

WISC-IV. Wechsler Intelligence Scale for Children – Fourth edition, 2. Auflage

STEFAN SCHMUKLE, Universität Leipzig

RALF SCHULZE, Methodenlehre und psychologische Diagnostik, Bergische Universität Wuppertal

Allgemeine Informationen

Die WISC-IV ist eine Übersetzung und Adaption des in englischer Sprache seit 2003 verfügbaren gleichnamigen Tests und war von 2007 bis 2011 unter der Bezeichnung HAWIK-IV erhältlich. Die Zielsetzung besteht in der Untersuchung der kognitiven Fähigkeiten von Kindern und Jugendlichen im Alter von 6;0 bis 16;11, inklusive der Beurteilung von individueller Hochbegabung, Intelligenzminderung sowie kognitiven Stärken und Schwächen. Er besteht aus zehn Kerntests (Dauer 65 bis 95 Minuten) und fünf optionalen Untertests (Dauer 15 bis 20 Minuten), wobei bei Einsatz der Kerntests die Bestimmung eines Gesamt-IQs sowie von vier Indexwerten (Sprachverständnis, wahrnehmungsgebundenes logisches Denken, Arbeitsgedächtnis und Verarbeitungsgeschwindigkeit) möglich ist. Die WISC-IV ist ein Einzeltest, für den keine Parallelförmigkeiten vorliegen. Weiterführende Informationen zur Anwendung des Tests für verschiedene diagnostische Fragestellungen finden sich im Fallbuch HAWIK-IV (Petermann & Daseking, 2009).

Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion

Die Konstruktion der WISC erfolgte auf Basis des Intelligenzkonzepts von David Wechsler und nicht einschlägiger Intelligenzstrukturtheorien. Die Einführung der Indexwerte mit der WISC-III und die Veränderung der Subtests im Rahmen der Testrevision der WISC-IV sind Indizien einer Anlehnung an das etablierte Cattell-Horn-Carroll-Modell (CHC). Die Korrespondenz der WISC-IV-Indexwerte mit den im CHC beschriebenen Fähigkeitskonstrukten wird aber nicht konsequent hergestellt und ist nicht in allen Fällen eindeutig. Einzelne Untertests können zwar auf Basis der einschlägigen Literatur im CHC verortet werden (z.B. Matrizen bei Fluid Reasoning, Zahlen-Symbol-Test bei Processing Speed), andere jedoch nicht (z.B. allgemeines Verständnis). Dennoch sind die Indexwerte eine zu begrüßende Weiterentwicklung gegenüber älteren Versionen der WISC, in denen die auf eher intuitiven Vorstellungen beruhende Trennung in Verbal- und Handlungs-IQ implementiert war, die empirisch nie gestützt werden konnte.

Objektivität

Die Materialien enthalten sehr ausführliche und gut nachvollziehbare schriftliche und videobasierte Beschreibungen, Vorgaben und Beispiele zur Durchführung und Auswertung des Tests. Die objektive Bewertung der offenen Antworten wird durch detaillierte Anweisungen für die Punktvorgabe im Handbuch erleichtert. An den Testleiter werden im Rahmen der intensiven Interaktion mit der Testperson extrem hohe Anforderungen durch die Vorgaben zur Durchführung und Auswertung gestellt. Einbußen der Durchführungsobjektivität sind hier plausibel (Lipsius, Petermann & Daseking, 2008) und empirisch klar belegt (McDermott, Watkins & Rhoad, 2014). Problematisch ist zudem, dass es keine eindeutigen Regeln dafür gibt, wann ein Kerntest für ungültig erklärt und durch einen der optionalen Untertests ersetzt werden darf. Die dadurch entstehenden Freiheiten bei der Untertestausswahl können die Auswertungsobjektivität des Tests einschränken. Durch die Normierung ist die Interpretationsobjektivität gegeben.

Normierung

Zur Normierung der WISC-IV in den Jahren 2005 und 2006 wurden N = 2600 Personen untersucht, von denen für Deutschland (N = 1440), Österreich (N = 84) und die Schweiz (N = 126) eine nach verschiedenen demografischen Merkmalen geschichtete Teilmenge verwendet wurde. Die Repräsentativität ist damit nicht gesichert, eine Verzerrung hinsichtlich weniger wichtiger Merkmale (Geschlecht, Schulabschluss) aber auch nicht plausibel. Eine ausführlichere und genauere Darstellung der Stichprobenauswahl und der Überprüfung ihrer Repräsentativität wäre aber wünschenswert gewesen. Normtabellen für die Bestimmung der Wertpunkte aus Rohwerten werden für die Stichprobe im Alter von 6;0 bis 16;11 in Viermonatsintervallen angeboten, wobei die so gebildeten Gruppen mit jeweils lediglich N = 50 Kindern und Jugendlichen besetzt sind, wodurch es zu Ungenauigkeiten bei der Schätzung der Normwerte insbesondere bei extremen Ausprägungen kommen kann.

Zuverlässigkeit

Die Reliabilitäten für Gesamt-IQ, Indexwerte sowie die Untertests werden unter Zugrundelegung der klassi-

Diese Testrezension wurde im Auftrag des Diagnostik- und Testkuratoriums der Föderation deutscher Psychologenvereinigungen (DGPs und BDP) gemäß den TBS-TK-Richtlinien (Testkuratorium, 2009, 2010) erstellt. Die beiden Autoren fungierten als gleichberechtigte Rezensenten; die Autorenreihenfolge wurde ausgelost.

WISC-IV. 2. Auflage	Die TBS-TK-Anforderungen sind erfüllt			
	voll	weit- gehend	teil- weise	nicht
Allgemeine Informationen, Beschreibung und diagnostische Zielsetzung	●			
Objektivität		●		
Zuverlässigkeit	●			
Validität			●	

schen Testtheorie geschätzt mit a) der Testhalbierungsmethode auf Basis der Normstichprobe sowie b) der Retest-Methode (N = 103), die insbesondere für die Speed-Tests von Bedeutung ist. Es gibt keine für die diagnostischen Zielsetzungen relevanten untergruppenspezifischen Reliabilitätsangaben. Die Reliabilitätsschätzungen liegen für die Indexwerte über alle Altersgruppen durchschnittlich zwischen .87 (Verarbeitungsgeschwindigkeit) und .94 (Sprachverständnis) und sind ebenso wie die Reliabilität des Gesamt-IQ mit durchschnittlich .97 als sehr gut zu bezeichnen. Für die Untertests werden niedrigere Werte zwischen .76 und .91 geschätzt, während für die Diskrepanzvergleiche im Rahmen der Prozessanalyse keine Reliabilitäten berichtet werden. Dies ist aber unproblematisch, solange Ergebnisse in einzelnen Untertests und den Prozessanalysen nicht auf individueller Ebene interpretiert werden.

Gültigkeit

Die faktorielle Validität wurde mittels exploratorischer Faktorenanalysen untersucht. Diese zeigen zwar plausible Ergebnisse, sind aber hinsichtlich der Strukturhypothese von vier (Index-)Faktoren nicht klar interpretierbar, da basale Merkmale der Faktorenanalysen nicht berichtet werden. Die Stabilität einer Vier-Faktoren-Struktur für verschiedene Altersgruppen wird lediglich impressionistisch betrachtet. Die für die Validierung der Faktorenstruktur besser geeignete konfirmatorische Faktorenanalyse wurde nur für die amerikanische Stichprobe durchgeführt. Die konvergente Validität wird lediglich mit dem HAWIK-III und dem WPPSI-III untersucht, was aufgrund der stark überlappenden Aufgabentypen schwer als Validitätsevidenz zu interpretieren ist. Konvergente Validitäten mit anderen Intelligenztests sowie diskriminante Validitäten werden nicht berichtet. Als kriteriumsbezogene Validität werden Testergebnisse für verschiedene klinische Gruppen sowie für hochbegabte Kinder berichtet.

Weitere Gütekriterien

Alle Untersuchungen an Kindern und damit auch die Ergebnisse sind prinzipiell störanfällig. Eine besondere Rolle bei allen Intelligenztests spielt dabei die Aufrechterhaltung der Motivation und der Konzentration des Kindes während der Untersuchung, damit aus der im Test gezeigten Leistung auch auf die zugrunde liegende Intelligenz geschlossen werden kann. Neben mangeln-

der Motivation und Konzentration kann auch bewusstes Faking Bad zu Minderleistungen führen. Faking Good ist hingegen nur nach vorheriger Kenntnis des Testmaterials möglich. Untersuchungen zur Skalierung der einzelnen Subtests werden im Manual nicht vorgestellt.

Abschlussbewertungen/Empfehlungen

In der Praxis gehören die WISC-IV und ihre Vorläufer zu den am häufigsten eingesetzten Verfahren zur Erfassung der kognitiven Fähigkeiten bei Kindern und Jugendlichen, was die Praktikabilität und Akzeptanz unterstreicht. Für die Einzelfalldiagnose im klinischen und pädagogischen Kontext ist die WISC-IV auch gewiss ein nützlicher Test. Zu den Stärken des Tests gehören unter anderem die äußerst anwendungsfreundlichen Durchführung- und Auswertungsanweisungen, die in Anbetracht der komplexen Durchführung auch als notwendig erachtet werden, sowie die hohen Reliabilitäten, die präzise Schätzungen der Fähigkeitswerte und -profile erlauben. Allerdings erfolgt die Intelligenzbestimmung in der WISC-IV vergleichsweise bildungsabhängig, wofür vor allem die zum Sprachverständnis gehörenden Untertests verantwortlich sind.

Zu den markanten Schwächen der WISC-IV gehört die eingeschränkte Validitätsevidenz. Die präsentierte faktorenanalytische Evidenz begründet keineswegs, dass »die vier Indexwerte und die dazugehörige psychologische Interpretation als empirisch gesichert gelten« (S. 75) können, und auch nicht, dass die an das CHC-Modell angelehnte Strukturhypothese für die WISC-IV korrekt ist. Zudem ist problematisch, dass auf den Auswertungsblättern eine Reihe von Analysen angeboten wird, die nicht auf aggregierten Werten, sondern auf den weniger reliablen Untertests beruhen. Dabei besteht die Gefahr, dass Anwender alle auf den Auswertungsblättern angebotenen Analysen unkritisch berechnen und interpretieren. Besonders problematisch sind dabei die auf Differenzwerten beruhenden Prozessanalysen, da die Evidenz unzureichend ist, um beispielsweise zu begründen, dass auf Basis der Prozessanalyse hilfreiche Information zur Beurteilung des spezifischen Informationsverarbeitungsstils von Kindern und Jugendlichen gewonnen werden kann (S. 41). Insgesamt sind somit Defizite bei der evidenzbasierten theoretischen Interpretation der Testwerte festzustellen.

TESTINFORMATIONEN

Weiß, R. H. (2006). CFT 20-R mit WS/ZF-R. Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – Revision (WS/ZF-R). Göttingen: Hogrefe.

Bezugsquelle

Testzentrale Göttingen, Robert-Bosch-Breite 25, 37079 Göttingen. Test komplett inklusive Auswertungs-CD 1487 Euro. 25 Protokollbogen 124,80 Euro, 25 Aufgabenhefte 1 (A + B) – Symbolsuche 166 Euro, 25 Aufgabenhefte 2 (Durchstreichtest) 101,50 Euro, 25 Aufgabenhefte 1A (6;0–7;11) oder 2B (8;0–16;11) 95,30 Euro. Bitte zitieren Sie diesen Artikel wie folgt:

Schmukle, S. & Schulze, R. (2016). TBS-TK Rezension: »WISC-IV. Wechsler Intelligence Scale for Children – Fourth Edition« 2016, report psychologie 2, 67–68.