

Intelligenzdiagnostik im Bereich Geistige Entwicklung

Angesichts der Fülle von Intelligenztestungen im sonderpädagogischen Kontext und stetig steigender Zahlen von Kindern mit sonderpädagogischen Unterstützungsbedarfen, welche maßgeblich durch Intelligenztestergebnisse begründet werden, sollen Einschränkungen bei der Anwendung von Intelligenztests bei kognitiv beeinträchtigten Kindern im Besonderen für den Bereich Geistige Entwicklung erinnert und ergänzt werden. Gängige Testverfahren werden skizziert und auf praktische Relevanz überprüft, sowie Hinweise zur Gestaltung der Testsituationen beschrieben.

In view of the abundance of intelligence tests in the special education context and the steadily increasing number of children with special education support needs, which are largely due to intelligence test results, limitations in the use of intelligence tests in cognitively impaired children, particularly in the area of intellectual development, should be recalled and supplemented. Common test procedures are outlined and checked for practical relevance, and tips on how to design the test situations are described.

Zitierempfehlung:

Joél, T. (2025). *Intelligenzdiagnostik im Bereich Geistige Entwicklung*.

<https://testseminare.de/Intelligenzdiagnostik-im-Bereich-Geistige-Entwicklung/>

Frage einer Teilnehmerin eines Seminars zur Anwendung von Intelligenztests: Ich habe einen IQ von 68 getestet, ist das dann Geistige Entwicklung?

Antwort: Hüten Sie sich vor Fachleuten, die Ihnen diese Frage beantworten.

Eine große Schnittmenge zu dem sonderpädagogischen Unterstützungsbedarf Geistige Entwicklung besteht zur klinischen Diagnose Intellektuelle Entwicklungsstörung (ICD-11), in der ICD-10 noch Intelligenzminderung genannt. Sowohl für eine klinische Diagnose, als auch zur Feststellung des Bedarfs Geistige Entwicklung, ist das Konstrukt Intelligenz zentral und eine entsprechende

Terminologie wird bereits in den zusammenfassenden Überschriften genutzt: geistig; intellektuell; Intelligenzminderung. Um also festzustellen, ob ein entsprechender Bedarf bzw. eine klinische Diagnose vorliegt, gehört die Einschätzung des Konstrukts Intelligenz zu den Regeln der Kunst.

Schulämter erwarten großteils die Anwendung von Intelligenztests und es besteht die Gefahr, dass ein attestierter Unterstützungsbedarf ohne Test abgelehnt wird; ebenso besteht die Möglichkeit, dass bei einer juristischen Auseinandersetzung ohne Testergebnisse der festgestellte Unterstützungsbedarf verworfen wird. In einer Studie zur Anwendung von

Intelligenztests in der Sonderpädagogik gaben von 929 Befragten 72 Prozent an, dass von Ihnen die Durchführung eines Intelligenztests erwartet wird (Joél, 2021, S. 145).

In den aktuellen Empfehlungen der Kultusministerkonferenz zum Schwerpunkt Geistige Entwicklung (KMK, 2021) stellt die Intelligenzdiagnostik lediglich einen von vielen diagnostischen Bausteinen dar (ebd., S.18). Die Empfehlungen korrespondieren kaum mit den Erwartungen von manchen Schulämtern oder Bundesländern, die die Ergebnisse aus Intelligenztests deutlich prominenter vertreten wissen wollen und nicht als ein Teilaspekt unter vielen.

Auf die Anwendung eines Intelligenztests in Brandenburg wird verpflichtend nur bei bereits vorliegenden Testergebnissen verzichtet (Handreichung Brandenburg, 2018). In Hessen regelt eine entsprechende Verordnung, dass die Ergebnisse eines Intelligenztestes vorliegen müssen. Die Testergebnisse belegen etwa die Hälfte des Anspruchs auf sonderpädagogische Förderung im Förderschwerpunkt Geistige Entwicklung (Hessisches Amtsblatt, 2020, S. 593). Diese problemlos ergänzbaren Beispiele unterstreichen die hohe Relevanz der Intelligenzdiagnostik im Bereich Geistige Entwicklung. Es liegt nahe: Zentral bei einem Kind mit einem vermuteten Bedarf Geistige Entwicklung ist die Intelligenz. Die Intelligenz soll eingeschätzt werden. Intelligenztests messen Intelligenz. Also werden Intelligenztests genutzt. Bei diesem sehr üblichen Vorgehen wird davon ausgegangen, dass Intelligenztests zuverlässig bei

vermeintlich intelligenzgeminderten Kindern die Intelligenz messen.

Das ist ein Irrtum.

Einschränkungen bei der Anwendung von Intelligenztests

Angenommen, eine Gruppe von erfahrenen Sonderpädagoginnen und Sonderpädagogen mit dem Arbeitsschwerpunkt Geistige Entwicklung würde gefragt werden, ob der IQ der wichtigste Bestandteil der Diagnostik wäre. Die meisten Lehrkräfte würden die Frage verneinen und auf wichtige diagnostische Aspekte im Rahmen der weiteren Begutachtung wie Motorik, Kommunikation, Sprache, Sozialkompetenz etc. verweisen. Angenommen, das Kind würde getestet werden und hätte einen gemessenen IQ von 56, wie wahrscheinlich wäre es, dass die wichtigen diagnostischen Aspekte an Bedeutung verlieren und das diagnostische „Urteil“ weitgehend feststeht. Selbst der Autor dieser Zeilen würde vermutlich von einer Intelligenzminderung ausgehen, diese beeindruckende Zahl vor Augen. Wieso ist das so? Die Reduzierung der Intelligenz auf eine einzelne Zahl entspricht dem menschlichen Wunsch nach Orientierung und Kategorisierung. Kann diese Zahl mit normativen Vorgaben verglichen werden, geht eine Einordnung leicht von der Hand. Kategorisierungen haben eine positive Funktion, da sie komplexe Sachverhalte vereinfachen und einordnen können, beinhalten aber auch die Gefahr einer Simplifizierung eines komplexen Sachverhalts. Ein Unterschied von einem IQ-

Punkt entscheidet z. B. in manchen Regionen über Zuständigkeiten der Ämter.

Obwohl das Konstrukt Intelligenz nicht nachweisbar ist, ist die wissenschaftliche Akzeptanz darüber ebenso weitestgehend Konsens wie die Akzeptanz von dessen valider Messbarkeit (Kuhl et al., 2022). Dennoch darf ein moderater Kenntnisstand über die mit der Anwendung von Intelligenztests verbundenen Kontroversen, Einschränkungen und Vorgaben (Dworschak et al., 2024; Renner, 2022; AWMF, 2021) vorausgesetzt werden, um sowohl die Ergebnisse der Tests richtig interpretieren zu können, sowie den Stellenwert der Intelligenzdiagnostik und die Gefahren einer überrepräsentativen Einordnung innerhalb eines Gutachtens zu erkennen.

Bodeneffekte

Sind für ein kognitiv schwaches Kind nicht genügend schwache Items vorhanden, können keine differenzierenden Aussagen über den gemessenen Entwicklungsstand getroffen werden. In der Reihe von Items eines Subtests steigt das Niveau kontinuierlich an. Würde z. B. ein fünfjähriges Kind getestet und das erste und somit einfachste Item lautete „was bedeutet Photosynthese“, alle weiteren Items wären schwieriger, würde das Kind vermutlich 0 Punkte erzielen und das Testergebnis wäre ohne Erkenntnisgewinn.

Ein Bodeneffekt würde auch vorliegen, wenn bereits bei einer richtig gelösten Aufgabe ein durchschnittliches Ergebnis berechnet werden würde oder zu wenige Items in einem Subtest

vorhanden sind. Würde ein Kind mit einem Rohwert von einem richtig gelösten Item ein weit unterdurchschnittliches, aber bereits mit zwei richtig gelösten Items ein unterdurchschnittliches und mit drei Items ein durchschnittliches Ergebnis erzielen, wäre die „ungünstige Schwierigkeitsstaffelung der Testitems“ (Wyschkon, 2019) die Ursache von Bodeneffekten. Renner (2022) beschreibt mehrere Beispiele für Testergebnisse, und fragt z. B., ob ein Prozentrang von 79 ein Ergebnis im durchschnittlichen Bereich wäre. Solange nicht geklärt ist, ob diese Ergebnisse von Bodeneffekten beeinflusst sind, müsste die Antwort „Keine Ahnung“ (ebd., S. 284) lauten. Tatsächlich ist ein Prozentrang von 79 auch bei einem Rohwert von 0 möglich in einem Entwicklungstest (ebd.).

Besonders Testergebnisse von kognitiv schwachen Kindern sind anfällig für Bodeneffekte, denn Normtabellen „fransen“ an den Rändern aus, sind also ungenauer als im durchschnittlichen Bereich. Im oberen Bereich wird jedoch nicht von Bodeneffekten sondern von Deckeneffekten gesprochen, dessen Auftreten im Bereich Geistige Entwicklung unwahrscheinlich ist.

Bodeneffekte sind teilweise bereits in den Testmanualen beschrieben. Der SON-R 2-8 hält bereits Normtabellen für 2;0 jährige Kinder bereit, er diskriminiert hingegen „optimal (...) von 2;9 – 6;0 Jahren“ (Tellegen et al., Heft 1 S.48). Die Formulierung bedeutet in der Konsequenz, dass Kinder besser nicht vor 2;9 Jahren mit diesem Test überprüft werden sollten. Bodeneffekte können dazu führen, dass schwache Kinder überschätzt werden, dass junge bzw. schwache Kinder kaum

am unteren Rand der Altersbereiche valide getestet werden können und dass divergierende Ergebnisse im Rahmen von Retestungen über einen längeren Zeitraum ursächlich auf Bodeneffekte zurückzuführen sind. Bei der Testung von kognitiv schwachen und/oder jungen Kindern wäre es nicht übertrieben, im Rahmen der Intelligenzdiagnostik Bodeneffekte als eines der größten Probleme zu bezeichnen. Um Testergebnisse richtig interpretieren zu können, gehören Kenntnisse über Bodeneffekte zur Stellenbeschreibung von jeder Testanwenderin bzw. jedem Testanwender; entsprechende Kenntnisse sind jedoch im Bereich Geistige Entwicklung unverzichtbar.

Eine anschaulich und mit Beispielen und Übungen belegte Auseinandersetzung zu Bodeneffekten findet sich bei Renner (2022).

Durchführungsobjektivität und Veränderungen der Testbedingungen

Wird ein Testergebnis mit Hilfe einer Normtabelle in einen vergleichbaren Wert übertragen, muss sichergestellt sein, dass die Vergleichbarkeit auch tatsächlich vorliegt. Viele hundert, oft mehrere tausend repräsentativ gewählte Kinder sind für die Normstichprobe eines Tests getestet worden. Sinn und Zweck eines normierten Tests besteht darin, das Testergebnis eines Kinds mit der repräsentativen Gesamtheit der Kinder zu vergleichen. So können durch den Vergleich zwischen Testergebnis und repräsentativem Durchschnitt Aussagen über den kognitiven Leistungsstand des Testkinds getroffen werden.

Daraus resultiert, dass das Testkind exakt so getestet wird, wie die Kinder aus der Normstichprobe getestet worden sind. Obwohl diese logische Konsequenz banal und einleuchtend klingt, darf die Umsetzung in testdiagnostisches Handeln bezweifelt werden.

Sie liegt nicht vor, wenn das Kind unter anderen Bedingungen getestet wurde als die Kinder der Normstichprobe. Das Testergebnis darf nicht von der Testerin abhängen, sondern ausschließlich von der Leistung des Kinds. Ein eindrückliches Negativbeispiel liefert die KABC-II (Kaufman & Kaufman, 2015): Bei dem Subtest Gestaltschließen darf das Kind auch in einer Fremdsprache antworten. Sofern die Testerin weiß, dass die in der Fremdsprache genannte Antwort richtig ist, wird die Antwort mit einem Punkt bewertet. Ansonsten würde die Antwort nicht bewertet werden können. Die Intelligenz des Kinds würde in diesem Fall also von der Fremdsprachenkompetenz der Testerin abhängen und nicht mehr von der Leistung des Kinds.

Absolute Durchführungsobjektivität läge vor, wenn es egal ist, welche Person ein Kind testet und die Ergebnisse in jedem Fall identisch wären.

Intelligenzgeminderte Kinder sind eine inhomogene Gruppe. Es ist fraglich, ob die im sonderpädagogischen Kontext meist angewandten Intelligenztests diese Heterogenität berücksichtigen. So wird z. B. eine vom Üblichen abweichende Art der kognitiven Verarbeitung bei Kindern mit Trisomie 21 diskutiert (Rauh, o. D.). Ein Vergleich mit den an der Norm orientierten und konstruierten Intelligenztests, die oft abstrakte Muster nutzen, könnte bei Kindern mit Trisomie

21 zu weniger Itemlösungen führen, gleichwohl die Kinder zu problemlösenden Kognitionen in der Lage wären, sofern die Stimuli eher anschaulich und konkret sind (ebd., S. 4). Manche Kinder haben zudem körperlich-motorische Beeinträchtigungen, so dass Tests unter Zeitdruck und mit motorischen Komponenten wie beim Zusammenlegen von mehrfarbigen Plättchen nach einer Vorlage unangemessen sein könnten. Der Vergleich mit der Normstichprobe könnte als nicht passend und unfair empfunden werden, was er auch ist. Es könnte versucht werden, diese Benachteiligung durch notwendig wirkende Anpassungen auszugleichen, wie es durchaus übliche Praxis ist. Doch das ist leider so gut wie nicht möglich. Die Forderung nach einer an die Bedürfnisse der Kinder angepassten Testdurchführung klingt humanistisch und gut, wird mit Ausnahme von Renner und Scholz (2022) aber meist nicht begleitet von Vorschlägen, wie Anpassungen gestaltet werden könnten (Bundschuh, 2010). Würde z. B. bei einem Kind mit motorischen Schwierigkeiten bei einem Speed-Test die Zeit verlängert werden, würde es sich nicht um eine angemessene Anpassung handeln, sondern um einen Kunstfehler. Testergebnisse, die auf begründeten, dennoch willkürlich vorgenommenen Veränderungen im Testablauf basieren, sind ohne Wert, da das Kind nicht so getestet wurde, wie die Kinder in der Normstichprobe getestet worden sind, somit keine Vergleichbarkeit vorliegt.

In einer Studie zur korrekten Anwendung von Intelligenztests in 14 Sonderschulen gaben nur 2,4 % der teilgenommenen 213 Sonderpädagoginnen

und Sonderpädagogen an, sich an die Anforderungen der Durchführungsobjektivität zu halten. Neben unerlaubten Hilfestellungen und Verlängerungen der Zeitvorgaben hat sogar ein Drittel während der Testungen „leichte Hinweise zu den Lösungswegen“ gegeben (Huber, 2000).

Wenn die Testdurchführung nicht angepasst werden kann an die besonderen Bedürfnisse der Kinder mit dem Bedarf Geistige Behinderung, was bedeutet dies für die Aussagekraft durchgeführter Testungen im Sinne der Durchführungsobjektivität? Es bedeutet, dass die Tests vermutlich formal korrekt durchgeführt worden wären, aber nicht passgenau verändert werden konnten. Die Tests blieben was sie waren: unfair. Die Unfairness kann meist nicht aufgehoben werden, denn die Durchführungsobjektivität schlägt notwendig scheinende Anpassungen. Ein unter unfairen Bedingungen durchgeführter Test ist jedoch weniger aussagekräftig, da die Güte von Tests auch mit dem Gütekriterium Testfairness belegt wird. Die meisten Tests haben weder bei der Konstruktion, noch bei der Normierung, noch bei den Instruktionvorgaben motorisch und/oder intelligenzgeminderte Kinder berücksichtigt (Renner und Mickley 2015).

Konfidenzintervall, fehlende Routine, Normtabellen und Verständlichkeit

Würde ein leicht verschnupftes und moderat schlechtgelauntes Kind getestet werden? Auf jeden Fall. Es ist anzunehmen, dass sich ebenso in der Normstichprobe nicht alle Kinder in einem

idealen Testzustand befanden und diese Aspekte in den Normtabellen inkludiert sind. Dennoch kann sich die Tagesform eines Kinds auf ein Testergebnis auswirken. Messfehler sind auf mehreren Ebenen möglich, auf der Ebene der Kinder, der Testerinnen und Tester und bei dem Test selbst. Auf der Ebene der Kinder wäre es möglich, dass das Kind die Testerin nicht mag, oder durch einen Streit vorher emotional aufgewühlt ist. Auf der Ebene der Testerinnen und Tester wäre es möglich, dass nicht genügend Zeit für die Vorbereitung und Auseinandersetzung mit den Testmaterialien und den Besonderheiten des Kinds zur Verfügung stand im Rahmen von Personalknappheit und verkürzten zeitlichen Ressourcen. Die Forschungslage ist eindeutig: es stellt sich nicht die Frage, ob Fehler bei der Anwendung gemacht werden, sondern wie viele (Afonso et al., 1998; Lipsius et al., 2008; Joél, 2021). Hinzu kommt die breite Verteilung der Gutachten auf viele Schultern. Gutachterinnen und Gutachter testen somit selten, teilweise einmal jährlich, anstatt die Tests von wenigen in der Testanwendung routinierten Personen die teilweise sehr komplexen und mehrdimensionalen Intelligenztests durchführen zu lassen, z. B. von den vielerorts bereits bestehenden Diagnostik-Teams. Auch bei der Konstruktion der Tests selber sind Messfehler möglich, ansonsten würde die Reliabilität 1 betragen, doch Tests mit einer hundertprozentigen Messgenauigkeit gibt es nicht. Die Ermittlung der wahren Intelligenz auf den Punkt genau ist nicht möglich, war nie möglich und wird niemals möglich sein. Dieser Fakt ist genauso wenig umstritten wie er nicht selten ignoriert wird durch die Angabe von exakten

Intelligenzwertangaben in Gutachten.

Konfidenzintervalle berücksichtigen die oben beschriebenen immer anzunehmenden Messfehler und sollen diese auffangen. Die Intervalle sind die Bereiche, in denen sich unter Berücksichtigung der Messfehler mit hoher – wenn auch nicht absoluter – Wahrscheinlichkeit die tatsächlichen Ergebnisse befinden.

Für den Bereich Geistige Entwicklung kommt erschwerend hinzu, dass mögliche Messungenauigkeiten noch verstärkt werden könnten. Die Kinder können z. B. im Verhalten ungewöhnlicher sein, in der Konzentration beeinträchtigt und es wäre denkbar, dass sie die Instruktionen nicht so gut verstehen wie Kinder mit einem höheren kognitiven Potential. Testerinnen und Tester benötigen nicht nur ein hohes Maß an testtheoretischer Kompetenz und Routine im Umgang mit den Materialien, sondern auch an pädagogischer Erfahrung im Umgang mit herausforderndem Verhalten.

Normierte Intelligenztests gehen von der Annahme aus, dass das Konstrukt Intelligenz gleichmäßig verteilt ist und mit Hilfe der Gaußschen Glockenkurve dargestellt werden kann. Demnach würden Kinder im Bereich Geistige Entwicklung im weit unterdurchschnittlichen Bereich (ca. 2% der Gesamtheit) getestet werden. Da die Gesamt-Normierungsstichprobe eines Tests geteilt werden muss durch die Anzahl vorhandener Normtabellen (meist im Dreimonatsabstand), verringert sich die Normstichprobe bei einigen Tests auf eine Größe, die gerade noch statistischen Mindeststandards entsprechen. Bei entsprechend kleinen

Normstichproben liegt nahe, dass die Angaben für weit unterdurchschnittliche Ergebnisse ungenauer werden, zumal intelligenzgeminderte Kinder in die Erstellung der Normtabellen häufig ausgeschlossen worden sind (Renner & Mickley, 2015) und mögliche Bodeneffekte die begrenzte Aussagekraft von Ergebnissen an den Rändern verstärken könnten. Bei der KABC-II z. B. basiert die Interpretation eines Testergebnisses auf den Vergleich mit der repräsentativ bestimmten Gesamtheit von durchschnittlich 37 Kindern, die dem Alter des Testkinds entsprechen. Beim WISC-V sind es sogar nur 33, beim WNV 34 Kinder (siehe Abbildung 3).

Die Nutzung eines Intelligenztests einzig zur Ermittlung von einigen wenigen Werten, evtl. sogar nur einem IQ-Wert und dessen Abgleich mit einer Einstufung in Bedarfe würde weder die testtheoretischen Einschränkungen ins besonders bei intelligenzgeminderten Kindern berücksichtigen, noch die unbestrittenen Möglichkeiten, die Intelligenztests für die Diagnostik bieten.

Praxis

Die Praxisleitlinie Intelligenzminderung, erstellt federführend von der Deutschen Gesellschaft für Kinder- und Jugendpsychiatrie (AWMF, 2021), empfiehlt bei der Anwendung von Intelligenztests folgende Vorgaben:

- Empfehlung von mehrdimensionalen Tests ab frühestens 5 Jahren unter Vermeidung von Bodeneffekten und einem Messbereich von unter IQ 60
- keine älteren Normen als 10 Jahren
- Orientierung am CHC-Intelligenzmodell unter Einschluss der Fluiden Intelligenz, der Kristallinen Intelligenz, des Arbeitsgedächtnisses und der Visuellen Verarbeitung
- Obergrenze des Konfidenzintervalls IQ 69
- Bei Bedarf wird die Durchführung von einzelnen Subtests empfohlen, wenn eine Gesamttestung nicht möglich ist
- unabhängig vom Alter sollen Normtabellen für 2-3 Jahre jüngere Kinder vorhanden sein
- Tests können „unbrauchbar“ sein (ebd., S. 32 ff.), wenn die körperlichen Voraussetzungen nicht gegeben sind, z. B. Sehfähigkeit, Hörfähigkeit, Mitteilungsfähigkeit und motorische Fertigkeiten
- Bei einer Retestung sollen mindestens 12 Monate zwischen den Testungen liegen
- Gesamtwert verliert an Bedeutung bei vorliegender Heterogenität

- Empfohlen werden KABC-II, WISC-V, WPPSI-IV, AID-3, RIAS, SON-R 2-8, SON-R 6-40 und der WNV

Neben der Beachtung der AWMF-Vorgaben empfiehlt sich eine primäre Orientierung an folgenden Fragen, nachdem die Indikation für die Anwendung eines Intelligenztests geklärt ist:

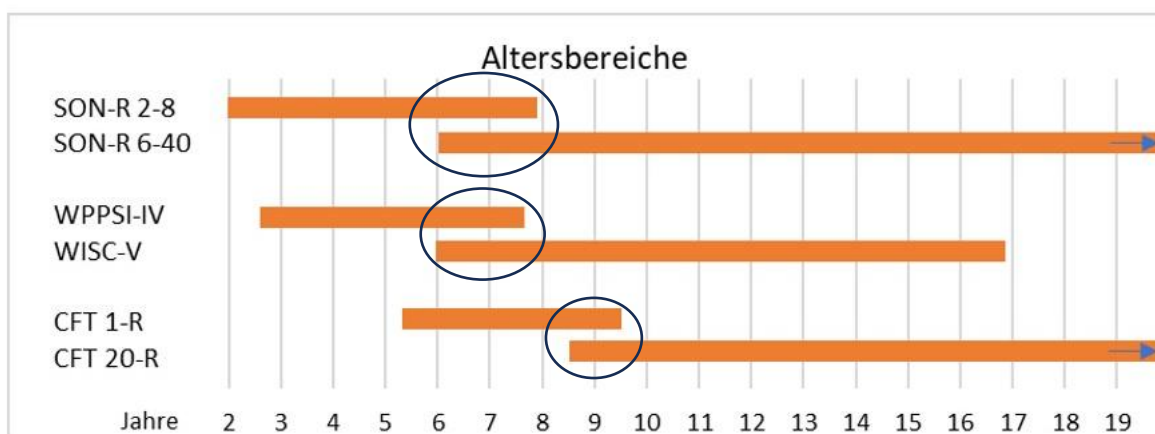
- Wie können Bodeneffekte vermieden werden?
- Welche Anpassungen sind möglich?
- Welche Qualifikationen sind notwendig, um die Tests angemessen durchführen zu können?
- Worauf ist bei der Testdurchführung und -auswahl zu achten?

Vermeidung von Bodeneffekten

Einige Testverfahren haben Varianten für jüngere und ältere Kinder (siehe Abbildung 1). Bei den beiden Tests aus der SON-Reihe gibt es z. B. eine Schnittmenge für den Altersbereich 6;0 – 7;11 Jahren. Fällt ein Kind in diese Schnittmenge,

sollte grundsätzlich der Test für jüngere Kinder gewählt werden, damit genügend Items für eine differenzierende Auswertung vorhanden sind. Wäre ein Kind z. B. 6 Jahre alt, könnte beim SON-R 6-40 das Anfangsitem bereits für ein vermeintlich intelligenzgemindertes Kind eine Herausforderung darstellen auf Grund der höheren Schweregrade der ersten Items. Beim SON-R 2-8 stünden hingegen bei einem sechsjährigen Kind auch die Items für die 2-5jährigen Kinder zur Verfügung, was Frustrationserlebnisse eher vermeiden hilft und die Wahrscheinlichkeit von Erfolgserlebnissen erhöht. Bodeneffekte können sich relativieren, je mehr Intelligenzdimensionen gemessen werden (Renner, 2022, S. 280), ein weiterer Vorteil des SON-R 2-8, der immerhin zwei Intelligenzdimensionen im Gegensatz zum SON-R 6-40 misst. Nicht nur zur Vermeidung von Bodeneffekten empfiehlt sich die Verwendung mehrdimensionaler Tests. Da komplexe Tests mehrere Intelligenzbereiche messen, bleiben Schwächen und Stärken weniger unentdeckt, was falsch negative Ergebnisse verhindern könnte.

Abbildung 1: Überschneidungen im Altersbereich ausgewählter Testreihen



Erzielt ein Kind in einem Subtest null Punkte, ist dies ein zu wertendes Ergebnis, das hinter dem Ergebnis stehende Konstrukt könnte eine Schwäche darstellen. Null Punkte sollten jedoch gültig nur interpretiert werden, wenn es sich nicht um einen Bodeneffekt handelt. Hier genügt oft ein kurzer Blick auf die entsprechende Alters-Normtabelle. Stellen null Punkte bereits ein Ergebnis nahe der Norm dar, ist es ohne Wert, da auf Bodeneffekte zurückzuführen. Ein weiterer kurzer Blick könnte prüfen, ob bereits eine Differenz von einem Rohwert zu rasanten Sprüngen innerhalb der Normtabellen führt. Ein letzter kurzer Blick gilt dem Formular: bestünde ein Subtest aus sehr vielen Items und beträgt die Altersbandbreite eines Tests aus wenigen Jahren, ist die Gefahr von Bodeneffekten geringer – wenn auch nicht ausgeschlossen – als bei Tests mit großen Altersbandbreiten und wenigen Items.

Ähnlich verhält es sich mit Referenzalter-Angaben. Bei den SON-Tests wird standardmäßig das Referenzalter bei der PC-Auswertung angegeben. Steht ein kleiner als Zeichen vor der Angabe, z. B. <5;6, sollte dieser irreführende Wert nicht interpretiert werden. Diese undifferenzierte Angabe könnte bedeuten, dass das Referenzalter 5;5 Jahre beträgt oder 3;4 Jahre oder 1;7 Jahre oder irgendein anderes Alter unter 5;6 Jahren.

Zusammengefasst kann festgestellt werden, dass Bodeneffekte bei der Testung intelligenzgeminderter Kinder eine bedeutsame Rolle spielen und nicht ignoriert werden dürfen. Gut verhindern lässt sich dieser Effekt, wenn jüngere Kinder getestet werden könnten mit den jeweiligen Subtests, mit denen das Testkind

getestet wird. Grundsätzlich gilt, dass im Testschrank Tests vorhanden sein sollten, die jüngere Kinder testen könnten, als es dem Alter der Kinder entspricht.

Anpassung versus

Durchführungsobjektivität

Nicht jede Anpassung verletzt die Durchführungsobjektivität so stark, dass eine Interpretation der Ergebnisse nicht mehr möglich ist. Einige Anpassungen sind durchaus auch aus Sicht der Testautorinnen und -autoren legitim und in den Manualen beschrieben. Es ist z. B. oft nicht vorstellbar, dass intelligenzgeminderte Kinder die teils komplexen Tests über einen Zeitraum von ein bis zwei Stunden konzentriert bearbeiten. Bei Bedarf wird die Durchführung verteilt auf mehrere Termine, damit weiter das Konstrukt Intelligenz getestet bleibt und nicht ein anderes Konstrukt wie Konzentration. Zwischen den Testterminen sollten allerdings nur wenige Tage liegen, damit die Normstichprobe sich nicht verändert, wenn das Kind bedeutsam älter ist am zweiten Termin.

Um beurteilen zu können, ob eine geplante Anpassung legitim ist oder die Durchführungsobjektivität zu stark gefährdet, wird ein Vorgehen in zwei Schritten vorgeschlagen. Im ersten Schritt wird geklärt, welches Konstrukt der betroffene Subtest misst. Bereits aus der Beantwortung dieser Frage ergibt sich meist, ob eine Anpassung möglich ist. Im zweiten Schritt wird entschieden, ob die Vergleichbarkeit mit der Normstichprobe weitestgehend erhalten bleibt oder die

Veränderung dazu führt, dass nicht mehr das postulierte Konstrukt gemessen wird.

Zwei Beispiele sollen das empfohlene Vorgehen verdeutlichen:

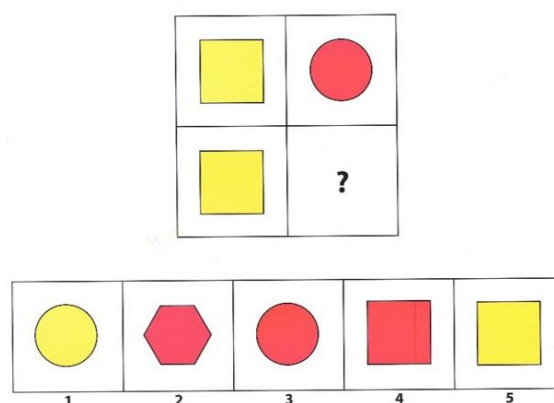
Beispiel 1: Geplant ist eine Vergrößerung der Stimuli beim Subtest Matrizen des WISC-V (siehe Abbildung 2), da das Kind eine starke Sehbeeinträchtigung hat. Matrizen-Tests messen die Fluide Intelligenz, das analytisch-abstrakt-logische Denken, konkret hier die Fähigkeit, Analogien bilden zu können, laut Manual auch die simultane Verarbeitung und Klassifikationsfähigkeiten (Wechsler, 2017). Wird die Größe der Stimuli verändert, wird dies kaum Einfluss nehmen auf die Durchführungsobjektivität, da unabhängig von der Größe der Stimuli weiterhin die Fähigkeit gemessen werden würde, Analogien bilden zu können im Rahmen der Messung der Fluiden Intelligenz.

Beispiel 2: Für den Zahlen-Symbol-Test des WISC-V soll die Zeitvorgabe verlängert werden, da ein motorisch beeinträchtigtes Kind einen Stift nicht angemessen nutzen kann. Der Subtest misst die Verarbeitungsgeschwindigkeit, wesentliches Merkmal ist das Arbeiten unter Zeitdruck. Gezählt

wird, wie viele Symbole entsprechend einer Vorlage in 120 Sekunden richtig zugeordnet werden können. Zentrales Merkmal von Subtests aus dem Bereich Verarbeitungsgeschwindigkeit ist – der Name deutet es an – die Geschwindigkeit. Werden Zeitvorgaben verändert, ist kein Vergleich mehr möglich mit den Kindern der Normstichprobe. Der Subtest würde nicht mehr das Konstrukt Verarbeitungsgeschwindigkeit messen, sondern Irgendetwas. Dieses Irgendetwas würde auf keinerlei Normstichprobe basieren oder auf ein von den Testautorinnen und -autoren definiertem Konstrukt. Tatsächlich würde bei diesem willkürlichen Vorgehen ein Ergebnis beschrieben werden, welches nichts mit dem beschriebenen Konstrukt Verarbeitungsgeschwindigkeit zu tun hätte. Immerhin 14 Prozent von 1077 Sonderpädagoginnen und -pädagogen gaben an, oft oder gelegentlich Zeitvorgaben verändert zu haben (Joél, 2021, S. 150).

Würde durch eine Veränderung die Vergleichbarkeit mit einer Normstichprobe entfallen, sollte die Veränderung unterbleiben, um nicht den Graubereich unseriösen diagnostischen Handelns zu betreten.

Abb. 2: Matrizen-Test WISC-V (Wechsler, 2017; mit freundlicher Genehmigung des Pearson Verlages)



Qualifikationen der Tester*innen

In einer Studie zur Anwendung von Intelligenztests im sonderpädagogischen Kontext gaben 39 Prozent der Befragten an ($n = 1077$), starke bis moderate Schwierigkeiten bei der Anwendung der Umkehrregel zu empfinden und 23 Prozent bei den Abbruchregeln (Joél, 2021, S. 151). Schwierigkeiten im Umgang mit den Regeln der Testverfahren werden bestätigt bei der Sichtung von Testformularen, die auf Korrektheit überprüft worden sind. Sechs Schulämter sendeten 248 Testprotokolle, die Tests wurden durchgeführt, um einen evtl. sonderpädagogischen Bedarf zu prüfen (ebd.). 151 Formulare (61%) waren fehlerhaft, insgesamt konnten 367 Fehler entdeckt werden. Neben der falschen Bestimmung der Rohwerte (43 % der Gesamtfehler) wendeten 21 Prozent die Abbruchregel, 13 Prozent die Umkehrregel falsch an. Bei eindimensionalen Tests wie CFT oder SON-R 6-40 wurden durchschnittlich 1,31 Fehler, bei den mehrdimensionalen Tests wie KABC-II durchschnittlich 1,76 Fehler festgestellt. Nur 20 Prozent der mehrdimensionalen Tests waren fehlerfrei. Fehler bei der Anwendung von Intelligenztests können teils vom Konfidenzintervall aufgefangen werden, würden ein Ergebnis allerdings kaschieren, dass überwiegend auf Fehlern basiert. Bei 48 Prozent der untersuchten 248 Formulare waren Auswirkungen durch eine fehlerhafte Anwendung möglich (ebd., S. 164). Renner merkt an, „Feststellungsverfahren sind keine Spielwiese für ungeübte Testleiter*innen“ (Scholz et al., 2019).

Doch was gehört zu einer angemessenen Qualifikation bei der Anwendung von Intelligenztests? Im Grunde sind dies drei Fähigkeiten: die Fähigkeit, einfühlsam, motivierend und verstehend mit Kindern umgehen zu können, die Fähigkeit, einen Test handwerklich sauber durchführen zu können unter Einbezug testtheoretischer Grundkenntnisse und die Fähigkeit, Ergebnisse auch unter Einbezug psychopathologischer Kenntnisse und Kenntnisse über die Zusammensetzung von Intelligenzdimensionen interpretieren zu können. Die erste Fähigkeit soll vorausgesetzt sein bei Menschen, die beruflich mit Kindern arbeiten.

Sind die üblicherweise im Studium vermittelten testtheoretischen Kenntnisse nicht vorhanden, ist deren Erwerb Bedingung. Könnten z. B. nicht ad hoc einminütige Impulsreferate z. B. zu den Begriffen Standardabweichung, Normalverteilung, Normstichprobe, Konfidenz- bzw. Vertrauensintervall, Standardisierte Werte, Prozentrang, Referenz- bzw. Äquivalenzalter, Flynn-Effekt, Generalfaktor, Boden- und Deckeneffekte, Umkehrregel und Abbruchregel gehalten werden, sollte eine Testdurchführung nicht erwogen werden. Herausforderndes Verhalten ist bei intelligenzgeminderten Kindern denkbar, Unsicherheiten im Umgang mit den Testmaterialien und der Vielzahl an teils mehreren hundert Durchführungsregeln könnten zu einer Verunsicherung bei den Kindern führen. Mehrere Probetestungen sind bei den mehrdimensionalen Tests Pflicht (KABC-II, WISC-V, WPPSI-IV), obwohl lediglich zweidimensional, ist der SON-R 2-8 erstaunlich kompliziert und sollte ebenfalls

geübt werden, besonders der Subtest Analogien. Kürzere Tests sind schneller zu lernen, z. B. WNV, CFT oder SON-R 6-40. Die Umkehrregeln der Tests sind von besonderer Bedeutung, da diese häufiger bei kognitiv schwachen Kindern Anwendung finden.

Oft werden noch alle Gutachten auf alle Mitarbeitenden in Teams, z. B. Förderzentren, gleichmäßig verteilt. Dies führt dazu, dass evtl. nur sehr selten jährlich getestet wird. Günstiger wäre die Durchführung der Tests durch erfahrene Diagnostikerinnen und Diagnostiker, z. B. „Diagnostikansprechpartner“ (Müller, 2009), denn die erworbene Routine führt dazu, dass die Testerinnen und Tester nicht durch einen Mangel an Praxiserfahrung eher mit dem Test als mit dem Kind in der Testsituation beschäftigt sind. Es wäre z. B. denkbar, dass einzelne Kolleginnen und Kollegen häufiger testen, andere dafür häufiger unterrichten.

Für die sichere Interpretation, besonders der mehrdimensionalen Testverfahren, ist eine Beschäftigung mit den in den Manualen erläuterten Intelligenztheorien bzw. -beschreibungen notwendig, auf denen die Tests basieren. Das Cattell-Horn-Carroll-Intelligenzmodell (CHC-Modell) deckt in diesem Zusammenhang einen großen Bereich zum Verständnis der Ergebnisse ab. Diskutiert werden bis zu 18 Intelligenzdimensionen bzw. „Breite Fähigkeiten“, wie z. B. Fluide und Kristalline

Intelligenz, Verarbeitungsgeschwindigkeit, Kurz- und Langzeitgedächtnis oder Visuelle Verarbeitung (McGrew, 2023), die mehr oder weniger gut erforscht und belegt sind, aber nicht alle relevant für den schulischen Kontext scheinen. Die „Breiten Fähigkeiten“ sind spezifiziert in über 90 „Enge Fähigkeiten“, z. B. die zur Fluiden Intelligenz zugeordneten Engen Fähigkeiten. Ein die gemessenen Fähigkeiten integrierender Generalfaktor soll einen Hinweis auf das allgemeine intellektuelle Potential eines Kinds ermitteln und wird meist als IQ dargestellt.

Eine angemessene Interpretation von Testergebnissen ist nicht denkbar ohne eine Auseinandersetzung mit den Intelligenzkonzepten, auf die sich in den Manualen berufen wird. Eine unangemessene Interpretation wäre die reduzierende Darstellung eines Gesamtwerts ohne Nennung der Konfidenzintervalle, durchaus übliche Praxis in Gutachten (Hennes et al., 2024).

Hinweise zur Gestaltung der Testsituation und zur Wahl der Tests

Abschließend sollen einige in der Praxis bewährte Hinweise zur Anwendung von Intelligenztests bei intelligenzgeminderten Kindern vorgestellt werden. Eine grafische Übersicht gängiger Testverfahren (Abbildung 3) stellt die Hinweise zusammen.

Abb. 3: Infos und Einschätzungen zu ausgewählten Intelligenztests

Alter	CHC Faktoren						Einschätzungen				Infos						Sprache				
	Fluide Intelligenz (gf)	Visuelle Verarbeitung (gv)	Langzeitgedächtnis (glr)	Kurzzeitgedächtnis (gsm)	Verarbeitungsgeschw. (gs)	Kristalline Intelligenz (gc)	Motivierend ¹	Anzahl von Items ¹	verständnis aus Sicht des Kindes ¹	Praktikabilität für Test-leitung ¹	Erhebung	Normstichprobe	Ø Kinder/Norm-stichprobe	Bodeneffekte	PC-Auswertung	Hochrechnen möglich	Substitution möglich	Subtests ohne Zeitdruck	sprachfrei	sprachfair	sprachgebunden
Normtabellen für ... Jahre	X ²	X ³					++	++	+	++	2015-17	Ø96	ja	ja	ja	ja	nein	3	X		
	X ⁴						+	++	+	++	2009-11	Ø84	ja	ja	ja	ja	nein	2	X		
CFT 1-R	X				X ⁵		++	+	+	++	2010	Ø309	ja	ja	ja	nein	nein ⁶	0		X	
CFT 20-R	X					X ⁷	+-	+	-	+	2015	Ø250 ⁸	ja	ja	ja	nein	nein	0		X	
WNV		X ⁹					++	++	+	+	2013	Ø34	kaum	nein	ja	ja	nein	2	X		
WPPSI-IV ¹⁰	X	X	X	X	X	X	+	++	+-	-	2015-17	Ø47	kaum	ja	ja	ja	ja	10		X ¹¹	X
WISC-V ¹⁰	X	X	X	X	X	X	+	++	+	-	2015-16	Ø33	ja	ja	ja	ja	ja	7		X ¹¹	X
IDS-2 ¹²	X	X	X	X	X	X	++	++	+	-	2015-17	n.b. ¹³	ja	nein	ja	ja	nein	7			X
KABC-II (IVI)		X	X	X			++	++	++	--	2013-14	Ø37	kaum	ja	ja	ja	ja	13 ¹⁴			X
KABC-II (IVI)	X	X	X	X			++	++	++	--	2013-14	Ø37	kaum	ja	ja	ja	ja	13 ¹⁴			X
KABC-II (FKI)		X	X	X	X	X	++	++	++	--	2013-14	Ø37	kaum	ja	ja	ja	ja	13 ¹⁴			X
KABC-II (FKI)	X	X	X	X		X	++	++	++	--	2013-14	Ø37	kaum	ja	ja	ja	ja	13 ¹⁴			X
KABC-II (SFI)	X ¹⁵	X ¹⁵					++	++	++	--	2013-14	Ø37	kaum	ja	ja	ja	nein	1-3 ¹⁶	X		

Anmerkungen. CHC = Cattell-Horn-Carroll-Intelligenzmodell. n.B. = nicht bekannt. 1) persönliche Einschätzungen; Einschätzungen fünfstufig: -- - - + + + laut Manual „Denkskala“. 3) laut Manual „Handlungsskala“. 4) gf laut Manual. 5) laut Manual „Wahrnehmungsgeschwindigkeit“. 6) motorische Schwierigkeiten: Auslassungen möglich 7) kristalline Zusatztests bei Bedarf. 8) ca. - Angabe (ungenau Beschreibung im Manual). 9) keine eigene Normtabelle für visuelle Verarbeitung. 10) keine explizite Bezugnahme zu den CHC Faktoren. 11) gilt nur für „Nonverbaler Index“. 12) nur Intelligenzpart. 13) Einsicht in Normtabellen nicht möglich, da nur online-Auswertung. 14) unabhängig vom Intelligenzmodell sind alle Subtests durchführbar. 15) begrenzt möglich. 16) abhängig vom Alter. FKI = Fluid-Kristallin-Index. IVI = Intellektueller Verarbeitungsindex. SFI = Sprachfrei-Index.

SON-R 6-40: Non-verbaler Intelligenztest. **SON-R 2-8:** Non-verbaler Intelligenztest (sic!). **CFT 1-R:** Grundintelligenztest Skala 1. **CFT 20-R:** Grundintelligenztest Skala 2. **WNV:** Wechsler Nonverbal Scale of Ability. **WISC-V:** Wechsler Intelligence Scale for Children – Fifth Edition. **WPPSI-IV:** Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition. **IDS-2:** Intelligence and Development Scales – 2. **KABC-II:** Kaufman Assessment Battery for Children – Second Edition.

Beziehungsgestaltung: Die Testsituation muss keine besondere Belastung für das Kind darstellen, im Gegenteil. Es handelt sich um eine Eins-zu-Eins Situation, was die Kinder als wohlthuend empfinden, das Kind erhält die ungeteilte Aufmerksamkeit und es wird dem generellen Bedürfnis von Menschen nach kognitiver Stimulation nachgekommen durch die Rätsel, die es zu lösen gilt. Die Tester*in sollte die Notwendigkeit des rigiden Einhaltens der Durchführungsobjektivität nicht mit einer generell rigiden Haltung in der Testsituation verwechseln. Stattdessen ist sie albern mit dem Kind, macht zwischendurch Quatsch und irgendwann bekommt das Kind zwischen zwei Subtests die Stoppuhr und die Tester*in wird getestet vom Kind, bevor es weitergeht. Der Fokus sollte vor Testbeginn auf die Interessen und Hobbys des Kinds gelegt werden. Mit Freude berichtet das Kind, wofür es sich interessiert und es ist eine gute Stimmung im Testraum im Vergleich zu einer schlechten Stimmung, die aus einer Fokussierung auf die Sorgen und Probleme im Warm-Up resultieren könnte. Zügig sollte der Test begonnen werden, da Ablenkung eine evtl. Aufregung vermindern hilft. Zu ausführliche Erklärungen zu Testbeginn könnten die Kinder kognitiv überfordern, die Stimmung im Raum zudem in eine unangemessen bedeutungsvolle Atmosphäre kippen lassen. Denkbar ist ebenso, dass vor Testbeginn zunächst das Kind zeitintensiv Vertrauen fassen und einen Zugang für die Testerin bzw. den Tester finden muss.

Intuitives Erklären: Für einige Subtests gestatten die Testverfahren einen Spielraum bei den

Instruktionen. Erklärt werden diese im besten Fall mit eigenen Worten so lange, bis das Kind verstanden hat. Günstig sind hier einfach zu erklärende Subtests. Ungünstig sind Lernaufgaben, die bereits bewertet werden, denn diese müssen im Sinne der Durchführungsobjektivität wortgetreu entsprechend dem Manual vorgegeben werden. Wenig komplex und oft intuitiv lassen sich die Subtests der KABC-II instruieren, komplexer und mit weniger Spielraum in den Erläuterungen wird der WISC-V durchgeführt. Bei den sprachfreien Tests (WNV, SON-R 2-8, SON-R 6-40, Sprachfrei Index KABC-II) ist ein intuitives Instruieren besonders gut möglich.

Körperlich motorische Beeinträchtigungen:

Kommen zu den kognitiven Defiziten motorische Beeinträchtigungen hinzu, die ein Nutzen der Hände verhindern, ist eine komplette Testdurchführung der üblicherweise angewendeten Intelligenztests nicht möglich. Um wenigstens einige Anhaltspunkte zu erhalten über das kognitive Potential, könnten einzelne Subtests durchgeführt werden, welche allerdings weniger aussagekräftig durch die eingeschränkte Reliabilität sind. Denkbare Ergebnisse resultieren aus Subtests, bei denen das Kind nicht einen Stift nutzen muss und die Aufgaben ohne Zeitdruck bewältigen kann. Bei der KABC-II sind dies immerhin 13 Subtests, beim WISC-V 7 Subtests.

Sprache: Bei nichtsprechenden bzw. nicht Deutsch sprechenden Kindern empfehlen sich sprachfreie (weder Kind noch Testleitung müssen

sprechen) bzw. sprachfaire Tests (Instruktionen werden verbal vorgegeben, Kind muss nicht sprechen). Sprachgebundene Tests (Kind und Testleitung müssen sprechen) sind vor allem dann ungünstig, wenn die Instruktionen sprachlich komplex sind. Aus diesem Grund wird die IDS-2 für den Bereich Geistige Entwicklung trotz der großen Aussagekraft bei erfolgreicher Durchführung nur bedingt empfohlen.

Motivation: Kinder mit vermutetem Unterstützungsbedarf haben häufig die unlustvolle Erfahrung mit Misserfolgen gemacht. Es bleibt Ihnen nicht verborgen, dass es in einem Test um richtige und falsche Lösungen geht und es ist denkbar, dass die Testsituation mit negativ erlebten Leistungssituationen assoziiert wird. Selbst wenn die Kinder ein Feedback über richtig oder falsch gelöste Items einfordern, dürfen diese nicht gegeben werden (Ausnahme: SON-Tests). Pädagogische Rückmeldungen über das Verhalten des Kinds („toll wie du dich anstrengst“) sind jedoch erwünscht und motivationsfördernd. Abbruchregeln vermeiden Frustrationserlebnisse, doch auch wenn nach drei falsch gelösten Items zum nächsten Subtest gewechselt wird, ahnt ein Kind evtl. den Misserfolg. Denkbar sind pädagogische „Tricks“ zur Erhaltung der Compliance. Wird eigentlich nach mehreren falsch gelösten Items zum nächsten Subtest gewechselt, könnten ähnlich dem gerade durchgeführten Subtest vergleichbar gestaltete Items erfunden werden, die so einfach sind, dass das Kind den Subtest mit einem Erfolgserlebnis beendet. Da die erfundenen Items nicht in die Bewertung einfließen, ist die Durchführungsobjektivität nicht

gefährdet. Die Tests selber können ebenfalls die Motivation beeinflussen, eine Einschätzung zur Attraktivität der Tests aus Sicht der Kinder ist Abbildung 3 zu entnehmen.

Substitution und Hochrechnen: Ist ein Subtest nicht durchführbar, ist die Berechnung eines Generalfaktors dennoch möglich, wenn der entsprechende Subtest ersetzt werden kann. Mehrdimensionale Tests wie KABC-II oder WISC-V (siehe Abbildung 3) unterscheiden in den zu beachtenden Vorgaben des Manuals zur Substitution zwischen durchzuführenden Kerntests und optionalen Ergänzungstests, die bei Bedarf für Substitutionen geeignet sind. Einige Tests berechnen zudem einen Generalfaktor bei gänzlich fehlenden Subtests.

Sowohl bei einer Substitution als auch beim Hochrechnen reduziert sich auf Grund schwächerer Testgütekriterien die Aussagekraft der Ergebnisse.

Ausblick:

Die Existenz eines Generalfaktors ist wissenschaftlich kaum umstritten, ebenso wenig der Zusammenhang zwischen erwartbarem Schulerfolg und Generalfaktor (Kuhl, 2022). Ein Irrtum liegt jedoch vor, wenn dies auch für Kindergruppen mit sehr besonderen Bedürfnissen oder vom Gros der Kinder abweichenden Besonderheiten gelten soll, im Besonderen intelligenzgeminderte Kinder.

Intelligenztests werden weiterhin vielfach durchgeführt werden. Zusammengefasst wird

dabei für die Berücksichtigung folgender Aspekte plädiert.

Es gibt keine gänzlich geeigneten Intelligenztests für intelligenzgeminderte Kinder. Richtig angewendet können vorhandene Tests dennoch im Rahmen einer sonderpädagogischen Begutachtung eine sinnvolle Ergänzung sein. Eine angemessene Anwendung berücksichtigt folgende Prämissen:

- Die Testergebnisse sind weniger bedeutsam, als es die Anwendungspraxis vermittelt, die Aussagekraft der Ergebnisse ist geringer, als es die Gewichtung in den Gutachten andeutet.
- Intelligenzdiagnostik ist eine Ergänzung der sonderpädagogischen Diagnostik, nicht die Grundlage zur Erkennung eines Unterstützungsbedarf.
- Intelligenztestergebnisse sind ungenau und messen lediglich Hinweise, nicht Beweise oder Nachweise. Intelligenztestergebnisse

bei intelligenzgeminderten Kindern sind noch ungenauer, zuweilen unbrauchbar.

- Die Verknüpfung von Testergebnissen mit Unterstützungsbedarfen ist eine schulpolitische Entscheidung. Intelligenztests messen keine sonderpädagogischen Unterstützungsbedarfe, sondern Intelligenz.
- Die Reduzierung des intellektuellen Potentials eines Kinds auf einen IQ ist trivial und ohne Berücksichtigung des Konfidenzintervalls in der Interpretation zudem falsch.
- Die Anwendung von Intelligenztests ist kompliziert und nachgewiesen fehleranfällig, eine angemessene Auseinandersetzung – besonders mit den mehrdimensionalen Tests – benötigt Zeit und Routine.

Dr. Torsten Joél

Europa-Universität Flensburg
Institut für Sonderpädagogik
Abteilung Sonderpädagogische Psychologie
torsten.joel@uni-flensburg.de

Literatur:

- Alfonso, V. C., Johnson, A., Patinella, L. & Rader, D. E. (1998). Common WISC-III examiner errors: Evidence from graduate students in training. *Psychology in the Schools*, 35, S. 119–125.
- AWMF. (2021). *Praxisleitlinie Intelligenzminderung*. Einsehbar unter: https://register.awmf.org/assets/guidelines/028-0421_S2k_Intelligenzminderung_2021-09.pdf [geprüft: 18.1.2024].
- Bundschuh, K. (2010). *Einführung in die sonderpädagogische Diagnostik* (7. Aufl.). München: E. Reinhardt.
- Dworschak, W.; Baysel, K. & Hackl, K. (2024). *Standards zur Feststellung des sonderpädagogischen Förderbedarfs im Förderschwerpunkt Geistige Entwicklung*. <https://osf.io/er9jb/files/osfstorage> [geprüft: 22.8.24].

- Eser, K. – H. (2007): Teilhabe-Barriere „*Sprachlich-Symbolische Enthinderung*“: Über den defizitären Umgang mit defizitären Lernprozessen. [http://www.sankt-nikolaus.de/web/st_nikolaus.nsf/gfx/26FB86E1E9A71896C12572D60043D8D0/\\$file/2007_05_07_eser_teilhabe_barriere.pdf](http://www.sankt-nikolaus.de/web/st_nikolaus.nsf/gfx/26FB86E1E9A71896C12572D60043D8D0/$file/2007_05_07_eser_teilhabe_barriere.pdf)
- Handreichung Brandenburg. (2018). *Handreichung zur Durchführung des sonderpädagogischen Feststellungsverfahrens*. https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/sonderpaedagogische_Foerderung_und_gemeinsamer_Unterricht/Handreichung_2020_zur_Durchfuehrung_sonderpaedagogisches_Feststellungsverfahren.pdf [geprüft: 22.8.24].
- Hennes, A.-K., Philippek, J., Dortants, L., Abel, M., Baysel, K., Dworschak, W., Fabel, L., Hövel, D., Jonas, K., Nideröst, M., Rösli, P., Schabmann, A., Stenneken, P., Wächter, J., Schmidt, B. M. (2024). Sonderpädagogische Diagnostik im Feststellungsprozess: Eine Ist-Stand-Analyse und der Blick nach vorn. *Zeitschrift für Heilpädagogik*, 7, S. 288 – 302.
- Hessisches Amtsblatt. (2020). Einsehbar: https://hessisches-amtsblatt.de/wp-content/plugins/pdf-viewer/stable/web/viewer.html?file=/wp-content/uploads/online_pdf/pdf_2020/10_2020.pdf%20 [geprüft: 22.8.24].
- Huber, C. (2000). Sonderpädagogische Diagnostik im Spannungsfeld traditioneller und gegenwärtiger Sichtweisen. Ergebnisse und kritische Reflexion einer Praxisuntersuchung an Schulen für Körperbehinderte. *Zeitschrift für Heilpädagogik*, 10, S. 411–416.
- Joél, T. (2021). *Die Anwendung von Intelligenztests im sonderpädagogischen Kontext. Eine empirische Untersuchung unter besonderer Berücksichtigung der Durchführungs- und Auswertungsobjektivität*. Weinheim: Beltz Juventa.
- Kaufman, A. S. & Kaufman, N. L. (2015). *KABC-II. Kaufman Assessment Battery for Children-II*. Deutschsprachige Fassung von P. Melchers und M. Melchers. Frankfurt: Pearson.
- KMK. (2021). *Empfehlungen zum Förderschwerpunkt Geistige Entwicklung*. Einsehbar unter: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_03_18-Empfehlungen-Schwerpunkt-Geistige-Entwicklung.pdf [geprüft: 2.9.22].
- Kuhl, J., Wittich, C. & Schulze, S. (2022). Intelligenz – Konstrukt und Diagnostik. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 131-146). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149>
- Lipsius, M., Petermann, F. & Daseking, M. (2008). Wie beeinflussen Testleiter die HAWIK-IV-Befunde? *Kindheit und Entwicklung* 17 (2), S. 107–117.
- McGrew, K. S. (2023). Carroll's three-stratum (3S) cognitive ability theory at 30 years: Impact, 3S-CHC theory clarification, structural replication, and cognitive-achievement psychometric network analysis. *Journal of Intelligence*, 11:32. Aktualisierter Forschungsstand: <https://www.themindhub.com/> [geprüft: 22.8.24].
- Müller, C. M. (2009). Schulinterne Diagnostikansprechpartner. Eine Praxiskonzeption zur Umsetzung systematischer Diagnostik im Schulalltag. *Zeitschrift für Heilpädagogik*, 5, S. 180–187.

- Rauh, H. (o. D.). *Kognitives Entwicklungstempo und Verhalten bei Kindern mit Down-Syndrom*. <https://downsyndrom-netzwerk.de/wp-content/uploads/2018/11/rauh-1.pdf> [geprüft: 22.8.24].
- Renner, G. & Mickley, M. (2015a). Berücksichtigen deutschsprachige Intelligenztests die besonderen Anforderungen von Kindern mit Behinderungen? *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 64, S. 88–103.
- Renner, G. & Scholz, M. (2022). Fair oder nicht fair, das ist hier die Frage!. Die Sicherung der Testfairness als Aufgabe der sonderpädagogischen Diagnostik. In Gebhardt M., Scheer D. & Schurig M. (Hrsg.). *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 259-274). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149> [geprüft: 22.8.24].
- Renner, G. (2022). Normtabellen analysieren und beurteilen I: Bodeneffekte erkennen und verstehen. In Gebhardt M., Scheer D. & Schurig M. (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 275-290). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149>
- Scholz, M., Renner, G. & Mickley, M. (2019). *Testinformation zur Kaufman Assessment Battery for Children – Second Edition (KABC-II)* (Dia-Inform Verfahrensinformationen 005-01). Ludwigsburg: Pädagogische Hochschule Ludwigsburg.
- Tellegen, P. J., Laros, J. A. & Petermann, F. (2018). *SON-R 2-8. Non-verbaler Intelligenztest*. Göttingen: Hogrefe.
- Wechsler, D. (2017). *WISC-V Durchführungs- und Auswertungsmanual*. Deutsche Fassung der WISC-V Wechsler Intelligence Scale for Children-Fifth Edition. Bearbeiter der deutschen Fassung F. Petermann. Frankfurt: Pearson.
- Wyschkon, A. (2019). *Normwerte in Leistungstests: Boden- und Deckeneffekte*. <https://www.hogrefe.com/de/thema/normwerte-in-leistungstests-boden-und-deckeneffekte> [geprüft: 22.8.24].