

# NEUERE TESTVERFAHREN

---

Petermann, F. (unter Mitarbeit von M. Lipsius) (Hrsg.) (2009). **WPPSI-III. Wechsler Preschool and Primary Scale of Intelligence-III**. Deutsche Version. Frankfurt: Pearson; Testkoffer komplett 850,- €, Verbrauchsmaterial pro Testung abhängig vom Alter des Kindes 1,80 € bzw. 4,- €, Ergänzungsset zum HAWIVA-III 175,- €.

---

Zwei Jahre nach Erscheinen des HAWIVA-III (Ricken et al., 2007) ist jetzt unter dem Titel WPPSI-III erneut eine deutsche Bearbeitung und Normierung der Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-III; Wechsler, 2002) veröffentlicht worden, was unter klinischen Kinderpsychologen teilweise Verunsicherung ausgelöst hat. Dabei scheinen es in erster Linie lizenztechnische Gründe zu sein, die zu einer Ablösung des HAWIVA-III durch den WPPSI-III geführt haben (F. Petermann, persönliche Mitteilung).

## Testaufbau

Die WPPSI-III basiert auf der Intelligenzkonzeption von David Wechsler und dient der Einzeluntersuchung von Kindern im Alter von 3;0 bis 7;2 Jahren. Der Testaufbau enthält, wie andere Wechsler-Tests auch, einen *Verbalteil*, mit dem sprachliches Denken erfasst wird, und einen *Handlungsteil*, in dem nichtsprachliche Anforderungen gestellt werden. Diese beiden Subskalen bilden im Wesentlichen den Gesamtwert, der als allgemeines Maß intellektueller Leistungsfähigkeit verstanden wird. Außerdem erlaubt es die WPPSI-III für alle Kinder eine *Allgemeine Sprachskala* anzuwenden, die aktiven und passiven Wortschatz erfasst. Bei Kindern ab einem Alter von 4;0 Jahren kann zusätzlich ein Index für *Verarbeitungsgeschwindigkeit* gebildet werden. Für dreijährige Kinder setzt sich die Gesamtskala aus fünf Untertests zusammen, ab 4 Jahre stehen 14 Untertests zur Verfügung, von denen 7 als Kernuntertests gelten und 7 weitere optional zusätzlich oder alternativ zu Kernuntertests eingesetzt werden können (s. Tab. 1). Die Testdurchführung dauert lt. Manual zwischen 20 und 120 Minuten. Nach eigenen Erfahrungen sind für die Durchführung der Kernuntertests maximal 40 bzw. 60 Minuten zu veranschlagen. Die Rohwerte werden anhand von Normtabellen (Altersumfang jeweils 3 Monate) in Wertpunkte ( $M = 10$ ,  $SD = 3$ ) umgewandelt, aus denen dann für Subskalen und Gesamtskala Standardwerte ( $M = 100$ ,  $SD = 15$ ) gebildet werden. Ein computergestütztes Auswertungsprogramm liegt bisher nicht vor.

Tabelle 1: Aufbau des WPPSI-III und Testreliabilitäten

| Altersbereich 3;0 bis 3;11                | Reliabilität WPPSI-III             |   | Reliabilität HAWIVA-III            |   |
|---|------------------------------------|---|------------------------------------|---|
|   | Reliabilität <sup>a,b</sup>        |   | Mittlere Reliabilität <sup>a</sup> | Reliabilitäten der Altersgruppen <sup>a,c</sup> |
| Gesamt-IQ                                 | .95                                |   | .94                                | .94   |
| <i>Verbalteil</i>                         | .90                                |   | .91                                | .90-.91   |
| Passiver Wortschatz (PW) <sup>f</sup>     | .84                                |   | .85                                | .82-.88   |
| Allgemeines Wissen (AW)                   | .84                                |   | .83                                | .82-.84   |
| <i>Handlungsteil</i>                      | .93                                |   | .91                                | .90-.92   |
| Mosaik-Test (MT)                          | .89                                |   | .89                                | .89   |
| Figuren legen (FL)                        | (.84 <sup>d,e</sup> )              |   | .85 <sup>d</sup>                   | .84-.86 <sup>d</sup>                            |
| <i>Allgemeine Sprachskala (AS)</i>        | .90                                |   | .91                                | .89-.93   |
| Aktiver Wortschatz (AK) <sup>e</sup>      | .85                                |   | .85                                | .85   |
| Altersbereich 4;0 bis 6;11                | Mittlere Reliabilität <sup>a</sup> | Reliabilitäten der Altersgruppen <sup>a,c</sup> | Mittlere Reliabilität <sup>a</sup> | Reliabilitäten der Altersgruppen <sup>a,c</sup> |
| Gesamt-IQ                                 | .95                                | .95   | .95                                | .94-.95   |
| <i>Verbalteil</i>                         | .93                                | .92-.94   | .93                                | .91-.94   |
| Allgemeines Wissen (AW)                   | .84                                | .81-.86   | .81                                | .77-.86   |
| Wortschatz-Test (WT)                      | .86                                | .84-.88   | .86                                | .84-.89   |
| Begriffe erkennen (BEN)                   | .82                                | .74-.88   | .86                                | .79-.89   |
| Allgemeines Verständnis (AV) <sup>f</sup> | .87                                | .81-.91   | .88                                | .87-.91   |
| Gemeinsamkeiten finden (GF) <sup>f</sup>  | .84                                | .75-.90   | .90                                | .87-.91   |
| <i>Handlungsteil</i>                      | .90                                | .89-.92   | .90                                | .87-.92   |
| Mosaik-Test (MT)                          | .87                                | .84-.90   | .86                                | .79-.91   |
| Matrizen-Test (MZ)                        | .81                                | .74-.88   | .80                                | .71-.86   |
| Bildkonzepte (BK)                         | .80                                | .76-.82   | .82                                | .78-.84   |
| Bilder ergänzen (BE) <sup>f</sup>         | .86                                | .74-.92   | .86                                | .69-.93   |
| Figuren legen (FL) <sup>f</sup>           | (.88 <sup>d,e</sup> )              | (.78-.88 <sup>d,e</sup> )                       | .73 <sup>d</sup>                   | .53-.85 <sup>d</sup>                            |
| <i>Verarbeitungsgeschwindigkeit</i>       | (.89 <sup>d,e</sup> )              | (.87-.91 <sup>d,e</sup> )                       | .84 <sup>d</sup>                   | .76-.90 <sup>d</sup>                            |
| Symbole Kodieren (SK)                     | (.84 <sup>d,e</sup> )              | (.79-.88 <sup>d,e</sup> )                       | .79 <sup>d</sup>                   | .76-.81 <sup>d</sup>                            |
| Symbol-Suche (SYS) <sup>f</sup>           | (.83 <sup>d,e</sup> )              | (.80-.86 <sup>d,e</sup> )                       | .75 <sup>d</sup>                   | .56-.87 <sup>d</sup>                            |
| <i>Allgemeine Sprachskala</i>             | .85                                | .80-.89   | .88                                | .82-.89   |
| Passiver Wortschatz (PW) <sup>f,g</sup>   | .74                                | .62-.82   | .79                                | .68-.85   |
| Aktiver Wortschatz (AK) <sup>f</sup>      | .78                                | .73-.84   | .81                                | .75-.85   |

Anmerkungen: <sup>a)</sup> Wenn nicht anders angegeben: Testhalbierungsreliabilität. Teilweise weichen die Angaben wegen anderer Zusammenstellungen der Altersgruppen von den Testmanualen ab  
<sup>b)</sup> Die Dreijährigen wurden bei der Reliabilitätsberechnung des WPPSI-III zu einer Altersgruppe zusammengefasst <sup>c)</sup> Minimale und maximale Reliabilitätsschätzung in den Altersgruppen <sup>d)</sup> Retestreliabilität <sup>e)</sup> eingeklammerte Werte beziehen sich auf die amerikanische Normstichprobe des WPPSI-III <sup>f)</sup> Optionaler Untertest <sup>g)</sup> Bildet zusammen mit AW die Allgemeine Sprachskala.

## Testanalyse und Normierung

Da HAWIVA-III und WPPSI-III sich in *Aufbau und Durchführung* nicht nennenswert unterscheiden, kann bzgl. einführender Darstellung u. a. auf die ausführliche Rezension zum HAWIVA-III in dieser Zeitschrift (Irblich, 2009) verwiesen werden. Diese Besprechung geht daher vor allem auf die Neuerungen der WPPSI-III, insbesondere hinsichtlich Normierung und Testgütekriterien, ein.

Durch die Umbenennung des Verfahrens besteht nunmehr eine engere begriffliche Anlehnung an das amerikanische Original und bei der Benennung der Untertests orientierten die Übersetzer sich, wo immer möglich, an der deutschen Version des HAWIK-IV (Petermann u. Petermann, 2007). So wurde „Begriffe erklären“ in „Wortschatz-Test“ umbenannt, aus „Klassen bilden“ wurde „Bildkonzepte“ und aus „Kodieren“ wurde „Symbole kodieren“, was auf allgemeinen Beifall bei Testbenutzern treffen dürfte.

Der Test wird jetzt mit im Prinzip identischem Testmaterial in einem schwarzen Lederkoffer geliefert („Pilotenkoffer“), der allerdings ca. doppelt soviel Platz beansprucht wie der graue Plastikkoffer des HAWIVA-III. Das Manual besteht nunmehr aus nur noch einem Band (234 Seiten, DIN A4, Hardcover). Die textliche Verschlangung geht mit einer Straffung der Durchführungshinweise und weitgehendem Weglassen von Angaben zur Testentwicklung einher. Die Formatverdoppelung des Manuals wird vermutlich wegen des größeren Platzbedarfs bei der Testdurchführung keine ungeteilte Zustimmung finden und die Buchbindung erweist sich für den Alltagsgebrauch als nicht stabil genug. Bei den Instruktionen und Durchführungshinweisen gibt es geringfügige Unterschiede gegenüber dem HAWIVA-III, die allerdings im Text nicht kenntlich gemacht werden, sodass „Umsteiger“ nicht umhin kommen, sich eingehend mit den neuen Formulierungen zu befassen. Einige nicht kindgemäße Begriffe in den Instruktionen wurden erfreulicherweise ausgetauscht („gerade Linie“ statt „Querlinie“ und „Zeichen“ statt „Symbol“), andere Veränderungen scheinen inhaltlich nicht begründet zu sein, erfordern aber nichtsdestotrotz eine Umstellung des Testanwenders. Auch bei den Bewertungsvorgaben für verbale Reaktionen sind kleine Unterschiede gegenüber dem HAWIVA-III festzustellen, die vermutlich nicht sonderlich ins Gewicht fallen, aber zur Vorsicht Anlass geben, wenn mit dem HAWIVA-III gewonnene Ergebnisse jetzt nachträglich in WPPSI-III-Normen umgerechnet werden sollen. Dem Manual liegt ein Blatt mit Fehlerkorrekturen bei, die den Matrizen-Test betreffen. Bei Item 14 sind im Stimulusheft die Lösungsalternativen 1 und 5 identisch. Im HAWIVA-III war diesbezüglich ein Aufkleber mit verändertem Stimulusmaterial enthalten. WPPSI-III behält den ursprünglichen Stimulus bei und fordert stattdessen, 2 Lösungsalternativen als richtig gelten zu lassen, was so aber im Protokollbogen nicht markiert ist. Außerdem wurden in der Normtabelle 4;0-4;2 im Manual die Werte für den Matrizen-Test vergessen, die jetzt auf einem Extrablatt beigelegt sind. Es ist zu hoffen, dass Fehler, soweit wie möglich, bei Neuauflagen ausgemerzt werden. Auch die Protokollbögen wurden geringfügig verändert. Für HAWIVA-III-Benutzer wird ein sog. Ergänzungssset angeboten, das Manual und Verbrauchsmaterial für je 25 Testdurchführungen für Dreijährige und ältere Kinder enthält (Kostensparnis gegenüber der Einzelbestel-

lung 69.- €). Dadurch kann beim Umstieg auf die WPPSI-III das Testmaterial weiterbenutzt werden. Ungünstig ist in diesem Zusammenhang aber, wenn bei der fehlerhaften Aufgabe 14 des Matrizen-Tests der Änderungsaufkleber bereits angebracht wurde.

Wesentlich ist der Hinweis, dass der *Altersbereich* der WPPSI-III gegenüber dem HAWIVA-III verschoben wurde. Das Mindestalter beträgt nunmehr 3;0 Jahre statt bisher 2;6 Jahre und die Testnormen reichen dafür bis 7;2 Jahre statt 6;11 Jahre. In Anbetracht der Bodeneffekte des HAWIVA-III bei den Zweieinhalbjährigen ist die Anhebung der Altersmindestgrenze sinnvoll. Die Ausdehnung der Normierung auf das erste Quartal der Siebenjährigen erweitert, wenn auch nur geringfügig, die Testanwendung bei entwicklungsverzögerten Kindern, die mit dem HAWIK-IV überfordert sind. Die Normgruppen wurden jetzt einheitlich auf 3 Altersmonate festgelegt. Beim HAWIVA-III waren bei den Sechsjährigen 3 Normgruppen à vier Monate gebildet worden.

Es wurde eine vollständige Neunormierung des Verfahrens im Winter und Frühjahr 2009 ausschließlich in Deutschland mit 710 Kindern durchgeführt. Pro Normgruppe nahmen daran 42 Kinder, je zur Hälfte Mädchen und Jungen teil. Damit ist die Normstichprobe nur etwa halb so groß wie beim HAWIVA-III und deckt das deutschsprachige Ausland nicht mit ab. Weitere Angaben zur Testnormierung sind dem Manual nicht zu entnehmen, weswegen eine abschließende Beurteilung der Testnormen für den Benutzer nicht möglich ist. Vom Herausgeber war ergänzend zu erfahren, dass die Stichprobe sich gleichmäßig über das gesamte Bundesgebiet verteilte und dass auf die Bildungsschichtung geachtet wurde (F. Petermann, persönliche Mitteilung). Unter den Eltern der getesteten Kinder scheinen aber, soweit man die Statistiken vergleichen kann, Akademiker deutlich überrepräsentiert zu sein (s. Tab. 2).

**Tabelle 2:** Bildungsstand der Väter in der Normstichprobe (persönliche Mitteilung des Herausgebers) und in der Bundesbevölkerung im Alter 20-50 Jahre (Quelle: Statistisches Bundesamt, 2009)

| Bildungsstand des Vaters         | Prozent in der Normstichprobe |        | Prozent i. d. Bevölkerungsstatistik |                    |
|----------------------------------|-------------------------------|--------|-------------------------------------|--------------------|
|                                  | Väter                         | Mütter | Väter                               | Mütter             |
| keine Angabe                     | 20,8                          | 12,5   | 0,3                                 | 0,3                |
| ohne Berufsabschluss             | 4,2                           | 5,9    | 20,1                                | 22,4               |
| mit Berufsabschluss <sup>a</sup> | 50,6                          | 57,2   | 62,8                                | 62,8               |
| mit Abitur, FH o. Uni-Abschluss  | 24,4                          | 24,4   | 16,7                                | 14,6               |
| Gesamt                           | 100,0                         | 100,0  | 100,0                               | 100,1 <sup>b</sup> |

Anmerkung: <sup>a</sup>inclusive Abitur, <sup>b</sup>Rundungsfehler

Die Itemgradienten sind für viele Untertests über die Bandbreite von 1 bis 19 Wertpunkten auffallend gleichmäßig. Deren Zustandekommen kann mangels Angaben zu Itemschwierigkeiten, Verteilungsmerkmalen und Strategien bei der Normenbildung im Manual nicht nachvollzogen werden. Bei der Skalenkonstruktion scheint darauf geachtet worden zu sein, dass bei durchgängig durchschnittlichen Untertestergebnissen auch genau durchschnittliche Skalenergebnisse erzielt werden.

Bei den Dreijährigen besteht ein Bodeneffekt des Handlungsteils, der insbesondere durch den Untertest „Figurenlegen“ verursacht wird. Weitere Bodeneffekte ergeben sich

im Alter von 4;0-4;2 Jahren, wo der geringstmögliche Gesamt-IQ bei 61 liegt (Verbalteil 64, Handlungsteil 63, Verarbeitungsgeschwindigkeit 73) und in 5 der 7 Kernuntertests 4 oder mehr Wertpunkte bei einem Rohwert von 1 erreicht werden. Gesamtwerte von 55 und weniger sind im WPPSI-III erst ab einem Alter von 4;6 Jahren möglich. In diesem Alter zeichnet sich dafür bereits ein Deckeneffekt im Untertest „Begriffe erkennen“ ab, ansonsten sind die Skalen für die Fünf- und Sechsjährigen bis auf die „Allgemeine Sprachskala“ weitgehend frei von Boden- und Deckeneffekte. Erst bei den Siebenjährigen häufen sich Deckeneffekte, doch lassen sich auch bei diesen noch Gesamt-IQs bis 138 ermitteln. Der Normanstieg zwischen den Altersklassen ist angemessen. Er beträgt bei den 4;0-7;2-Jährigen für die Subskalen und den Gesamt-Test meist zwischen  $\frac{1}{4}$  und  $\frac{1}{2}$  Standardabweichung, bei den 3;0-3;11-Jährigen fallen die Werte erwartungsgemäß etwas höher aus.

Vergleicht man die Normierung der Untertests beider Testversionen in den Altersgruppen 3;0-3;2, 4;0-4;2, 5;0-5;2 und 6;0-6;2, so sind die Unterschiede zwischen WPPSI-III und HAWIVA-III meist auf 1 bis 2 Wertpunkte begrenzt. In der Altersgruppe 3;0-3;2 liegen die WPPSI-III Wertpunkte tendenziell etwas höher als im HAWIVA-III. In der Altersgruppe 4;0-4;2 sind die Bewertungsunterschiede zwischen beiden Verfahren in etwa gleichmäßig verteilt, während in den Altersgruppen 5;0-5;2 und 6;0-6;2 der Trend eher zur Vergabe höherer Wertpunkte in den Untertests des HAWIVA-III geht. Bei den Sechsjährigen reichen die Differenzen zugunsten des HAWIVA-III vereinzelt bis zu 4 Wertpunkten (Begriffe erkennen). Insgesamt sind die Unterschiede aber eher gering. Beispielhaft ist in Abbildung 1 der Wertpunktevergleich bzgl. des Mosaik-Tests für die Normgruppe 5;0-5;2 dargestellt.

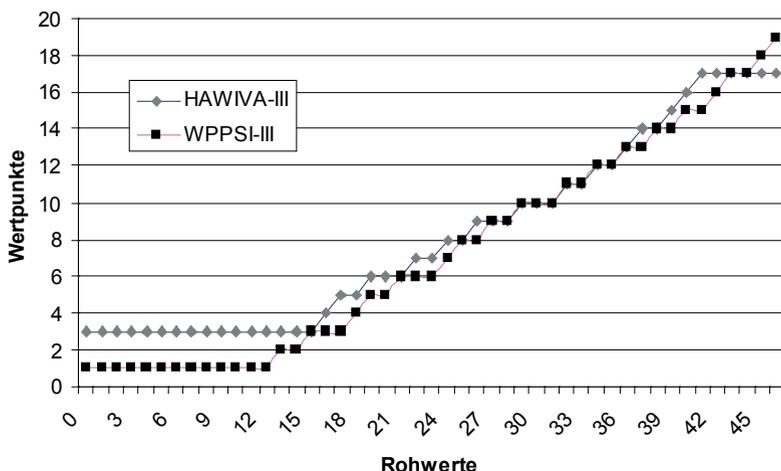


Abbildung 1: Vergleich der Bewertung des Mosaik-Tests in WPPSI-III und HAWIVA-III in der Normierungsgruppe Alter 5;0-5;2

Um die Messeigenschaften der Subskalen von WPPSI-III und HAWIVA-III miteinander vergleichen zu können, wurde die Verrechnung fiktiver Testprotokolle im unterdurchschnittlichen, durchschnittlichen und überdurchschnittlichen Bereich anhand der beiden Testnormierungen vorgenommen. Dabei wurden in ausgewählten Altersgruppen die Rohwerte so festgelegt, dass sie durchgängig bei allen Untertests bestimmten Wertpunkten im HAWIVA-III (WP 4, 7, 10 oder 14) entsprachen und anschließend wurden die korrespondierenden Untertest- und Skalenwerte für die WPPSI-III ermittelt.

Für die Altersgruppe der 3;0 bis 3;2-Jährigen sind die Differenzen der Skalenwerte in Abbildung 2 abzulesen. Dabei ergeben sich insbesondere bei unterdurchschnittlichen Testergebnissen erhebliche Auswertungsunterschiede. Im Handlungsteil werden Diskrepanzen von bis zu 8 Wertpunkten zugunsten der WPPSI-III gefunden. Im Verbalteil sind die Unterschiede auf maximal 3 Wertpunkte begrenzt, dafür treten im Gesamttest Differenzen bis zu 5 Wertpunkten auf, was immerhin  $\frac{1}{3}$  Standardabweichung entspricht. Bis auf den Wertebereich von -1 SD (7 Wertpunkte) im Handlungsteil werden in dieser Altersgruppe mit der WPPSI-III gleichgroße oder höhere Skalenwerte erzielt.

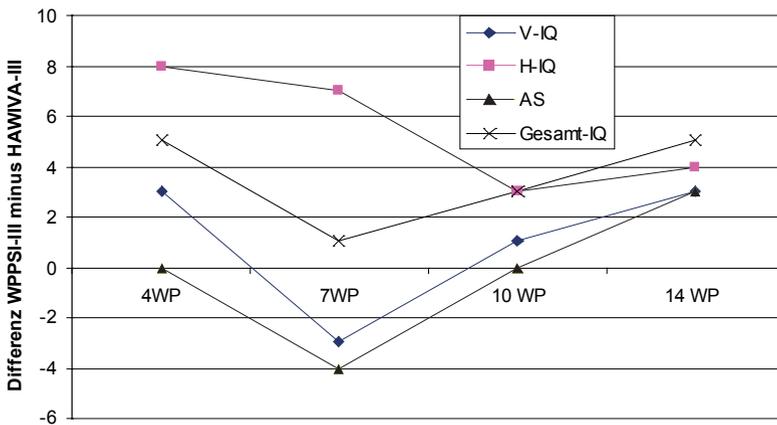


Abbildung 2: Skalendifferenzen zu ausgewählten Testprotokollen der Altersgruppe 3;0-3;2 V-IQ = Verbal-IQ; H-IQ = Handlungs-IQ, AS = Allgemeine Sprachskala. Differenzwerte über Null bedeuten bei gleichen Rohwerten einen höheren Skalenwert in der WPPSI-III als im HAWIVA-III (Weitere Erklärung s. Text)

Bei den Kindern der Altersgruppe 3;6-3;8 liegen dagegen bis auf geringe Ausnahmen die Skalenwerte des HAWIVA-III über denen der WPPSI-III. Im Verbalteil beträgt die Differenz für die ausgewählten Testergebnisse mindestens 3 Wertpunkte. Für den Gesamttest werden in dieser Altersgruppe maximal 3 IQ-Punkte mehr im HAWIVA-III erzielt (s. Abb. 3).

Ähnlich uneinheitlich fallen die Differenzen zwischen den Testnormierungen in den Altersgruppen der älteren Kinder aus. Die größten Abweichungen ergeben sich für die Subskalen Verarbeitungsgeschwindigkeit und Allgemeine Sprachskala, wo in den Ex-

trembereichen Differenzen von bis zu 19 IQ-Punkten zugunsten des HAWIVA-III bei identischen Rohwertkonstellationen erreicht werden. In den drei „Hauptskalen“ Verbalteil, Handlungsteil und Gesamtskala sind die Differenzen im mittleren Bereich gering (0 bis 3 IQ-Punkte) nehmen aber in den Extrembereichen deutlich zu. Der größte ermittelte Unterschied liegt bei 13 IQ-Punkten im Verbalteil bei den Vierjährigen im überdurchschnittlichen Bereich. Meist werden mit dem WPPSI-III die höheren Gesamt-IQs erzielt (s. Abb. 4). Im mittleren Intelligenzbereich sind die Differenzen gering, in den Extrembereichen ergeben die Modellrechnung Unterschiede von bis zu 13 Wertpunkten.

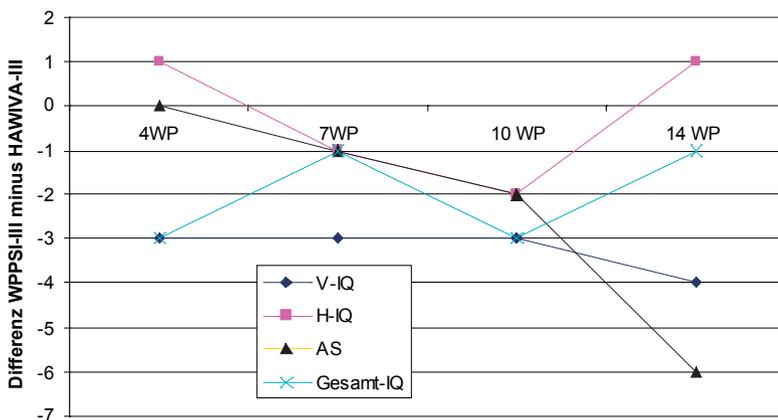


Abbildung 3: Skalendifferenzen zu ausgewählten Testprotokollen der Altersgruppe 3;6-3;8 V-IQ = Verbal-IQ; H-IQ = Handlungs-IQ; AS = Allgemeine Sprachskala (Weitere Erläuterungen s. im Text und bei Abb. 2)

Es findet sich über die Altersgruppen und das Intelligenzband hinweg kein durchgängiger Trend zugunsten der einen oder anderen Normierung. Teilweise sind die Unterschiede aber so beträchtlich, dass es für die Interpretation der Ergebnisse von Belang ist, welche Normen zur Anwendung kommen. Am deutlichsten macht sich dies im weit unterdurchschnittlichen und im überdurchschnittlichen Bereich bemerkbar. Solange die erforderlichen Angaben zur Normierung der WPPSI-III fehlen, ist Zurückhaltung bei der Interpretation von Testergebnissen beider Testversionen insbesondere in den Extrembereichen angebracht.

Die *Testreliabilität* der WPPSI-III wurde jahrgangsweise durch Testhalbierungsmethode unter Berücksichtigung der Spearman-Brown-Korrektur ermittelt und ergab für die einbezogenen Untertests befriedigende bis sehr gute Werte (s. Tab. 1). Die Unterschiede zu den Reliabilitätsangaben des HAWIVA-III sind gering, sodass die Reliabilität des Verfahrens als weitgehend gesichert angesehen werden kann. Keine eigenen Reliabilitäten wurden für die geschwindigkeitsabhängigen Untertests „Figurenlegen“, „Symbolsuche“ und „Symbole kodieren“ ermittelt, die eine Retestung erforderlich gemacht hätten. Hier

wird, ebenso wie für die Subskala *Verarbeitungsgeschwindigkeit*, auf die amerikanischen Gütekriterien verwiesen, die somit zumindest teilweise in den entsprechenden Reliabilitätskoeffizient des Handlungsteils der 3;0 bis 3;11-Jährigen mit eingehen. Die Reliabilität des Gesamt-Tests fällt mit .95 sehr gut aus. Auch die Koeffizienten für Verbal- und Handlungsteil liegen über .90. Die Reliabilität der Allgemeinen Sprachskala liegt bei den 4;0-7;2-Jährigen unter .90 und fällt insbesondere bei den Sechs- und Siebenjährigen nochmals deutlich ab. Sie stellt in diesem Altersbereich daher kein verlässliches Messinstrument dar. Das Manual enthält Angaben zu signifikanten Skalen- und Untertestdifferenzen auf dem 5%- und 15%-Niveau sowie zu statistischen Häufigkeiten von Wertdifferenzen.

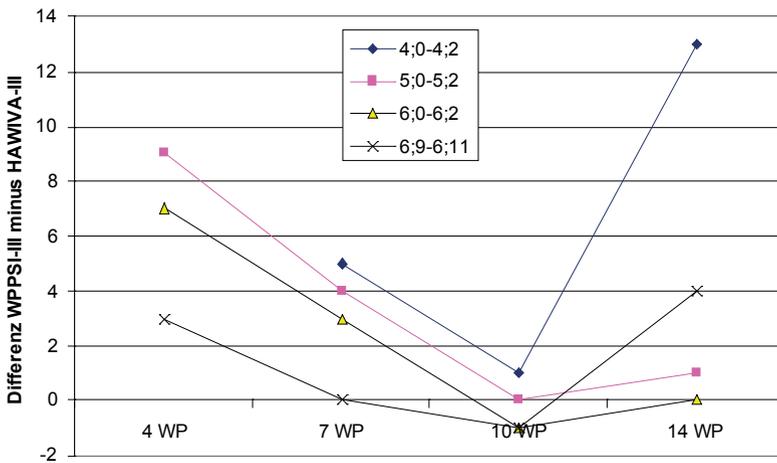


Abbildung 4: Skalendifferenzen des Gesamt-IQ bei ausgewählten Wertpunkten (Erläuterung s. Text) in verschiedenen Altersgruppen der älteren Kinder. Für die Altersgruppe 4;0-4;2 wurde wegen vorhandener Bodeneffekte keine Berechnung für 4 WP vorgenommen

Die *Inhaltliche Validität* ist, ebenso wie beim HAWIVA-III, gegeben. Die *Konstruktvalidität* wurde durch Interkorrelationen und Faktorenanalysen mit der deutschen Normierungsstichprobe untersucht. Bei den Dreijährigen bildet sich lt. Manual die Skalenstruktur des Tests in den Interkorrelationen ab. Die Untertests des Handlungsteils korrelieren untereinander höher als mit den verbalen Untertests und umgekehrt. Eine Unterscheidung zwischen Verbalteil und Allgemeiner Sprachskala aufgrund der Interkorrelationen ist dagegen nicht möglich. Die Interkorrelationen der Untertests bei den 4;0-7;2-Jährigen bestätigen ebenfalls die Homogenität des Verbalteils und in abgeschwächter Form auch die des Handlungsteils, allerdings korreliert der Untertest „Bildkonzepte“ durchgängig niedrig mit allen Untertests unabhängig von deren Subskalenzugehörigkeit. Dies gilt auch für „Figuren legen“. Eine exploratorische Faktorenanalyse, die mit der deutschen Normierungsstichprobe gerechnet wurde, ergab eine zweifaktorielle Lösung für die Dreijährigen, wobei die Untertests sich in Übereinstimmung mit der Testkonstrukt-

on einem verbalen und einem nichtverbalen Faktor zuordnen ließen. Bei den 4;0-7;2-Jährigen konnten Faktorenanalysen, die mit verschiedenen Untertestkombinationen gerechnet wurden, die Struktur des Verfahrens ebenfalls weitgehend bestätigen. Lediglich der Untertest „Bildkonzepte“ fiel, wie schon im amerikanischen Original, aus der Reihe. Bei der deutschen Version der WPPSI-III lud der Untertest vornehmlich auf dem Faktor, der der Verarbeitungsgeschwindigkeit zugeordnet wird. Beim amerikanischen Original tendieren die „Bildkonzepte“ faktoriell dagegen stärker zum Verbalteil (Lichtenberger u. Kaufman, 2004; Sattler u. Dumont, 2004). Somit bestätigt sich erneut die faktorielle Heterogenität des Handlungsteils.

Zur *Kriteriumsvalidität* liegen keine Ergebnisse vor, die mit der deutschen Version der WPPSI-III ermittelt worden wären. Stattdessen referiert das Manual einige amerikanische Studien, die für akzelerierte („hochbegabte“) Kinder und solche mit Intelligenzminderung signifikante Zusammenhänge mit WPPSI-III-Ergebnissen nachweisen konnten. Auch für andere Diagnosegruppen liegen erwartungskonforme Validitätsergebnisse vor. Das Fehlen empirischer Studien zur deutschen WPPSI-III-Version stellt ein erhebliches Manko dar. Da auch die Befundlage zur Validität des HAWIVA-III sehr spärlich war (s. Irblich, 2009), ist festzuhalten, dass überzeugende Belege für die Kriteriumsvalidität der deutschen Version der WPPSI-III bisher noch fehlen. Auch der Hinweis auf eine Vergleichsstudie, die für sechsjährige Kinder mit dem HAWIVA-III (sic!) und dem HAWIK-IV durchgeführt wurde und zu hohen Übereinstimmungswerten führte, gibt keine befriedigenden Hinweise auf das Messverhalten der WPPSI-III insbesondere im unteren Intelligenzspektrum. Es wäre zu begrüßen gewesen, wenn im Manual der WPPSI-III auch die Validitätsstudien referiert würden, die mit dem HAWIK-III durchgeführt wurden, da dessen Manual nicht mehr erhältlich ist.

## Bewertung

Die WPPSI-III ist objektiv durchführbar, weist aber dieselben Schwächen wie der HAWIVA-III in der Auswertungsobjektivität auf, die auch durch die Straffung der Auswertungsrichtlinien nicht beseitigt wurden. Die Qualität der Testnormierung ist kaum zu beurteilen, da die dazu erforderlichen Angaben im Manual fast vollständig fehlen. Da das Zustandekommen der HAWIVA-III-Normen bedeutend besser nachzuvollziehen ist, besteht derzeit keine Veranlassung, die Normen der WPPSI-III für vertrauenswürdiger zu halten und Testinterpretationen nachträglich zu revidieren. Die Normen weichen vor allen Dingen im unter- und überdurchschnittlichen Bereich z. T. erheblich voneinander ab, sodass Vorsicht bei der Interpretation walten sollte.

Die Reliabilitätskennwerte der WPPSI-III sind, soweit vorhanden, befriedigend bis sehr gut. Lediglich die *Allgemeine Sprachskala* weist bei den ältesten Kindern Schwächen auf. Es fehlen jedoch erforderliche Reteststudien und Untersuchungen zur Auswerterübereinstimmung. Die inhaltliche Validität des Verfahrens gilt als gesichert. Angaben zur Konstruktvalidität fallen weitgehend zufriedenstellend aus, lediglich der Untertest „Bild-

konzepte“ fügt sich nicht in die postulierte Faktorenstruktur ein, was die diagnostische Aussagekraft des Handlungsteils einschränkt. Es fehlen Studien zur Kriteriumsvalidität der deutschen Version der WPPSI-III.

Wie schon der HAWIVA-III ist die Eignung der WPPSI-III als intelligenzdiagnostisches Basisverfahren eingeschränkt. Es fehlen Aufgaben zum rechnerischen Denken und zum Arbeitsgedächtnis, die ergänzend aus anderen Testbatterien hinzuzufügen sind. Wegen der faktoriellen Heterogenität des Handlungsteils ist diese Subskala zur Erfassung der fluiden Intelligenz nicht empfehlenswert. Für die diagnostische Praxis wichtig und nahezu alternativlos ist dagegen die Verbalskala, auch wenn deren Auswertungsobjektivität Mängel aufweist und die Durchführung bei Kindern mit erheblichen sprachlichen Schwierigkeiten Motivationsprobleme aufwerfen kann.

Verglichen mit dem HAWIVA-III stellt die WPPSI-III (deutsch) bislang keine erkennbare Verbesserung dar. Testanwender, die um Aktualität ihrer Diagnostik bemüht sind und bei der Intelligenzmessung von Vorschulkindern einen Wechsler-Test einsetzen wollen, werden dennoch nicht umhin kommen, sich auf die WPPSI-III umzustellen, da HAWIVA-III aus dem Handel genommen wurde. Durch das Ergänzungsset wird dabei zumindest keine komplette Neuanschaffung des Testkoffers erforderlich, allerdings weicht das Testmaterial nunmehr beim Matrizen-Test an einer Stelle von dem des HAWIVA-III ab. Die Abwandlungen der Testinstruktionen sind vor allem für routinierte Tester gewöhnungsbedürftig. Durch fehlende Hinweise auf Veränderungen bei den Instruktionen und den Auswertungsrichtlinien wird die Umstellung von HAWIVA-III auf WPPSI-III unnötig erschwert. Neueinsteiger können sich über den etwas niedrigeren Anschaffungspreis freuen und die günstigeren Kosten für Verbrauchsmaterialien kommen allen Anwendern zugute.

## Literatur

- Irblich, D. (2009). Rezension zu HAWIVA-III. Hannover-Wechsler-Intelligenztest für das Vorschulalter – III. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 58, 467-476.
- Lichtenberger, E. O., Kaufman, A. S. (2004). *Essentials of WPPSI-III Assessment*. Hoboken, NJ: Wiley.
- Petermann, F., Petermann, U. (Hrsg.). *HAWIK-IV. Hamburg-Wechsler-Intelligenztest für Kinder*. Bern: Huber.
- Ricken, G., Fritz, A., Schuck, K. D., Preuß, U. (Hrsg.) (2007). *HAWIVA-III. Hannover-Wechsler-Intelligenztest für das Vorschulalter – III*. Bern: Huber.
- Sattler, J. D., Dumont, R. (2004). *Assessment of children. WISV-IV and WPPSI-III Supplement*. San Diego, CA: Sattler.
- Statistisches Bundesamt (2009). *Bildungsstand der Bevölkerung*. Ausgabe 2009. Wiesbaden: Statistisches Bundesamt.
- Wechsler, D. (2002). *WPPSI-III. Wechsler Preschool and Primary Scale of Intelligence – III*. San Antonio, TX: The Psychological Corporation.

Dieter Irblich, Auel