönlichkeits-Inventar – revidierte Form
Paul T. Costa und Robert R. McCrae, Fritz Ostendorf und Alois Angleitner 575

OLMT
Objektiver Leistungsmotivations-Test
Lothar Schmidt-Atzert, Markus Sommer, Markus Bühner, Astrid Jurecka 583

OMT
Operanter Motiv-Test
Julius Kuhl, David Scheffer, Jan Eichstaedt 591

OPQ32
Occupational Personality Questionnaire - 32 Scales
SHL Group 597

OSM
Obermann + Schiel Motivationsfragebogen
Christof Obermann 603

OSP
Obermann + Schiel Persönlichkeitsprofil
Christof Obermann 607

PAM
Potenzialanalyse für Management- und Fachfunktionen
Ingrid Stratemann, Hans-Georg Sonnenberg, Heinrich Wottawa 611

PAPI™
Personality and Preference Inventory
Max M. Kostick, PA Consulting Group 619

PC-OFFICE
Georg P. Fennekels 627

PC-Postkorb
Scharley & Partner 631

PI
The Predictive Index® System
Arnold Daniels 637

PIA
Persönlichkeitsinventar zur Integritätsabschätzung
S & F Personalpsychologie Managementberatung 643

PMC
Perspectives on Management Competencies –
Managemententwicklung mit 360°-Feedback
SHL Group 649

POTENTIAL
Jean-Michel Fourgous 655

PRF
Deutsche Personality Research Form
Heinrich Stumpf, Alois Angleitner, Theresia Wieck,
Douglas N. Jackson, Heidrun Beloch-Till 661

pro facts 360°-Assessment
Stefan Etzel, Anja Etzel 667

PSSI
Persönlichkeits-Stil-und-Störungs-Inventar
Julius Kuhl, Miguel Kazén 677

!Response
360°-Feedback
Martin Scherm 683

RIS
Rechnen in Symbolen
Christoph Schmotzer, Klaus D. Kubinger, Christian Maryschka 691

RST
Rechtschreibtest - Neue Rechtschreibregelung
Stephan Bulheller, Hartmut O. Häcker 695

RT
Rechtschreibungstest
Martin Kersting, Klaus Althoff 699

SAM
Fragebogen zur Erfassung dispositionaler Selbstaufmerksamkeit
Sigrun-Heide Filipp, Elke Freudenberg 705

SCG Selbstprofil
Mechtild Laumen-Schiel 711

shapes
adaptive Messung berufsbezogener Kompetenzen
Andreas Lohff, Achim Preuß 717

SIGNAL DETECTION
Gernot Schuhfried 723

SIMKAP
Simultankapazität/Multi-Tasking
Oswald Bratfisch, Eva Hagman 727

SKASUK
Skalen zur Service- und Kundenorientierung
Hans-Georg Sonnenberg 731

SKI
Selbstkonzept-Inventar
Richard von Georgi, Dieter Beckmann 735

SMK
Sensomotorische Koordination
Herbert Bauer, Giselher Guttmann, Michael Leodolter, Ulrich Leodolter 741

SPMPLS
Raven's Standard Progressive Matrices Plus
John C. Raven 745

Sprinter AG
e-Mail-Postkorb
Walter Lieberei 749

STRATOS-00
René Hirsig, Astrid E. de With 753

SYNPRO-FAI
Führungs-Analyse-Instrument
Patricia Simon, Andreas Donaubauer 759

SYNTEX
Computersimuliertes Unternehmensplanspiel
Patricia Simon, Alexander Thomas 765

TAVTMB
Tachistokopischer Verkehrsauffassungstest Mannheim –Bildschirmvorgabe
Bernd Biehl 771

TMPF
Team Management Profil Fragebogen
Charles Margerison, Dick McCann 777

TOM
Test operationeller Mehrfacharbeit (Test of Multiple Task Performance)
Deutsches Zentrum für Luft- und Raumfahrt 783

UTOPIA
PC-Planspiel
Scharley & Partner, Konstanz 787

VERGED
Verbaler Gedächtnistest
Stefan Etzel, Lutz F. Hornke 793

views
adaptive Messung berufsbezogener Wertvorstellungen
Andreas Lohff, Achim Preuß 801

ViQ
Visual Questionnaire
Dieter Loerwald, David Scheffer, Alexander Ribowski 807

VISGED
Visueller Gedächtnistest
Stefan Etzel, Lutz F. Hornke 815

VISION
Videobasierte Identifikation Sozialer Intelligenz – Online
Bernd Runde, Stefan Etzel 823

WBT
Wirtschaftskundlicher Bildungs-Test
Klaus Beck,Volker Krumm, Rolf Dubs 831

WII
Werte- und Interessen-Inventar
Joyce C. Hogan , Robert T. Hogan, Ansfried B. Weinert 837

WPBF
The Workplace Big Five®
CentACS, PiMedia 843

KAPITEL II: Testsysteme

ELIGO-System 851
Hogrefe TestSystem 865
Jobmatcher 873
PERLS 879
pro facts-Testsystem 893
Wiener Testsystem 905

REGISTER

REGISTER I: Getestete Merkmale 917
REGISTER II: Vorgesehene Einsatzbereiche
1. geordnet nach Zweck 943
2. geordnet nach Zielgruppe 951
3. geordnet nach Inhalt 959
REGISTER III: Inhaltsgruppen von Tests 969

KAPITEL I

Tests

16PA

16-Persönlichkeits-Adjektiv-Skalen

Autor

Hermann Brandstätter

Verlag, Erscheinungsort und -jahr

Bis zur Veröffentlichung sind Fragebogen, Interpretationshinweise, Normen und Auswertungsprogramme vom Autor zu beziehen.

Einsatzbereiche

Beratung (insbesondere Studien- und Berufsberatung), Personalauswahl, Personalbeurteilung, Marktsegmentierung nach Persönlichkeitstypen, Forschung. Ab 16 Jahren in allen Bildungsstufen zur Selbstbeurteilung und zur Beurteilung durch andere verwendbar.

Konstruktionsgrundlagen

Die Cattellschen Globaldimensionen (Deutsche Version von K.A. Schneewind, G. Schröder und R.B. Cattell, 1983) sollen mit den 16-Persönlichkeits-Adjektiv-Skalen möglichst genau geschätzt werden, wenn nicht genügend Zeit für einen Einsatz des 16PF oder eines anderen umfangreichen Persönlichkeitsfragebogens zur Verfügung steht. Ein individueller Konsistenzkoeffizient (Korrelation der Profile von Form A und Form B) soll eine personspezifische Einschätzung der Vertrauenswürdigkeit der Daten ermöglichen.

Es wird angenommen, dass Menschen ein im allgemeinen treffendes Bild von der eigenen Persönlichkeit aus den Erfahrungen mit sich selbst in den verschiedensten Lebenssituationen und aus den Reaktionen ihrer Interaktionspartner auf ihr Verhalten gewinnen und dies auch mitteilen können. Dabei ist zu beachten, dass die Mitteilung immer ein Kompromiss zwischen privatem Eindruck und Opportunität der Aussage ist. Daraus folgt, dass die Interpretation der 16PA-Selbstbeurteilungen, nicht anders als die Antworten in traditionellen Persönlichkeitsfragebogen, vom Kontext der Erhebungssituation (anonyme Forschungssituation, Beratungssituation, Selektionssituation) abhängig ist und auf kontextspezifische Vergleichsnormen und Validitätsstudien angewiesen ist.

Inhalte

Erfasst werden die 16 Primärdimensionen mit je zwei Eigenschaftspolaritäten. Diese Dimensionen heißen in der revidierten Fassung des 16PF:

A	Wärme
B	Logisches Schlussfolgern
C	Emotionale Stabilität
E	Dominanz
F	Lebhaftigkeit
G	Regelbewusstsein
H	Soziale Kompetenz
I	Empfindsamkeit
L	Wachsamkeit
M	Abgehobenheit
N	Privatheit
O	Besorgtheit
Q1	Offenheit für Veränderung
Q2	Selbstgenügsamkeit
Q3	Perfektionismus
Q4	Anspannung

Regressionsanalytisch werden die fünf Globaldimensionen (Normgebundenheit, Belastbarkeit, Unabhängigkeit, Entschlussbereitschaft, Kontaktbereitschaft in der 16PF-Version von 1983 bzw. Extraversion, Unabhängigkeit, Ängstlichkeit, Selbstkontrolle, Unabhängigkeit in der revidierten Fassung des 16 PF-R) geschätzt. Die fünf 16PA-Globalfaktoren lassen sich wie folgt kurz charakterisieren, wobei jeweils die starke Ausprägung (ein hoher Skalenwert) erläutert wird:

Normgebundenheit

Orientiert sich an den Forderungen des eigenen Gewissens und/oder an den sozialen Regeln und Verpflichtungen. Selbstdisziplin und Beständigkeit.

Belastbarkeit

Lässt sich nicht leicht aus der Fassung bringen. Stabiles Selbstvertrauen.

Unabhängigkeit

Geht eigenen Weg, behauptet sich und setzt sich durch; kritisch gegenüber anderen.

Entschlussbereitschaft

Auf nüchterne Zweckmäßigkeit eingestellt, verstandesbetont, handlungsorientiert.

Kontaktbereitschaft

Sozial aufgeschlossen, sucht die Nähe von anderen Menschen, achtet auf soziale Beziehungen.

Frage- und Antwort-Formate

Bipolare Eigenschaftsliste mit einer neun-stufigen Skala der Form 4 3 2 1 0 1 2 3 4.

Subtests

Es macht wenig Sinn, den Gesamttest in Subtests aufzuspalten.

Gütekriterien

Reliabilität

Retest-Reliabilität der Primärskalen (Zeitabstand drei Monate) r = .72 (Median), der Globalskalen r = .73 (Median) in einer Stichprobe von 113 Teilnehmern und Teilnehmerinnen an Studienberatungstests.

Retest-Reliabilität der Primärskalen (Zeitabstand durchschnittlich 4 Jahre) r = .48 (Median), der Globalskalen (Zeitabstand durchschnittlich 4 Jahre) r = .51 (Median) in einer Stichprobe von N = 1.340 Teilnehmerinnen und Teilnehmern an Studienberatungstests.

Validität

1. Korrelation zwischen 16PA- und 16PF-Primärfaktoren $r_{(228)}$ = .41 (Median). Korrelation zwischen 16PA- und 16PF-Globaldimensionen $r_{(228)}$ = .61 (Median) (Brandstätter, 1988).
2. Multiple Korrelation (N = 279) zwischen den fünf Dimensionen des NEO-FFI (in diesem Band) als abhängige Variablen und den 16PA-Primärdimensionen als unabhängige Variablen R = .67 (Median) (Studierende verschiedener Fakultäten als Teilnehmer/innen an sozialpsychologischen Experimenten). Auf der Ebene der Globalfaktoren korreliert Normgebundenheit mit Offenheit für Erfahrung und Gewissenhaftigkeit ($r_{(279)}$= –.30 bzw. $r_{(279)}$=.29), Belastbarkeit mit Neurotizismus und Gewissenhaftigkeit (–.58 bzw. .29), Unabhängigkeit mit Extraversion, Neurotizismus und Verträglichkeit (.44, –.32, –.33), Entscheidungsbereitschaft mit Neurotizismus und Offenheit für Erfahrung (–.34, –.21), Kontaktbereitschaft mit Extraversion, Verträglichkeit und Gewissenhaftigkeit (.44, .21, –.22). In einer gemeinsamen Hauptkomponentenanalyse der fünf PA-Globalfaktoren und der fünf NEO-FFI-Faktoren bilden Neurotizismus (umgepolt) gemeinsam mit Belastbarkeit und Entschlussbereitschaft den ersten Faktor, Extraversion mit Kontaktbereitschaft und Unabhängigkeit den zweiten Faktor, Offenheit für Erfahrung (umgepolt) mit Normgebundenheit den dritten Faktor, Gewissenhaftigkeit mit Normgebundenheit den vierten Faktor und Verträglichkeit (umgepolt) mit Unabhängigkeit den fünften Faktor.
3. Übereinstimmung der Selbstbeurteilung mit durchschnittlicher Beurteilung durch 4 nach Zufall ausgewählte Mitschüler/innen $r_{(120)}$=.38 (Median bei geringer individueller Konsistenz) und $r_{(123)}$=.50 (Median bei hoher individueller Konsistenz) (Brandstätter, Filipp & Drescher, 1992).
4. Übereinstimmung zwischen zwei Fremdbeurteilern auf den Globaldimensionen $r_{(243)}$ = .41 (Median) (Brandstätter, Filipp & Drescher, 1992).
5. 120 von 183 (66 %) der Unternehmer/innen des Persönlichkeitsclusters „wenig normgebunden, belastbar, unabhängig“ gehören zu den Unternehmensgründern, 63 von 183 (34 %) dieses Clusters gehören zu den Unternehmenserben; im Persönlichkeitscluster „normgebunden, wenig belastbar, wenig unabhängig“ finden sich nur 39 von 87 (45 %) Unternehmensgründer/innen, dagegen 48 von 87 (55 %) Unternehmenserben (Brandstätter, 1993).
6. In einer repräsentativen holländischen Stichprobe von 48 selbständigen (self-employed) Männern ist das durchschnittliche Nettoeinkommen (in z-Werten dargestellt) bei hoher Be-

lastbarkeit ($z_{[Belastbarkeit]}$=+1.00) und geringer Extraversion ($z_{[Extraversion]}$ = –1.00) $z_{[Nettoeinkommen]}$ = .68, bei geringer Belastbarkeit ($z_{[Belastbarkeit]}$ = –1.00) und geringer Extraversion ($z_{[Extraversion]}$ = –1.00) $z_{[Nettoeinkommen]}$ = –.68. Bei hoher Belastbarkeit und hoher Extraversion (jeweils z = +1.00) liegt das Einkommen im Durchschnitt bei z = .24, bei geringer Belastbarkeit und hoher Extraversion ist das durchschnittliche Einkommen z = .16 (Paneldaten des Jahres 1993 aus dem VSB-CentER Savings Project, Tilburg University).

7. Bei geringer Belastbarkeit (mediangeteilt) und geringer Kontaktbereitschaft (geringer Extraversion) berichten Teilnehmer/innen an Befindensstudien (Zeitstichproben im Verlauf von 4 Wochen) positive Gefühle gleich häufig wie negative, dies sowohl in der Freizeit außer Haus wie bei der Arbeit außer Haus. Bei hoher Belastbarkeit und hoher Kontaktbereitschaft ist das Verhältnis der Häufigkeiten positiver Gefühle zur Häufigkeit negativer Gefühle in der Freizeit 1.73, in der Arbeit 0.80. Zwischen diesen Extremgruppen liegen die beiden übrigen Konfigurationen (wenig belastbare Extravertierte und belastbare Introvertierte), wobei die wenig belastbaren Extravertierten mit einer Verhältniszahl von 0.62 die ungünstigste Gefühlsbilanz bei der Arbeit haben. Jeder der hier berichteten Werte basiert auf mindestens 1.000 Momentaufnahmen des Befindens in Alltagssituationen (Brandstätter, 2001).
8. Bei Maturanten (Teilnehmer/innen an Studienberatungstests) korreliert Kontaktbereitschaft (Extraversion) negativ mit Schulleistungen (r = –.23). In dieser Gruppe finden sich auch die theoretisch erwarteten Korrelationen mit den Hollandschen Interessenskalen (in der AIST-Version) zwischen Normgebundenheit und Interessen für administrative (conventional) Interessen (r = .24), zwischen Unabhängigkeit und unternehmerischen Interessen (r = .39), zwischen Entschlussbereitschaft (besser: tough-mindedness) auf der einen und künstlerischen sowie sozialen Interessen auf der anderen Seite (r = –.29 bzw. –.23), schließlich zwischen Kontaktbereitschaft (Extraversion) und wissenschaftlichen Interessen (r = –.26).
9. In einer Längsschnittstudie aus dem Jahre 1997 (unveröffentlicht) mit 66 unverheirateten Paaren korreliert die Dauer der Beziehung mit Normgebundenheit (r = .27 bei Frauen und r = .16 bei Männern).
10. In einer Reihe von Experimenten zum Verhalten in Gruppen (Ultimatumspiele und Assessment Center-Simulationen) konnten mit den 16PA-Globalskalen Hypothesen zu individuellen Differenzen in den Reaktionen auf Bedingungsvariationen geprüft und bestätigt werden (z.B. Brandstätter & Farthofer, 1997; Brandstätter & Güth, 2000, 2002).

Normen

Der 16PA wurde 1996 von einem Marktforschungsinstitut in einer repräsentativen Stichprobe der Bundesrepublik Deutschland (N = 1.089; Alter 15 bis 85) geeicht. Es liegen nach Altersgruppen und Geschlecht differenzierte Normen vor, die selbstverständlich nur für Anwendungssituationen gültig sind, die der Situation der Testeichung (der Fragebogen wurde von den Interviewten schriftlich beantwortet und anonym dem Interviewer übergeben) hinreichend ähnlich sind. Es gibt außerdem Normen (Daten in den letzten 10 Jahren erhoben) für Studienberatung von Maturanten (Abiturienten), Normen für Verwendung des Fragebogens in Bewerbungssituationen (Bewerber/innen des gehobenen und höheren Verwaltungsdienstes) und Normen für Bewerber/innen um Studienplätze an Fachhochschulen.

Bearbeitungsbedingungen und -dauer

Gruppen- und Einzeltestung möglich. Testdauer im Durchschnitt 7 Minuten. Der Fragebogen ist zwar durch Nicht-Psychologen administrierbar, die Verantwortung für den Einsatz des Fragebogens und die Schlüsse, die aus den Ergebnissen gezogen werden, muss jedoch professionellen Psychologen vorbehalten bleiben.

Ergebnisrückmeldung

Die Ergebnisrückmeldung erfolgt computerunterstützt in Profilform mit schriftlichen Erläuterungen.

Publikationen

Brandstätter, H. (1988). Sechzehn Persönlichkeits-Adjektivskalen (16PA) als Forschungsinstrument anstelle des 16PF. *Zeitschrift für experimentelle und angewandte Psychologie*, *35*, 370-391.

Brandstätter, H. (2001). The time-sampling diary: An ecological approach to the study of emotions in everyday life situations. In H. Brandstätter & A. Eliasz (Eds*.), Persons, situations, and emotions. An ecological approach* (pp. 20-52). New York: Oxford University Press.

Brandstätter, H. & Farthofer, A. (1997). Personality in social influence across tasks and groups. *Small Group Research, 28*, 146-163.

Brandstätter, H., Filipp, G. & Drescher, P. (1992). Zur Validität der 16 Persönlichkeits-Adjektivskalen (16PA). *Zeitschrift für Differentielle und Diagnostische Psychologie, 13,* 53-67.

Brandstätter, H. & Güth, W. (2002). Personality in dictator and ultimatum games. *Central European Journal of Operations Research, 10,* 191-215.

Schneewind, K.A., Schröder, G. & Cattell, R.B. (1983). *Der 16-Persönlichkeits-Faktoren-Test (16PF)*. Bern: Huber.

Erfahrungen in der Wirtschaftspraxis

Bewährung des Verfahrens in

- Verlaufsstudien zum Erleben und Verhalten am neuen Arbeitsplatz
- Studien zur Arbeitszufriedenheit in der theoretischen Perspektive von Person-Umwelt-Korrespondenz
- Studien zur Evaluation eines Auswahlverfahrens für Schuldirektoren
- Beratung von Unternehmensgründern
- Marktsegmentierung (im Bereich von Sportgeräten und Versicherungen) nach Persönlichkeitstypen

Sonstiges

Der 16PA (Copyright beim Autor) sollte in der Erprobungsphase – vor seiner Veröffentlichung als Test – nur nach Rücksprache und im Einvernehmen mit dem Autor eingesetzt werden.

Preise

Für Forschungszwecke kann der 16PA nach Rücksprache mit dem Autor kostenlos eingesetzt werden, wenn die mit dem Verfahren erhobenen Daten dem Testautor zur Ergänzung der Datenbasis nach Abschluss der Studie zur Verfügung gestellt werden. Voraussetzung einer kommerziellen Verwendung ist eine Lizenzvereinbarung mit dem Autor.

Bezugsadresse

Prof. em. Dr. Hermann Brandstätter
Auerweg 25
A – 4203 Altenberg
Fon: +43 (0)7230/8449
Email: h.brandstaetter@jk.uni-linz.ac.at

(Verfasser des Beitrags: Hermann Brandstätter)

16PF-R

Der 16 Persönlichkeits-Faktoren-Test – Revidierte Fassung

Autoren

Klaus A. Schneewind und Johanna Graf

Verlag, Erscheinungsort und -jahr

Huber-Verlag, Bern, 1998

Einsatzbereiche

Erwachsene ab 18 Jahren. Verwendung in der Arbeits-, Betriebs- und Organisationspsychologie, z.B. zur Eignungsdiagnostik (Auswahl, Platzierung, Potentialanalyse), zur Berufs- und Karriereberatung, zu Training und Coaching. Zudem in der Pädagogischen und Klinischen Psychologie (z.B. zur Therapieplanung) sowie in der Forschung.

Konstruktionsgrundlagen

Der 16PF-R ist ein objektiver Fragebogentest zur mehrdimensionalen Persönlichkeitsdiagnostik im Erwachsenenalter. Im anglo-amerikanischen Sprachraum ist dieses von R.B. Cattell entwickelte Verfahren seit mehr als 45 Jahren ein bewährtes und weit verbreitetes Instrument zur Erfassung der Persönlichkeitsstruktur. Die vorliegende deutschsprachige Fassung basiert auf der neuesten englischsprachigen Ausgabe des 16PF (4. amerikanische Revision, 1994). Primäres Anliegen bei der Testkonstruktion war es, ein umfassendes, psychometrisch abgesichertes, ökonomisches und für die psychologische Praxis brauchbares Persönlichkeitsinventar zu entwikkeln. Die deutsche Vorläuferversion (16PF, Schneewind, Schröder & Cattell, 3. korrigierte Auflage, 1994) ist in Forschung und Anwendungspraxis weit verbreitet. Ziele der Revision waren die Formulierung zeitgemäßer und geschlechtsinsensitiver Items, eine separate Erfassung des logischen Schlussfolgerns, die Einbeziehung von Skalen zur Überprüfung von Antworttendenzen, eine vereinfachte Handhabbarkeit der Primärdimensionen durch leicht verständliche unipolare Bezeichnungen (z.B. „Wärme"), die Normierung an einer neueren, hinsichtlich Geschlecht, Alter, Bildung, Region, Bundesland und Gemeindegröße repräsentativen Eichstichprobe sowie die Gewährleistung von faktorieller und Konstruktvalidität. Die Übereinstimmung zwischen 16PF und 16PF-R liegt durchschnittlich bei r = .67.

Inhalte

Der Fragebogen erfasst 16 Primärdimensionen und (daraus abgeleitet) 5 Globalfaktoren der Erwachsenenpersönlichkeit (vgl. das Big-Five-Modell der Persönlichkeit nach Costa & McCrae).

Der Test erlaubt somit auf unterschiedlichen Analyseebenen eine differenzierte Erfassung des Persönlichkeitsprofils.

Primärfaktoren

A: Wärme, B: Logisches Schlussfolgern, C: Emotionale Stabilität, E: Dominanz, F: Lebhaftigkeit, G: Regelbewusstsein, H: Soziale Kompetenz, I: Empfindsamkeit, L: Wachsamkeit, M: Abgehobenheit, N: Privatheit, O: Besorgtheit, Q1: Offenheit für Veränderung, Q2: Selbstgenügsamkeit, Q3: Perfektionismus, Q4: Anspannung (detaillierter in der nachfolgenden Übersicht).

Wärme (A+)	**Reserviertheit (A-)**
- warmherzig - aufmerksam für die Gefühle anderer - interessiert am Wohlergehen anderer gerne in der Nähe anderer	- reserviert - unpersönlich - distanziert - vorsichtig beim Eingehen von Beziehungen - gerne allein

Hohes Logisches Schlussfolgern (B+)	**Niedriges Logisches Schlussfolgern (B-)**
- hohe Denkfähigkeit	- niedrige Denkfähigkeit

Emotionale Stabilität (C+)	**Emotionale Instabilität (C-)**
- emotional stabil - ausgeglichen, zufrieden - Gefühl, mit Alltagsanforderungen zurechtzukommen	- emotional instabil - stimmungslabil - empfindlich, leicht aus dem seelischen Gleichgewicht zu bringen

Dominanz (E+)	**Nachgiebigkeit (E-)**
- dominant - durchsetzungsfähig - sich selbst behauptend - willensstark - die eigene Meinung vehement vertretend - überzeugend	- nachgiebig - kooperativ - bereit, selbst zurückzustecken entgegenkommend - wohlwollend - konfliktvermeidend

Lebhaftigkeit (F+)	**Ernsthaftigkeit (F-)**
- lebhaft, ausgelassen - spontan, impulsiv - gesellig	- ernst - zurückhaltend, gemessen - bedacht, vorsichtig
- begeisterungsfähig - liebt Beachtung	- ruhig, schweigsam

Regelbewusstsein (G+)	Unangepasstheit (G-)
- regel-, pflichtbewusst - konventionelle kulturelle Werte (Disziplin, Zuverlässigkeit)	- unangepasst - nonkonformistisch - leichtsinnig

Soziale Kompetenz (H+)	Schüchternheit (H-)
- sozial kompetent - kontaktstark	- scheu, schüchtern - gehemmt, unsicher

Empfindsamkeit (I+)	Sachlichkeit (I-)
- empfindsam - gefühlvoll - sentimental - ästhetisch, feinsinnig	- sachlich - nüchtern - unsentimental - rational

Wachsamkeit (L+)	Vertrauen (L-)
- wachsam, misstrauisch - argwöhnisch - skeptisch	- vertrauensvoll - arglos - gutgläubig

Abgehobenheit (M+)	Bodenständigkeit (M-)
- träumerisch - phantasievoll - ideenreich, kreativ	- praktisch, vernünftig - auf dem Boden der Tatsachen stehend

Privatheit (N+)	Selbstöffnungsbereitschaft (N-)
- verschlossen, diskret - sich nicht öffnend - vertraut sich anderen nicht an	- geradeheraus - echt, natürlich - selbstöffnungsbereit - erzählt offen von sich

Besorgtheit (O+)	Selbstsicherheit (O-)
- besorgt, sorgenvoll - grüblerisch - selbstzweifelnd - verletzlich	- selbstsicher - selbstbejahend - zuversichtlich - selbstzufrieden

Offenheit für Veränderung (Q1+)	Traditionalismus (Q1-)
- experimentierfreudig - aufgeschlossen für neue Ideen, Sichtweisen, Außergewöhnliches	- am Gewohnten, am Altbewährten haftend - traditionalistisch - hinterfragt nicht

Selbstgenügsamkeit (Q2+)	Soziale Orientierung (Q2-)
- selbstgenügsam - zurückgezogen - einzelgängerisch - gerne allein	- sozial orientiert - anschlussfreudig - sucht Kontakt und Nähe zu anderen

Perfektionismus (Q3+)	Flexibilität (Q3-)
- planvoll, systematisch - selbstdiszipliniert - sorgfältig, ordentlich, ordnungsliebend	- flexibel - unorganisiert, ungenau - legt wenig Wert auf Ordnung/Disziplin

Anspannung (Q4+)	Entspannung (Q4-)
- angespannt - getrieben - nervös, reizbar - regt sich schnell auf	- entspannt - ruhig - gelassen, geduldig - bewahrt Ruhe

Globalfaktoren

- Extraversion: Wärme, Dominanz, Lebhaftigkeit, Soziale Kompetenz, Privatheit (-), Selbstgenügsamkeit (-)
- Unabhängigkeit: Wärme (-), Dominanz, Empfindsamkeit (-), Besorgtheit (-)
- Ängstlichkeit: Emotionale Stabilität (-), Besorgtheit, Anspannung
- Selbstkontrolle: Regelbewusststein, Perfektionismus
- Unnachgiebigkeit: Abgehobenheit (-), Offenheit für Veränderung

Validitätsskalen

IM Impression Management (Tendenz zur sozialen Erwünschtheit), AK Akquieszenz (Tendenz zur Zustimmung), IF Infrequenz (Tendenz zu Zufallsantworten). Zudem Angaben zu Vertrauensintervallen, kritischen Differenzen und Profilvergleichen.

Frage- und Antwort-Formate

Der Test besteht aus 184 unterschiedlich gepolten Items, wobei drei Antwortmöglichkeiten zur Verfügung stehen („stimmt“, „?“, „stimmt nicht“).

Beispiel:

Ich erledige Dinge besser, wenn ich alleine arbeite als mit einer Arbeitsgruppe.

a) stimmt

b) ?

c) stimmt nicht

Subtests

Der Gesamttest ist nicht zerlegbar, und es existieren keine Kurzformen.

Gütekriterien

Reliabilität

Die internen Konsistenzkoeffizienten (Cronbachs Alpha) der Primärdimensionen betragen im Schnitt .74 (Range .66-.89); bei den Globaldimensionen .81 (Range .73-.87), und können somit als zufriedenstellend angesehen werden. Die Testwiederholungsreliabilitäten liegen bei einem Untersuchungsintervall von einem Monat für die Primärdimensionen im Schnitt bei .83 (Range .60-.92) und für die Globaldimensionen bei .86 (Range .78-.90).

Validität

Der 16PF-R verfügt über eine hohe faktorielle Gültigkeit. Zudem lässt sich die Konstruktvalidität des Tests durch Beziehungen zu anderen Instrumenten der mehrdimensionalen Persönlichkeitsdiagnostik bestätigen (z.B. NEO-FFI, FPI-R, PRF-KA, TPF-2, MMPI-II, siehe Testmanual; BIP und MBTI, siehe weitere Publikationen). Kriteriumsvaliditäten wurden beispielsweise erhoben in Hinblick auf Maße Arbeitszufriedenheit, Gesundheit und Stress oder Paarklima. Zudem ist der Test in der Lage, bei einer Klassifikationsgenauigkeit von 84 % korrekt zwischen Führungskräften und Studierenden zu diskriminieren.

Normen

Es liegen nach Geschlecht und Alter (18-29; 30-49; 50 Jahre und älter) differenzierte Normen in Form von Sten-Werten vor (Gesamtstichprobe N = 1.209).

Testmaterial

Der Test besteht aus Testheft, Manual, Antwortbogen, Auswertungsschablone, Auswertungsbogen für die Primär- und Globalskalen sowie einem Profilblatt.

Zudem ist der 16PF-R Bestandteil des Hogrefe-TestSystems (Durchführung und automatische Auswertung am Computer).

Auch gibt es ein computergestütztes Auswerteprogramm.

Bearbeitungsbedingungen und -dauer

Die Bearbeitungsdauer beträgt etwa 30 bis 45 Minuten. Der Test ist auch für Gruppentestungen geeignet und durch Nicht-Psychologen administrierbar. Die manuellen Auswertungsarbeiten können ebenso von eingewiesenen Hilfskräften durchgeführt werden.

Ergebnisrückmeldungen

Auf Grundlage der Normaldaten kann ein Ergebnisprofil mit allen Primär- und Globaldimensionen erstellt werden. Ein Vorteil des 16PF-R besteht darin, dass das Profilblatt direkt mit dem

Klienten besprochen werden kann, da die Skalenbezeichnungen hinsichtlich ihrer sozialen Wertigkeit neutral formuliert sind. Im Manual sowie bei Karson, Karson und O´Dell (1999) finden sich ausführliche Interpretationshinweise, die zur Rückmeldung der 16 PF-R Ergebnisse genutzt werden können.

Publikationen

Brück, S. & Schwarz, C. (2002). 16-Persönlichkeits-Faktoren-Test (16 PF-R). (Rezension). In U.P. Kanning & H. Holling (Hrsg.), *Handbuch personaldiagnostischer Instrumente* (S. 392-401). Göttingen: Hogrefe.

Graf, J. & Schneewind, K.A. (zur Publikation eingereicht). *16 PF-R und BIP. Zwei Persönlichkeitsinventare im Vergleich.*

Haupt, T.C. & Schneewind, K.A. (in Druck). Die Beziehungen des deutschen 16PF-R zum deutschen MBTI. *Zeitschrift für Differentielle und Diagnostische Psychologie.*

Karson, M., Karson, S. & O'Dell, J. (1999). *16PF-R. Interpretation in der klinischen Praxis. Ein Leitfaden für den 16PF-R* (übers. u. bearb. v. J. Graf & J. Kruse). Bern: Huber.

Rodenhausen, T. (1999). Der 16-Persönlichkeits-Faktoren-Test – Revidierte Fassung (16 PF-R) (Rezension). In E. Fay (Hrsg.), *Tests unter der Lupe II* (S. 116-129). Lengerich: Pabst.

Schneewind, K.A., Graf, J. & Kruse, J. (1998). Modernisierung des deutschsprachigen 16PF: der 16PF-R. *Zeitschrift für Differentielle und Diagnostische Psychologie, 19*, 57.

Erfahrungen in der Wirtschaftspraxis

Der 16PF-R hat auch in seiner revidierten Fassung zahlreiche Anwender in Berufs- und Karriereberatung, in Personalplanung und –entwicklung (z.B. Potentialanalyse, Training und Coaching) gefunden. Übereinstimmend wird von einer hohen Akzeptanz berichtet. Es finden regelmäßig Workshops zur Anwendung des 16PF-R in der Wirtschaftspraxis statt.

Preise

Test komplett (Manual, 5 Testhefte, 1 Auswertungsschablone, 20 Antwortbögen, 20 Testprofilbögen, 20 Auswertungsbögen für Primärskalen, 20 Auswertungsbögen für Globalskalen):	226 €
Testauswertprogramm:	250 €
Computerversion (incl. 100 Durchführungen und Manual):	850 €

Bezugsadresse

Testzentrale Göttingen
Robert-Bosch-Breite 25
D – 37027 Göttingen
Fon: +49 (0)551/506 88 -0/-14/-15
Fax: +49 (0)551/506 88 -24
Email: testzentrale@hogrefe.de
Internet: www.testzentrale.de

(Verfasser des Beitrags: Johanna Graf und Klaus A. Schneewind)

A3DW

Adaptiver Dreidimensionaler Würfeltest

Autor

Georg Gittler

Verlag, Erscheinungsort und -jahr

Verlag Dr. Schuhfried, Mödling (Österreich), 1999

Einsatzbereiche

Der A3DW ist ein Test zur Messung des räumlichen Vorstellungsvermögens, anwendbar für Jugendliche ab 13 Jahren und Erwachsene beiderlei Geschlechts.

Für zahlreiche Berufe, bei denen „Dreidimensionalität" und „Raum-Zeit-Koordinationen" eine Rolle spielen, ist ein gut ausgeprägtes Raumvorstellungsvermögen unabdingbar (etwa technische und Industrieberufe; Ärzte, besonders in operativen Fächern; Architekten; Fluglotsen, Fernfahrer etc.), in Extremfällen sogar eine Frage des Überlebens (Piloten; Hochbauarbeiter etc.). Spätestens seit Thurstone (1938) gilt das räumliche Vorstellungsvermögen als eine der primären Intelligenzdimensionen, welche durch die zunehmende Technisierung in unserer „modernen Welt" mehr und mehr an praktischer Bedeutung gewinnt. Diese originär nicht-sprachliche Fähigkeit zur mentalen Repräsentation und Transformation räumlicher Relationen versetzt uns Menschen nicht nur bei konkret-anschaulichen Problemstellungen, sondern auch während abstrakter Denkvorgänge, in die Lage, Handlungen zu planen und ihre Konsequenzen vorherzusehen.

Der A3DW kann daher folgenden Zwecken dienen: Personalauslese und Personalentwicklung (insbesondere in technischen und Industrieberufen); Berufs-, Studien-, Bildungs- und Karriereberatung; Fähigkeitsdiagnostik (u.a. von Hochbegabung); gesundheitspsychologische Diagnostik (Unfallverhütung); verkehrspsychologische Diagnostik (insbesondere Luft- und Raumfahrt); Evaluation von Trainings- und Unterrichtsprogrammen; transkulturelle Diagnostik (cross cultural testing).

Konstruktionsgrundlagen und Inhalte

Der A3DW ist ein eindimensionaler (Rasch-homogener) Computertest. Die Aufgabenpräsentation erfolgt adaptiv, d.h. im Verlauf einer Testung werden jeweils genau jene Aufgaben präsentiert (entnommen aus einem „Aufgabenpool"), die – in einem testtheoretisch definierten Sinn – „optimal" zum individuellen Fähigkeitsniveau der Testperson passen. Daraus ergibt sich

erhöhte Präzision der Messung bei zumeist geringerer Testdauer verglichen mit dem herkömmlichen Testvorgabemodus.

In der psychologischen Fachliteratur werden vielfach die testtheoretischen Eigenschaften von Raumvorstellungstests bemängelt. Die Probleme beruhen zu einem Gutteil auf Unzulänglichkeiten der klassisch-testtheoretischen Skalenkonstruktionsmethoden: So ist etwa nach einem halben Jahrhundert faktorenanalytischer Raumvorstellungsforschung nicht klar, welche Aufgabentypen als geeignete Indikatoren für Raumvorstellungsleistungen gelten können (Problematik der Validität, insbesondere der Konstruktvalidität), und inwieweit durch derartige Tests bei verschiedenen Probandengruppen (männliche vs. weibliche; jüngere vs. ältere etc.) dieselbe Fähigkeitsdimension erfasst wird (Homogenitätsproblematik).

Die Entwicklung eines neuen Raumvorstellungstests (3DW; Gittler, 1990) wurde daher auf Basis einer modernen Testtheorie (Item Response Theory: IRT) vorgenommen, insbesondere weil die IRT Messmodelle zur Verfügung stellt, die - zumindest teilweise - als Lösungsansätze für die genannten Probleme gelten können. Die Testentwicklung des 3DW darf inzwischen als gelungen bezeichnet werden, da zahlreiche Nachfolgestudien die Eindimensionalität (Homogenität) nach dem Modell von Rasch und die Validität des Tests bestätigten (Gittler, 1992; vgl. auch die unter „Gültigkeit“ angeführten Zitate). Mit der nun vorliegenden Erweiterung des 3DW zu einem adaptiven Verfahren (A3DW) wurde der Weg hin zu einer modernen Testentwicklung konsequent fortgesetzt.

Adaptives Testen

So wie man die Gewichtszunahme eines Babys nicht mit einer normalen Badezimmerwaage, sondern mit einer im unteren Gewichtsbereich feiner differenzierenden Babywaage messen würde, besteht die Idee des adaptiven Testens darin, den Testpersonen verschieden schwierige, bezüglich ihres individuellen Fähigkeitsniveaus „maßgeschneiderte“ Testaufgaben aus einem Itempool vorzugeben – und nicht, wie in herkömmlichen Tests, allen Personen dieselben Aufgaben. Der Itempool des A3DW umfasst 142 Rasch-homogene Aufgaben.

Das adaptive Testen ist unverzichtbar auf die Erkenntnisse der Item Response Theory angewiesen, weil die Bestimmung eines fairen Testwerts bei unterschiedlichen Aufgaben je Testperson anders nicht möglich ist. Die Auswahl geeigneter, leistungsadäquater (= informativer) Aufgaben aus dem Itempool kann nur mittels Computer bewerkstelligt werden, da während einer individuellen Testung mehrfach numerisch aufwendige Schätzungen des aktuellen Personenfähigkeitsparameters (Rasch Modell) erforderlich sind. Es handelt sich also um einen genuinen Computertest.

Parameterschätzung (Rasch-Modell)

Die Itemschwierigkeitsparameter der Poolaufgaben wurden mittels bedingter Maximum-Likelihood-Schätzung aus einer Kalibrierungsstichprobe von insgesamt 7.148 Vpn bestimmt. Die jeweils aktuellen Personenfähigkeitsparameter werden im Verlauf der Testung mittels der von Warm (1989) beschriebenen „Weighted Likelihood Estimation“-Methode (Bayes-Ansatz) geschätzt.

Frage- und Antwort-Formate

Multiple-Choice-Format; 6 Würfelabbildungen („Antwortwürfel“) sind mit einem „Testwürfel“ zu vergleichen; die Antwort erfolgt durch Markieren eines Antwortwürfels bzw. einer von zwei Verbalantworten („Kein Würfel richtig“ bzw. „Ich weiß die Lösung nicht“). Mit sieben „nicht-trivialen“ Distraktoren – die letztgenannte Verbalantwort ist nicht dazuzuzählen – wird die Ratewahrscheinlichkeit gering gehalten.

Subtests

Keine.

Gütekriterien

Die Angaben zu den klassischen Testgütekriterien stammen aus dem Handbuch des 3DW (Gittler, 1990). Da der A3DW nur eine Ausweitung (mehr Items im Itempool) und Verbesserung (präzisere Messung aufgrund der adaptiven Testvorgabe) des 3DW darstellt, und zudem beide Verfahren laut Rasch-Modell nachweislich dieselbe Fähigkeitsdimension messen (Piringer, 1997), können die folgenden Testkennwerte als „untere Schranke“ für den A3DW gewertet werden.

Objektivität

Durchführung, Auswertung und Interpretation können nicht nur beim 3DW, sondern auch bei A3DW (Computertest) als objektiv gelten.

Zuverlässigkeit

Die innere Konsistenz ist aufgrund der Geltung des Rasch-Modells gegeben (gilt für *alle* Poolaufgaben). Die numerischen Werte der in verschiedenen Stichproben und nach unterschiedlichen Methoden (Split-half; Cronbachs Alpha) bestimmten Reliabilitätskoeffizienten liegen zwischen .82 und .91. Die Retest-Reliabilität beträgt bei 231 Schülern, die im Abstand von 21 Monaten getestet wurden (Erhebungszeitpunkt 1: 1984; Zeitpunkt 2: 1986), r_{tt}=.61.

Gültigkeit

Zahlreiche Ergebnisse von korrelationsstatistischen Analysen und Intergruppenvergleichen sprechen für konvergente und diskriminante Validität des Tests und lassen eine differenzierte Beurteilung des 3DW bezüglich verschiedener Aspekte der Validität zu (vgl. Gittler, 1990). Auch in diversen neuen Vergleichsuntersuchungen wird die Validität des 3DW immer wieder bestätigt (Gittler, 1992; Gittler, & Vitouch, 1994; Tanzer, Gittler & Sim, 1994; Tanzer, Gittler & Ellis, 1995; Broer, 1996; Vitouch, Bauer, Gittler, Leodolter & Leodolter, 1997; Arendasy, 1998; Gittler & Glück, 1998; Gittler, 1999).

Als Beispiele für einen modernen Validitätsaspekt sei die sog. inkrementelle Validität des 3DW im Vergleich zum IST-70 (Subtest 8, Würfelaufgaben) bzw. zu den Schlauchfiguren (Stumpf & Fay, 1983) explizit angeführt (vgl. Gittler, 1990): Der 3DW vermag mit einer Effektgröße von d = .82 (starker Effekt) besser zwischen Technikstudenten und Nicht-Technik-Studenten zu trennen als die Würfelaufgaben des IST-70 mit d = .58 (mittlerer Effekt). Eine Kurzform des 3DW (9 Testitems) korreliert - im Sinne von Übereinstimmungsvaliditäten - mit den Kennwerten

(Richtungsmaß: Rm und Abstandsmaß: Am) eines praxisnahen Orientierungstests (sog. Planzeichentest; kleinere Werte bedeuten bessere Lokalisationsleistungen) deutlich höher als die Schlauchfiguren (SF): r(3DW/Rm) = -.50 ($p < .001$), r(SF/Rm) = -.02 ($p = .444$), r(3DW/Am) = -.32 ($p = .023$), r(SF/Am) = +.04 ($p = .393$).

Normen

Arrer (1992) konnte zeigen, dass die Papier-Bleistift-Normen des 3DW auch für die Computerfassung des Tests Gültigkeit haben. Da der 3DW und der A3DW auf einer gemeinsamen Fähigkeitsskala verrechnet werden können (dies ist aufgrund der Geltung des Rasch-Modells möglich), konnten die Normen des 3DW – nach adäquater Umrechnung – auch für den A3DW verfügbar gemacht werden: Das Computerprogramm enthält somit für Österreich repräsentative, geschlechtsunspezifische und -spezifische Schulnormen (N = 4.064), aufgegliedert nach verschiedenen Schulstufen (=Altersäquivalent) und Schultypen. Zudem sind (nicht repräsentative) Normen für Studenten – differenziert in technische und nicht technische Universitäten – und Vergleichswerte für Testpersonen von 30-39, 40-49 und 50-72 Jahren vorhanden. Dass die Schülernormen auch für Deutschland Geltung beanspruchen dürfen, konnte aufgrund einer in Deutschland durchgeführten Vergleichsuntersuchung gezeigt werden.

Testmaterial

Computertest mit automatisierter Auswertung; Die Bearbeitung erfolgt mittels Probandentastatur, Lichtgriffel, PC-Keyboard, Maus oder Touchscreen.

Bearbeitungsbedingungen und -dauer

Einzeltestung pro Computerarbeitsplatz. Die Bearbeitungsdauer kann nicht exakt angegeben werden, weil es sich erstens um einen Niveau- (Power-) Test handelt, und zweitens das Erreichen eines Abbruchkriteriums bei adaptiver Testvorgabe vom individuellen Testverlauf abhängt. Aus Erfahrung kann jedoch gesagt werden, dass für die Standardform (vgl. im folgenden unter „Testformen“) mit einer Testdauer von ca. 15 bis 30 Minuten zu rechnen ist.

Ergebnisrückmeldungen

Die automatisierte Auswertung in Form eines Computerausdrucks umfasst folgende Informationen:

- zentraler kriteriumsorientierter Testkennwert: Personenfähigkeitsparameter (Rasch-Modell), der einen gruppenunspezifischen, absoluten Vergleich zwischen Testpersonen ermöglicht;
- zentraler normorientierter Testkennwert: Prozentrang (bzw. T-Wert), der einen gruppenspezifischen, relativen Vergleich in Abhängigkeit von der gewählten Vergleichsnorm (Referenzpopulation) ermöglicht;
- aufgabenspezifische Kennwerte: Anzahl der bearbeiteten Items, Gesamtbearbeitungszeit, mittlere itemspezifische Bearbeitungszeit und deren Streuung;
- grafische Profile auf Basis jeder ausgewählten Vergleichsnorm;
- ausführliches Testverlaufsprotokoll.

Zur Ergebnisrückmeldung an den Testanden – falls nötig und gewünscht – bedürfen diese Daten der mündlichen Erläuterung durch einen geschulten Testanwender.

Publikationen

Fay, E. (1996). Dreidimensionaler Würfeltest 3DW (Rezension). In E. Fay, *Tests unter der Lupe: Aktuelle psychologische Testverfahren – kritisch betrachtet* (Bd 1; S. 21-27). Heidelberg: Asanger.

Gittler, G. (1990). *Dreidimensionaler Würfeltest (3DW). Ein Rasch-skalierter Test zur Messung des räumlichen Vorstellungsvermögens* (Testmappe inklusive Manual, Testheft und Antwortbogen). Weinheim: Beltz.

Gittler, G. (1999). Sind Raumvorstellung und Reasoning separierbare Fähigkeitsdimensionen? Dimensionalitätsanalysen zweier Rasch-skalierter Tests: 3DW und WMT. *Diagnostica, 45,* 69-81.

Jenko, D.M. (2002). Dreidimensionaler Würfeltest (3 DW) (Rezension). In U.P. Kanning & H. Holling (Hrsg.), *Handbuch personaldiagnostischer Instrumente* (S. 139-142). Göttingen: Hogrefe.

Tanzer, N. K., Gittler, G. & Ellis, B.B. (1995). Cross-cultural validation of item complexity in a LLTM-calibrated spatial ability test. *European Journal of Psychological Assessment, 11*, 170-183.

Vitouch, O., Bauer, H., Gittler, G., Leodolter, M. & Leodolter, U. (1997). Cortical activity of good and poor spatial test performers during spatial and verbal processing studied with Slow Potential Topography. *International Journal of Psychophysiology, 27,* 183-199.

Erfahrungen in der Wirtschaftspraxis

Bei zahlreichen Einsätzen des 3DW wurden gute Erfahrungen in folgenden Bereichen der Bewerberauswahl berichtet: Berufs- und Militärpiloten; Fluglotsen; CAD-Techniker; Facharbeiter bzw. Techniker für halbautomatisierte Fertigungsstraßen; Zulassung zu Facharbeiterintensivausbildungen in unterschiedlichen Industriebereichen; Fahrschullehrer.

Sonstiges

Die automatisierte Aufbereitung der Testergebnisse in einer Datenbank bietet die Möglichkeit, sowohl betriebsspezifische Diagnosekriterien als auch unternehmensweit einsetzbare Personalauswahlstrategien zu entwickeln.

Das Verfahren ist in vielen verschiedenen Sprachen erhältlich.

Testformen

Der A3DW kann je nach diagnostischer Fragestellung und je nach ökonomischen Rahmenbedingungen in einer Kurzform (Screening-Version), der Standardform sowie einer Langform mit erhöhter Messgenauigkeit durchgeführt werden.

Preise

50 Testvorgaben:	172 €
7 Jahre Nutzungsrecht:	1.032 €

Bezugsadresse

Dr. Schuhfried GmbH
Hyrtlstraße 45
A – 2340 Mödling
Fon: +43 (0)2236/423 15
Fax: +43 (0)2236/465 97
Email: info@schuhfried.at
Internet: www.schuhfried.at

(Verfasser des Beitrags: Gerald Schuhfried)

ABAT-R

Revidierter Allgemeiner Büroarbeitstest

Autoren

Gustav A. Lienert und Heinz Schuler

Verlag, Erscheinungsort und -jahre

Hogrefe-Verlag, Göttingen, 1994, 2003 (aktualisierte Auflage)

Einsatzbereiche

Zielgruppen

Interessenten und Bewerber für Büroberufe, Schüler (ab 14 Jahre), Auszubildende und Berufstätige in Büroberufen.

Einsatzzwecke

Auswahl von Mitarbeitern für Bürotätigkeiten, insbesondere in kaufmännischen Ausbildungsberufen, für Sekretariats- und andere Verwaltungstätigkeiten; Identifikation von Personalentwicklungsbedarf; Beratung von Berufswählenden, Umschülern und Rehabilitanden.

Konstruktionsgrundlagen

Der ABAT-R verbindet den Konstruktansatz mit dem Simulationsansatz der Berufseignungsdiagnostik (s. hierzu Schuler & Funke, 1995). Auf der „Oberfläche" der Messung werden Fertigkeiten erfasst, wie sie für Bürotätigkeiten anforderungsentsprechend sind (z.B. Kategorisieren). „Hinter" diesen Prozessen stehen generelle kognitive Fähigkeiten, wie sie auch mit klassischen Intelligenztests gemessen werden. Während die Simulationsebene des ABAT-R die analoge Testwertinterpretation unterstützt, die für Arbeitsproben charakteristisch ist und Transparenz sowie Akzeptanz gewährleistet, stellt die Konstruktebene sicher, dass gleichzeitig allgemeine, in ihrer Validität generalisierbare Fähigkeitsfaktoren erfasst werden.

Inhalte

Sowohl bei konventioneller wie bei EDV-gestützter Arbeitsweise bestehen die Grundanforderungen in Büroberufen in rascher und qualitätsorientierter Arbeitsweise, wobei ein Großteil der Arbeitszeit mit dem Sichten, Vergleichen, Kategorisieren und Kontrollieren von Informationen in verbaler und numerischer Form verbracht wird. Verbale und numerische Fähigkeiten so-

wie Bearbeitungsgeschwindigkeit und, insbesondere bei komplexeren Aufgaben, Verarbeitungskapazität stellen die wichtigsten Fähigkeitsanforderungen erfolgreicher Tätigkeit in Büroberufen dar.

Auf der Simulationsebene der arbeitsprobenartigen Gestaltung (s. Konstruktionsgrundlagen) lassen sich die erfassten Fertigkeiten gut mit den Bezeichnungen der Subtests identifizieren: Kundenbriefe-Sortieren, Adressen-Prüfen, Summen-Prüfen, Rechtschreibung-Korrigieren, Textaufgaben-Lösen, Zeichen-Setzen. Auf der Ebene der psychologischen Konstrukte ließen sich durch die Validitätsstudien insbesondere von Schmidt (1993) vier Faktoren nachweisen, die im Berliner Intelligenzstrukturmodell (Jäger, 1984) zu den Kernfaktoren geistiger Leistung zählen: Verarbeitungskapazität, Bearbeitungsgeschwindigkeit, verbale Fähigkeiten und numerische Fähigkeiten (vgl. hierzu auch den Beitrag zum BIS-4; in diesem Band).

Der ABAT-R kann deshalb sowohl als Testverfahren wie auch als Arbeitsprobe gelten, was seine hohe Validität verständlich macht.

Frage- und Antwort-Formate

Die Fragevorgabe erfolgt entweder in unmittelbarer Entsprechung zur Gestalt der entsprechenden beruflichen Aufgaben oder in symbolischer Analogie dazu. Beispielsweise bedeutet die Aufgabe „Hack, Flugzeugbau, 04.01.“, dass ein Brief den drei Karteien Alphabet, Branche und Datum richtig zuzuordnen ist. Die Antwortformate sind als Mehrfachwahl vorgegeben oder es sind kurze Eintragungen, Fehlerkorrekturen oder Richtigkeitsvermerke vorzunehmen.

Subtests

Stanine-Werte werden gesondert für jeden Subtest ermittelt; Normen sind allein für den Gesamtwert (Stanine-Summe) verfügbar.

Gütekriterien

Objektivität

Bei Beachtung der Anweisungen für Durchführung, Auswertung und Interpretation ist vollständige Objektivität gegeben.

Reliabilität

Der Paralleltestkoeffizient für die Stanine-Summe beträgt .79, die Retest-Reliabilität nach einem Intervall von acht Wochen wurde mit .83 ermittelt. Der Übungsgewinn bei Testwiederholung beläuft sich auf ca. .4 bzw. .7 Standardabweichungen bei Vorgabe der Parallelversion bzw. der gleichen Testversion.

Validität

Da im ABAT-R Fähigkeiten erhoben werden, die wichtigen Anforderungsmerkmalen für Bürotätigkeiten entsprechen, kann für das Verfahren inhaltliche Validität konstatiert werden. Bezüglich der faktoriellen Validität konnten neben einem verbalen und einem numerischen Faktor auch die Intelligenzfaktoren Bearbeitungsgeschwindigkeit und Verarbeitungskapazität gefunden

werden. Korrelationen mit verwandten Verfahren, darunter auch Intelligenztests, sind mittel bis hoch, mit dem Leistungs-Prüf-System (Horn, 1983) beträgt sie .67.

Angaben zur prognostischen Validität sind derzeit noch weitgehend auf Studien angewiesen, die mit dem Vorgängerverfahren ABAT durchgeführt wurden. Hier ergaben sich für zwei Gruppen von kaufmännischen Berufsschülern Korrelationen mit dem Abschneiden in den Abschlussprüfungen vor der Industrie- und Handelskammer in Höhe von .44 und .46. Für drei Bewerberstichproben (kaufmännische Berufe in der Industrie und im öffentlichen Dienst) ergaben sich Zusammenhänge der ABAT-Ergebnisse mit den aus allen übrigen Auswahlverfahren resultierenden kombinierten Eignungsbeurteilungen zwischen .65 und .71. Bei N =171 kaufmännischen Berufsschülern betrug die Korrelation zwischen ABAT-R-Ergebnissen und Berufsschul-Gesamtnoten .49. Unter den Einzelnoten ergaben sich für „Rechnen“ (.45) und „Datenverarbeitung (.42) die höchsten Werte.

Normen

Der Test wurde an einer Stichprobe von 1.659 Testteilnehmern geeicht, darunter befanden sich 420 Realschüler, 765 Berufsschüler, 197 berufliche Rehabilitanden, 139 Ratsuchende, 138 Bewerber um kaufmännische Lehrstellen. Gesonderte Normen liegen vor für vier Altersgruppen von 14 bis 25 Jahren über alle Teilstichproben, für kaufmännische Berufsschüler, für Realschüler (9. Klasse), für Realschüler (10. Klasse), Rehabilitanden, Ratsuchende, Gesamtstichprobe Bewerber, Teilstichprobe „Bewerber mit Abitur“, Teilstichprobe „Bewerber Realschule“.

Testmaterial

Testhefte mit Rasterschema für die Auswertung, Auswertungsschablonen, Handanweisung. Zusätzlich benötigt werden Schreibstifte und Stoppuhr.

Bearbeitungsbedingungen und -dauer

In Einzeldurchführung oder als Gruppentest. Es liegen zwei Formen A und B vor. Die Instruktionen werden für jeden Untertest mit Beispielen laut vorgelesen. Die Bearbeitungszeit ist begrenzt und beträgt für alle Untertests insgesamt 28 Minuten. Die Durchführungszeit inklusive der Anweisungen beträgt etwa 45 Minuten. Die Durchführung kann von Hilfskräften übernommen werden.

Ergebnisrückmeldungen

Nach Maßgabe des Diagnostikers können die Testwerte als Stanine-Profil oder als Prozentwerte dargestellt werden. Informativ für Testteilnehmer ist die Rückmeldung und gegebenenfalls Erörterung v.a. der Subtestwerte.

Publikationen

Fay, E. (1996). Revidierter Allgemeiner Büroarbeitstest ABAT-R (Rezension). In E. Fay, *Tests unter der Lupe: aktuelle psychologische Testverfahren kritisch betrachtet* (Bd. 1, S. 81-91). Heidelberg: Asanger.

Jäger, A.0. (1994). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau, 35,* 21-35.

Mattke, K. (2002). Revidierter Allgemeiner Büroarbeitstest (ABAT-R) (Rezension). In U.P. Kanning & H. Holling (Hrsg.), *Handbuch personaldiagnostischer Instrumente* (S.210-219). Göttingen: Hogrefe.

Schmidt, J.U. (1993). Der Allgemeine Büroarbeitstest (ABAT) - mehr als ein Bürotest? *Diagnostica, 39,* 151-168.

Schuler, H. & Funke, U. (1995). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (2. Aufl.; S. 235-283). Bern: Huber.

Erfahrungen in der Wirtschaftspraxis

Der ABAT-R gehört zu den in Unternehmen und Verwaltungsorganisationen am häufigsten eingesetzten Verfahren für Büroarbeitstätigkeiten. Eine Großorganisation der Kreditwirtschaft hat mit einer Parallelform des Vorgängerverfahrens (ABAT) bis 1999 ca. 1 Mio. Testanwendungen bei Bewerbern um Ausbildungsplätze durchgeführt. Das Verfahren wird aufgrund seines arbeitsprobenartigen Charakters und der abwechslungsreichen Aufgaben gut akzeptiert. Die revidierte Fassung ABAT-R kann auch für anspruchsvollere Bürotätigkeiten eingesetzt werden.

Preise

Mappe komplett (bestehend aus: Handanweisung, je 5 Testheften A + B, Auswertungsschlüssel A + B und Lösungsschablonen A + B):	192 €
Handweisung:	30 €
Testheft A:	2 €
Testheft B:	2 €
2 Auswertungsschlüssel A:	8 €
2 Auswertungsschlüssel B:	8 €
4 Lösungsschablonen A:	18 €
4 Lösungsschablonen B:	18 €
Mappe leer:	4 €

Bezugsadressen

Deutschland

Testzentrale Göttingen
Robert-Bosch-Breite 25
D – 37027 Göttingen
Fon: +49 (0)551/506 88 -0/-14/-15
Fax: +49 (0)551/506 88 -24
Email: testzentrale@hogrefe.de
Internet: www.testzentrale.de

Schweiz

Testzentrale der Schweizer Psychologen
Länggass-Str. 84
CH − 3000 Bern 9
Fon: +41 (0)31/300 45 -45
Fax: +41 (0)31/300 45 -90
Email: testzentrale@hanshuber.com
Internet: www.hanshuber.com

(Verfasser des Beitrags: Heinz Schuler)

ABLE – BDA

Business Decision Analysis

Editor

Psychometric Research & Development (PRD)

Publishing company, place and year of publication

OPP, Oxford, 1996

Fields of application

This is a high-level test aimed at candidates who will have to show a considerable degree of intellectual independence and curiosity in their work, finding effective concepts to operate with rather than preset patterns and routines. High-flier graduate entry and promotion to fast-track positions are two significant areas where BDA will find appropriate use. The main cognitive skills employed are concept formation and hypothesis testing, although, verbal comprehension and numeracy skills are also employed.

Base of conceptualization

By the end of the eighties, Blinkhorn and his colleagues at PRD had recognised the limits of traditional approaches to ability testing. Schmidt and Hunter's (1977) ground breaking work on validity generalisation had freed ability testing from the problem of situational specificity only a little more than a decade before. What researchers at PRD detected however were the first stirrings of discontent to emerge from a psychometric market that was becoming more sophisticated and questioning.

The increasing sophistication of the typical test purchaser meant that, for the first time, they were able to articulate concerns with the range of products on offer as failing to meet completely the real world needs of organisations. Asked by test purchasers for ability measures that could tap into 'integrated skills and occupational flexibility' Blinkhorn and his colleagues returned to the first principles and Structured Learning Exercises (SLE's) represented the highly innovative response that PRD gave.

It is important to emphasize that the BDA concept (in fact the whole ABLE concept) is not based on any espoused theory of mental abilities. As applied practitioners, PRD Psychologists are concerned with the utility of their measures, and not at all with offering support for any particular model of the intellect. That said, based as they are on work by PRD developing SLE's from mid-eighties on, the BDA approach can be identified with relatively modern theories of

intellectual abilities. This is especially true in respect of the conceptualisation of intelligence as an adaptive construct rather than a relatively fixed characteristic of the individual.

Contents

BDA involves no teaching. Applicants have to discover and learn for themselves the concepts which underlie business decisions presented to them, and then apply those concepts to new instances. The questions cover a wide range of business processes all set within the context of the confectionery industry. The main cognitive skills employed are concept formation and hypothesis testing, although, verbal comprehension and numeracy skills are also employed.

Stimulus- and Response-Formats

In an information booklet the candidate is given concepts which underlie business decisions. To apply those concepts, the candidate has to answer 12 questions. For each question, the candidate has to answer 4 sub-questions. The candidate can answer a sub-question with a multiple-choice response “a” “b” or “c”. For each question the labels for a, b and c are different, depending on the underlying business concepts that are used.

Subtests

None.

Psychometric properties

Internal consistencies

BDA does not fit well with the usual assumptions for the calculation of an internal consistency coefficient. The values for alpha as calculated may significantly underestimate the effective reliability of the test.

The main problem is that the method for calculating alpha makes the assumption that the test is homogeneous, but this assumption is clearly broken since the content, mode of presentation and type of response category change from block to block throughout the test.

For this reason, alpha is likely to be depressed if calculated on an item-by-item basis. As an alternative, each block of four items can be scored as a short scale taking the within-block variance as an estimate of the true variance and the between-block variance as an estimate of the covariance. This approach results in an estimated reliability coefficient of .75. We suggest that the estimate of .75 should be adopted for the reliability, with a corresponding standard error of measurement of 5.30 raw score points.

Face validity

The contextualised setting of the means that this exercise is seen as highly face valid.

Construct validity

BDA is not a conventional ability test. It does not set out to measure individual mental abilities, and would not be expected to show extremely high correlations with conventional measures of ability. Indeed it would be expected that BDA would show:

- High correlations with other ABLE exercises measuring similar skills
- High correlations with direct job simulations exercises, where available
- Moderate correlations with other ABLE exercises measuring different skills
- Moderate correlations with conventional ability tests.

And: Within any given group, the intercorrelation between conventional ability tests will usually be higher than that between the BDA and any one conventional test.

In a number of real life selection and development situations in which the BDA exercise has been used, other tests have also been administered. This data do provide some support for the above hypotheses, as shown in the table below.

Correlation of BDA with other tests

	External test	Pearson r	Sample Size	Sample description
BDA English	GMA-V	.18	140	Senior managers in an assessment centre
	GMA-A	.24	140	
	GMA-A	.19	212	Airline employees and sponsored students
	CIT4	.26	101	
	GMA-A	.26	229	Graduate applicants
	VMG	.37	462	Graduate applicants
	NMG	.43	462	
	AH6	.36	26	Managers in a large consultancy
BDA Dutch	Raven's APM	.29	112	Dutch managers
BDA French	Raven's APM	.20	57	French managers

Criterion-related validity

Managers (N =135) of a large utility business completed the BDA and were in addition rated on 17 competencies relating to their performance in the job. BDA did show a significant correlation with "Reasoning" (.28, $p<.01$), with "Analytical Reasoning" (.24, $p<.01$) and with "Stability under Pressure" (.25, $p<.01$). No significant correlation was found with "Conceptual Thinking", and with "Interpersonal Competencies". This may suggest that the BDA can predict success in competencies related to reasoning skills, but not in competencies relating to interpersonal skills.

Norms

British normgroup

The trial sample for the BDA consisted of 111 individuals. Of these individuals, 67 were graduates, 24 were undergraduates and 20 were non-graduates undergoing managerial training. There were 73 females and 38 males. Ages ranged from 19 to 47 with an average of 26.

An additional 14 British normgroups are available.

Belgian normgroup

The sample consisted of 175 individuals, all were graduates, participating on a Recruitment Event. There were 74 females and 101 males, and 115 Dutch speaking and 60 French speaking participants. Age ranged from 22 to 25 with an average of 23 years.

Test material

All candidates receive a preview information booklet, a question and answer booklet, and an information file. Administrators have separate and administration instructions, a scoring template and of course a test manual.

Requirements and time of testing

40 minutes are allowed for completion of the test (individual or group testing).

Feedback of results

The raw score can be translated into a T-score (Great Britain) or Stanine (Belgium).

References

None.

Experiences in the field of business

The BDA has been used very often in recruitment & selection procedures and in assessment & development centers.

Note

In addition to BDA, there are a further 11 tests in the ABLE Series. These include Critical Business Planning (CBP), a sibling test to BDA.

Prices

Question & Answer Booklet (10):	60 £	(86 €)
Information file (10):	82 £	(118 €)
Manual & User's guide (1):	21 £	(30 €)
Administration Instructions (1):	8 £	(11 €)
Scoring Template (1):	14 £	(20 €)
Reference Set (1):	58 £	(83 €)

Ordering addresses

Oxford Psychologists Press Ltd.
Elsfield Hall
15-17 Elsfield Way
Oxford, OX2 8EP
United Kingdom
Fon: +44 (0)1865/40 45 00
Fax: +44 (0)1865/31 03 68
Email (purchases): orders@opp.co.uk
Email (enquires) : enquiry@opp.co.uk
Internet : www.opp.eu.com

German language contact telephone (national rate): 01 803 000 768

(Verfasser des Beitrags: John Hackston)

ABLE – FA

Financial Appraisal

Editor

Psychometric Research & Development (PRD)

Publishing company, place and year of publication

OPP, Oxford, 1997

Fields of application

The test makes wide-ranging demands on cognitive abilities. The verbal comprehension required is fairly straightforward but candidates must not be thrown by the need to learn technical terms. The arithmetic required is fairly simple except that candidates have to be comfortable making rough estimates. Other numerical skills involved include the use of proportions, percentages, ratios, table reading, graph reading, simple equations, probabilities and averages.

Candidates with good numerical skills get through the first half of the test relatively quickly and can be expected to answer most of the questions correctly. The second half of the test is, therefore, important for differentiating amongst candidates being considered for graduate level jobs with significant numerical/financial content.

Base of conceptualization

By the end of the eighties, Blinkhorn and his colleagues at PRD had recognised the limits of traditional approaches to ability testing. Schmidt and Hunter's (1977) ground breaking work on validity generalisation had freed ability testing from the problem of situational specificity only little more than a decade before. What researchers at PRD detected however were the first stirrings of discontent to emerge from a psychometric market that was becoming more sophisticated and questioning.

The increasing sophistication of the typical test purchaser meant that, for the first time, they were able to articulate concerns with the range of products on offer as failing to meet completely the real world needs of organisations. Asked by test purchasers for ability measures that could tap into 'integrated skills and occupational flexibility' Blinkhorn and his colleagues returned to the first principles and Structured Learning Exercises (SLE's) represented the highly innovative response that PRD gave.

It is important to emphasize that the FA concept (in fact the whole ABLE concept) is not based on any espoused theory of mental abilities. As applied practitioners, PRD Psychologists are concerned with the utility of their measures, and not at all with offering support for any particular

model of the intellect. That said, based as they are on work by PRD developing SLE's from mid-eighties on, the FA approach can be identified with relatively modern theories of intellectual abilities. This is especially true in respect of the conceptualisation of intelligence as an adaptive construct rather than a relatively fixed characteristic of the individual.

Contents

In the first half of the test, candidates are taught a set of rules for calculating and interpreting some fictitious financial ratios, indicators and concepts. At first, these rules are entirely mechanical. Although they demand a reasonable degree of numeracy the emphasis is on speed of learning and accuracy of application of the rules. Later, but still in the first half of the test, some extrapolation is required from the rules that are being taught. In the second half of the test there is no further teaching. Candidates have to discover and learn for themselves some new financial concepts which affect the rules they have been taught earlier. Hence, the exercise requires the skills of concept formation and hypothesis testing but set within a numerical/financial context.

Stimulus- and Response-Formats

In the guidance booklet, the candidates have to learn a set of rules for analysing financial and statistical information concerning applicants for venture capital and then answer questions to see how well they can apply them. The question and answer booklet contains information of various sorts about individual applicants. This includes business plans, ratings on the key criteria and, in the second part of the test, additional information.

Subtests

None.

Psychometric properties

Internal consistencies

Because there is a possibility of dependencies between two answers given for each question, the reliability estimate is calculated on the basis of there being 21 separate questions with two marks available for each as opposed to 42 questions each with one mark. This means that the estimate of .84 is conservative but still clearly acceptable.

Face validity

The contextualised setting of FA means that this exercise is seen as highly face valid.

Construct validity

FA is not a conventional ability test. It does not set out to measure individual mental abilities, and would not be expected to show extremely high correlations with conventional measures of ability. Indeed it would be expected that FA would show:

- High correlations with other ABLE exercises measuring similar skills
- High correlations with direct job simulations exercises, where available

- Moderate correlations with other ABLE exercises measuring different skills
- Moderate correlations with conventional ability tests.

And: Within any given group, the intercorrelation between conventional ability tests will usually be higher than that between the FA and any one conventional test.

In a number of real life selection and development situations in which the FA exercise has been used, other tests have also been administered. This data does provide some support for the above hypotheses, as shown in the table below.

Correlation of FA with other tests

	External test	Pearson r	Sample Size	Sample description
FA English	VMG	.23	53	Graduate applicants
	NMG	.33	53	
FA Dutch	Raven's APM	.36	113	Dutch managers
FA French	Raven's APM	.20	57	French managers

Norms

British normgroup

The trial sample for the FA consisted of 133 individuals. Of these individuals, 24 were graduates already in employment, 53 were applicants for graduate vacancies in a major IT company and 56 were undergraduates who had not yet entered the job market. Ages in the sample ranged from 19 to 55 with an average of 24. There were 61 males and 72 females.

Belgian normgroup

The sample consisted of 365 individuals, all were graduates, participating on a Recruitment Event. There were 161 females and 204 males, and 219 Dutch speaking and 145 French speaking participants. Age ranged from 22 to 25 with an average of 23 years.

Test material

All candidates receive a preview information booklet, a question and answer booklet, and an information file. Administrators have separate and administration instructions, a scoring template and of course a test manual.

Requirements and time of testing

45 minutes is allowed for completion of the test (individual or group testing).

Feedback of results

The raw score can be translated into a T-score (Great Britain) or Stanine (Belgium).

References

None.

Experiences in the field of business

FA has been used very often in Recruitment & Selection procedures and in Assessment & Development centers.

Note

In addition to FA, there are a further 11 tests in the ABLE Series.

Prices

Question & Answer Booklet (10):	60 £	(86 €)
Guidance Booklet (10):	82 £	(118 €)
Manual and User's Guide (1):	21 £	(30 €)
Administration Instructions (1):	8 £	(11 €)
Scoring Template (1):	14 £	(20 €)
Reference Set (1):	58 £	(83 €)

Ordering addresses

Oxford Psychologists Press Ltd.
Elsfield Hall
15-17 Elsfield Way
Oxford, OX2 8EP
United Kingdom
Fon: +44 (0)1865/40 45 00
Fax: +44 (0)1865/31 03 68
Email (purchases): orders@opp.co.uk
Email (enquiries): enquiry@opp.co.uk
Internet: www.opp.eu.com

German language contact telephone (national rate): 01 803 000 768

(Verfasser des Beitrags: John Hackston)

ACIL

Adaptive Computergestützte Intelligenz-Lerntestbatterie

Autoren/Herausgeber

Jürgen Guthke (Hrsg.)
Jens F. Beckmann, Henrik Stein, Silke Rittner und Hans Vahle (Autoren)

Verlage, Erscheinungsorte und -jahre

Verlag Dr. Schuhfried, Mödling (Österreich), 1994
Hogrefe-Verlag, Göttingen, 1998

Einsatzbereiche

Zunächst wurde das Verfahren entwickelt für Jugendliche im Alter von 12-18 Jahren. Es soll für Beratungs- und Auswahlzwecke nutzbar sein, ähnlich wie die herkömmlichen Intelligenzstatustests. Dabei wird davon ausgegangen, dass eine Lerntestbatterie gegenüber herkömmlichen rein konstatierenden Intelligenzstatustests eine Reihe von Vorteilen (s.u.) aufweist und diagnostische Zusatzinformationen zur Verfügung stellt.

Insgesamt gesehen sind die Hauptzielgruppen von Lerntests Menschen, die unter irregulären Lernbedingungen aufgewachsen sind und deren Intelligenzpotenz (Lernfähigkeit) daher häufig unterschätzt wird, wenn man nur in herkömmlichen Intelligenztests ihren gegenwärtigen Status prüft, der bekanntlich auch wesentlich das Ergebnis einer „Lerngeschichte“ ist.

Zusatzinformationen sind zur Sicherung einer faireren Intelligenzdiagnostik vor allem für folgende Adressatengruppen, bei denen der „reine“ Intelligenzstatus über die wahre intellektuelle Potenz hinwegtäuschen kann, von besonderer Bedeutung:

- Untersuchung von „Schul- und Berufsversagern“ aus ungünstigem häuslichen Milieu
- Jugendliche mit „Sinnesstörungen“ und minimaler cerebraler Dysfunktion
- lernbehinderte Jugendliche und Jugendliche mit schlechtem Schulabschluss
- Immigrantenkinder und „Gastarbeiter“
- besonders testängstliche, neurotische und impulsive Testanden
- ältere und ausländische Arbeitnehmer
- Erwachsene, die wegen körperlichen Erkrankungen auf Büroberufe umgeschult werden müssen

Konstruktionsgrundlagen

Der Test soll den „Kernfaktor der Intelligenz“ (Thurstone), das sog. schlussfolgernde Denken (Reasoning) überprüfen. Vornehmlich wird schlussfolgerndes Denken hier als induktives Den-

ken untersucht, wobei Regelmäßigkeiten und Relationen auf der Basis von erkannten Gemeinsamkeiten und Unterschieden zwischen „Objekten“ (Zahlen, Wörtern, Figuren) vom Testanden gefunden werden müssen. „Reasoning“ erweist sich nicht nur in allen faktoranalytischen Modellen der Intelligenz als der wichtigste Intelligenzfaktor (vor allem der sog. fluiden Intelligenz), sondern darüber hinaus zeigen Validitätsuntersuchungen im Schul- und Berufsbereich, dass Reasoning-Tests unter den Intelligenztests sich als die besten Prädiktoren von Schul- und Berufserfolg erweisen. Insbesondere stehen sie auch im engen Zusammenhang mit der Leistungsfähigkeit im mathematisch-technisch-naturwissenschaftlichen Bereich. Abschätzungen der „Gesamtintelligenz“ lassen sich wegen der hohen Korrelationen der Reasoning-Tests mit dem Gesamtergebnis umfangreicher Testbatterien und mit sog. Außenkriterien der Intelligenz noch am ehesten auf Grund von Reasoning-Tests vornehmen.

Die in der ACIL benutzten Aufgabentypen – nämlich Zahlen- und Figurenfolgen, verbale Analogien – sind Bestandteile fast aller bewährten Intelligenztests und auch Gegenstand vieler kognitionspsychologischer Grundlagenarbeiten, die wir der Konstruktion der einzelnen Testaufgaben (Beckmann & Guthke, 1999; s.u.) zugrunde legten. Die ACIL verwirklicht in ihrem Aufbau die Verbindung von folgenden Prinzipien und aktueller Anforderungen an die Intelligenzdiagnostik:

Kognitionspsychologische Fundierung der Itempools

Die Items werden auf der Grundlage kognitionspsychologischer Erkenntnisse konstruiert und nach ihrer objektiven Komplexität im Test gestaffelt. Das heißt, die ansonsten in der „klassischen Testtheorie“ üblichen rein statistischen Schwierigkeitsbestimmungen der Items werden ergänzt durch eine „objektive Schwierigkeits- bzw. Komplexitätsbestimmung“. Ein solcher kontentvalider Testaufbau gestattet eher eine adaptive Testung, fehlerorientierte Hilfen während der Testung (s.u.) und auch Prozessanalysen.

Adaptive Testvorgabe

Hierunter wird bei der ACIL verstanden, dass sehr leistungsfähige Testanden schnell durch das Programm „springen“ können, während leistungsschwächere Testanden zusätzliche Aufgaben und Hilfen bekommen. Das Programm enthält sog. Targetitems in den verschiedenen Komplexitätsstufen, die alle Testanden bearbeiten müssen. Durch die Reaktion bei den Targetitems „erkennt“ das Programm, ob der Testand nicht zunächst erst noch leichtere Aufgaben benötigt, bevor er wiederum zum Targetitem geführt wird. Die fehlerorientierte Auswertung bei einzelnen Übungsaufgaben verweist den Testanden auf „Denkhilfen“, die ihn dazu führen soll, beim nächsten „Anlauf“ das Item zu lösen. Es handelt sich hierbei also um eine Verquickung der Technik des adaptiven Testens (allerdings nicht in der „klassischen Form“) mit dem Lerntestkonzept.

Lerntestkonzept

Die ACIL ist ein sog. Kurzzeit-Lerntest vom Typ des Diagnostischen Programms (Beckmann & Guthke, 1999), der das „Veränderungspotential“ (bzw. die sog. Zone der nächsten Entwicklung im schlussfolgernden Denken sensu Wygotski) untersuchen soll. Damit ist die Potenz des Testanden gemeint, von Feedbacks und Denkhilfen im Testprozess zu profitieren, so dass er seine Leistung während der Testung verbessern kann. Geprüft wird also, welchen Nutzen jemand aus standardisierten Lernanregungen im Test für den weiteren Testverlauf zieht. Lerntests versuchen

der Forderung nach einer „dynamischen Testdiagnostik“ (Guthke & Wiedl, 1996) zu entsprechen, die jetzt auch international immer stärker als eine innovative Weiterentwicklung der herkömmlichen Intelligenzdiagnostik betrachtet wird (z.B. Grigorenko & Sternberg, 1998). Ursprünglich wurde das Lerntestkonzept (Guthke, 1977) durch sog. Langzeit-Lerntests realisiert, die nach dem Schema Prätest-Trainingsphase-Posttest aufgebaut waren. Der Lerntest „Schlussfolgerndes Denken“ (LTS von Guthke, Jäger & Schmidt, 1983), dessen inhaltlicher Aufbau auch der ACIL zugrunde liegt, ist ein solcher Langzeit-Lerntest. Die Kurzzeit-Lerntests sind beträchtlich zeitökonomischer, da sie wie der übliche Intelligenztest nur aus einer Sitzung bestehen, in die die Trainingsphase durch Denkhilfen und Feedbacks eingebaut ist (s.o.).

Inhalte

Nach modernen faktoranalytischen Theorien (vgl. Beckmann & Guthke, 1999) zum Aufbau der Intelligenz (bzw. auch des Reasoning-Faktors) müssen unter der inhaltlichen Facette zumindest drei Materialbereiche beachtet werden. Es handelt sich hierbei um den verbalen, den mathematisch-numerischen und den figuralen Bereich. Diesen Erkenntnissen trägt die ACIL Rechnung. Sie besteht aus einem verbalen Analogientest (Adaptiver Analogien-Lerntest, ADANA), einem Zahlenfolgentest (Adaptiver Zahlenfolgen-Lerntest, AZAFO) und einem Figurenfolgentest (Adaptiver Figurenfolgen-Lerntest, ADAFI). Die Tests können einzeln ausgewertet und angewandt werden. Man kann aber durch Summierung der Einzelwerte einen Gesamtwert bilden. Nach dem bimodalen Intelligenzstrukturmodell von Jäger (Jäger, Süß & Beauducel, 1997) erfasst die ACIL in der operativen Facette die sog. Verarbeitungskapazität (entspricht dem Reasoning-Faktor) und in der inhaltlichen Facette alle drei Materialbereiche, allerdings im Unterschied zum Berliner Intelligenz-Struktur-Test (vgl. BIS-4, in diesem Band) in einer Lerntestvariante.

Frage- und Antwort-Formate

Der ADAFI und der ADANA sind Multiple-Choice-Verfahren; der AZAFO erfordert Ergänzungsantworten (Fortsetzen der Zahlenfolge)

Subtests

Die drei Untertests ADANA, AZAFO und ADAFI können auch einzeln vorgegeben werden.

Gütekriterien

Objektivität/Konkordanz

Durch die computergestützte Darbietung und Auswertung des Tests ist die Objektivität der Testdurchführung und formalen Testauswertung gesichert. Die Testinterpretation wird durch die Vorgabe von Testnormwerten und Auswertungsbeispielen standardisiert.

Zuverlässigkeit

Das klassische Konzept der Reliabilitätsmessung ist für Lerntests nur bedingt tauglich. Schließlich werden ja im Testprozess Veränderungen der individuellen Rangplätze während des Testablaufs direkt angezielt, so dass ein Testschwacher in der ersten Hälfte des Tests durchaus ein

zumindest „Durchschnittlicher“ in der zweiten Hälfte werden kann (vgl. Konzept der Testhalbierung als Schätzer für Zuverlässigkeit). Eigentlich wäre eine Vierfachtestung notwendig, um die Reliabilität von Lerntests exakter zu bestimmen, was aus untersuchungsökonomischen Gründen nicht möglich war. Die trotz dieser Bedenken berechneten Zuverlässigkeitskennwerte mit der Testhalbierungsmethodik und mit Cronbachs Alpha (Datenerhebungen bis 1999) betragen beim

- ADANA: N = 886 Schüler der Klassen 5-9, Testhalbierung: .68 (nach Aufwertung auf die Gesamtlänge .81) Cronbachs Alpha: .70,
- AZAFO: N = 692 Schüler der Klassen 5-9, Testhalbierung: .84 (aufgewertet:.93) Alpha: .77 und beim
- ADAFI: N = 772 Schüler der Klassen 5-9, Testhalbierung: .68 (aufgewertet: .85) Alpha: .61).

Gültigkeit (Validität)

Zunächst ist im Konzept einer externen Validitätsprüfung zu fragen, ob die ACIL mit herkömmlichen, bewährten Verfahren zur Erfassung von Reasoning korreliert. Zumindest sollte nachgewiesen werden, dass die ACIL höher mit Reasoning-Tests korreliert als mit Intelligenztests, die andere Faktoren prüfen. Dieser Nachweis ist durch Einsatz des Raven-Tests und des Kognitiven Fähigkeitstests (KFT, Heller et al., 1985) gelungen, wobei alle relevanten Tests auf einem Faktor laden (Beckmann & Guthke, 1999).

Im Rahmen einer mehr internen Validitätsprüfung ist aber die Frage wichtig, ob das Antwortverhalten in Lerntests möglicherweise allein durch den Status bestimmt wird und die zusätzliche Annahme eines „Lernfähigkeitsparameters“ daher überflüssig ist. Exploratorische, konfirmatorische Faktoranalysen und vor allem auch die Anwendung spezieller Rasch-Modelle ließen erkennen (vgl. ausführlich hierzu Beckmann & Guthke, 1999), dass hypothesengemäß Status- und Lerntests zum Reasoning inhaltlich das Gleiche messen (eben schlussfolgerndes Denken), dass aber die prozeduralen Unterschiede von Status- und Lerntests deutliche Auswirkungen auf das Antwortverhalten haben. Dieses lässt sich nur durch ein Antwortmodell genügend gut erklären, das neben einem Intelligenzstatusfaktor das Wirken eines „Lernfaktors“ annimmt (also ein Zweifaktorenmodell darstellt).

Externe Validierungen an „Außenkriterien“ wie Zensuren und Lehrereinschätzungen ergaben sowohl bei der konkurrenten als auch bei der prognostischen Validierung (zwei Jahre nach Testaufnahme) eher mäßige Korrelationen (vergleichbar mit üblichen Intelligenztests). Allerdings wird von den Autoren sehr infrage gestellt, ob es sinnvoll ist, Intelligenztests bzw. Lerntests - wie bisher üblich - an solch fragwürdigen Außenkriterien für Intelligenz wie Zensuren und Lehrerurteilen zu validieren. Schließlich werden ja Intelligenztests entwickelt und angewandt, da man die Validität dieser Außenkriterien anzweifelt. Favorisiert werden daher von den Testautoren experimentelle Validierungen, bei denen die Ergebnisse von Intelligenzstatus- und Lerntests den Lernerfolg unter standardisierten Bedingungen in „Lernversuchen“ (z.B. zwei Jahre nach der Testung) vorhersagen sollen. Bei Verwendung von Computerszenarios mit curriculumsfernen Lerninhalten (das Funktionieren einer fiktiven Maschine musste erlernt werden; siehe Beckmann, 1994) und curriculumsnahen Lernstoffen (aus Geografie und Mathematik) zeigte sich, dass die Varianzaufklärung dieser experimentell gewonnenen Außenkriterien bei Benutzung von Lerntests deutlich verbessert wird. Kommunalitätenanalysen ergaben, dass die Informationen von Statustests auch im Lerntest enthalten sind, aber umgekehrt nicht die diagnostische Information des Lerntests im Statustest. Einzelangaben über die Vielzahl von externen und

internen Validitätskennwerte aus unterschiedlichen Untersuchungen findet der Testanwender in der Monografie zum Test von Beckmann und Guthke (1999).

Normen

Die Handanweisungen enthalten Normen für die Klassenstufen 5-9 (getrennt auch nach Schultypen). Der Test ist aber bereits bei mehreren Hundert Testanden im Arbeitsamt und in beruflichen Reha-Einrichtungen (1998) angewandt worden. Es handelt sich hierbei meist um Personen mit körperlichen und psychischen (schlechter Schulabschluss) Beeinträchtigungen. Mittelwerte und Standardabweichungen für diese Gruppen können bei Bedarf bei den Testautoren angefordert werden. Auch für den Personenkreis von Studenten liegen erste Normwerte vor.

Testmaterial

Die ACIL ist ein vollständig (Instruktion, Durchführung und Auswertung) computergestütztes Verfahren.

Bearbeitungsbedingungen und -dauer

Die ACIL ist im Rahmen von Einzelsitzungen durchführbar. Der Einsatz in Kleingruppen (bis ca. 10 Testanden) hat sich ebenfalls bewährt. Während der Testung muss – den Richtlinien für den Einsatz computergestützter Diagnostik des Testkuratoriums (1996) entsprechend – ein Ansprechpartner zugegen sein. Diese Funktion kann durch Nicht-Psychologen wahrgenommen werden.

Ergebnisrückmeldung

Die Rückmeldung der Ergebnisse muss von einem Psychologen vorgenommen werden. Neben der angemessenen Interpretation der Normwerte in den einzelnen Testparametern (Menge benötigter Hilfen und Übungsaufgaben), ist eine Interpretation der Leistungsparameter in den einzelnen Komplexitätsstufen der Tests möglich. Des Weiteren kann anhand des Testprotokolls – welches bis auf Itemebene rekonstruierbar ist – z.B. anhand der Latenzzeiten die Aufgabenorientiertheit im gesamten Testverlauf kontrolliert werden. Aufgrund aktueller Forschungsergebnisse ist die Zuordnung des individuellen Testverhaltens zu sog. Verlaufstypen möglich (prozessorientierte Auswertung).

Publikationen

Beckmann, J.F. (1994). *Lernen und komplexes Problemlösen*. Ein Beitrag zur Konstruktvalidierung von Lerntests. Bonn: Holos.

Beckmann, J.F. & Dobat, H. (2000). *Zur Validierung der Diagnostik intellektueller Lernfähigkeit*. Zeitschrift für Pädagogische Psychologie, 14, 97-105.

Beckmann, J.F. & Guthke, J. (1999). *Psychodiagnostik des schlußfolgernden Denkens*. Göttingen: Hogrefe.

Guthke, J. & Wiedl, K.-H. (1996). *Dynamisches Testen. Zur Psychodiagnostik der intraindividuellen Variabilität*. Göttingen: Hogrefe.

Guthke, J., Beckmann, J.F., Stein, H., Rittner, S. & Vahle, H. (1995). *Adaptive Computer-gestützte Intelligenz-Lerntestbatterie* (ACIL). Mödling: Schuhfried.

Erfahrungen in der Wirtschaftspraxis

Das Verfahren wurde bisher vor allem im Arbeitsamt und in Berufsförderungswerken angewandt. Zurzeit sind die Validitätsuntersuchungen noch nicht abgeschlossen, so dass über die Anwendbarkeit in der Wirtschaftspraxis und im Arbeitsamt noch keine Aussagen getroffen werden können. Erhofft wird allerdings von den Autoren, dass die ACIL insbesondere bei Testanden mit Leistungsschwächen (aber nicht bei geistig Behinderten, dafür ist sie zu schwierig) aus irregulären Lernbedingungen zusätzliche diagnostische Informationen liefert (s.o.). In der Arbeitsgruppe von Sarges (Mitherausgeber dieses Bandes) wurde die ACIL mit Erfolg auch bei Studenten der Pädagogik und des Maschinenbaus eingesetzt. Nach unseren eigenen Erfahrungen mit Testungen bei Hochschulabsolventen wird beim ADAFI und AZAFO selten die Testdecke erreicht (hoch Leistungsfähige durchlaufen den Test allerdings wie einen Statustest, da sie keine Hilfen benötigen und mit wenigen Targetitems ganz schnell am Ziel sind).

Da in der Wirtschaft immer mehr die „Lernfähigkeit“ als Schlüsselqualifikation betont wird, dürfte es interessant sein zu sehen, ob in künftigen wirtschaftsnäheren Anwendungen der ACIL auch bei Studierenden und erwachsenen Arbeitnehmern Hinweise auf diese Schlüsselqualifikation zu erhalten sind. In allen modernen Untersuchungen über die Vorhersage des Berufserfolges in den verschiedensten Berufsfeldern haben sich die Tests zum schlussfolgernden Denken als die besten Prädiktoren auf der Intelligenzseite erwiesen, so dass auch von der inhaltlichen Ausrichtung die ACIL ein interessantes Verfahren für die Berufseignungsdiagnostik werden könnte.

Preise

50 Testvorgaben :		237 €	
7 Jahre Nutzungsrecht (exkl. MwSt.):		1.422 €	
einzelne Verfahren:			
	ADANA	ADAFI	AZAFO
50 Testvorgaben:	112 €	121 €	112 €
7 Jahre Nutzungsrecht:	672 €	726 €	672 €

Bezugsadressen

Dr. G. Schuhfried GmbH
Hyrtlstraße 45
A – 2340 Mödling Fon: +43 (0)2236/423 15
Fax: +43 (0)2236/465 97

Email: info@schuhfried.at
Internet: www.schuhfried.at

Testzentrale Göttingen
Robert-Bosch-Breite 25
D – 37027 Göttingen
Fon: +49 (0)551/506 88 -0/-14/-15
Fax: +49 (0)551/506 88 -24
Email: testzentrale@hogrefe.de
Internet: www.testzentrale.de

(Verfasser des Beitrags: Gerald Schuhfried)

AHA

Arbeitshaltungen

Autoren

Klaus D. Kubinger und Johannes Ebenhöh

Verlage, Erscheinungsorte und -jahre

Swets Test Services, Frankfurt, 1996
Verlag Dr. Schuhfried, Mödling (Österreich), 1998

Einsatzbereiche

Berufliche Eignungsdiagnostik; klinische Fragestellungen in Bezug auf Arbeitsbelastung, burnout.

Konstruktionsgrundlagen

Die „Arbeitshaltungen" zählen zu den sog. „Objektiven Persönlichkeitstests" sensu R.B. Cattell und dienen zur Erfassung des Arbeits- und Kontaktverhaltens einer Person bei Leistungsanforderung. Wesentlich ist dabei nicht, was oder wie viel verarbeitet wird, sondern mit welchem persönlichen Verhaltensstil dies geschieht. Gemessen wird Impulsivität/Reflexivität, Anspruchsniveau, Frustrationstoleranz und Leistungsmotivation.

Zu den am häufigsten zitierten kognitiven Stilen zählt die Polarisierung in die Typen „impulsiv" und „reflexiv". Im wesentlichen wird dabei zwischen Personen zu differenzieren versucht, die in Problemsituationen entweder langsam und fehlerarm oder schnell und fehlerreich arbeiten, wobei diese Polarisierung nicht wertend in Richtung „besser" oder „schlechter" zu verstehen ist. Das Konstrukt „Leistungsmotivation" ist in der Testbatterie AHA an die Definition von McClelland als eine „Auseinandersetzung mit einem Tüchtigkeitsmaßstab" angelehnt. Im reinen Fall ist Leistungsgüte oder Leistungsmenge ein Ziel um seiner selbst willen, wobei in der Regel auch bestimmte Nebenziele mitspielen, die nicht leistungsthematischer Natur sind – wie etwa Ansehen erringen oder materiellen Gewinn erzielen. Um Fähigkeitsaspekte bei der Messung dieses Konstrukts auszuschalten, muss die als Test indizierte Aufgabenstellung ziemlich leicht und nicht ge-„speedet" sein; somit ist die gemessene Eigenschaft aber auch (konzentrierte) „Ausdauer" oder „Willensanstrengung" zu nennen. Modernere Betrachtungen zum Konstrukt „Leistungsmotivation" berücksichtigen auch situationsbedingende Faktoren bzw. schließen die beiden generalisierten, d.h. dispositionalen Motive „Hoffnung auf Erfolg" und „Furcht vor Misserfolg" mit ein. In den Arbeitshaltungen wird dies mit der Variable Anspruchsniveau erfasst. Unter „Frustrationstoleranz" ist die Eigenschaft einer Person zu verstehen, Frustrationen

über einen längeren Zeitraum auszuhalten, ohne das ursprüngliche Ziel der Motivbefriedigung, die Motivation, aufzugeben. In der Testbatterie AHA interessiert lediglich der leistungsbezogene Aspekt der Frustration, also die Kollision von (vermeintlichem) Leistungsniveau und Anspruchsniveau.

Ausgangspunkt der Entwicklung der Testbatterie Arbeitshaltungen war der Versuch von Kubinger (1996), den Nachweis zu führen, dass herkömmliche Persönlichkeitsfragebögen regelmäßig verfälscht sind und dass die mit ihnen zu messen beabsichtigten Eigenschaften valide mittels sog. „Objektiver Persönlichkeitstests" zu erfassen sind.

Inhalte

Folgende Dimensionen werden erfasst: Impulsivität/Reflexivität, Anspruchsniveau, Frustrationstoleranz, Leistungsmotivation (detaillierter unter „Subtests").

Frage- und Antwort-Formate

Der erste Untertest „Flächengrößen Vergleichen" erlaubt drei Antwortmöglichkeiten: Auswählen der linken oder rechten Fläche sowie keine Entscheidung treffen. Im zweiten Untertest muss die Testperson Symbole zuordnen und die Anzahl der Kodierungen prognostizieren, welche sie in der nächstfolgenden Etappe zu leisten erwartet. Im dritten Untertest hat die Testperson aus einer Reihe von vier geometrischen Figuren jene auszuwählen, die sich von den anderen unterscheidet.

Subtests

Flächengrößen Vergleichen

Der Untertest „Flächengrößen Vergleichen" dient der Erfassung von Impulsivität/Reflexivität und besteht aus 20 Items. Es werden jeweils zwei Bilder mit unregelmäßigen Flächen nebeneinander dargeboten. Innerhalb der Bearbeitungszeit von insgesamt 30 Sekunden soll beurteilt werden, ob das linke oder das rechte Bild die größere Fläche beinhaltet. Als dritte Alternative steht eine Schaltbutton „keine Entscheidung" zur Verfügung. Die Bildpaare sind so konstruiert, dass eine Entscheidung oft schwer fällt und insbesondere Personen impulsiven Typs zum Raten (und somit zu Fehlern) verleitet werden. Für diesen Speed-Untertest werden drei Testkennwerte berechnet:

- *Exaktheit*: der relative Anteil richtiger Antworten.
- *Entschlussfreudigkeit*: die Anzahl getroffener Entscheidungen unabhängig davon, ob diese richtig oder falsch war.
- *Impulsivität/Reflexivität*: eine hierarchische Verschlüsselung der Anzahl richtiger und falscher Entscheidungen sowie der Wahl, keine Entscheidung zu treffen. Die Unterscheidung in Impulsivität versus Reflexivität wird dominiert von der absoluten Fehleranfälligkeit einer Versuchsperson. Die Tendenz, keine Entscheidung zu treffen, spricht für den reflexiven Typ.

Symbole Kodieren

Der Untertest „Symbole Kodieren" erfasst die Variablen Anspruchsniveau und Frustrationstoleranz. Vier abstrakten, schwarzweißen Symbolen sind nach einem (permanent dargebotenen und nicht variierten) Kodierschlüssel vier abstrakte, farbige Symbole zuzuordnen. Nach der ersten Etappe erfolgt eine Rückmeldung über die Zahl der bearbeiteten Symbole sowie die Aufforderung, eine Prognose abzugeben, wie viele Symbole in der nächsten Etappe bearbeitet werden. Nach der zweiten Etappe erfolgt zusätzlich eine Frustration, die darin besteht, dass fälschlich darüber informiert wird, dass andere Versuchspersonen eine höhere Leistung erbringen. Die genannte Leistung anderer Testpersonen wird individuell so gewählt, dass sie um ca. 10% über der Leistung der Testperson liegt. Es werden fünf Testkennwerte für diesen Speed-Untertest berechnet:

- *Leistungsniveau*: Anzahl der richtigen Kodierungen innerhalb der zweiten Etappe
- *Anspruchsniveau*: die Differenz von erster Prognose und darauf folgender zweiter Leistung; Erfolgsmotivierte sollen einen absolut kleinen Testwert erreichen, Misserfolgsmotivierte entweder hoch positive oder hoch negativ Testwerte.
- *Frustrationstoleranz*: Differenz zwischen fünfter und zweiter Prognose; das Vorzeichen dieser Differenz soll anzeigen, ob sich die Testperson durch die negativen Rückmeldungen gravierend beeindrucken lässt (und ihre Prognosen dementsprechend anpasst) oder nicht.
- *Zeitpunkt der Leistungsspitze*: diejenige Etappe, in der die Versuchsperson die größte Anzahl richtiger Kodierungen erreicht.
- *Zieldiskrepanz*: die Summe der vier absoluten Abweichungen zwischen Prognose und darauf folgender Leistung

Figuren Unterscheiden

Der Untertest „Figuren Unterscheiden" misst die Leistungsmotivation der Testperson. Pro Bildschirmseite werden sechs Items, bestehend aus vier einfachen geometrischen Figuren, dargeboten. Je einer der Figuren fehlt ein beliebiges Strichelement, die Figur ist durch Mausklick zu markieren. Die Testvorgabe erfolgt so lange, bis die Testperson sie beendet. Die Instruktion fordert dazu auf, den Test dann zu beenden, wenn der Eindruck entsteht, dass nicht mehr konzentriert gearbeitet wird. Als Testkennwert wird die Leistungsmotivation bestimmt, d.h. die Anzahl richtig bearbeiteter Items.

Gütekriterien

Objektivität

Wie alle normierten Computertests weist die Testbatterie Arbeitshaltungen maximale Objektivität, d.h. Testleiterunabhängigkeit, Verrechnungssicherheit und Interpretationseindeutigkeit auf.

Reliabilität

Da die Bestimmung eines Standardmessfehlers für die Arbeitshaltungen nicht möglich bzw. sinnvoll ist, kann auch der Messfehler im Einzelfall nicht mittels Konfidenzintervall abgeschätzt werden.

Validität

Von der inhaltlichen Gültigkeit der erhobenen Kennwerte kann ausgegangen werden. Durchgeführte Validierungsstudien zeigten beispielsweise, dass die Untertests „Flächengrößen Vergleichen" und „Symbole Kodieren" signifikant zur Diskrimination von Mitarbeitern beitragen; dem Charakteristikum eines kognitiven Stils entsprechend zeigt sich, dass je nach beruflicher Anforderung unter den „erfolgreichen" Mitarbeitern einmal die eher impulsiven Persönlichkeiten zu finden sind, das andere Mal eher die reflexiven. Die Ergebnisse decken sich mit einer Studie von Frebort (2002), in der die Testkennwerte der AHA signifikant und hypothesenkonform zwischen erfolgreichen und weniger erfolgreichen Auszubildenden trennen. Wagner-Menghin (2003) konnte anhand zweier unabhängiger Stichproben mittels KFA übereinstimmend zeigen, dass sich mit Hilfe der AHA drei Typen von Testpersonen unterscheiden lassen, die sich auch signifikant in Leistungstests unterscheiden: der Typus „Hoffnung auf Erfolg", „Furcht vor Misserfolg" und „Meiden von Misserfolg".

Unverfälschbarkeit

Die hier beschriebenen objektiven Persönlichkeitstests der AHA erweisen sich als praktisch nicht verfälschbar: Kontroll- und Versuchsgruppe bei Kubinger unterscheiden sich nicht signifikant in ihren durchschnittlichen Testwerten, obwohl letztere im Gegensatz zu ersterer instruiert wurde, „sich ganz genau in die Situation eines Bewerbers zu versetzen: ...Sie werden sich daher sehr bemühen, gut abzuschneiden". Die Verfälschungsresistenz konnte von Hofmann und Kubinger (2001) sowie Benesch (2003) ebenfalls nachgewiesen werden. Dieses Ergebnis ließ sich auch in einer Studie von Greifes (2002) replizieren, in der auch gezeigt werden konnte, dass sich die Testkennwerte der Arbeitshaltungen nach unterschiedlichen Feedbacks über die Messinstruktion nicht signifikant verändern.

Normen

Es liegt eine repräsentative Normierungsstichprobe (N = 231) von 99 (42,9%) Männern und 132 (57.1%) Frauen im Alter von 18 bis 84 Jahren vor. Die Normen stehen auch getrennt für zwei Altersgruppen zur Verfügung. Darüber hinaus liegt eine Evaluierungsstichprobe von N = 314 (großteils Studenten) vor. Zusätzlich gibt es noch eine Erwachsenen Stichprobe (20-60 Jahre; N = 134). An einer Weiternormierung des Verfahrens wird gearbeitet.

Testmaterial

Das Testverfahren wird computergestützt vorgegeben und ausgewertet. Die Bearbeitung erfolgt mittels Maus.

Bearbeitungsbedingungen und -dauer

Die Durchführungsdauer des Untertests „Flächengrößen Vergleichen" und „Symbole Kodieren" beträgt ca. 15 Minuten. Für den Untertest „Figuren Unterscheiden" ist durchschnittlich mit einer Bearbeitungszeit von 20 Minuten, in Einzelfällen bis zu 45 Minuten, zu rechnen.

Ergebnisrückmeldungen

Die Darstellung der Ergebnisse erfolgt am PC-Bildschirm und kann zudem ausgedruckt werden. Sie umfasst eine Ergebnistabelle mit Roh- und Standardwerten (Prozentrangwerte und T-Werte), ein Testprofil, je ein Verlaufsprotokoll und ein Testprotokoll für jeden Subtest. Das Testprofil ermöglicht einen Vergleich der einzelnen Skalenwerte auf der Basis von Prozentrangwerten und T-Werten. Der Verlauf des Tests über die einzelnen Durchgänge wird anhand der Variablen „Richtige", „Falsche" und „Prognose" in einer Tabelle dargestellt. Die Testprotokolle für die Subtests geben Auskunft über alle Einzelantworten. Zur Ergebnisrückmeldung an den Probanden – falls nötig und gewünscht – bedürfen diese Daten der mündlichen Erläuterung durch einen geschulten Testanwender.

Publikationen

Benesch, M. (2003). *Zur Verfälschbarkeit von Persönlichkeitsfragebögen und Objektiven Persönlichkeitstests: Reaktionszeiten (Dissertation).* Wien: Universität.

Frebort, M. (2002). *Evaluation einer Batterie Objektiver Persönlichkeitstests zur Auswahl von TierpflegeschülerInnen* (Dipl. Arbeit). Wien: Universität.

Greifes, U. (2002). *Eine experimentelle Untersuchung über den Einfluss der Vorstellung der Testpersonen zur Messintension der kurzen Testbatterie Arbeitshaltungen* (Dipl. Arbeit). Wien: Universität.

Hofmann, K. & Kubinger, K.D. (2001). *Herkömmliche Persönlichkeitsfragebogen und Objektive Persönlichkeitstests im Wettstreit um (Un-) Verfälschbarkeit*. Report Psychologie, 26, 298-304.

Wagner-Menghin, M. (2003). *Identifying achievement-motivation-types with CFA: Interpreting the test battery „working style" as a new instrument in assessing the achievement motive*. In E. Lautsch & A. von Weber (Eds.), Configural frequency analysis (CFA) – problems and applications (Supplement: Psychology Science).

Erfahrungen in der Wirtschaftspraxis

Die Testbatterie AHA eignet sich im Bereich der Personalselektion dort, wo der Erfolg vor allem davon abhängt, ob der Mitarbeiter an sich hohe oder weniger hohe Ansprüche in Bezug auf die berufliche Leistung stellt. Ein hohes Anspruchsniveau ist zum Beispiel vor allem bei relativ frei wählbarer Arbeitsleistung etwa im Bereich Außendienst, Kundenbetreuung etc. günstig. Die Arbeitshaltung wird mit großem Erfolg in den Bereichen Auswahl von Call Center Mitarbeitern, Außendienstmitarbeitern, Filialleitern und Mitarbeitern in der Tourismusbranche, Sachbearbeitern und gewerblichen Mitarbeitern angewendet.

Sonstiges

Der Test ist mehreren Sprachen erhältlich.

Preise

50 Testvorgaben:	81 €
7 Jahre Nutzungsrecht:	486 €

Bezugsadressen

Harcourt Test Services
Baseler Straße 35-37
D – 60329 Frankfurt/M.
Fon: +49 (0)69/75 61 46 11
Fax: +49 (0)69/75 61 46 10
Email: ralfhorn@harcourt.de
Internet: www.harcourt.de

Dr. G. Schuhfried GmbH
Hyrtlstraße 45
A – 2340 Mödling
Fon: +43 (0)2236/423 15
Fax: +43 (0)2236/465 97
Email: info@schuhfried.at
Internet: www.schuhfried.at

(Verfasser des Beitrags: Gerald Schuhfried)

AIST-R

Allgemeiner Interessen-Struktur-Test

Autoren

Christian Bergmann und Ferdinand Eder

Verlag, Erscheinungsort und -jahr

Beltz Test, Göttingen, 2004

Einsatzbereiche

Berufs- und Laufbahnentscheidungen von Jugendlichen und Erwachsenen (Berufsorientierung; Ausbildungs-, Studien- und Berufsberatung; innerbetriebliche Personalentwicklung).

Konstruktionsgrundlagen

Der Test basiert auf dem Person-Umwelt-Modell von J. Holland (1973, 1985, 1997); siehe im Folgenden unter „Inhalte".

Inhalte

Nach der berufsbezogenen Persönlichkeitstheorie von J. Holland gibt es in unserem Kulturkreis sechs grundlegende Interessen- bzw. Persönlichkeitsorientierungen. Der AIST erfasst diese sechs Interessendimensionen, die als Persönlichkeitsorientierungen interpretiert werden können:

(1) „Praktisch-technische Orientierung" (R - realistic)

Personen dieses Typs bevorzugen Tätigkeiten, die Kraft, Koordination und Handgeschicklichkeit erfordern und zu konkreten, sichtbaren Ergebnissen führen. Sie weisen Fähigkeiten und Fertigkeiten vor allem im mechanischen, technischen, elektrotechnischen und landwirtschaftlichen Bereich auf, während sie erzieherische oder soziale Aktivitäten eher ablehnen.

(2) „Intellektuell-forschende Orientierung" (I - investigative)

Personen dieses Typs bevorzugen Aktivitäten, bei denen die Auseinandersetzung mit physischen, biologischen oder kulturellen Phänomenen mit Hilfe systematischer Beobachtung und Forschung im Mittelpunkt steht. Sie weisen Fähigkeiten und Fertigkeiten vor allem im mathematischen und naturwissenschaftlichen Bereich auf.

(3) „Künstlerisch-sprachliche Orientierung“ (A - artistic)

Personen dieses Typs bevorzugen offene, unstrukturierte Aktivitäten, die eine künstlerische Selbstdarstellung oder die Schaffung kreativer Produkte ermöglichen. Ihre Fähigkeiten liegen vor allem im Bereich von Sprache, Kunst, Musik, Schauspiel und Schriftstellerei.

(4) „Soziale Orientierung“ (S - social)

Personen dieses Typs bevorzugen Tätigkeiten, bei denen sie sich mit anderen in Form von Unterrichten, Lehren, Ausbilden, Versorgen oder Pflegen befassen können. Ihre Stärken liegen im Bereich der zwischenmenschlichen Beziehungen.

(5) „Unternehmerische Orientierung“ (E - enterprising)

Personen dieses Typs bevorzugen Tätigkeiten und Situationen, bei denen sie andere Menschen mit Hilfe der Sprache oder anderer Mittel beeinflussen, zu etwas bringen, führen, auch manipulieren können. Ihre Stärken liegen im Bereich der Führungs- und Überzeugungsqualität.

(6) „Konventionelle Orientierung“ (C - conventional)

Personen dieses Typs bevorzugen Tätigkeiten, bei denen der strukturierte und regelhafte Umgang mit Daten im Vordergrund steht, z.B. Aufzeichnungen führen, Daten speichern, Dokumentationen führen, mit Büromaschinen arbeiten u.ä. (ordnend-verwaltende Tätigkeiten). Ihre Stärken liegen im Bereich rechnerischer und geschäftlicher Fähigkeiten.

Das Berufsregister des Testmanuals enthält etwa 300 Berufe sowie Ausbildungsmöglichkeiten, die nach diesen Dimensionen (RIASEC) charakterisiert sind. Sie können als Grundlage für die Überprüfung der Person-Umwelt-Kongruenz bzw. für die Auswahl persongemäßer (kongruenter) beruflicher Umwelten herangezogen werden.

Frage- und Antwort-Formate

Der AIST umfasst pro Dimension je 10, insgesamt 60 Items mit Aktivitäten, zu denen auf einem 5-stufigen Antwortformat das Ausmaß des Interesses angegeben wird.

Subtests

Der AIST erfasst zwar 6 gut unterscheidbare Dimensionen, die aber wegen ihres Modellzusammenhanges sinnvollerweise als Ganzes erhoben werden sollen.

Gütekriterien

Das Verfahren weist die für Fragebogenverfahren zu erwartende Objektivität und gute Kennwerte für die Reliabilität (Cronbachs Alpha in der Konstruktionsstichprobe zwischen .82 und .87) auf. Es liegen umfangreiche Validitätsanalysen hinsichtlich der Gültigkeit der zentralen Konzepte (Kongruenz, Differenziertheit) sowie zur prognostischen Validität des Verfahrens mit überzeugenden Ergebnissen vor (vgl. Handanweisung bzw. Literatur s.u.).

Normen

Es existieren Gesamtnormen (N = 2.496) und Geschlechtsnormen.

Testmaterial

Als Papier-Bleistift-Verfahren umfasst das Verfahren 20 AIST-Testexemplare einschließlich Auswertungsblatt, Auswertungsschablone und Handanweisung. Zusätzlich enthält die Testmappe unter der Bezeichnung Umweltstrukturtest (UST) 20 UST-Testexemplare, mit denen – anhand derselben Items mit geänderter Instruktion – berufliche Umwelten erfasst bzw. charakterisiert werden können.

Computer-Versionen des AIST existieren im „Hogrefe TestSystem“ sowie im „Wiener Testsystem“ (beide in diesem Band)

Bearbeitungsbedingungen und -dauer

Der AIST ist als Einzel- und Gruppenverfahren administrierbar (Bearbeitungsdauer ca. 10 min). In den Computerversionen ist er auch durch Nichtpsychologen bzw. weitgehend ohne externe Unterstützung administrierbar.

Ergebnisrückmeldungen

Die Standardauswertung ergibt ein Interessen- (bzw. Umwelt-) profil, das bei individueller Anwendung im Auswertungsgespräch kommunikativ validiert werden kann. (durch Überprüfung der Übereinstimmung mit selbst wahrgenommenen Interessenschwerpunkten). Darüber hinaus werden Angaben gemacht zur Differenziertheit des Interessenprofils, zur Konsistenz (Stimmigkeit) der dominierenden Personorientierungen sowie zur Interessengemäßheit (Kongruenz) der erwogenen Ausbildungs- bzw. Berufsalternativen.

Die Computerversionen erlauben die maschinelle Erstellung eines Interessenprofils mit grafischem Output. Die Version im „Wiener Testsystem“ erlaubt optional den Ausdruck von Hinweisen auf mit dem Interessenprofil korrespondierende Berufe.

Publikationen

Bergmann, C. (2001). Personality and vocational interests: Evaluation of the correspondence between the Five-Factor Model of personality and Holland’s Six Vocational Orientations. In K.W. Kallus, N. Posthumus & P. Jimenez (Eds.), *Current psychological research in Austria* (pp. 189-192). Graz: Akademische Druck- und Verlagsanstalt.

Bergmann, C. (2003). Berufliche Interessentests – Wider die Anwendung geschlechtsspezifischer Normen. *Zeitschrift für Personalpsychologie, 2,* 66-77.

Bergmann, C & Eder, F. (2004*). Allgemeiner Interessen-Struktur-Test R (AIST-R) /Umwelt-Struktur-Test (UST). Manual*. Göttingen: Beltz-Test-Verlag.

Bergmann, C & Eder, F. (2000*).* Geschlechtsspezifische Interessen in der Sekundarstufe II*. Empirische Pädagogik, 14,* 255-286.

Eder, F. (1998). Differenziertheit der Interessen als Prädiktor der Interessenentwicklung. In J. Abel & C. Tarnai (Hrsg.). *Pädagogisch-psychologische Interessenforschung in Studium und Beruf* (S. 63-77). Münster: Waxmann.

Hammerschmidt, J. (2002). Allgemeiner Interessen-Struktur-Test/Umwelt-Struktur-Test (AIST/UST) (Rezension). In U.P. Kanning & H. Holling (Hrsg.), *Handbuch personaldiagnostischer Instrumente* (S. 304-309). Göttingen: Hogrefe.

Erfahrungen in der Wirtschaftspraxis

Neben der Verwendung in der Ausbildungs- und Berufsberatung bei der beruflichen Ersteinmündung (in Berufsinformationszentren, Maturanten-/Abiturienten- und Studienwahlberatung) auch Einsatz im Zusammenhang mit Berufswechsel bzw. beruflicher Umschulung (Berufsorientierungsmaßnahmen) sowie bei Seminaren für Unternehmensgründer.

Sonstiges

Fremdsprachige Versionen des AIST: Tschechisch (PC-Version), Englisch (PC-Version), Italienisch (PC-Version), Portugiesisch (PC- und Papier-Bleistift-Version), Schwedisch (PC-Version).

Preise

Testmappe komplett:	68 €

Computerversion (Hogrefe TestSystem)

100 Testvorgaben:	750 €
weitere 100 Testvorgaben jeweils:	80 €

Computerversionen (Wiener Testsystem)

50 Testvorgaben:	136 €
7 Jahre Nutzungsrecht:	816 €

Bezugsadressen

Testzentrale und Apparatezentrum Göttingen
Robert-Bosch-Breite 25
D – 37079 Göttingen
Fon: +49 (0)551/50688 -0/-14
Fax: +49 (0)551/50688 -24
Email: testzentrale@hogrefe.de
Internet: www.testzentrale.de.

Dr. G. Schuhfried GmbH
Hyrtlstrasse 45
A – 2340 Mödling

Fon: +43 (0)2236/42315
Fax: +43 (0)2236/46597
Email: schuhfried@adis.at
Internet: www.schuhfried.co.at/deu/wts/aist.htm

englische Version:

Lafayette instrument europe incorp. campden instruments
4 Park Road, Sileby
GB-Loughborough, Leics., LE12 7TJ
United Kingdom & Northern Ireland
Fon: +44 (0)1509/817700
Fax: +44 (0)1509/817701
Email: gprescott@lafayetteinstrument.com
Internet : www.lafayetteinstrument.com

(Verfasser des Beitrags: Christian Bergmann)

ALS

Arbeitsleistungsserie

Autor

Gernot Schuhfried (revidierte Fassung für PC des Pauli-Tests von 1936)

Verlag, Erscheinungsort und -jahr

Verlag Dr. Schuhfried, Mödling (Österreich), 1990

Einsatzbereiche

Allgemeiner Leistungstest zur Messung von Konzentrationsfähigkeit, psychischer Sättigung und Ermüdbarkeit in der angewandten Psychologie, insbesondere in der Eignungsdiagnostik.

Konstruktionsgrundlagen

Bei der ALS, einer Weiterentwicklung des Kraepelinschen Arbeitsversuchs und des Pauli-Tests, wird die Methode fortlaufenden Addierens verwendet, um den Verlauf der Arbeitsleistung eines Probanden zu ermitteln: Er hat die Aufgabe, über einen längeren Zeitraum jeweils zwei einstellige Zahlen möglichst rasch und fehlerfrei zu addieren. Der Test ist in Zeitabschnitte (Teilzeiten) unterteilt, so dass ein Leistungsverlauf ermittelt werden kann.

Es können auch Modifikationen des Tests appliziert werden. Sie ergeben sich durch Rechenaufgaben unterschiedlicher Schwierigkeit und die Einbeziehung einer zusätzlichen Kurzzeitgedächtnisanforderung. Dadurch werden Aspekte der fluiden Intelligenz des Probanden zugänglich.

Frage- und Antwort-Formate

Die computergestützte Durchführung des ALS beinhaltet nach einer standardisierten Instruktion und Übungsphase eine Testphase von 20 Minuten, in der möglichst viele Additionen von jeweils zwei Zahlen vorzunehmen sind. Sie werden übereinander auf dem Bildschirm dargeboten. Die Ergebnisse der Rechnungen sind vom Probanden über die Tastatur des Probandenpanels einzugeben.

Bei Aufgaben mit Kurzzeitgedächtnisanforderung wird die untere Ziffer nach jeder Eingabe eines Resultats nach oben gesetzt und dort verdeckt. Der Proband muss sich daher stets vor der Eingabe einer Lösung die untere Ziffer merken, um die nachfolgende Aufgabe lösen zu können.

Inhalte

Es werden folgende Variablen erfasst:

- „Gesamtzahl der bearbeiteten Aufgaben“ als Maß für die Arbeitsgeschwindigkeit
- „Anstieg der Bearbeiteten“ als Maß für den Leistungsanstieg oder Abfall des Probanden
- „Prozentualer Anteil fehlerhafter Lösungen und Gesamtzahl der Korrekturen“ als Maße für die Genauigkeit der Arbeit des Probanden

Die zu Diagrammen zusammengefassten Teilzeitergebnisse geben Auskunft über den Leistungsverlauf.

Eine hohe Gesamtzahl bearbeiteter Aufgaben bei einem geringen Anteil fehlerhafter Lösungen sowie die Vornahme von Korrekturen sprechen für gute Konzentrationsfähigkeit und Aufmerksamkeitsleistungen sowie geringe Ermüdbarkeit des Probanden.

Subtests

Keine.

Gütekriterien

Objektivität

Aufgrund der computergestützten Durchführung und Auswertung ergibt sich ein hohes Maß an Objektivität. Alle Probanden werden in gleicher Weise unabhängig vom Testleiter über die Testanforderung instruiert, auch die Datenregistrierung und der Normwertvergleich erfolgen automatisch und ohne Mitwirkung von Auswertern.

Reliabilität

Eigene Untersuchungen zeigen bei verschiedenen Stichproben für die „Gesamtzahl der bearbeiteten Aufgaben“ und die „Fehler“ Split-half-Reliabilitäten zwischen .91 und .99 (N = 187 bzw. 662; Datenerhebung: 1992-1996).

Sie stimmen weitgehend mit den Angaben für die Papier-Bleistift-Version überein: Dort wurden für die Gesamtzahl der bearbeiteten Aufgaben durchgängig Retest- und Split-half-Reliabilitäten von über .95, für Fehler und Korrekturen zwischen .68 und .88 gefunden.

Validität

Das Messprinzip der ALS wird in der Literatur als „anhaltende Konzentration bei geistiger Tempoarbeit“ umschrieben. Hierzu zählen auch Persönlichkeitsmerkmale wie „Festigkeit“ gegen Ablenkungen und Störungen, Interferenzen, Motiviertheit und Willenseinsatz sowie die Fähigkeit, die Aufmerksamkeit für eine bestimmte Aufgabe aufrechtzuerhalten.

Die zur Testform S1 vorliegenden Untersuchungsergebnisse von 662 Stellenbewerbern, 130 Hirnorganikern und 187 Arbeitssuchenden wurden einer differenzierten statistischen Analyse unterzogen (Neuwirth, 1996). Die Testleistung der Stellenbewerber unterscheidet sich in der Variablen Gesamtzahl Bearbeitete signifikant von der der beiden anderen Stichproben. Sie bearbeiten im Mittel mehr Aufgaben und begehen weniger Fehler.

Normen

Die Testergebnisse können mit folgenden Eichstichproben verglichen werden:

S1 - Stellenbewerber

Im Zeitraum 1992-95 wurden 662 Stellenbewerber verschiedener Berufsgruppen mit mittlerem und höherem Schulabschluss, Facharbeiterausbildung bzw. Abschluss der HTL/Fachhochschule, Universität/Hochschule durch ein österreichisches Personalvermittlungsunternehmen dem Test unterzogen.

S1 - Arbeitssuchende

187 Arbeitssuchende aus Österreich und Schweden wurden mit der Testform S1 untersucht (1996).

S1 - Organiker

Aus dem klinischen Bereich standen die Ergebnisse der Diagnostik von 130 Organikern, Patienten der Reha-Klinik Klausenbach, Nordrach, aus den Jahren 1990 und 1991 zur Verfügung.

S1 – Ältere Menschen

95 Ältere Menschen wurden in Forschungslabors der Firma Dr. Schuhfried im Zeitraum von 2001-2002 erhoben.

S2 - Stellenbewerber

Es wurden 105 Stellenbewerber verschiedener Berufsgruppen mit mittlerem und höherem Schulabschluss, Facharbeiterausbildung bzw. Abschluss der HTL/Fachhochschule, Universität/ Hochschule durch ein österreichisches Personalvermittlungsunternehmen dem Test unterzogen.

S2 - Arbeitssuchende

103 Arbeitssuchende aus Österreich und Schweden wurden mit der Testform S2 untersucht.

S2 – Sicherheitskräfte Macao

Für diese Testform stehen nun auch Normen einer Stichprobe von N =168 Sicherheitskräften aus Macao zur Verfügung.

S7 - Arbeiter

Mit der Testform S7 wurden 2.907 Beschäftigte eines luxemburger Stahlwerkes, die vor allem Tätigkeiten als Anlagenfahrer, Lok- und Kranführer ausübten und Personal für Steuer- und Überwachungstätigkeiten, sowie Auszubildende in den Jahren 1988-95 untersucht.

Testmaterial

Computerversion; die Bearbeitung erfolgt mittels PC-Keyboard oder speziellem Panel mit ergonomischen Ziffernblock.

Bearbeitungsbedingungen und -dauer

Die Testdauer beträgt für die Testformen S1 (Pauli-Test) und S2 ca. 25 Minuten, für die Testform S7 ca. 15 Minuten.

Ergebnisrückmeldungen

Die Darstellung der Ergebnisse erfolgt am PC-Bildschirm und kann zudem ausgedruckt werden. Sie umfasst eine Ergebnistabelle mit Roh- und Standardwerten (Prozentrangwerte und T-Werte), ein Testprofil, eine Verlaufsdarstellung und ein Testprotokoll. Das Testprofil ermöglicht einen Vergleich der einzelnen Skalenwerte auf der Basis von Prozentrangwerten und T-Werten. Der Verlauf des Tests über die Teilzeiten wird in einem Koordinationssystem abgebildet und das ausführliche Testprotokoll gibt Auskunft über alle Einzelantworten pro Teilzeit. Zur Ergebnisrückmeldung an den Probanden – falls nötig und gewünscht – bedürfen diese Daten der mündlichen Erläuterung durch einen geschulten Testanwender.

Publikationen

Arnold, W. (1975). *Der Pauli-Test.* München: Barth.

Bartenwerfer, H. (1983). Allgemeine Leistungsdiagnostik. In K.-J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik.* Göttingen: Hogrefe.

Kallenbach, K. (1978). Auswirkungen unterschiedlicher Verfahren zum Pauli-Arbeitsversuch – ein Vergleich von Abwandlungsformen. *Psychologie und Praxis, 22*, 157-163.

Neuwirth, W (1996). *Handanweisung ALS*. Mödling: Schuhfried.

Erfahrungen in der Wirtschaftspraxis

Es liegen gute Erfahrungsberichte aus der Personalselektion für Kran- und Staplerfahrer sowie für Personen mit Überwachungstätigkeiten vor.

Sonstiges

Es stehen folgende drei Testformen zur Verfügung:

- S1: Pauli Test

Es ist die Summe einer oberen und einer unteren Zahl zu errechnen. Die Form umfasst 20 Teilzeiten zu je einer Minute.

- S2: Pauli Test mit Kurzzeitgedächtnisanforderung

Es ist die Summe einer oberen und einer unteren Zahl zu errechnen, wobei die untere Ziffer nach jeder Eingabe eines Resultats nach oben gesetzt und dort verdeckt wird. Der Proband muss sich daher stets vor der Eingabe einer Lösung die untere Ziffer merken, um die nachfolgende Aufgabe lösen zu können. Die Form umfasst 20 Teilzeiten zu je einer Minute.

- S7: Luxemburger Form

Beide Zahlen sind zu addieren bzw. es ist die Differenz zu bilden. Eine Korrekturlogik unterbindet Aufgaben, die negative Resultate ergeben. Die Form umfasst 10 Teilzeiten zu je einer Minute.

Der Zahlenraum der Lösungen reicht in jeder Testform von 1 bis 17.

Preise

50 Testvorgaben:	120 €
7 Jahre Nutzungsrecht:	720 €

Bezugsadresse

Dr. G. Schuhfried GmbH
Hyrtlstraße 45
A – 2340 Mödling
Fon: +43 (0)2236/423 15
Fax: +43 (0)2236/465 97
Email: info@schuhfried.at
Internet: www.schuhfried.at

(Verfasser des Beitrag: Gerald Schuhfried)

AMT

Adaptiver Matrizen Test

Autoren

Lutz F. Hornke, Stefan Etzel und Klaus Rettig

Verlag, Erscheinungsort und -jahr

Verlag Dr. Schuhfried, Mödling (Österreich), 1997

Einsatzbereiche

Personalauswahl und -entwicklung; Berufs-, Studien- und Bildungsberatung; Flug- und Verkehrspsychologie; Forschung

Konstruktionsgrundlagen

Verschiedensten Modellen zur allgemeinen Intelligenz ist gemeinsam, dass das Erkennen von Regeln im Sinne schlußfolgernden Denkens bzw. die Fähigkeit „eine Ordnung in der Unordnung erkennen zu können“ ein zentraler Bestandteil ist. Der Versuch, dies sprachfrei zu erfassen, hat eine lange Tradition in Form der figuralen Testaufgaben der Raven-Matrizen. Auf dieser Forschungstradition wurde das vorliegende Testverfahren aufgebaut. Das Neuartige liegt in

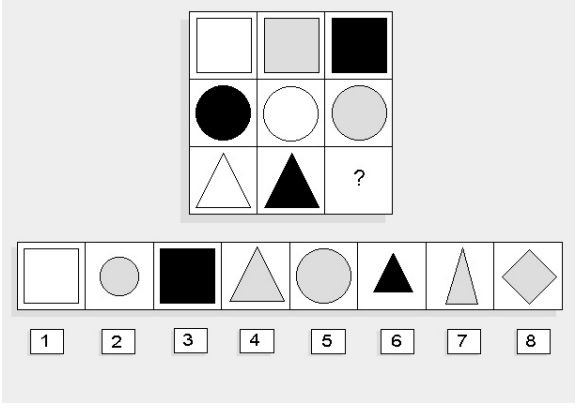
- der Testkonstruktion (rationale Itemkonstruktion) basierend auf kognitionspsychologischen Erkenntnissen,
- der Testauswertung (Rasch-Skalierung) und
- der Teststrategie (adaptives Testen).

Inhalte

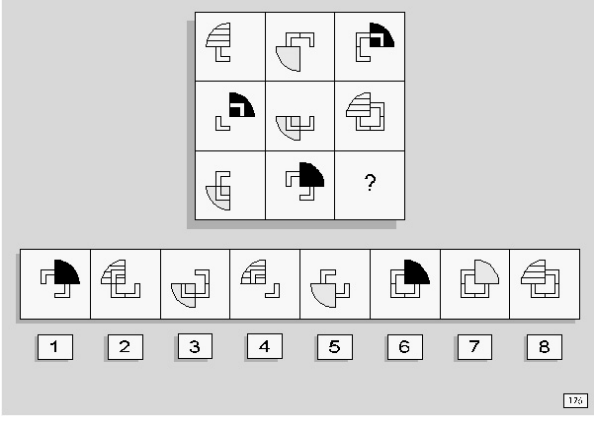
Der Test erfasst mit 266 homogenen Items die allgemeine Intelligenz und hier insbesondere schlussfolgerndes Denken. Dabei wird schlussfolgerndes Denken über das Erkennen von Regeln operationalisiert. Ausgehend von auf figurales Material anwendbaren kognitiven Operationen, wurden anhand der vorliegenden Literatur und der Kritik an bestehenden Matrizentests explizite Regeln formuliert, nach denen Items erstellt werden können. In den Aufgaben sind u.a. verschiedene Regeln umgesetzt: Identität der Elemente, Addition, Subtraktion, Einzelelementaddition. Dabei werden sie sowohl für einzelne Figuren oder Figurelemente miteinander waagerecht, senkrecht oder beidem kombiniert. Die Figuren sind entweder als geschlossene oder offene Gestalten oder als Figuren mit separierbaren Komponenten repräsentiert.

Frage- und Antwort-Formate

Jede Testaufgabe besteht aus einer Matrix mit neun Feldern; acht Felder sind jeweils vorgegeben. Die Aufgabe des Probanden ist es dann, diejenige von acht Antwortalternativen auszuwählen, die aufgrund einer abzuleitenden Regel eindeutig zu den vorgegebenen Elementen ‘passt’. Die nachfolgende Abbildung illustriert eine Aufgabe.



Die Aufgaben sind nach verschiedenen Regeln aufgebaut. Im nachfolgenden Beispiel ist Antwortalternative 2 die richtige Lösung, die sich durch Addition der Quadratelemente und Variation der Muster der Viertelkreise in den Zeilen der Matrix ergibt.



Die in einer Aufgabe angebotenen Distraktoren sind ebenfalls regelgeleitet konstruiert, so dass in einigen Antwortalternativen auch teilrichtige Lösungen enthalten sind.

Subtests

Die Vorgabe folgender, auf unterschiedliche Genauigkeit bzw. Anwendungszwecke abzielender Testformen ist möglich:

Screening, Testform S1

Mit Hilfe dieser Testform kann ein kurzer Grobüberblick über die Leistungsfähigkeit des untersuchten Kandidaten ermittelt werden. Besonders ist dabei, dass im Gegensatz zu traditionellen Testverfahren hier der adaptive Test den gleichen Standardmessfehler für alle Niveaus und Probanden einhält. Der Standardmessfehler liegt bei dieser Variante bei einem Wert von SEM = .63

Standard, Testform S2

Der voreingestellte Standardmessfehler liegt bei dieser Testform bei SEM = .44

Precision, Testform S3

Der Standardmessfehler ist auf einen Wert von SEM = .39 festgelegt. Wegen der resultierenden höheren Messgenauigkeit ist es notwendig, eine größere Anzahl von Testaufgaben vorzugeben.

Gütekriterien

Objektivität

Aufgrund der standardisierten, computergestützten Instruktions- und Testdarbietung sowie der rechnergestützten Auswertung der Probandendaten können sowohl die Durchführungs- als auch die Auswertungs- und Interpretationsobjektivität als gegeben betrachtet werden.

Reliabilität

Das Verfahren ist adaptiv, d.h. die Testung wird solange fortgesetzt, bis ein voreingestellter Standardmessfehler erreicht ist. Unterstellt man Einheitsvarianz so ergibt sich die an allen Orten erzielbare Reliabilität aus r_{tt}=1-SEM. Diese beträgt für die Testform S1 .60, für Testform S2 .80 und für Testform S3 .85.

Validität

Im Unterschied zu herkömmlichen Matrizenaufgaben wurden die Aufgaben aufgrund eines expliziten Konstruktionsrationals erstellt. Durch diese regelgeleitete Konstruktion ist die inhaltliche Gültigkeit des Tests gewährleistet. Zudem trägt dieses Vorgehen dazu bei, eine hinreichende Rasch-Homogenität der Items sicherstellen zu können. Es liegen noch keine Validitätsstudien vor.

Ökonomie

Der Test ist in Bezug auf die erforderliche Testleiterzeit höchst ökonomisch, da sowohl die Testinstruktion als auch die Auswertung computergestützt erfolgen. Bezüglich der Durchführungsdauer gibt es eine Abhängigkeit zwischen Messgenauigkeit und Anzahl der vorzugebenden Items und damit der Probandenzeit: Die Erhöhung der Messgenauigkeit macht die Be-

arbeitung einer größeren Zahl von Items erforderlich. Im Vergleich mit einem klassischen Test fester Länge ist – bei vergleichbarer Messgenauigkeit – die durchschnittliche Anzahl zu bearbeitender Items beim adaptiven Testen jedoch stets geringer.

Testfairness

Nach den bisherigen Erfahrungen kann der AMT als fairer Test bezeichnet werden: Insbesondere werden Personen ohne Computererfahrung nicht benachteiligt, da für die Testdurchführung sehr wenige Computerkenntnisse erforderlich sind, und die Instruktionsphase ausreicht, die Antworteingabe auch bei Computerneulingen einzuüben.

Normen

Die Normierung basiert auf den Daten von 1.356 Probanden (580 Männer, 776 Frauen im Alter von 15 bis 77 Jahren, die an verschiedenen Orten (Kattowitz, Polen; Moskau, Russland und Wien, Österreich) erhoben wurden. In der gegenwärtigen Version werden maximum-likelihood-geschätzte θ-Werte und zugehörige Prozentrangwerte verwendet. Aufgrund der demographischen Daten kann die Untersuchungsstichprobe als hinreichend repräsentativ angesehen werden. Die Normierung wird in gewissen Abständen aufgrund neuerer Untersuchungen entsprechend angepasst. Zudem stehen nun auch Normen einer Normierungsstichprobe von 208 Probanden (75 Männer, 133 Frauen) im Alter von 18-81 Jahren zur Verfügung. Diese Stichprobe ist hinsichtlich Bildungsgrad und Alter repräsentativ für Deutschland und Österreich.

Testmaterial

Itembank mit insgesamt 270 Items, nur als Computerversion einzusetzen. Die Bearbeitung erfolgt mittels Probandentastatur, Lichtgriffel, PC-Keyboard, Maus oder Touchscreen.

Bearbeitungsbedingungen und -dauer

- Computergestützte adaptive, sich an das Leistungsniveau anpassende Vorgabe der Test-Items.
- Auch ohne Psychologen durchführbar; die Interpretation hat aber durch Psychologen zu erfolgen.
- Der Proband kann aufgrund der einfachen Handhabung und standardisierten Instruktion den Test selbständig und ohne weitere Hilfe bearbeiten.
- Die Testdauer variiert von Proband zu Proband je nach Itemumfang. Erfahrungsgemäß werden pro Item 60 Sekunden als mittlere Beantwortungszeit benötigt (plus eine 10-minütige Instruktionszeit). In der Regel kommen 95% der Pbn mit 35 (Testform S3) bis 20 Minuten (Testform S1) Testzeit aus. Dies ergibt eine geschätzte mittlere Durchführungsdauer für die Testform S1 von 25 Minuten, für die Testform S2 von 35 Minuten und für die Testform S3 von 45 Minuten.

Gruppentestungen sind möglich, sofern auf eine hinreichend ruhige Arbeitsumgebung geachtet wird.

Ergebnisrückmeldungen

Ergebnisse auf Aufgabenebene werden dem Probanden ausschließlich während der Instruktionsphase direkt am Bildschirm rückgemeldet. In der anschließenden Testphase erhält der Kandidat keine Informationen über die Richtigkeit seiner Lösungen.

Eine ausführliche schriftliche Ergebnisauswertung steht dem Anwender im Anschluß an die Testung zur Verfügung. Hier gibt es zum einen die Option, sich einen Grobüberblick über die Resultate zu verschaffen (Personenparameter, Konfidenzintervall, Prozentrangwert, benötigte Itemzahl, Bearbeitungsdauer etc.). Zum anderen können die Ergebnisse aber auch in detaillierter Form aufbereitet werden. Das bedeutet, es können zu jedem Item nicht nur die Itemparameter, sondern selbstverständlich auch die Beantwortung durch den Probanden, die Lösungszeit etc. betrachtet werden.

Publikationen

Gierschmann, E. (2000). AMT – Adaptiver Matrizentest (Rezension). In E. Fay (Hrsg.), *Tests unter der Lupe III* (S. 21-33). Lengerich: Pabst.

Hornke, L.F. (1998). Benefits from computerized adaptive testing as seen in simulation studies. *European Journal of Psychological Assessment, 15,* 91-98.

Hornke, L.F., Küppers, A. & Etzel, S. (2000). Konstruktion und Evaluation eines adaptiven Matrizentests. *Diagnostica, 46,* 182-188.

Erfahrungen in der Wirtschaftspraxis

Es liegen gute Erfahrungen im Bereich der Pilotenselektion beim Österreichischen Bundesheer vor.

Sonstiges

- Der AMT ist nahezu sprachfrei (nur Instruktion) und daher „culture fair".
- Eine hinreichende Skalenhomogenität der 266 Items im Sinne der probabilistischen Testtheorie ist gegeben.
- Der AMT bietet aufgrund seiner adaptiven Teststrategie erhebliche ökonomische und testpraktische Vorteile gegenüber konventionellen Tests: Den Probanden werden nur solche Aufgaben vorgegeben, die ihrem Leistungsniveau entsprechen, d.h. sie werden nicht durch zu leichte Aufgaben gelangweilt bzw. durch unangemessen schwierige Aufgaben frustriert. Des Weiteren kann die Messgenauigkeit mit Blick auf die Entscheidungs- oder Anwendungszwecke entsprechend erhöht oder der Test und die Testzeit verkürzt werden.
- Außerdem wird mit der Vorgabe jedes Items stets die maximale Information über den Probanden gewonnen und es gilt für alle Probanden der gleiche Standardmessfehler.

Preise

50 Testvorgaben:	172 €
7 Jahre Nutzungsrecht:	1.032 €

Bezugsadresse

Dr. G. Schuhfried GmbH
Hyrtlstraße 45
A – 2340 Mödling
Fon: +43 (0)2236/423 15
Fax: +43 (0)2236/465 97
Email: info@schuhfried.at
Internet: www.schuhfried.at

(*Verfasser des Beitrags: Lutz F. Hornke und Stefan Etzel)*

APM

Advanced Progressive Matrices

Autoren

Kurt A. Heller, Heinrich Kratzmeier und Angelika Lengfelder: Matrizen-Test-Manual, Band 2. Handbuch mit deutschen Normen zu den Advanced Progressive Matrices (APM) von J.C. Raven

Verlage, Erscheinungsorte und -jahre

Hogrefe-Verlag, Göttingen & Beltz-Test-Verlag, Weinheim, 1998
Verlag Dr. Schuhfried, Mödling (Österreich), 2000

Einsatzbereiche

Die APM können bei Jugendlichen ab dem 12. Lebensjahr eingesetzt werden. Sie sind zur Testung überdurchschnittlich begabter Jugendlicher und Erwachsener gedacht und werden u.a. in der Schulberatung, der klinischen Diagnostik, der Eignungs- und Laufbahnberatung sowie in experimentellen Studien als Grundlagendiagnostikum verwendet.

Konstruktionsgrundlagen

Die Ausgangsbasis für die Entwicklung der Advanced Progressive Matrices (APM von Raven, 1943, 1947) war, ebenso wie für die etwas früher entwickelten Standard Progressive Matrices (SPM von Raven, 1938), Spearmans Intelligenzmodell. Dieses Modell postuliert einen Generalfaktor „g" der Intelligenz, welcher in eine reproduktive Komponente (d.h. die Fähigkeit, erworbenes Wissen wiederzugeben) und eine eduktive Komponente (d.h. die Fähigkeit zur Abstraktion bezüglich Ähnlichkeit und Vergleich entsprechend der „fluid ablity" im Sinne Cattells) unterteilt wird. Zur Erfassung der letztgenannten eduktiven Fähigkeit wurden die Matrizen-Tests konzipiert. Von diesen Tests weisen die APM ein höheres Anspruchsniveau auf als die SPM und die noch einfacheren Coulored Progressive Matrices (CPM). In Klauers (1993 a/b) Theorie des induktiven Denkens spielen Matrizenaufgaben eine prominente Rolle zur Operationalisierung der „Systembildung" z.B. bei der Unterscheidung von gleichen vs. verschiedenen Objektrelationen, also der Messung induktiver Denkkompetenzen. Im Vergleich mit den SPM besitzen die APM eine besser Differenzierungsfähigkeit im oberen Leistungsbereich.

Inhalte

Die Aufgaben der SPM- und der APM-Skalen lassen sich grob in perzeptuelle und logisch-konzeptuelle (vgl. Carlson et al., 1974) unterteilen. Dabei setzen die perzeptuellen Aufgaben das

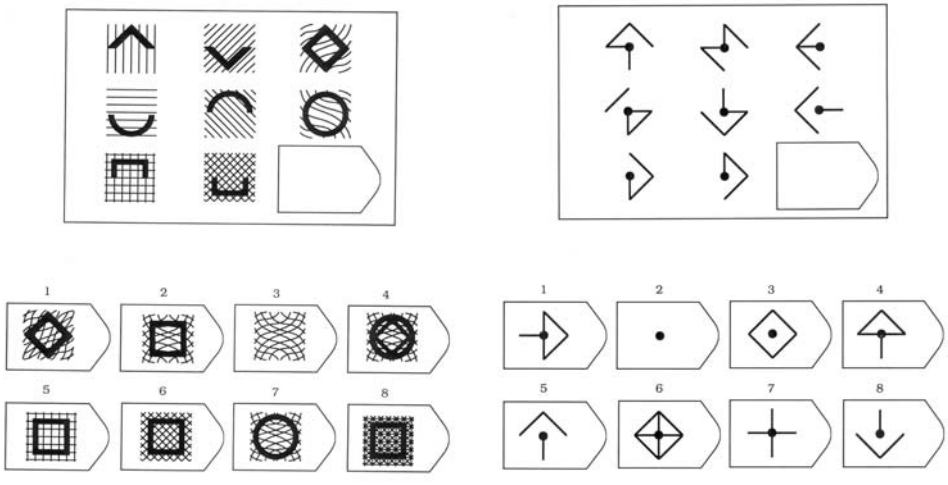
Erkennen von Gestaltprinzipien voraus, während die logisch-konzeptuellen Aufgaben neben Wahrnehmungsleistungen abstraktes, analogisierendes und induktives Denken erfordern.

Nach Angaben Ravens befinden sich die zur Lösung der Aufgaben erforderlichen kognitiven Prozesse auf einem Kontinuum von Wahrnehmungsprozessen bis hin zu induktivem Denken und Schlussfolgern, wobei Wahrnehmung als „*conceptual* activity" (Raven, Raven & Court, 1995, S. G 56) verstanden wird.

Frage- und Antwort-Formate

Der Test besteht aus einem Übungsteil (Set I: 12 Matrizen) und einem Hauptteil (Set II: 36 Matrizen). Die Matrizenaufgaben (Beispiele siehe Abbildung 1) bestehen jeweils aus 3 x 3 geometrischen Mustern, wobei stets ein Muster in der rechten unteren Ecke fehlt. Dieses soll unter Berücksichtigung des horizontalen und vertikalen Ordnungsprinzips aus je acht Lösungsvorschlägen ausgesucht werden.

Abbildung 1: Zwei Beispielmatrizen aus Set II



15 36

Testmaterial

Das Testmaterial besteht aus Durchschreibe-Antwortbögen sowie zwei Testheften, Set I und Set II, welche pro Seite je eine Matrizenaufgabe enthalten. Die richtige Lösung ist auf dem Antwortbogen einzutragen.

Neben der klassischen Paper-Pencil-Version wird auch eine computergestützte Fassung angeboten.

Subtests

Keine.

Erfasst werden soll nur ein Merkmal, nämlich die eduktive Komponente des Faktors „g". Der Test besteht aus zwei Teilen. Die Aufgaben sind nach Angabe des Testautors nach aufsteigender Schwierigkeit geordnet. Set I mit seinen 12 Aufgaben ist als Trainingstest anzusehen. Wegen der geringen Reliabilität dieses Testteils ist Set I keinesfalls als Kurz-Screening-Test zu empfehlen. Die 36 Aufgaben im Hauptteil (Set II) sind schwieriger und komplexer als jene in Set I und in den SPM-Skalen. Für einige APM-Items gilt aber, dass sie nicht die gleichen Schwierigkeitsränge auf allen Leistungsstufen aufweisen und somit nicht den Anforderungen des Rasch-Modells entsprechen.

Gütekriterien

In der vorliegenden Form erfüllen die APM-Skalen die Anforderungen hinsichtlich der Durchführungs-, Auswertungs- und Interpretations*objektivität*.

Reliabilitätstestungen für die APM erbrachten recht zufriedenstellende Ergebnisse. Für die innere Konsistenz der APM (Set II) werden Koeffizienten zwischen .69 und .88 (Medianwert: .81) sowie Split-Half-Koeffizienten zwischen .67 und .95 (Medianwert: .86) berichtet. Die Reliabilitätskoeffizienten liegen bei Retestungen nach zwei Wochen bzw. drei Monaten zwischen .58 und .90 (Medianwert: .75). Bei der Bewertung der Retest-Reliabilitäten wurden die jeweiligen Medianwerte zu Grunde gelegt, der niedrigste Wert .75 erscheint als Retestergebnis zufriedenstellend, wohingegen die Medianwerte bzgl. der internen Konsistenz sowie der Split-Half-Methode auf eine gute Reliabilität hinweisen.

Die APM weisen mittlere Korrelationen mit anderen (inhaltlich vergleichbaren) Intelligenztests (z.B. CFT20, CFT3) sowie mit Verfahren auf, die Informationsverarbeitungsgeschwindigkeit als Indikator der Intelligenz betrachten (z.B. ZVT, Coding-Test). Zur *externen Validierung* herangezogene Schulleistungsmaße korrelieren insbesondere im Fachbereich Mathematik mit den APM-Leistungen zwischen .35 und .50, wobei die Zusammenhänge mit sprachlichen Leistungen (auch in internationalen Studien) durchweg niedriger ausfallen (ausführlicher vgl. Heller und Perletz, 2000, S.111 ff.).

Die faktorenanalytische Validierung der Eindimensionalität der APM-Skala ist dagegen umstritten.

Die APM erscheinen vor allem differentialdiagnostisch geeignet zwischen Personen mit unterschiedlichen Leistungsniveaus (z.B. Schülern verschiedener Schularten oder Erwachsenen verschiedener Ausbildungsniveaus) zu differenzieren.

Normen

Zur Interpretation der errechneten Rohwerte sind im Manual alters- und klassenstufenbezogene Normtabellen für Schüler verschiedener Schularten (ab Klasse 6), für hochbegabte Gymnasiasten, für Studierende verschiedener Altersgruppen und Studienrichtungen (Alter: 19-35 Jahre) sowie Normen für Personen über 60 Jahre enthalten. Im deutschsprachigen Raum liegen bislang für die Population der 19- bis 60-jährigen nur Normierungsdaten Studierender vor. Darüber hinaus werden für die Anwender, die an den Testergebnissen dieser Altersgruppen interessiert sind, in fünf Tabellen des Anhangs Median- und Quartilwerte erwachsener Normierungsgruppen aus