

**Testbesprechung:
PSYNDEX Test-Dokument (<http://www.zpid.de>):
WIT-2. Wilde-Intelligenz-Test 2.
(PSYNDEX Tests Review)**

Preis: Euro 328,00 (Test komplett; Stand: 23.10.2009)

Bezugsquelle: Testzentrale; URL: <http://www.testzentrale> / Stand: 23.10.2009

Autoren des Tests: Kersting, Martin; Althoff, Klaus; Jäger, Adolf Otto

Publikationsjahr 2008

Testbesprechung: WIT-2. Wilde-Intelligenz-Test 2 von Kersting, M., Althoff, K. & Jäger, A.O. (2008).

Testbesprechung von: Dr. Ludwig Kreuzpointner (Universität Regensburg)

1. Testkonzept

1.1 Theoretischer Hintergrund

Der Wilde-Intelligenz-Test-2 (WIT-2; Kersting, Althoff & Jäger, 2008) ist ein auf der klassischen Testtheorie basierender, mehrdimensionaler Intelligenztest vornehmlich entworfen für die Eignungsdiagnostik. Er stellt eine Revision des Wilde-Tests (Jäger, 1963) bzw. Wilde-Intelligenz-Tests (WIT; Jäger & Althoff, 1983, 1994; "WIT-1") dar. Es wurden acht der 15 Subtests übernommen, bei sechs Subtests wurden die besten Items ausgewählt und nur geringfügige Änderungen vorgenommen, zwei Subtest wurden deutlich modifiziert. Vier Subtests wurden völlig neu entwickelt. Die 11 Subtests des WIT-2 werden in acht Module (Dimensionen) aufgeteilt, die prinzipiell unabhängig voneinander administriert werden können.

Basierte der WIT-1 noch auf dem Strukturmodell nach Thurstone (z.B. 1938), wird für den WIT-2 ein den aktuellen Erkenntnissen der Intelligenzforschung angepasstes Modell vorgestellt: Das modifizierte Modell der Primary Mental Abilities (MMPMA). Dieses Modell ergänzt Thurstones Strukturmodell (PMA) um "(1) Facettenansatz, (2) Hierarchieannahme und (3) Kognitive Korrelate Ansatz" (Kersting et al, 2008, S. 23). Als Facettenansatz beschreiben die Autoren die Zusammenfassung von drei Subtests, die sich zwar inhaltlich unterscheiden (figural, verbal, numerisch), operational sich allerdings im Sinne schlussfolgernden Denkens entsprechen (entsprechend dem Vorgehen im BIS von Jäger, 1982). Hierarchisch wird eine Ebene über und eine Ebene unter den PMA postuliert und zu Ehren von John Carroll Stratum genannt. Die Ebene III über den PMA, die in Ebene II teilweise nochmals hierarchisch angeordnet werden, stellt den höchsten Generalisierungsgrad dar. In ihr werden die fluide und kristallisierte Intelligenz im Sinne von Cattell und Horn (z.B. Cattell, 1971) abgebildet. Dem "Kognitive Korrelate-Ansatz" entspricht die Arbeitshypothese, "dass die fluide Intelligenz (als die dem schlussfolgernden Denken übergeordnete Fähigkeit) und das Arbeitsgedächtnis hochgradig überlappende Konstrukte sind" (Kersting et al., 2008, S. 25) und entsprechend

als derselbe Faktor modelliert werden. Im Stratum I werden spezifischere Faktoren angenommen. Auf Basis des WIT-2 werden zwei Wissensdomänen und ein Faktor Arbeitseffizienz operationalisiert. Die Faktoren des Stratum I sollen von mehreren Faktoren der beiden höheren Ebenen beeinflusst werden können. Eine Darstellung und grafische Aufbereitung des Modells findet sich außer im Testmanual bisher nur online (URL: http://kersting-internet.de/wit/iq_modell.html, Stand: 22.10.2009).

Der WIT-2 erfasst folgende Dimensionen (Kersting et al., 2008, S. 29, Kasten 1):

- Modul 1: Sprachliches Denken. Die Fähigkeit, mit sprachlichen Konzepten umzugehen, wobei Wortschatz, Sprachverständnis und sprachlogisches Denken eine Rolle spielen.
- Modul 2: Rechnerisches Denken. Die Fähigkeit, einfache Rechenoperationen der Addition, Subtraktion, Multiplikation und Division korrekt auszuführen.
- Modul 3: Räumliches Denken. Die Fähigkeit zur Vorstellung räumlicher Relationen.
- Modul 4: Schlussfolgerndes Denken. Die Fähigkeit, bestimmte logische Regeln und Gesetzmäßigkeiten zu erkennen und zweckentsprechend anwenden zu können.
- Modul 5: Merkfähigkeit. Die Fähigkeit, sich kurz zuvor eingeprägte Informationen und Assoziationen zu merken und wieder zu erkennen.
- Modul 6: Arbeitseffizienz. Die Leistung, komplexe Informationen eines vergleichsweise geringen Schwierigkeitsgrades mit andauernder Konzentration effizient zu verarbeiten und anzuwenden. 'Effizienz' wird dabei durch die gleichzeitige Optimierung von Arbeitsgeschwindigkeit und -genauigkeit bei der Selektion und Extraktion relevanter Informationen und deren anschließender Kombination erzielt.
- Modul 7: Wissen Wirtschaft. Kenntnisse auf dem Gebiet Wirtschaft.
- Modul 8: Wissen Informationstechnologie. Kenntnisse auf dem Gebiet Informationstechnologie.

1.2 Testaufbau

Der WIT-2 besteht aus 243 fortlaufend nummerierten Items, die auf 11 Subtests verteilt den acht Modulen (Dimensionen) zugeordnet werden können, wobei einige Subtests in zwei Dimensionen verrechnet werden (nach Kersting et al., 2008, S. 30, Tabelle 3):

- (AL) Analogien (Sprachliches Denken und schlussfolgerndes Denken, 20 Items): Auf der linken Seite eines Gleichheitszeichens sind zwei Wörter vorgegeben, die in einer bestimmten Beziehung zueinander stehen, auf der rechten Seite ein Wort. Von fünf Wahlwörtern ist dasjenige auszuwählen, das auf der rechten Seite des Gleichheitszeichens eine analoge Beziehung herstellt.
- (AW) Abwicklungen (Räumliches Denken und schlussfolgerndes Denken, 20 Items): Zu einer aus mehreren Flächen mit verschiedenen Zeichen bestehenden Faltvorlage ist aus fünf vorgegebenen Körpern derjenige herauszufinden, der sich aus der Faltvorlage herstellen lässt.
- (EM) E-Mails Bearbeiten (Arbeitseffizienz, 42 Items): Die Aufgabe simuliert die Bearbeitung des E-Mail-Posteingangs. Zur Bearbeitung des Posteingangs stehen

sechs Möglichkeiten zur Verfügung, z.B. "Weiterleiten der E-Mail", "Speichern der E-Mail", "Beantworten der E-Mail" usw., wobei bestimmte Bedingungen zu beachten sind. Beispielsweise sind Datum und Dringlichkeit der Nachricht, das Projekt, um welches es in der E-Mail geht, sowie die Firma, aus welcher die E-Mail stammt, für die Wahl der richtigen Lösung entscheidend.

- (ER) Eingekleidete Rechenaufgaben (Rechnerisches Denken, 20 Items): Verbal eingekleidete Rechenaufgaben sind vorgegeben und auf dem Antwortbogen sind die Ziffern der Lösungen einzutragen.
- (GR) Grundrechnen (Rechnerisches Denken, 20 Items): Zu Rechenaufgaben der vier Grundrechenarten sind auf dem Antwortbogen die Ziffern der Lösungen durchzustreichen.
- (GW) Gleiche Wortbedeutungen (Sprachliches Denken, 20 Items): Zu einem vorgegebenen Wort ist aus fünf anderen Wörtern das sinnähnlichste herauszusuchen.
- (MF) Merkfähigkeit (Merkfähigkeit, 21 Items): In einem Text befinden sich verbal, numerisch und figural kodierte Informationen. Einer Einprägungszeit von 4 Minuten folgen für ca. 17 Minuten andere Subtests (Störphase). In dem anschließenden Reproduktionstest sollen Einzelheiten unterschieden werden. Die richtige Lösung ist aus einer Reihe von sechs Alternativen auszuwählen.
- (SP) Spiegelbilder (Räumliches Denken, 20 Items): Von fünf Strichfiguren lassen sich vier durch einfaches Verschieben in der Ebene, die fünfte dagegen nur durch Umklappen in der Vorstellung zur Deckung bringen; diese Figur ist herauszufinden.
- (WI) Wissen Informationstechnologie (Wissen Informationstechnologie, 20 Items): Es werden Fragen zum Thema Informationstechnologie gestellt. Die richtige Lösung ist aus vier Alternativen auszuwählen.
- (WW) Wissen Wirtschaft (Wissen Wirtschaft, 20 Items): Es werden Fragen zum Thema Wirtschaft gestellt. Die richtige Lösung ist aus vier Alternativen auszuwählen.
- (ZN) Zahlenreihen (Schlussfolgerndes Denken, 20 Items): Vorgegeben ist eine Folge von Zahlen, die nach einer Regel aufgebaut ist; auf dem Antwortbogen sind die Ziffern der Zahl durchzustreichen, die als nächstes Glied der Reihe folgen müsste.

1.3 Auswertungsmodus

Jedes Item kann klar als richtig oder falsch beurteilt werden. Pro Subtests wird die Anzahl richtig gelösten Items bestimmt. Die Rohwerte werden per Normierungstabelle in Standardwerte transformiert. Zur Bestimmung der Dimensionswerte, die alleinig zur Interpretation herangezogen werden sollen, werden die entsprechenden Subtest-Standardwerte ungewichtet gemittelt. Die Werte können anschließend in eine Profildarstellung übertragen werden, aus der neben den Standardwerten IQ-Werte und Prozentränge ersichtlich werden. Von einer Profilinterpretation wird jedoch abgeraten. Zur Bestimmung eines Gesamtwertes wird eine den Anforderungen der diagnostischen Fragestellung angepasste gewichtete Mittelwertbildung der relevanten Standardwerte empfohlen.

1.4 Auswertungshilfen

Für die Auswertung per Hand liegen Schablonen bereit. Das Profilblatt ermöglicht eine übersichtliche Zusammenstellung der Ergebnisse. Durch die Darstellung und Schattierungen wird eine falsche Verrechnung der Subtests zu den Dimensionen nahezu ausgeschlossen. Neben der manuelle Auswertung kann ein Testauswerteprogramm käuflich erworben werden.

1.5 Auswertungszeit

Die Auswertungszeit variiert aufgrund der möglichen Auswahl einzelner Module. Für die Auswertung des Gesamttests wird im Manual (Kersting et al., 2008, S. 41) eine Bearbeitungszeit von 30 Minuten pro Teilnehmer angegeben.

1.6 Itembeispiele

Anmerkung: Als Itembeispiele werden, soweit darstellbar, die Beispielitems der Instruktionen angeführt. Anderweitig sind sie unter "Testaufbau" beschrieben.

- (AL) Analogien
Schaf : Wolle = Vogel : ?
Flügel B) Nest C) Adler D) Federn E) Tier

- (ER) Eingekleidete Rechenaufgaben
Aus Zinseinkünften erhält eine Kontoinhaberin monatlich 30 Euro. Wie viele Euro bekommt sie dann in 7 Monaten?

- (GR) Grundrechnen
53
+ 44

= ?

- (GW) Gleiche Wortbedeutungen
Kopf
Haar B) Haupt C) Körper D) Mensch E) Hut

- (ZN) Zahlenreihen
5 8 11 14 17 ?

2. Durchführung

2.1 Testformen

Der WIT-2 kann als Einzel- und Gruppentest durchgeführt werden. Es liegen zwei pseudoparallele Testformen vor, die sich in Itemreihenfolge und teilweise in der Reihenfolge der Darstellung der Distraktoren unterscheiden. Da jedes der acht Module separat durchgeführt werden kann, existiert prinzipiell eine Vielzahl von Testformen. Bei entsprechendem Bedarf können den Bedürfnissen angepasste Testhefte produziert werden (Kersting et al., 2008, S. 33). Der Test liegt bisher nur in Deutsch vor, fremdsprachige Testformen werden jedoch angestrebt (S. 85).

2.2 Altersbereiche

Ab 14 Jahre. Normierungen liegen bis 42 Jahre vor. Eine Datenbasis älterer Teilnehmer wird angestrebt.

2.3 Durchführungszeit

Abhängig von den durchgeführten Modulen variiert die Durchführungszeit. Für die Gesamttestdauer werden inklusive Instruktionen und Pause 2:23 Stunden angegeben. Für die einzelnen Module werden folgende Gesamtzeiten veranschlagt (Kersting et al., 2008, S. 35, Tabelle 5):

- Modul 1 (Sprachliches Denken): 12:00 min
- Modul 2 (Rechnerisches Denken): 26:30 min
- Modul 3 (Räumliches Denken): 22:00 min
- Modul 4 (Schlussfolgerndes Denken)
- Als eigenständiges Modul (Subtests AL, ZN und AW): 34:30 min
- Falls auch Module 1 und 3 eingesetzt werden (nur Subtest ZN): 14:00 min
- Modul 5 (Merkfähigkeit einschließlich "Störaufgabe"): 26:00 min
- Modul 6 (Arbeitseffizienz): 18:30 min
- Modul 7 (Wissen Wirtschaft): 5:30 min
- Modul 8 (Wissen Informationstechnologie): 5:30 min

2.4 Material

Die Testmappe enthält das Manual, ein Instruktionsheft und das vollständige Testmaterial (zwei Testhefte pro Testform, Schablonen). Daneben ist seitens des Teilnehmers nur ein Stift erforderlich. Der Testleiter benötigt zudem eine Stoppuhr. Verbrauchsmaterial (Antwortbögen, Testprofilbögen, Notizblätter) für 10 Testungen liegen der Mappe bei. Weiteres Material kann separat bezogen werden. Die Testhefte sind wieder verwendbar. Für die separat zu erwerbende Computerversion des WIT-2 wird das Hogrefe Testsystem benötigt, das zusätzlich vom Verlag gekauft werden muss. Die PC-Testung kann dann lokal oder auch online durchgeführt werden. Je nach Testungsart entstehen unterschiedliche Kosten pro Durchführung. Alle Teile sind separat erhältlich.

2.5 Instruktion

Die ausführliche Instruktion erfolgt zum einen schriftlich im Testheft. Hierfür ist pro Subtest eine extra Seite vorgesehen. Der Testleiter liest zudem jede Instruktion aus einem Instruktionsheft vor, in dem weitere Anmerkungen hinsichtlich der Durchführung, aber auch bezüglich der Reaktionen auf Nachfragen angegeben sind.

2.6 Durchführungsvoraussetzungen

"Die Testanweiser sollten intellektuell und sozial kompetent, durchsetzungsfähig und verantwortungsbewusst sein" (Kersting et al., 2008, S. 36). Einer speziellen Vorbildung bedürfte es nicht, da im Instruktionsheft alle relevanten Informationen aufgeführt seien. Bei Gruppentestung sollte bei einem Testleiter die Gruppe nicht größer als 30 Personen und nie größer als 50 Personen sein. Für den Testraum werden ausreichende Größe und Arbeitsplätze, geeignete Beleuchtung, Akustik und Rahmenbedingungen (z.B. Ruhe, Ungestörtheit) gefordert (Kersting et al., 2008, S. 36).

3. Testkonstruktion

Der WIT-2 stellt die auf der Klassischen Testtheorie basierende Überarbeitung des WIT (Jäger & Althof, 1983, 1994) dar. Neben seinen überalterten Normen (Kersting et al., 2008) wurden auch seine Ökonomie und Zumutbarkeit in Frage gestellt (Jäger, 1997; Kubinger, 2001). Auf Basis der vielen Ergebnisse zu Konstrukt- und Kriteriumsvalidität wurden die Dimensionen Sprachgebundenes Denken, Rechnerisches Denken, Räumliches Vorstellen und Gedächtnis beibehalten. An der Dimension Schlussfolgerndes Denken wurde aufgrund ihrer Wichtigkeit trotz inkonsistenter Nachweise festgehalten. Die konstruktvalide Erhebung der Dimensionen Wortflüssigkeit und Wahrnehmungsgeschwindigkeit wurden als zu aufwändig gesehen und ausgeschlossen. Die beibehaltenen Subtests wurden aus den besten Items der beiden parallelen Versionen zusammengestellt. Für einige Subtests wurden zudem neue Items konstruiert. Basierend auf den Anforderungen der Eignungsdiagnostik und dem modifizierten theoretischen Modell der MMPMA (siehe unter "Theoretischer Hintergrund") wurden vier neue Subtests entwickelt, so dass nach Entfernen von acht Subtests des alten WIT 11 Subtests zur Erhebung von acht Dimensionen zusammengestellt wurden. Da die Dimensionen Gedächtnis und Merkfähigkeit nicht konsistent erhoben werden konnten, wurde der Subtest Merkfähigkeit neu konstruiert. Für die Dimensionen Arbeitseffizienz, Wissen IT und Wissen Wirtschaft, die auf Stratum I lokalisiert werden, wurde je ein Subtest konstruiert bzw. aufgenommen. Weitere Entwicklungen auf dieser Ebene I, also weitere Wissenstests und Arbeitsproben, werden als wünschenswert gesehen (Kersting et al., 2008, S. 85).

4. Gütekriterien

4.1 Objektivität

Die Durchführungsobjektivität wird durch umfassende und klar verständliche Anweisungen sowohl im Vorfeld durch das Manual als auch durch das extra vorliegende Instruktionsheft für den Testleiter sicherzustellen versucht. Die extra erhobene Akzeptanz des Verfahrens hat gezeigt, dass die Versuchspersonen "jederzeit wussten, was von ihnen bei der Aufgabenbearbeitung verlangt wurde" (Kersting et al., 2008, S. 56). Die Auswertungsobjektivität ist aufgrund der klaren Beurteilung der Items als richtig oder falsch durch Schablonen und den gut strukturierten Ergebnisbogen gewährleistet. Aufgrund der Darstellung der Dimensionen und diverser Normtabellen ist Interpretationsobjektivität gegeben.

4.2 Reliabilität

Die Werte der internen Konsistenz (Cronbach Alpha) der Subtests liegen zwischen .77 und .95; sie wurden an unterschiedlichen Stichproben mit $2234 \leq n \leq 10024$ Teilnehmern erhoben. Für die Dimensionen, die aus mehreren Subtests konstruiert sind, werden Werte zwischen .87 und .94 (berechnet nach dem stratifizierten Alpha; Feldt & Brennan, 1989) bei $n = 839$ angegeben. Die Retestreliabilität wurde an zwei Stichproben erhoben, wobei jeweils nur eine Auswahl der Subtests benutzt wurde. Der Delay beträgt neun Monate ($n = 63$) bzw. sechs Wochen ($n = 30$), die Kennwerte liegen zwischen .59 und .93 (vgl. Tabelle 1). Für den Gesamtscore wurde ein stratifiziertes Alpha von .98 und eine Retestreliabilität (9 Monate) von .88 ermittelt.

Tabelle 1

Reliabilitätskennwerte der Subtests und Dimensionen des WIT-2

Subtests/Module	interne Konsistenz		Retestreliabilität		
	N	alpha	N	delay	r
Sprachliches Denken (AL,GW)	839	.87	63	9 Monate	.80
Analogien	9219	.79	63	9 Monate	.75
Gleiche Wortbedeutungen	10024	.78	63	9 Monate	.67
Rechnerisches Denken (ER,GR)	839	.88	63	9 Monate	.79
Eingekleidete Rechenaufgaben	2714	.81	63	9 Monate	.76
Grundrechnen	2461	.77	63	9 Monate	.62
Räumliches Denken (AW,SP)	839	.94	63	9 Monate	.76
Abwicklungen	3092	.84	63	9 Monate	.84
Spiegelbilder	2432	.91	29	6 Wochen	.73
Schlussfolg. Denken (ZN,AL,AW)	839	.94	63	9 Monate	.86
Zahlenreihen	2488	.87	63	9 Monate	.72
Merkfähigkeit	2234	.78	29	6 Wochen	.67
Arbeitseffizienz (E-Mails Bearb.)	2440	.95	63	9 Monate	.59
Wissen IT	2285	.89	30	6 Wochen	.93
Wissen Wirtschaft	2440	.81	30	6 Wochen	.84
Gesamt	839	.98	63	9 Monate	.88

Anmerkungen. Alle Korrelationen einseitig $p \leq .01$.

4.3 Validität

Kriteriumsvalidität des WIT-2 wird von den Autoren auf zweierlei Arten aufgezeigt. Zum einen werden Korrelationen mit relevanten Außenkriterien angeführt, zum anderen wird die Gültigkeit ähnlicher Verfahren sowie der Vorläuferversion aufgezeigt. Eine metaanalytische Aufbereitung der übernommenen Subtests des WIT-1 zeigt reliabilitätskorrigierte Korrelationen mit nicht näher spezifizierten Außenkriterien zwischen .20 und .37 (unkorrigiert zwischen .15 und .24, bei 12 bis 21 Studien mit $1009 \leq n \leq 1781$ Teilnehmern).

Die statistisch signifikanten Korrelationen mit Schulnoten entsprechen in den meisten Fällen den Erwartungen ($n = 129$ Schüler und Abiturienten zwischen 14 und 22 Jahren). Mit der Mathematiknote korrelieren Rechnerisches Denken (.24), Schlussfolgerndes Denken (.23) und Sprachliches Denken (.15), mit der Deutschnote Sprachliches Denken (.23), Merkfähigkeit (.18), Arbeitseffizienz (.23) und Wissen Wirtschaft (.22). Die Korrelationen mit den Durchschnittszeugnisnoten liegen zwischen .15 (Merkfähigkeit) und .37 (Sprachliches Denken). In einer Untersuchung mit $n = 367$ Erwachsenen (vorwiegend ohne Abitur) zeigen sich Korrelationen mit der abgefragten Abschlussnote zwischen .08 (räumliches Denken) und .25 (sprachliches Denken).

Bei $n = 104$ Teilnehmern eines Rehabilitationsvorbereitungslehrgangs wurden die WIT-2 Werte mit den Schlussnoten in Mathematik und Deutsch in Zusammenhang gesetzt. Während die Schulabschlussnote nur gering mit den beiden Bereichen korreliert (.18 für Mathematik und .01 für Deutsch), zeigen sich bedeutsame Einzelkorrelationen für fast alle Dimensionen des WIT-2: Mathematik korreliert zwischen .14 (Arbeitseffizienz) und .44 (Räumliches Denken); Deutsch korreliert zwischen .22 (Merkfähigkeit) und .46 (Sprachliches Denken). Die multiple Korrelation aller Dimensionen mit der Mathematiknote betrug .58. Die Deutschnote zeigt eine multiple Korrelation mit Abschlussnote, Sprachlichem Denken und Rechnerischem Denken von .51. In beiden Fällen führen die WIT-2-Dimensionen zu einer signifikanten Steigerung der Notenvorhersage, wenn zuerst nur die Abschlussnote berücksichtigt wurde. Beim Vergleich der WIT-2-Werte mit dem Ausbildungserfolg einer Erstausbildung (vorwiegend zum Fachinformatiker) bei $n = 51$ Teilnehmern sind die Korrelationen der Dimensionen (.11 mit Räumlichem Denken bis .39 mit Rechnerischem Denken) geringer als die Schulabschlussnote (.49). Die geringste Korrelation besteht interessanterweise mit Wissen IT (.10).

In einer Studie mit $n = 49$ Studierenden wurden die beiden Dimensionen Wissen IT und Wissen Wirtschaft mit der mittels Fragebogen erhobenen Erfahrung in Informationstechnologien und finanzwirtschaftlichen Fragen verglichen. Für den Bereich IT zeigte sich eine Korrelation von .77, für den Bereich Wirtschaft ein Wert von .31.

Zur Beleuchtung der Konstruktvalidität werden basierend auf $n = 839$ Datensätzen vier Strukturgleichungsmodelle angegeben, deren Modellfit die Modellannahmen des MMPMA jeweils rechtfertigen. Zwei Modelle zeigen, dass die Zuordnung der Subtests zu den Dimensionen und die Annahme übergeordneter Faktoren in Form von fluider und kristalliner Intelligenz gerechtfertigt sind. Ein drittes Modell zeigt, dass die Annahme des den anderen Dimensionen übergeordneten Faktors Schlussfolgerndes Denken beibehalten werden kann. Das vierte Modell ($n = 641$) berücksichtigt neben den WIT-2-

Subtests konstruktähnliche Subtests anderer Verfahren. Die Berücksichtigung dieser weiteren Verfahren entsprechend der postulierten Inhalte führt wiederum zu einem akzeptablen Modellfit.

Die WIT-2-Dimensionen wurden zudem mit einer Vielzahl anderer Verfahren zur Bestimmung der konvergenten und divergenten Validität verglichen. Im Vergleich mit den Skalen des IST 2000 R (Amthauer, Brocke, Liepmann & Beauducel, 2001) liegen Korrelationen in erwartbaren Höhen vor (vgl. Tabelle 2). Lediglich die Korrelationen mit Schlussfolgerndem Denken resp. Reasoning sind insgesamt deutlich höher, was jedoch, da sie aus in den anderen Dimensionen/Skalen bereits berücksichtigten Subtests gebildet werden, zu erwarten ist. Die Korrelationen mit dem Figure Reasoning Test (FRT; Daniels & Booth, 2004) fallen in diesem Zusammenhang etwas niedriger aus (.53 mit Schlussfolgerndem Denken).

Tabelle 2

Korrelationen einiger WIT-2-Dimensionen mit den Skalen des IST 2000 R und dem FRT (n = 78; Kersting, Althoff & Jäger, 2008, S. 75)

WIT-2 ... Denken	IST- verbal	IST- numerisch	IST- figural	IST- Reasoning	FRT
Sprachliches	.79	.56	.45	.74	.33
Rechnerisches	.57	.81	.33	.71	.44
Räumliches	.37	.56	.60	.62	.58
Schlussfolgerndes	.61	.75	.62	.81	.53

Anmerkungen. Alle Korrelationen einseitig $p \leq .01$.

Die Korrelationen der WIT-2-Dimensionen mit dem CFT-3 (Cattell & Weiß, 1971) sind entsprechend der Implikationen für eine Stichprobe von n = 79 Erwachsenen für Sprachliches Denken am niedrigsten (.49) und für Schlussfolgerndes Denken am höchsten (.69). Im Vergleich zum Leistungsprüfsystem (LPS; Horn, 1983) zeigen ebenso die jeweiligen Faktoren Schlussfolgerndes Denken (im LPS die Subtests 3 und 4) die höchste Korrelation (.66). Allerdings korreliert das räumliche Denken des LPS (Subtests 8 und 9) auch am höchsten mit Schlussfolgerndem Denken des WIT-2 (.58), der Wert bezüglich Räumlichen Denkens im WIT-2 ist jedoch nur geringfügig niedriger (.56). Mit der mittels Mannheimer Rechtschreib-Test (M-R-T; Jäger & Jundt, 1981) erhobenen Rechtschreibleistung korreliert interessanterweise das Rechnerische Denken am höchsten (.39), der Zusammenhang mit dem Sprachlichen Denken ist etwas geringer (.30), noch geringer der Zusammenhang zum Schlussfolgernden Denken (.24). Die Korrelation mit dem Räumlichen Denken ist hingegen 0. Die Zusammenhänge sowohl der einzelnen Subtests des WIT-2 als auch der Dimensionen mit dem Rechtschreibtest START-R (Kersting, in Vorb.; n = 290) liegen demgegenüber allesamt im Bereich von .34 (Wissen Wirtschaft) und .55 (Schlussfolgerndes Denken).

Die internationale Einordnung des WIT-2 erleichtert den Angaben der Autoren nach der Vergleich mit dem in den USA verbreiteten Wonderlic Intelligenztest (Wonderlic, 1992). Das Schlussfolgernde Denken korreliert mit diesem Verfahren zu .66 (die drei konstruierenden Subtests Analogien, Abwicklungen und Zahlenreihen korrelieren mit .50, .43 und .56).

Die Zusammenhänge zur Skala Gewissenhaftigkeit des NEO-PI-R (Ostendorf & Angleitner, 2004) wurde für das Schlussfolgernde Denken (.18) und die Arbeitseffizienz (.23) an $n = 111$ Erwachsenen untersucht.

Die Dimension Merkfähigkeit wurde am Lern- und Gedächtnistest (LGT-3; Bäumler, 1974) validiert ($n = 76$). Die Gesamtscores korrelieren mit .65, werden hingegen nur die räumlichen Items berücksichtigt, zeigt sich kein signifikanter Zusammenhang (.17).

4.4 Normierung

Es liegen drei bildungsspezifisch (gesamt, ohne und mit Abitur) und sechs altersspezifisch (gesamt, 14-17, 18, 19-22, 23-27 und älter 28 Jahre) aufgeteilte, also insgesamt 18 Normgruppentabellen vor (Standardwerte $M = 100$, $SD = 10$). Die Größe der Stichproben schwankt je nach Subtest zwischen 2234 und 10024. Ziel bei der Normerstellung war die "Repräsentativität für eignungsdiagnostisch relevante Gruppen (Bewerber, Rehabilitanden)" (Kersting et al., 2008, S. 43). Auf geschlechtsspezifische Normen wurde verzichtet, da Geschlechtereffekte "häufig mit entwicklungspsychologischen Effekten und Bildungseffekten sowie Effekten der Selbstselektion konfundiert" (S. 48) seien. Zudem weisen die Autoren darauf hin, dass sich in den meisten Subtests keine geschlechtsspezifischen Unterschiede zeigen, die Effekte für Rechnerisches Denken nur gering sind, der Effekt bei Merkfähigkeit mittel zugunsten der Frauen ist und der hohe Effekt bei Wissen IT zugunsten der Männer in Übereinstimmung mit anderen Ergebnissen steht.

5. Anwendungsmöglichkeiten

Der WIT-2 wurde explizit für die Verwendung in der beruflichen Eignungsdiagnostik, Personalauswahl und Personalentwicklung konzipiert, was nicht zuletzt an der entsprechenden Normierung deutlich wird. Die acht Dimensionen decken ein breites Spektrum der interessierenden kognitiven Leistungsfähigkeiten ab, die bei der Beurteilung beruflicher Eignung relevant sein können. Durch die Modularisierung können einzelne Dimensionen separat erhoben werden, so dass die Anwendung für gezielte Fragestellungen entsprechend den Dimensionen Schlussfolgerndes, Sprachliches, Rechnerisches sowie Räumliches Denken, Merkfähigkeit, Arbeitseffizienz, Wissen Wirtschaft und Wissen Informationstechnologie ebenso möglich ist.

6. Kurzfassung

Diagnostische Zielsetzung:

Mit dem WIT-2 kann die kognitive Leistungsfähigkeit in den Dimensionen Schlussfolgerndes, Sprachliches, Rechnerisches sowie Räumliches Denken, Merkfähigkeit, Arbeitseffizienz, Wissen Wirtschaft und Wissen Informationstechnologie insbesondere im Kontext der beruflichen Eignungsdiagnostik erhoben werden.

Aufbau:

Die acht genannten Dimensionen werden durch die 11 Subtests Analogien (AL), Abwicklungen (AW), E-Mails Bearbeiten (EM), Eingekleidete Rechenaufgaben (ER), Grundrechnen (GR), Gleiche Wortbedeutungen (GW), Merkfähigkeit (MF), Spiegelbilder (SP), Wissen Informationstechnologie (WI), Wissen Wirtschaft (WW) und Zahlenreihen (ZN) erhoben, die aus insgesamt 243 Items bestehen.

Grundlagen und Konstruktion:

Der WIT-2 ist die Überarbeitung des Wilde-Intelligenz-Tests (Jäger, 1963; Jäger & Althoff, 1983, 1994). Basierend auf einer Erweiterung der Primary Mental Abilities (Thurstone, 1938), dem sogenannten modifizierten Modell der Primary Mental Abilities (MMPMA), und den aktuellen Erkenntnissen der Intelligenzforschung wurden acht Subtests entfernt und drei Subtests neu konstruiert. Für die verbliebenen acht Subtests wurden aus den beiden Parallelversionen die psychometrisch besten Items ausgewählt und teilweise überarbeitet. Modelltests mittels LISREL bestätigen die Struktur der neuen Zusammenstellung und die angenommenen Faktoren höherer Ordnung (fluide und kristallisierte Intelligenz).

Empirische Prüfung und Gütekriterien:

Reliabilität: Die Werte der internen Konsistenz der Subtests liegen zwischen .77 und .95, die der Dimensionen zwischen .87 und .94 sowie der Gesamtscore bei .98. Die Retestreliabilität liegt zwischen .59 und .93 bei einem Delay von 9 Monaten (n = 63) bzw. 6 Wochen (n = 30); für den Gesamtscore wurde ein Wert von .88 (9 Monate) ermittelt.

Validität: Es liegt eine Vielzahl von empirischen Untersuchungen zur Validität der WIT-2 vor, die dem Verfahren zum einen Kriteriumsvalidität bezüglich Schulnoten und Ausbildungserfolg bescheinigen und zum anderen konvergente und divergente Validität zu vielen etablierten Verfahren (IST 2000 R, CFT-3, FRT, LPS uvm.) belegen.

Normen: Es liegen bildungsspezifische und altersspezifische Normen vor (Standardwerte), die je nach Subtest an zwischen 2234 und 10024 Testteilnehmern erhoben wurden. Ziel der Normierung war die "Repräsentativität für eignungsdiagnostisch relevante Gruppen (Bewerber, Rehabilitanden)" (Kersting et al., 2008, S. 43).

7. Bewertung

Der WIT-2 kann als gelungene Revision des Wilde-Intelligenz-Tests gesehen werden. Das Verfahren wurde durch Kürzungen ökonomischer, ist durch die Modularisierung flexibel anwendbar und im Hinblick auf das vorgestellte MMPMA theoretisch verortbar. Aufgrund der Nähe der Autoren zur DIN 33430 war zu erwarten, dass die aktuell formulierten Standards der Qualitätssicherung in der psychologischen Diagnostik eingehalten werden. Der Sicherstellung der Objektivität wird große Aufmerksamkeit geschenkt. Ein Testleiter kann sie eigentlich nur mehr grob fahrlässig verletzen. Die Reliabilitäten der Dimensionen entsprechen den Anforderungen an einen Intelligenztest; lediglich die interne Konsistenz der Dimension Merkfähigkeit ist mit .78 etwas niedrig. Die Retestreliabilitätskennwerte der Dimensionen Merkfähigkeit (.67, 6 Wochen) und Arbeitseffizienz (.59, 9 Monate) beschädigen das Bild eines psychometrisch annähernd perfekten Tests ein wenig. Ein zweiter Wehrmutstropfen ist die fehlende Überprüfung der Äquivalenz zur bestehenden PC-Version im Hogrefe Testsystem. Zwar bestehen bereits gesonderte Normen für die PC-Version, diese können aber aufgrund ihrer Fallzahlen noch nicht denselben Anspruch auf Repräsentativität erheben, wie die Papier-Bleistift-Version des WIT-2. Die Validität des Verfahrens wird bereits mit vorliegendem Manual durch eine Vielzahl von Studien für einige Bereiche der Eignungsdiagnostik und ebenso für die Anwendung im Kontext der Intelligenzdiagnostik belegt. Hervorzuheben ist noch das Ziel, Neuerungen und Erkenntnisse zum WIT-2 über die Website www.kersting-internet.de/wit allgemein zeitnah zugänglich zu machen. Insgesamt kann der Test aber ohne Vorbehalte für verschiedenste Anwendungen empfohlen werden.

8. Literatur

Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R. Göttingen: Hogrefe.

Bäumler, G. (1974). Lern- und Gedächtnistest LGT-3. Göttingen: Verlag für Psychologie Dr. C.J. Hogrefe.

Cattell, R.B. (1971). Abilities: Their structure, growth, and action. Boston: Houghton Mifflin.

Cattell, R.B. & Weiß, R.H. (1971). Grundintelligenztest Skala 3 (CFT-3). Göttingen: Hogrefe.

Daniels, J.C. & Booth, J. (2004). FRT/FRT-J. Figure Reasoning Test. Neubearbeitung des "klassischen" Tests durch Neunormierung und die Erstellung der Parallelförm. Frankfurt: Harcourt.

Feldt, L.S., & Brennan, R.L. (1989). Reliability. In R.L. Linn (Ed.), Educational Measurement (3rd ed. pp. 105-146). New York: Macmillan.

Horn, W. (1983). Leistungsprüfsystem. L-P-S (2., erweiterte Auflage). Göttingen: Hogrefe.

- Jäger, A.O. (1963). Der Wilde-Test. Ein neues Intelligenzdiagnostikum. Zeitschrift für experimentelle und angewandte Psychologie , 10, 260–278.
- Jäger, A.O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. Diagnostica, 28 (3), 195-225.
- Jäger, R.S. (1997). WILDE-Intelligenz-Test (WIT). Zeitschrift für Differentielle und Diagnostische Psychologie, 18 (1/2), 62-65.
- Jäger, A.O. & Althoff, K. (1983). Der Wilde-Intelligenz-Test (WIT). Ein Strukturdiagnostikum. Herausgegeben von der Deutsche Gesellschaft für Personalwesen e.V. Göttingen: Hogrefe.
- Jäger, A.O. & Althoff, K. (1994). Der Wilde-Intelligenz-Test (WIT). Ein Strukturdiagnostikum. Herausgegeben von der Deutschen Gesellschaft für Personalwesen e.V. (2., revidierte Auflage). Göttingen: Hogrefe.
- Jäger, R. & Jundt, E. (1981). Mannheimer Rechtschreib-Test (M-R-T). Göttingen: Hogrefe.
- Kersting, M. (in Vorb.). Testbatterie für Berufseinsteiger Rechtschreibung (START-R). Göttingen: Hogrefe.
- Kersting, M., Althoff, K. & Jäger, A.O. (2008). WIT-2. Wilde-Intelligenz-Test 2. Göttingen: Hogrefe.
- Kubinger, K.D. (2001). Zur Zumutbarkeit in der Psychologischen Diagnostik. In J.F. Beckmann & P.Y. Herzberg (Hrsg.), Dynamik im Testen (S. 95-107). Landau: Verlag Empirische Pädagogik.
- Ostendorf, F. & Angleitner, A. (2004). NEO-PI-R. NEO Persönlichkeitsinventar nach Costa und McCrae. Revidierte Fassung. Göttingen: Hogrefe.
- Thurstone, L.L. (1938). Primary Mental Abilities. Chicago: University of Chicago Press.
- Wonderlic, E.F. (1992). Wonderlic Personnel Test. Libertyville, IL: E.F. Wonderlic & Associates, Inc.