

Testbesprechung

Kersting, M., Althoff, K. & Jäger, A. O. (2008). WIT-2. Wilde-Intelligenz-Test 2. Göttingen: Hogrefe. Preis: € 328,-.

Der WIT-2 dient der differenzierten Erfassung intellektueller Leistungsfähigkeit von Jugendlichen und Erwachsenen ab 14 Jahren. Er stellt die Revision und Neukonstruktion des Wilde-Intelligenz-Tests (WIT) von Jäger und Althoff (1983, 1994) dar. Der WIT galt lange Zeit als eines der am häufigsten eingesetzten und in der eignungsdiagnostischen Praxis beliebtesten Verfahren. Davon abgesehen entspricht er heutzutage jedoch nicht mehr den modernen testpsychologischen Qualitätsstandards. Schon vor mehr als 10 Jahren beschrieb Jäger (1997) die Anwendung des WIT aufgrund einer mangelhaft elaborierten, theoretischen Fundierung, der bruchstückhaft angeführten Ausführungen zu den grundlegenden Gütekriterien und der extrem veralteten Normierung als nicht gerechtfertigt. Demzufolge war es das Ziel der Testautoren, auf Basis des ihrer Ansicht nach gleichermaßen revisionsbedürftigen wie auch revisionswürdigen WIT ein modernes, psychometrisch hochwertiges und zugleich anwenderfreundliches Instrument zur berufsbezogenen Diagnostik zu entwickeln. Die diagnostische Zielsetzung des Tests bezieht sich auf die Anwendung innerhalb der Eignungsdiagnostik, insbesondere der Personalauswahl, Personalentwicklung sowie der Berufs- und Studienlaufbahnberatung (Kersting, Althoff & Jäger, 2008).

Theoretischer Hintergrund

Die theoretische Grundlage des Tests ist ein modernisiertes und erweitertes Thurstone-Modell, das von den Autoren als „Modifiziertes Modell der Primary Mental Abilities“ (MMPMA) bezeichnet wird. Während sich nach dem ursprünglichen Modell der Primary Mental Abilities (PMA) von Thurstone (1938) intellektuelle Leistungen auf die sieben Primärfaktoren (1) Schlussfolgerndes Denken, (2) Räumliche Vorstellung, (3) Rechnerisches Denken, (4) Sprachliches Denken, (5) Merkfähigkeit, (6) Wortflüssigkeit und (7) Wahrnehmungsgeschwindigkeit zurückführen lassen, wurde das MMPMA um einige Annahmen der modernen Intelligenzforschung erweitert. Die Neuerungen beruhen den Herausgebern zufolge auf moderneren statistischen Analysetechniken und der Weiterentwicklung von Intelligenzmodellen, welche das Primärfaktorenmodell nicht als gegensätzlich zu hierarchischen Modellen sehen, sondern als sich einander ergänzende Konstrukte offenbaren. Kombiniert wurde das Primärfaktorenkonzept mit dem hierarchischen Generalfaktorenkonzept, der Anwendung des Facettenansatzes und einigen Faktoren mit niedriger Generalität, also sehr spezifischen Fertigkeiten und Kenntnissen, die durch die übergeordneten Faktoren beeinflusst werden. Das MMPMA kann demzu-

folge in drei unterschiedliche Generalitätsebenen mit sich überlappenden Faktoren gegliedert werden. Der hierarchisch obersten Ebene, dem Stratum III, werden *Fluide Intelligenz/Arbeitsgedächtnis* und *Kristallisierte Intelligenz* zugeordnet. Das zweite Stratum beinhaltet *Schlussfolgerndes Denken*, *Sprachliches*, *Rechnerisches* und *Räumliches Denken* sowie *Merkfähigkeit*, *Wahrnehmungsgeschwindigkeit* und *Wortflüssigkeit*, wobei es auch innerhalb dieser Ebene zu Generalitätsunterschieden kommt. So ist das *Schlussfolgernde Denken* als einer dem *Verbalen*, *Rechnerischen* und *Räumlichen Denken* übergeordneter Faktor konstruiert. Das erste Stratum auf unterster Ebene umfasst sehr spezifische Fertigkeiten, wobei es sich den Autoren zufolge um mehr als 70 spezifische Faktoren handeln könnte, welche von den übergeordneten Faktoren maßgeblich beeinflusst werden. Beispielhaft für das noch nicht vollständig ausgearbeitete erste Stratum werden *Berufsspezifisches Wissen* und *Arbeitseffizienz* angeführt.

Im Rahmen einer WIT-2-Testung können einige, jedoch nicht alle der im MMPMA postulierten Fähigkeiten erfasst werden. Der Test repräsentiert mit acht Testmodulen in erster Linie Fähigkeiten aus dem zweiten Stratum, welche auch fünf der sieben Primärfaktoren von Thurstone abbilden: *Sprachliches Denken*, *Rechnerisches Denken*, *Räumliches Denken*, *Schlussfolgerndes Denken* und *Merkfähigkeit*. Auf die Untertests zur Erfassung der *Wortflüssigkeit* und der *Wahrnehmungsgeschwindigkeit* aus dem WIT wurde aufgrund mangelnder Kriteriumsvalidität, Augenscheinvalidität und Akzeptanz sowie einem erheblichen Aufwand im Falle einer Neukonstruktion von Items in der neuen Version verzichtet. Zudem wurde die Trennung von *Gedächtnis* und *Merkfähigkeit* wegen mangelnder konzeptueller Grundlagen aufgehoben und in einen Faktor integriert. Dahingegen wurden drei Untertests neu hinzugefügt und dem ersten Stratum zugeordnet. Sie sollen sehr spezifische, jedoch für viele Berufsfelder besonders relevante Fertigkeiten erfassen und somit auch eine außerordentlich hohe Kriteriumsvalidität aufweisen: *Arbeitseffizienz*, *Wissen Wirtschaft* und *Wissen Informationstechnologie*.

Testaufbau, Material und Durchführung

Der WIT-2 besteht aus 11 Subtests mit insgesamt 243 Items. Er ist folgendermaßen aufgebaut: *Analogien*, *Zahlenreihen*, *Abwicklungen*, *Gleiche Wortbedeutungen*, *Grundrechnen*, *Merkfähigkeit*, *Eingekleidete Rechenaufgaben*, *Spiegelbilder*, *E-Mails Bearbeiten*, *Wissen-Informationstechnologie*, *Wissen-Wirtschaft*.

Ähnlich wie im WIT die einzelnen Subtests Skalen zugeordnet werden können, werden auch im WIT-2 die 11

Subtests acht Modulen/Dimensionen zugeordnet, die dann Kennwerte zur Interpretation bzw. ein differenziertes Leistungsprofil liefern. Von einer unmittelbaren Interpretation auf Subtestebene wird mit der Begründung abgesehen, dass erst durch die Aggregation von Subtests das erforderliche Generalitätsniveau erreicht wird. Modul 1 *Sprachliches Denken* erfasst demnach mit den zwei Subtests *Analogien* und *Gleiche Wortbedeutung* die Fähigkeit, mit sprachlichen Konzepten umzugehen, also Wortschatz, Sprachverständnis und sprachlogisches Denken. Das Modul 2 *Rechnerisches Denken* erfasst mathematische Kenntnisse und beinhaltet einfache Rechenoperationen der Addition, Subtraktion, Multiplikation und Division mit den Subtests *Grundrechnen*, und *Eingekleidete Rechenaufgaben*. Das *Räumliche Denken* (= Modul 3) beinhaltet die Fähigkeit zur Vorstellung und Bearbeitung räumlicher Relationen und wird wie auch bereits im WIT mit den Subtests *Abwicklungen* und *Spiegelbilder* erhoben. Die Subtests *Analogien* und *Abwicklungen* werden mehreren Modulen zugeordnet. Sie tragen neben den bereits erwähnten Dimensionen auch zur Messung der operativen Fähigkeiten schlussfolgernden Denkens bei und dienen gemeinsam mit dem Subtest *Zahlenreihen* der Erfassung der Dimension *Schlussfolgerndes Denken*, dem Modul 4. Dem Modul 5 *Merkfähigkeit* ist der gleichnamige Subtest *Merkfähigkeit* zuzuordnen. Er erfordert die Fähigkeit, sich Informationen und Assoziationen einzuprägen und wieder zu erkennen. Auch Modul 6 beinhaltet nur einen Subtest (*E-Mails Bearbeiten*) zur Erfassung der *Arbeitseffizienz*. Die Module *Wissen-Wirtschaft* (7) und *Wissen-Informationstechnologie* (8) beinhalten die jeweils gleichnamigen Subtests und erfassen sehr spezifisches berufsbezogenes Faktenwissen. Modul 7 und 8 können als Indikatoren für *kristallisierte Intelligenz* gesehen werden, während alle anderen Module dem Faktor *Fluide Intelligenz/Arbeitsgedächtnis* zugehörig sind. Eine Interpretation auf Ebene des dritten Stratum findet jedoch nicht statt.

Gegenüber dem WIT wurde im WIT-2 auf die Subtests *Buchstabenreihen*, *Gedächtnis*, *Satzergänzungen*, *Schätzen*, *Sprichwörter*, *Zahlen-Merken*, *Wortgewandtheit* und *Beobachtungen* aus Gründen mangelnder Konstrukt- und Kriteriumsvalidität bzw. niedriger Akzeptanz durch die TestteilnehmerInnen verzichtet. Zudem wurde er um die nicht in die Auswertung miteinbezogenen Warming-up-Aufgaben des Subtests *Satzergänzungen* gekürzt. Dieser Subtest wies nachweislich nicht den erwünschten Effekt der Eingewöhnung auf, sondern brachte lediglich den Nachteil der Testverlängerung mit sich.

Bei den verbleibenden Subtests wurden auf Basis umfassender Testanalysen die Anzahl und Art der Items modifiziert. Während es in den Subtests *Abwicklungen*, *Grundrechnen*, *Spiegelbilder* und *Zahlenreihen* lediglich zu einer Reduktion der Anzahl der Items durch die Auswahl der besten WIT-Items kam, wurden die Subtests *Analogien*, *Eingekleidete Rechenaufgaben* und *Gleiche Wortbedeutung* deutlicheren Modifikationen unterzogen. Diese Items bedurften teilweise einer neuen semantischen Einkleidung aufgrund veralteter Begriffe bzw. Währungen

und einer kulturell und gesellschaftlich nicht mehr adäquaten Ansicht über die Stellung der Frau.

Vier Subtests wurden ganz neu erstellt. Der Subtest *E-Mails Bearbeiten* zur Erfassung der Arbeitseffizienz stellt die Simulation einer Büroroutinetätigkeit in Form einer Arbeitsprobe dar. Er verlangt die Bearbeitung eines E-Mail Posteingangs in Form von 42 Items, wobei nach einer 10-minütigen Instruktion ein Zeitlimit von 8,5 Minuten zur Bearbeitung vorgesehen ist. Der Untertest bedarf laut Autoren keiner Vorerfahrungen und soll die Fähigkeit messen, komplexe Informationen mit geringem Schwierigkeitsgrad effizient zu verarbeiten und anzuwenden. Neu konstruiert ist auch das Itemmaterial des Subtests *Merkfähigkeit*. Hier hat die Testperson 4 Minuten Zeit, sich einen Text mit verbalen, numerischen und figuralen Inhalten einzuprägen. Nach einer ca. 17-minütigen Störphase, in der andere Subtests zu bewältigen sind, werden die zu merkenden Inhalte in Form eines Multiple-Choice-Formats (6 Alternativen) abgefragt. Die Autoren betonen hierzu insbesondere, dass durch die zwischengeschaltete Störphase die Merkfähigkeit und nicht die unmittelbare Behaltensleistung erfasst wird, anders als es in „anderen Tests“ üblich sei (Kersting et al., 2008, S. 22). Die Subtests *Wissen-Wirtschaft* und *Wissen-Informationstechnologie* überprüfen mit jeweils 20 Items die Kenntnisse der Testperson auf dem Gebiet Wirtschaft und Informationstechnologie. Die richtige Lösung ist hier aus insgesamt 4 Alternativen auszuwählen.

Die acht Module des WIT-2 können alle einzeln oder in beliebigen Kombinationen vorgegeben werden. Die Module sind auf zwei Aufgabenhefte verteilt, wobei Heft 2 den Subtest *Arbeitseffizienz* und die zwei Wissenstests umfasst. Die Subtests sollten in der vorgesehenen Reihenfolge vorgegeben werden. Bei der Vorgabe von Einzelteilen sollten lediglich die entsprechenden Subtests übersprungen werden. Im Manual findet der/die AnwenderIn sehr umfangreiche und detaillierte Angaben über die Testzeiten der einzelnen Module, der Modulkombinationen und des Gesamttests. Die Vorgabe des Gesamttests erfordert laut Autoren inklusive Instruktionen ca. 2 Std. und 23 Min., was im Vergleich zum WIT als enormer ökonomischer Fortschritt zu sehen ist. Die Durchführung des WIT umfasste je nach Lang- oder Kurzforum zwischen drei und vier Stunden.

Der WIT-2 besteht aus einem Manual, einem Instruktionshft für den/die TestleiterIn, einer Informationsbroschüre, zwei Aufgabenheften 1 (Form A und Form B), zwei Aufgabenheften 2 (Form A und Form B), zwei Antwortbögen, einem Notizbogen, zwei Sets Auswertungsschablonen (Form A und Form B) und einem Profilblatt. Der Test ist als Einzel- und Gruppentest sowie als Computertest durchführbar. Bei Form A und Form B handelt es sich um Pseudoparallelförmige, bei denen identisches Itemmaterial bzgl. der Reihenfolge unterschiedlich gestaltet wurde.

Der Test enthält in der Handanweisung sehr detaillierte und gut verständliche Informationen zur Durchführung sowie ein sehr klar strukturiertes Instruktionshft für

den/die TestleiterIn, das wichtige Informationen zur Durchführung sowie die wortwörtlichen Instruktionen beinhaltet. Dies gewährleistet eine möglichst standardisierte Vorgabe des Tests. Eine Besonderheit stellt die Informationsbroschüre für die TestteilnehmerInnen dar. Sie beinhaltet allgemeine Informationen zur Vorbereitung auf den Test, eine Kurzbeschreibung der Fähigkeiten, die erhoben werden sowie Beispiel- und Übungsaufgaben. Die Autoren zielen damit auf eine erhöhte Transparenz und Akzeptanz der psychologischen Testung durch die Testpersonen ab.

Die Auswertung kann manuell oder mit Hilfe eines computergestützten Auswertungsprogramms erfolgen. Sehr detailliert und umfangreich sind auch die Anweisungen zur Auswahl der Normstichprobe und die Interpretationshinweise. Vorgesehen ist eine anforderungsorientierte Auswertung. Dabei wird den einzelnen Dimensionen, orientiert an einer vorausgehenden Anforderungsanalyse, unterschiedliche Gewichtung beigemessen und somit ein Gesamtwert gebildet. Als Indikator für die allgemeine Intelligenz (*Fluide Intelligenz*) kann das Modul 4 *Schlussfolgerndes Denken* genutzt werden. Von Profilitätsinterpretationen raten die Autoren aufgrund von Instabilitäten vehement ab, die Anforderungen zur Nutzung von Differenzwerten seien jedoch erfüllt.

Testanalyse und Normierung

Die Testanalyse erfolgte nach klassischen Testgütekriterien. Die Anforderungen an Durchführungs- und Auswertungsobjektivität sind durch die detailreichen und genauen Verfahrenshinweise (z. B. wortwörtliche Instruktion, Eingehen auf Nachfragen) und durch das standardisierte Test- bzw. Auswertungsmaterial ausreichend erfüllt. Interpretationsobjektivität gewährleisten umfassende Informationen zu den Normdaten sowie eine klare Definition der erfassten Fähigkeiten. Zwei Fallbeispiele geben Anhaltspunkte für die Interpretation der anforderungsorientierten Auswertung des WIT-2.

Das Manual umfasst Angaben zur internen Konsistenz und zur Retestreliabilität auf Ebene der Subtests und der Module. Die internen Konsistenzen der einzelnen Module liegen zwischen $\alpha = .78$ und $\alpha = .95$, die Retestreliabilitäten (Intervall: 6 Wochen oder 9 Monate) zwischen $r_{tt} = .59$ und $r_{tt} = .93$. Zusätzlich sind für jeden Subtest Trennschärfen und Itemschwierigkeiten angegeben. Zudem finden sich im Manual Angaben zu den Standardmessfehlern und Vertrauensbereichen auf Modulebene sowie für den Gesamttest. Die Reliabilitäten auf Modulebene können als zufriedenstellend klassifiziert werden, die Reliabilität des Gesamttests ist mit $\alpha = .98$ sehr hoch.

Im Manual liegen ausführliche Validitätsinformationen vor. Als Nachweis der Kriteriumsvalidität wird einerseits auf Validitätsstudien verwiesen, die den „alten“ WIT als einen Intelligenztest mit sehr hoher Kriteriumsvalidität auszeichnen, andererseits werden auch mehrere aktuelle Untersuchungen angeführt, die den Zusammenhang der einzelnen Module des WIT-2 mit den externen Kriterien

Selbsteinschätzung, Schulnoten, Erfolg in der beruflichen Weiterbildung, Erfolg in der beruflichen Erstausbildung sowie biografische Daten belegen. Die Struktur des Tests wurde mittels konfirmatorischer Faktorenanalyse untersucht. Die Ergebnisse sprechen für eine gute Übereinstimmung der durch die Normstichprobe erhobenen Daten mit dem MMPMA. Alle Subtests laden zufriedenstellend hoch auf den ihnen zugeordneten Faktoren. Zudem lassen sich alle Subtests mit Ausnahme der Wissenstests, die auf dem Faktor *kristallisierte Intelligenz* laden, dem Faktor *Fluide Intelligenz/Arbeitsgedächtnis* zuordnen. Für eine zufriedenstellende Konstruktvalidität sprechen mittlere bis hohe Korrelationen von $r = .65$ bis $r = .90$ zwischen den WIT-2 Modulen und den jeweiligen Subtests konvergenter Tests wie z. B. dem *IST 2000 R*, *BIS-r-DGP*, *CFT-3*, *FRT*, *KLT-R*, *LGT-3* und dem *Wonderlic Intelligenztest*. Es liegen Angaben zu allen Modulen wie auch zu den einzelnen Subtests vor. Die diskriminative Validität soll durch sehr niedrige Korrelationen mit Fragebögen zur Leistungsmotivation (*AMS-R*), Testängstlichkeit (*TAI*) und der Skala Gewissenhaftigkeit des *NEO-PI-R* belegt werden. Insgesamt werden zahlreiche empirische Indikatoren der konvergenten und diskriminativen Validität angeführt, die in einem für Intelligenztests üblichen Bereich liegen. Wegen der Verwendung von Verfahren, die auch in den USA vermehrt eingesetzt werden (*AMS-R*, *Wonderlic Test*), weisen die Herausgeber dem WIT-2 neben seiner hohen Praxisrelevanz auch eine beachtliche Bedeutung in der Intelligenzforschung zu.

Es liegen Normdaten von insgesamt fast 42.000 Personen vor, die laut den Testentwicklern aus eignungsdiagnostisch relevanten Gruppen (SchülerInnen, Auszubildende, BewerberInnen, RehabilitandInnen) stammen und somit in erster Linie keine Bevölkerungsrepräsentativität, sondern eine Repräsentativität für eignungsdiagnostische Fragestellungen darstellen. Die Rohwerte der einzelnen Subtests werden in Standardwerte ($\mu = 100$, $\sigma = 10$) umgerechnet. Die Skala bietet der diagnostischen Zielsetzung entsprechend eine ausreichende Differenzierungsfähigkeit. Zur Ermittlung der Dimensionswerte stehen keine spezifischen Bezugsgruppen zur Verfügung. Die entsprechenden Standardwerte werden aus dem Mittel der einzelnen Subtests gebildet (z. B.: Standardwert für die Dimension *Räumliches Denken* = Standardwert *Abwicklungen* + Standardwert *Spiegelbilder/2*).

Insgesamt stehen zur Interpretation der Rohdaten 18 (3×6) verschiedene Normtabellen gestaffelt nach Bildungsniveau (Gesamtgruppe, mit Abitur und ohne Abitur) und Altersgruppen (Gesamtgruppe, 14–17 Jahre, 18 Jahre, 19–22 Jahre, 23–27 Jahre, ab 28 Jahre) sowie die Kombination von bildungsspezifischen und altersspezifischen Normen zur Verfügung. Die Größe der Teilstichproben variiert in Abhängigkeit von Alter, Bildungsniveau und Subtest. Die Stichprobengröße reicht von 51 (Subtest *Zahlenreihen*, mit (Fach-)Hochschulreife, 28 Jahre und älter) bis 10.025 (Subtest *Gleiche Wortbedeutung*, Gesamt). Beste Differenzierungsfähigkeit ist für den Altersbereich 14 bis 42 Jahre gegeben. Es wurden nur wenige Normdaten bei über 43-Jährigen erhoben ($N = 59$ bis 108), das

Durchschnittsalter der Untersuchungsgruppen liegt bei einem Median von 17 bis 20 Jahren.

Es liegen keine getrennten Normen für die Papier-Bleistift-Version und für die Computerversion vor. Auch gibt es diesbezüglich keine Angaben zu Äquivalenzstudien. Angeführt ist jedoch eine Äquivalenzstudie zu geschlechtsspezifischen Leistungsausprägungen. Es ergaben sich in vier der elf Subtests signifikante Mittelwertsunterschiede (*Eingekleidete Rechenaufgaben, Grundrechnen, Merkfähigkeit, Wissen-Informationstechnologie*), wobei vor allem im Subtest *Wissen-Informationstechnologie* Frauen deutlich niedrigere Werte aufwiesen als Männer. Getrennte Geschlechternormen liegen nicht vor. Die Autoren weisen jedoch darauf hin, die Mittelwertsunterschiede bei der Interpretation der Werte zu berücksichtigen.

Die Erhebung der Normdaten sowie der Studien zu den Testanalysen fand zwischen Februar 2002 und Mai 2005 in den Bundesländern der Bundesrepublik Deutschland statt. Im Sinne der DIN 33430, die vorschreibt, die Angemessenheit der Norm-, Reliabilitäts- und Validitätswerte alle 8 Jahre zu überprüfen, verweisen die Autoren auf die entsprechende Homepage zum WIT-2: www.kersting-internet.de/wit. Dort sollen aktuelle Informationen zum Test veröffentlicht werden.

Den Gütekriterien „Störanfälligkeit“ und „Unverfälschbarkeit“ soll durch ausführliche Informationen zur Testdurchführung, durch die Gestaltung der Testmaterialien und durch Anleitungen bezüglich der zu schaffenden Rahmenbedingungen genüge getan werden. Zur Frage der Skalierung findet der/die TestanwenderIn den Hinweis im Manual, dass im Rahmen der klassischen Testtheorie eine Überprüfung der Skalierung nicht möglich sei. Indirekte Hinweise auf die Skalierung sehen die Autoren jedoch durch die hohen internen Konsistenzen, die hohen Trennschärfen sowie durch die Ergebnisse der Strukturanalysen gegeben. Ökonomie und Zumutbarkeit sind als sehr gut einzuschätzen. Die Akzeptanz des WIT-2 wurde mit einem Fragebogen erhoben, der darauf abzielt, die Einschätzung von Kontrollierbarkeit, Messqualität, Augenscheinvalidität und Belastungsfreiheit bezüglich bestimmter Verfahren von TestteilnehmerInnen zu erheben. Dabei zeigte sich eine hohe Akzeptanz bezüglich der Kontrollierbarkeit und Belastungsfreiheit.

Plus/Minus

Mit der Neuentwicklung des WIT-2 ist es den Autoren gelungen, ein Verfahren hervorzubringen, das moderne Qualitätsstandards und Anforderungen an psychologische Tests hinreichend erfüllt. Der diagnostischen Zielsetzung entsprechend eignet der Test sich hervorragend für eignungsdiagnostische Fragestellungen und für die Schullaufbahn-, Berufs- und Studienberatung im Altersbereich der 14- bis 28-Jährigen. Den zu Recht geäußerten Kritikpunkten am WIT wurde Rechnung getragen. Der WIT-2 basiert zwar auf dem altbewährten Primärfaktoren-

modell nach Thurstone (1938), wurde jedoch um einige Erkenntnisse der modernen Intelligenzforschung erweitert. Die Auswahl der besten Items und Subtests, die Modifikation bzw. semantische Neueinkleidung veralteter Items, die Neukonstruktion von Modulen sowie die vollständig neuen Berechnungen der Normen und Testgütekriterien haben dazu geführt, dass der WIT-2 mehr als eine Revision des alten Tests darstellt.

Das Verfahren zielt auf eine klar definierte Personengruppe ab und ist darüber hinaus nur mit Einschränkungen anwendbar. Die optionalen Module 6 bis 8, die dem noch sehr unzureichend ausgearbeiteten dritten Stratum des MMPMA zuzuordnen sind, erfassen sehr berufsspezifische Kenntnisse und Fertigkeiten, die betreffenden Normen hierzu differenzieren jedoch bei domänenspezifisch gut gebildeten TestteilnehmerInnen nur unzureichend. Die Normstichproben umfassen vor allem sehr junge Testpersonen. So liegen z.B. in der Normstichprobe des Subtests *Analogien* Mittelwert und Median bei 19,4 bzw. 17 Jahren. Mehr als 50% der Personen der Normstichprobe repräsentieren die Altersgruppe der 14- bis 17-Jährigen, während die Altersgruppe der ab 43-Jährigen nur 1,1% ausmacht. Einige Normstichproben zu den Subtests fallen sehr gering aus. Da eignungsdiagnostische Fragestellungen auch die über 40-Jährigen betreffen, sind Normdaten auch für diese Altersgruppen wünschenswert. Nicht erwähnt ist die regionale Verteilung der Stichprobe. Um eine Repräsentativität der Daten für den gesamten deutschen Sprachraum zu gewährleisten, ist auch eine Einbeziehung von österreichischen und Schweizer Jugendlichen und Erwachsenen angebracht, zumal der Subtest *Wissen-Wirtschaft* auch zwei deutschlandspezifische Fragen beinhaltet. Eine Studie zum WIT (Bettinville, Rabenberg & Haensgen, 2005) zur Vergleichbarkeit der Papier-Bleistift-Version und der Computerversion zeigte eine strukturelle, jedoch keine normative Äquivalenz der beiden Medien. Signifikante Mittelwertsunterschiede sprechen für eine gesonderte Normierung. Getrennte Normen für Papier-Bleistift-Version und computergestützte Datenerhebung oder zumindest eine Äquivalenzüberprüfung zum WIT-2 stehen hier noch aus.

Wie bei vielen kultur- und bildungsabhängigen Konstrukten ist auch hier zu befürchten, dass die Items der Wissenstests – bedenkt man den raschen Wandel wirtschaftlicher und informationstechnologischer Fortschritte und Erkenntnisse – sehr bald an Aktualität und dadurch auch an Akzeptanz durch die TestteilnehmerInnen verlieren werden. Weitere Subtests als Indikatoren für kristallisierte Intelligenz wären günstig, um Diagnosen bezüglich der fluiden und kristallisierten Intelligenz ableiten zu können.

Die beiden Testformen A und B wurden als Pseudoparallelförmigkeiten konzipiert, um unterschiedliche Aufgabenschwierigkeiten zu vermeiden, was allerdings bei Testwiederholungen Übungs- und Lerneffekte mit sich bringt. Kritisch zu sehen ist auch die fehlende Überprüfung des Gütekriteriums der Skalierung mittels Raschmodell-Analyse. Eine adäquate Abbildung der Verhaltensrelationen und

die Verrechnungsfairness kann für diesen Test daher nicht angenommen werden (Kubinger, 2003).

Zusammenfassend stellt der WIT-2 ein sehr gutes und modernes Testverfahren mit Entwicklungspotenzial dar. Es ist theoretisch gut begründet und bezüglich aller Gütekriterien ausreichend. Besonders hervorzuheben ist die hohe Transparenz des Verfahrens für TestanwenderInnen und TestteilnehmerInnen. Auf der bereits erwähnten Website zum WIT-2 sind neben sehr ausführlichen Hinweisen zum Testaufbau, Testanalysen und sonstigen Informationen auch aktuelle Studien, neue Entwicklungen und Änderungen bzgl. der Verfahrenshinweise nachzulesen. Ein ständiger Austausch mit PraktikerInnen und eine potenzielle Verbesserung des Tests sind erwünscht. Für Testpersonen steht eine Informationsbroschüre zur Vorbereitung zur Verfügung. Das Manual orientiert sich stark an den Anforderungen an Verfahrenshinweisen der DIN 33430 und erfüllt diese auch. Auf der letzten Seite befindet sich bereits eine Checkliste mit der Angabe der Seiten im Testmanual, auf denen die jeweiligen Informationen zu finden sind.

Literatur

- Bettinville, A., Rabenberg, H. & Haensgen, K. D. (2005). Untersuchung zum WILDE-Intelligenz-Test (WIT): Vergleichbarkeit der Anwendung der Papier-Bleistift-Fassung vs. computerbasierten Durchführung an Daten des Berufsförderungswerkes Leipzig. *Die Rehabilitation*, 44, 237–243.
- Jäger, A. O. & Althoff, K. (1983). *Der Wilde-Intelligenz-Test (WIT)*. Göttingen: Hogrefe.
- Jäger, A. O. & Althoff, K. (1994). *Der Wilde-Intelligenz-Test (WIT)* (2. rev. Auflage). Göttingen: Hogrefe.
- Jäger, R. S. (1997). Testrezension zu WILDE-Intelligenz-Test (WIT). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, 62–63.
- Kersting, M., Althoff, K. & Jäger, A. O. (2008). *Wilde-Intelligenz-Test 2 (WIT-2)*. Göttingen: Hogrefe.
- Kubinger, K. D. (2003). Skalierung. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 387–390). Weinheim: Beltz PVU.
- Testkuratorium. (2006). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. *Psychologische Rundschau*, 58, 25–30.
- Thurstone, L. L. (1938). The perceptual factor. *Psychometrika*, 3, 1–17.

Hannelore Koch, Wien

DOI: 10.1026/0049-8637.40.4.208