

Preusche, Ingrid; Koller, Margot; Kubinger, Klaus D.

Sprachfreie Administration von Intelligenztests nicht ohne Äquivalenzprüfung - am Beispiel des AID 2

Praxis der Kinderpsychologie und Kinderpsychiatrie 55 (2006) 7, S. 559-569

urn:nbn:de:0111-opus-9784

Erstveröffentlichung bei:



www.v-r.de

Nutzungsbedingungen

pedocs gewährt ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit dem Gebrauch von pedocs und der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Kontakt:

peDOCS

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Informationszentrum (IZ) Bildung

Schloßstr. 29, D-60486 Frankfurt am Main

eMail: pedocs@dipf.de

Internet: www.pedocs.de

1
2
3
5

Praxis der Kinderpsychologie und Kinderpsychiatrie

Ergebnisse aus Psychoanalyse,
Psychologie und Familientherapie

10
55. Jahrgang 2006

15
Herausgeberinnen und Herausgeber

Manfred Cierpka, Heidelberg – Ulrike Lehmkuhl, Berlin –
Albert Lenz, Paderborn – Inge Seiffge-Krenke, Mainz –
Annette Streck-Fischer, Göttingen

20
Verantwortliche Herausgeberinnen

25
Ulrike Lehmkuhl, Berlin
Annette Streck-Fischer, Göttingen

30
Redakteur

Günter Presting, Göttingen

35
40
42
43
44
45
V&R Verlag Vandenhoeck & Ruprecht in Göttingen

Sprachfreie Administration von Intelligenztests nicht ohne Äquivalenzprüfung – am Beispiel des AID 2

Ingrid Preusche, Margot Koller und Klaus D. Kubinger

Summary

An experiment for testing the psychometric equivalence of the non verbal instruction of the Adaptive Intelligence Diagnosticum

The Adaptive Intelligence Diagnosticum's manual additionally offers an optional non verbal instruction for seven of the fourteen subtests. There has, however, been no empirical evidence with regard to the psychometric equivalence of the tests with both versions of instruction. Therefore, an experiment was carried out. A total of 120 children (aged 6–8 years) were randomized into control (verbal instruction) and experimental groups (non verbal instruction). In the one subtest "Competence in Realism" nearly 50 percent of the children could not understand the relevant non verbal instruction. However, when only those children who did understand the non verbal instruction of a subtest are taken into account, no significant difference is observed, neither with respect to the mean of the scores nor to their standard deviation. In this respect, the test of equivalence has insofar been successful.

Key words: psychometric equivalence – non verbal instruction – fairness – Adaptive Intelligence Diagnosticum

Zusammenfassung

Im Manual des AID 2 wird für insgesamt sieben Unter- bzw. Zusatztests auch eine sprachfreie Instruktion angeboten, die optional anstatt der originalen sprachlichen Instruktion eingesetzt werden kann. Allerdings begründen sich diese nicht auf empirischen Ergebnissen, die die psychometrische Äquivalenz der Tests bei beiden Instruktionen belegen würden. Das hier vorgestellte Experiment unternimmt nun eine solche Äquivalenzprüfung. 120 sechs- bis achtjährige Kinder der „Normal“-Population wurden in randomisierter Weise einer Kontroll- (gesprochene sprachliche Instruktion) und Versuchsgruppe (sprachfreie Instruktion) zugewiesen. Für einen der Tests, 2 Realitätssicherheit, musste beobachtet werden, dass fast die Hälfte der Kinder die sprachfreie Instruktion nicht begreifen. Davon abgesehen konnte für nur diejenigen Kinder, welche die Instruktion je Test unmittelbar verstanden, festgestellt werden, dass zwischen Versuchs- und Kontrollgruppe keine signifikanten Leistungsunterschiede bestehen. So gesehen kann die Äquivalenzprüfung als gelungen bezeichnet werden.

Schlagwörter: psychometrische Äquivalenz – sprachfreie Instruktion – Fairness – AID 2

1 Einleitung

Im Zuge kinderpsychologischer Untersuchungen werden zur Abklärung der kognitiven Fähigkeiten standardmäßig Intelligenztestbatterien vorgegeben. Probleme entstehen in der Praxis jedoch regelmäßig bei der Entscheidung für das konkrete Verfahren, wenn einzelne Bereiche schon vor der Testung als beeinträchtigt bekannt sind. Dementsprechend ist z. B. von einem Einsatz des HAWIK-III bei motorischen Behinderungen abzuraten (Preusche u. Leiss 2003). Bei hörgeschädigten Personen ist der Einsatz eines Verfahrens, welches (auch) ohne Sprache zu artikulierten vorgegeben werden kann, zwingend; analog ergeben sich bei Testungen von Kindern mit einer anderen Muttersprache als Deutsch Grenzen des Einsatzes der deutschen Sprache. Die dafür vermeintliche Lösung, einen (Gebärden-)Dolmetscher zur Testung heranzuziehen, scheidet in der Praxis oftmals an den verfügbaren Ressourcen bzw. ist der Einsatz einer Begleitperson (z. B. eines Elternteils) als Dolmetscher psychologisch gesehen höchst problematisch. Abgesehen davon ergäben sich bei (frei) übersetzten Items bzw. auch nur der Instruktion so genannter „sprachfreier“ Tests dieselbe Problematik, wie sie im Folgenden angesprochen wird: die Frage der psychometrischen Äquivalenz beider Vorgabeweisen.

Gleichzeitig besteht seitens der Praxis und Forschung bei Kinderpsychologen und Kindertherapeuten der Wunsch nach einem möglichst universell einsetzbaren Messinstrument, um die Vergleichbarkeit der Ergebnisse am besten gewährleisten zu können. Dabei ist einerseits an die Vergleichbarkeit von Testpersonen mit verschiedenen Graden der Beeinträchtigung zu denken (also auch grundsätzliche Vergleichbarkeit mit unbeeinträchtigten Testpersonen), andererseits an die Vergleichbarkeit der Testergebnisse in Hinblick auf Prä-/Posttestungen (z. B. bei fortschreitenden Erkrankungen oder postoperativen Überprüfungen der Leistungsfähigkeit).

Die Intelligenz-Testbatterie AID 2 (Adaptives Intelligenz Diagnostikum, Version 2.1; Kubinger u. Wurst 2000) besitzt nun gegenüber ihrer ursprünglichen Version aus dem Jahre 1985 auch eine sprachfreie Instruktion – und zwar für die Unter- bzw. Zusatztests zur Messung manuell-visueller Fähigkeiten. Damit sollen insofern die Einsatzmöglichkeiten des AID 2 auch im Vergleich zu konkurrierenden Verfahren erweitert werden, als nunmehr fremdsprachige Testpersonen ebenfalls untersuchbar sind. Weniger intendiert, aber möglich soll auch die Testung Gehörloser mittels der sprachfreien Instruktion werden.

Allerdings bietet das Manual des AID 2 sprachfreie Instruktionen für sämtliche dafür infrage kommende fünf Unter- bzw. zwei Zusatztests zwar an, die psychometrische Äquivalenz mit der originalen, sprachlich artikulierten Instruktion wird dort aber nicht belegt. Das heißt, streng genommen ist ungeklärt, ob die sprachfreien Instruktionen in der Fallbehandlung eingesetzt werden können, ohne die betreffende Testperson damit gegenüber der geeichten Fassung der Testbatterie systematisch zu benachteiligen (oder zu bevorteilen). Der Klärung dieser Frage widmet sich die vorliegende Studie.

2 AID 2

Kurz gefasst handelt es sich um eine Intelligenz-Testbatterie für Kinder und Jugendliche zwischen 6;0 und 15;11 Jahren „zur Prüfung komplexer und basaler Kognitionen“ (Kubinger u. Wurst 2000, S. 8). Vorzugeben ist sie als Einzelverfahren, eine Gruppenvorgabe ist ausgeschlossen. Thematisch in Bezug auf etliche Untertests angelehnt an David Wechsler, ist diese Testbatterie geprägt durch ihre testtheoretische Fundierung gemäß Probabilistischer Testtheorie (vgl. z. B. Kubinger 2003): So konnte im AID 2 zum einen das Gütekriterium Skalierung (Verrechnungsfairness) erfüllt, zum anderen das adaptive Testen realisiert werden. Eine Besonderheit ist auch die Möglichkeit zum Screening von Teilleistungsstörungen. Die Überprüfung der Äquivalenz von sprachfreier und artikulierter sprachlicher Instruktion tangiert folgende Unter- bzw. Zusatztests:

- 2 Realitätssicherheit: Geprüft wird, inwieweit die Wirklichkeit um Dinge des Alltags verstanden wird bzw. kontrolliert werden kann – der Testperson wird auf Bildkarten jeweils ein Gegenstand mit einem fehlenden Detail gezeigt, das sie entdecken muss. Dabei genügt es, auf den fehlenden Teil hinzuzeigen.
- 4 Soziale und Sachliche Folgerichtigkeit: Geprüft wird die Fähigkeit, die Abfolge sozialen Geschehens bzw. alltäglicher Sachgegebenheiten zu verstehen und zu kontrollieren – die Testperson erhält ungeordnete Bildfolgen jeweils einer anderen Geschichte; die Bilder muss sie logisch richtig ordnen.
- 7 Kodieren und Assoziieren: Geprüft werden zwei voneinander partiell unabhängige Fähigkeiten, die Informationsverarbeitungsschnelligkeit und die Fähigkeit zum inzidentellen Lernen – die Testperson hat zuerst nach einer Vorlage Symbole zu kodieren, später aus dem Gedächtnis heraus.
- 8 Antizipieren und Kombinieren – figural: Geprüft wird schlussfolgerndes Denken in der Hinsicht, Teile eines (konkreten) Ganzen erkennen und dieses Ganze gestalten zu können – der Testperson werden Teile jeweils einer Figur vorgelegt, die sie zusammensetzen hat.
- 10 Analysieren und Synthetisieren – abstrakt: Geprüft wird die Fähigkeit, komplexe (abstrakte) Gestalten durch eine geeignete Strukturierung reproduzieren zu können – der Testperson werden geometrische Muster vorgegeben, die sie mit den verschiedenen Seiten von Würfeln nachlegen muss.
- 5a Unmittelbares Reproduzieren – figural/abstrakt: Geprüft wird die Kapazität der seriellen Informationsverarbeitung (im visumotorischen Bereich) – der Testperson werden jeweils Bilder einer Bildertafel in bestimmter Reihenfolge vorgegeben, was sie komplett und in derselben Aufeinanderfolge nachzumachen hat.

Der Zusatztests 10a Strukturieren – visumotorisch soll laut Manual erst ab 8 Jahren vorgegeben werden. Aufgrund der Altersverteilung der in dieser Studie verwendeten Stichprobe (ein Großteil der Testpersonen war jünger als 8 Jahre und bearbeitete diesen Test nicht) wird dieser Test – obwohl eine nonverbale Instruktion vorliegt – nicht in die Äquivalenzprüfung miteinbezogen.

Für all diese Tests gibt es also im Manual eine eigene sprachfreie Instruktion – das heißt eine Erklärung der Aufgabenstellung, ohne Verwendung von Schrift- oder gar gesprochener Sprache. Exemplarisch seien zwei davon hier wiedergegeben, und zwar in Gegenüberstellung zur sprachlichen Instruktion für die beiden Tests 2 Realitätssicherheit und 8 Antizipieren und Kombinieren – figural (Tab. 1).

Tab. 1: Sprachliche (Original-)Instruktion und sprachfreie Instruktion im Test 2 Realitätssicherheit (aus Kubinger u. Wurst 2000, S. 100, S. 161).
Sprachliche (Original-)Instruktion und sprachfreie Instruktion im Test 8 Antizipieren und Kombinieren – figural (aus Kubinger u. Wurst 2000, S. 129, S. 164)

sprachliche Instruktion UT 2	„Ich werde dir jetzt einige Bilder zeigen. Bei diesen stimmt irgend etwas nicht, es fehlt etwas. Wenn du genau schaut und ein wenig nachdenkst, findest du es oft sehr leicht. Zeige also jedes Mal hin oder sage mir, was auf dem Bild fehlt.“
sprachfreie Instruktion UT 2	Benötigt wird ... ein Blatt Papier und zwei Bleistifte; die Mine eines Bleistifts muß extra abgebrochen sein. Vorgehen ist wie folgt: Der Bleistift mit Mine ist vor der Tp [Testperson] vorläufig versteckt aufzubewahren. Den Bleistift ohne Mine legt der/die Tl [Testleiter/in] vor die Tp. Er/sie deutet auf die Bruchstelle, das heißt auf die Bleistiftspitze, an der die Mine fehlt. Nun nimmt er/sie diesen Bleistift in die Hand, versucht damit auf dem Blatt Papier zu schreiben und demonstriert, daß das Schreiben unmöglich ist; er/sie schüttelt dazu (verneinend) den Kopf. Nun holt er/sie den anderen Bleistift hervor, berührt mit dem eigenen Zeigefinger die Spitze der Mine und schreibt ein bis zwei Worte oder zeichnet ein Strichmännchen auf das Papier. Dann räumt der/die Tl diesen Bleistift wieder weg, legt nochmals den Bleistift ohne Mine vor die Tp und klopft mit dem Zeigefinger zwei- bis dreimal an die Stelle der fehlenden Mine. Jetzt wird der Tp das Spiralheft vorgelegt; mit offener, auf das erste Bild gerichteter Handfläche ist sie zur Aufgabenbearbeitung aufzufordern.
sprachliche Instruktion UT 8	Die Teile der ersten Figur (Aufgabe) sind in Reichweite der Tp gemäß Vorlage (im Testhandbuch) aufzulegen, und zwar mit dem „Anker“-Teil (als Mittelstück) unmittelbar vor ihr. „Wenn du (alle) diese Teile richtig aneinander setzt, wird eine Figur daraus, etwas, das du sicher gut kennst. Bevor du das jetzt probierst, sage ich dir noch, daß an diesen Teil ...“ – jetzt ist auf den „Anker“-Teil zu zeigen – „... alle anderen anstoßen“ – dabei ist auf die Begrenzungslinien des „Anker“-Teils hinzudeuten – „Versuche nun, diese Teile richtig zusammensetzen.“
sprachfreie Instruktion UT 8	Die Teile der ersten Figur (Aufgabe) sind in Reichweite der Tp gemäß Vorlage (im Testhandbuch) aufzulegen, und zwar mit dem „Anker“-Teil (als Mittelstück) unmittelbar vor ihr. Sodann legt der/die Tl die eigene linke Hand flach auf den „Anker“-Teil und dreht sie danach zwei- bis dreimal leicht nach links, zurück in die Ausgangslage und schließlich leicht nach rechts; beim Drehen der Handfläche nach links und rechts hebt er/sie gleichzeitig die rechte geöffnete Hand in die Höhe (☺) und schüttelt diese zum Zeichen für „nein“ kräftig um die Achse des Handgelenks. Danach hebt er/sie die linke flache Hand etwa 5cm hoch, um sie sofort darauf wieder kräftig auf den „Anker“-Teil zu pressen; dieses Hochheben und Festpressen ist einmal zu wiederholen. Unter weiterem Festhalten des „Anker“-Teils mit der linken Hand tippt er/sie nun mit dem Zeigefinger der rechten Hand nacheinander die übrigen Teile der ersten Figur an; anschließend tippt er/sie mit demselben Zeigefinger drei- bis fünfmal auf den Tisch, und zwar rund um die linke Hand. Gleich darauf tippt er/sie nochmals einen beliebigen Teil an, hebt den Zeigefinger von diesem Teil betont langsam in die Höhe und fährt mit dem Zeigefinger in der Luft in einem deutlichen Bogen zur linken Hand, um dort wieder drei- bis fünfmal rund um die linke Hand auf den Tisch zu tippen. Dieses Vorgehen soll für einen beliebigen zweiten Teil wiederholt werden. Nun ist die linke Hand vom Ankerteil zu nehmen und die Tp mit offener, auf die vor ihr liegenden Teile gerichteter Handfläche zur Aufgabenbearbeitung aufzufordern.

In dem im Folgenden vorgestellten Experiment wurden alle Kinder auch mit dem verbal-akustische Fähigkeiten messenden Test 1 Alltagswissen getestet – zur Abschätzung der „Allgemeinen Intelligenz“, um mögliche Unterschiede im Leistungsniveau zwischen den durch Randomisierung gewonnenen beiden experimentellen Gruppe a posteriori kontrollieren zu können:

- 1 Alltagswissen: Geprüft wird die Fähigkeit, sich Sachkenntnisse über Inhalte anzueignen, die in der heutigen Gesellschaft alltäglich sind – der Testperson werden (mündlich) Fragen gestellt, die sie (mündlich) zu beantworten hat.

Einige Kinder erhielten aus Gründen, die weiter unten ausgeführt werden, auch noch den ebenfalls verbal-akustische Fähigkeiten messenden Test 9 Funktionen abstrahieren:

- 9 Funktionen abstrahieren: Geprüft wird die Fähigkeit, durch Abstraktion zu einer Begriffsbildung zu gelangen – der Testperson werden jeweils zwei Objekte genannt, deren gemeinsame wesentliche Funktion sie erkennen und beschreiben muss.

3 Äquivalenzprüfung

Unter Äquivalenzprüfung versteht man die Überprüfung der (psychologischen) Gleichwertigkeit zweier Versionen ein und desselben psychologisch-diagnostischen Verfahrens (vgl. z. B. Wagner-Menghin 2003). Es gibt dafür zwei Herangehensweisen: Einerseits die Beurteilung durch Testpersonen vorrangig mittels qualitativer Verfahren, wie Interview und der Methode des „lauten Denkens“ („erfahrungsbezogene Äquivalenz“). Andererseits die statistische Analyse in Bezug auf wesentliche Gütekriterien („psychometrische Äquivalenz“). Die Evaluierung letzterer zielt vor allem darauf ab, ob die zu vergleichenden Testversionen beide dieselbe Eigenschaft (Trait) messen und ob die Verteilung der Testwerte identisch ist – das heißt, zumindest Mittelwert und Standardabweichung für beide Versionen übereinstimmen.

In jüngster Zeit wurde zwar häufig die Äquivalenz der Computerversion eines psychologisch-diagnostischen Verfahrens mit seiner ursprünglichen, originalen Papier-Bleistift-Version geprüft (vgl. z. B. Kindsvater u. Sturm 2003; Klinck 2002); schon seltener findet man Äquivalenzprüfungen bei Versionen, die für spezielle Populationen modifiziert wurden (vgl. Bühner u. Schmidt-Atzert 2004 mit der Überprüfung der Äquivalenz der DIN A3-Version des Tests d2 bei älteren Testpersonen). In der Literatur lassen sich jedoch keine Untersuchungen finden, die grundsätzlich oder wenigstens beispielhaft die psychometrische Äquivalenz der sprachfreien Instruktion eines Verfahrens mit seiner (originalen) sprachlichen Instruktionversion prüfen.

4 Experiment

Zur Äquivalenzprüfung beider Instruktionen des AID 2, der originalen sprachlichen – für die die Testbatterie geeicht wurde – und der sprachfreien, wurde folgendes Experiment geplant. Eine Stichprobe von Kindern mit Deutsch als Muttersprache und ohne

diagnostizierte Hörbeeinträchtigung sollte zwei Versuchsbedingungen in randomisierter Form zugeteilt werden; der Versuchsgruppe mit der sprachfreien Instruktion und der Kontrollgruppe mit der auszusprechenden sprachlichen Instruktion.

Obwohl die sprachfreie Instruktion ihrer Intention nach für fremdsprachige (und eventuell für gehörlose) Testpersonen gedacht ist, sind im Experiment für die angestrebte Äquivalenzprüfung explizit deutschsprachige bzw. unbeeinträchtigt hörende Kinder heranzuziehen. Nur so ist ein aussagekräftiger Vergleich zwischen beiden Versionen möglich.

4.1 Versuchspersonen

Die der Stichprobe zugrunde liegende Population wurde dazu auf 6- bis 8-Jährige beschränkt. Insgesamt 120 unauffällige Kinder aus Regelschulen konnten so getestet werden. Davon wurden zufällig 58 Kinder der Kontrollgruppe (KG, sprachliche Instruktion) und 62 der Versuchsgruppe (VG, sprachfreie Instruktion) zugeordnet. Die Altersverteilung ist unter beiden Versuchsbedingungen weitestgehend gleich und entspricht jeweils etwa einem Drittel pro Altersstufe (VG: 6 Jahre: $n = 23$; 7 Jahre: $n = 21$; 8 Jahre $n = 18$; KG: 6 Jahre: $n = 17$; 7 Jahre: $n = 21$; 8 Jahre: $n = 20$). Auch das Geschlechterverhältnis ist in beiden Versuchsbedingungen ziemlich ausgeglichen; insgesamt nahmen 63 männliche (VG: 32 Kinder, KG: 31 Kinder) und 57 weibliche Testpersonen (VG: 30 Kinder, KG: 27 Kinder) am Experiment teil.

4.2 Durchführung

Die Kinder erhielten die oben angeführten sechs Tests, die über eine sprachfreie Instruktion verfügen, entweder mit der sprachlichen Instruktion oder mit der sprachfreien. Der Test 1 Alltagswissen wurde am Beginn der Testung vorgegeben. Diejenigen Kinder, welche die sprachfreie Instruktion vorgegeben bekamen, erhielten zusätzlich eine standardisierte Erklärung dafür, dass bei den nächsten Aufgaben nicht mehr gesprochen werden würde.

Zum Einsatz kamen vier Testleiterinnen. Diese wurden intensiv in die Anwendung des AID 2 eingeschult, und zwar sowohl hinsichtlich der originalen als auch speziell hinsichtlich der sprachfreien Vorgabe. Um Reihenfolgeeffekte über die beiden Versuchsbedingungen zu vermeiden, wurde bei der Testorganisation strikt darauf geachtet, dass jede Testleiterin abwechselnd ein Kind mit der sprachlichen und das nächste Kind mit der sprachfreien Instruktion testete.

Die vier Testleiterinnen wurden auch über ihre Eindrücke und Beobachtungen beim Einsatz der sprachfreien Instruktion befragt, und zwar zu zwei Zeitpunkten, nach der Hälfte jeweils getesteter Kinder und am Ende der Untersuchung.

Schon bei der Testungen der ersten zehn Kinder mit der sprachfreien Instruktion offenbarten sich bei einigen, und zwar bei den jüngeren Kindern, beachtliche Verständnisschwierigkeiten, vor allem beim Test 2 Realitätssicherheit. Sie verstanden nicht, was zu tun ist, das heißt sie konnten nicht mit der Aufgabenbearbeitung beginnen. Die Verständnisschwierigkeiten bestehen vermutlich darin, die notwendige Generalisierung vorzunehmen, vom Aufgabenprinzip („auf den fehlenden Teil hin-

deuten“) – illustriert am physisch vorliegenden Gegenstand Bleistift – auf andere, nur abgebildete Gegenstände. Selbst die in der Folge (nach den ersten 10 Testungen) eingeführte Modifikation wurde regelmäßig nicht verstanden, nämlich bei Nichtverständnis der sprachfreien Instruktion dem Kind zusätzlich noch einen Bleistift mit fehlender Mine ebenfalls nur als bildliche Darstellung vorzugeben. Einige Kinder glaubten zum Beispiel, etwas mit dem Bleistift zeichnen zu müssen. Letztlich wurden die Kinder mit massiven Verständnisschwierigkeiten ausnahmsweise bei diesem Test doch mit der sprachlichen Instruktion getestet, um den Testablauf der vorgesehenen Testbatterie nicht zu ändern – die dabei erzielten Ergebnisse gehen in die Auswertung des Experiments nicht ein. Zusätzlich wurden diese Kinder auch mit dem Test 9 Funktionen abstrahieren getestet, um den Zusammenhang zwischen Nichtverständnis der sprachfreien Instruktion beim Test 2 Realitätssicherheit und der Fähigkeit zum Abstrahieren zu prüfen.

4.3 Statistische Analyse

Um die Vergleichbarkeit der beiden Gruppen zu prüfen, wurde deren mittleres Niveau im Test 1 Alltagswissen (bezüglich der T-Werte) verglichen. Anschließend erfolgte die Prüfung auf psychometrische Äquivalenz je vorgegebenem Test (T-Werte).

5 Ergebnisse

Weil der Mann-Whitney-U-Test genau genommen die Gleichheit zweier beliebiger stetiger Verteilungen prüft (vgl. z. B. Rasch u. Kubinger 2006) und nicht nur deren Lage wie vor allem der klassische t-Test zum Mittelwertsvergleich, wurde ersterem generell der Vorzug gegeben. Ergänzend wurde allerdings der F-Test eingesetzt, um auch explizit auf Streuungsunterschiede hin zu testen. Als Signifikanzniveau wurde $\alpha = 0.05$ gewählt.

5.1 Vergleichbarkeit der Gruppen

Die Randomisierung in Versuchs- und Kontrollgruppe ist, gemessen am Test 1 Alltagswissen als dem intelligenzmäßigen „Anker“, gelungen. Es besteht diesbezüglich kein Verteilungsunterschied (VG: $n = 62$, $M (SD) = 52.27 (10.21)$; KG: $n = 58$, $M (SD) = 53.26 (7.97)$; $z = -.77$; $p = .457$). Auch konkret die Varianzen unterschieden sich nicht: $F (57, 61) = 1.64$, $p > .05$. Somit sind die folgenden Ergebnisse hinsichtlich der Äquivalenzprüfung (s. Tab. 3) schlüssig zu interpretieren.

5.2 Befragung der Testleiterinnen

Die Befragung ergab, dass alle vier Testleiterinnen zu beiden Zeitpunkten übereinstimmten, die sprachfreie Instruktion im Test 2 Realitätssicherheit sei generell schwer verständlich. Mit einer Ausnahme gilt dies auch für den Test 7 Kodieren und Assoziieren. Besonders gut verständlich war nach Meinung aller die sprachfreie Instruktion beim Test 8 Antizipieren und Kombinieren – figural.

Die Einschätzung der Testleiterinnen spiegelt sich auch in der Tabelle 2 wieder, in der pro Test zusammengestellt ist, wie viele der 62 mit der sprachfreien Version getesteten Kinder die Instruktion offensichtlich insofern verstanden haben, als sie mit der Aufgabenbearbeitung ohne weiteres beginnen konnten.

Tab. 2: Anzahl der Kinder ohne Instruktionsverständnisschwierigkeit je Test

<i>Test</i>	<i>Instruktion verstanden</i>
2 Realitätssicherheit	34 (54.8 %)
4 Soziale und Sachliche Folgerichtigkeit	56 (90.3 %)
7 Kodieren und Assoziieren	45 (72.6 %)
8 Antizipieren und Kombinieren – figural	61 (98.4 %)
10 Analysieren und Synthetisieren – abstrakt	59 (95.2 %)
5a Unmittelbares Reproduzieren – figural/abstrakt	60 (96.8 %)

5.3 Psychometrischen Äquivalenz

Die Prüfung auf psychometrische Äquivalenz konzentriert sich auf die Hypothese, dass in Bezug auf die laut Eich Tabellen resultierenden T-Werte je Test keine Verteilungsunterschiede zwischen Versuchs- und Kontrollgruppe bestehen.

Die Hypothesenprüfung erfolgte – dem angestrebten U-Test entsprechend – univariat, also je Test. Da das α -Niveau in diesem Fall nicht vergleichs- sondern versuchsbezogen gewählt werden muss, wurde auch eine α -Adjustierung durchgeführt. Die Ergebnisse zeigt Tabelle 3.

Im Test 8 Antizipieren und Kombinieren – figural ergeben sich beim nominellen $\alpha = .05$ zwar laut U-Test signifikante Unterschiede (und zwar zugunsten der sprachlichen Instruktion), nicht jedoch nach erfolgter α -Adjustierung ($\alpha^* = 1 - \sqrt[7]{1 - \alpha}$; $\alpha = .05$, $\alpha^* = .0073$). Unabhängig davon ist der beobachtete Unterschied der Mittelwerte gering, jedenfalls mit 2,76 T-Werten inhaltlich nicht relevant. Konkret die Varianzen unterscheiden sich bei allen Tests nicht signifikant. Somit spricht alles für eine äquivalente Verteilung der Testwerte in beiden Instruktionsversionen – vorausgesetzt die sprachfreie Version wurde verstanden. Darüber hinaus wurde noch Folgendes ausgewertet. Es interessierte, ob sich diejenigen Kinder der Versuchsgruppe, welche in mindestens einem Test die Instruktion nicht verstanden hatten, von denjenigen in Bezug auf den Test 1 Alltagswissen unterscheiden, welche in keinem Test Schwierigkeiten mit der sprachfreien Instruktion hatten. Das Ergebnis war nicht signifikant – VG „Instruktion immer verstanden“, $n = 30$, $M (SD) = 51.10 (11.31)$; VG „Instruktion nicht verstanden“: $n = 32$, $M (SD) = 53.38 (9.09)$; $z = -.635$; $p = .525$). Des Weiteren interessierte, ob ein Zusammenhang zwischen Instruktionsverständnis und Leistungsniveau im Test 2 Realitätssicherheit, dem Test mit der besonders schwer verständlichen sprachfreien Instruktion, besteht. Die Varianzanalyse für die drei Gruppen VG und „Instruktion verstanden“, VG und „Instruktion nicht ver-

standen und daher mit sprachlicher Instruktion getestet“ sowie KG in Bezug auf den Test 2 Realitätssicherheit erbrachte ein signifikantes Ergebnis – VG „Instruktion verstanden“: $n = 34$, $M (SD) = 52.00 (5.61)$; VG „Instruktion nicht verstanden“: $n = 25$, $M (SD) = 45.75 (7.02)$; KG: $n = 58$, $M (SD) = 51.69 (6.97)$; $p < .001$. Nahezu 6 T-Werte niedriger ist der durchschnittliche Testwert bei den Kindern, die die sprachfreie Instruktion nicht verstanden haben und deshalb mit der sprachlichen Instruktion getestet wurden. Schließlich interessierte, ob Kinder, die im Test 2 Realitätssicherheit Verständnisschwierigkeiten hatten, auch Schwierigkeiten in der Abstraktionsfähigkeit haben. Der Einstichproben-t-Test in Bezug auf den Test 9 Funktionen abstrahieren ergab hinsichtlich des Mittelwerts der Eichstichprobe kein signifikantes Ergebnis – bei VG, „Instruktion nicht verstanden“: $n = 23$, $M (SD) = 49.65 (10.58)$; Eichstichprobe: $M = 50$, $SD = 10$; $t(22) = -.158$; $p = .876$.

Tab. 3: Vergleich der Verteilungen hinsichtlich der T-Werte je Test bei Versuchs- und Kontrollgruppe – Gruppengröße (n), Mittelwert und Standardabweichung (SD), Teststatistik im U-Test (z), p -Wert sowie die Angaben zum F-Test

Test	Versuchsgruppe (sprachfrei)		Kontrollgruppe (sprachlich)		z-Wert	p	F (df_1, df_2)	F ₉₅ (2-seitig)	F-Test (sign.)
	n	M (SD)	n	M (SD)					
1 Alltagswissen	62	52.27 (10.21)	58	53.26 (7.97)	-.744	.457	1.64 (57,61)	1.68	n. s.
2 Realitätssicherheit	34	52.00 (5.61)	58	51.69 (6.97)	-.152	.880	1.54 (33,57)	1.90	n. s.
4 Soziale und Sachliche Folgerichtigkeit	56	54.36 (10.68)	58	54.34 (11.50)	-.760	.447	1.16 (55,57)	1.70	n. s.
7 Kodieren und Assoziieren – Kodiermenge	45	49.76 (8.77)	58	50.17 (7.42)	-.639	.523	1.40 (44,57)	1.74	n. s.
7 Kodieren und Assoziieren – Assoziationen	45	49.29 (11.29)	58	49.55 (9.38)	-.417	.677	1.45 (44,57)	1.74	n. s.
8 Antizipieren und Kombinieren – figural	61	48.15 (9.21)	58	50.91 (8.98)	-2.231	.018	1.05 (57,60)	1.68	n. s.
10 Analysieren und Synthetisieren – abstrakt	59	47.90 (10.02)	58	49.78 (7.93)	-1.079	.281	1.60 (57,58)	1.68	n. s.
5a Unmittelbares Reproduzieren – figural/abstrakt	60	51.00 (8.82)	58	52.55 (10.93)	-.896	.370	1.54 (57,59)	1.68	n. s.

6 Diskussion

Die im Manual des AID 2 implizierte psychometrische Äquivalenz der betroffenen Tests hinsichtlich der beiden Versionen (artikulierte sprachliche und sprachfreie Instruktion) erweist sich mit dem vorliegenden Experiment weitgehend, jedoch nicht vollständig gegeben. Immerhin konnte einer der Tests nach sprachfreier Instruktion von nahezu jedem zweiten Kind nicht bearbeitet werden. Bei den anderen Tests ist der entsprechende Prozentsatz deutlich niedriger bzw. praktisch vernachlässigbar. Verstanden die Kinder die sprachfreie Instruktion, das heißt begannen sie mit der Aufgabenbearbeitung ohne weiteres, zeigten sich keine inhaltlich relevanten Verteilungsunterschiede in den Testwerten. Das gilt für alle Tests, auch für den mit der sprachfreien Instruktion am schwierigsten zu verstehenden Test 2 Realitätssicherheit. Die nicht signifikanten Unterschiede im Test 1 Alltagswissen und Test 9 Funktionen abstrahieren bei Kindern mit und ohne Instruktionsverständnisschwierigkeit sind dahingehend zu interpretieren, dass solche mit Schwierigkeiten ausschließlich im Test 2 Realitätssicherheit und der damit gemessenen Fähigkeit Defizite aufweisen – sie liegen durchschnittlich eine halbe Standardabweichung unter der „Norm“, wenn sie letztlich doch mit der sprachlichen Instruktion getestet werden. Das heißt, dass gerade der Mangel an der Fähigkeit, die Wirklichkeit um Dinge des Alltags zu verstehen bzw. zu kontrollieren, dafür verantwortlich ist, wenn aus der gegebenen sprachfreien Instruktion nicht genügend Information geschöpft werden kann. Trotzdem ist auch noch ohne Verbesserung der sprachfreien Instruktion beim Test 2 Realitätssicherheit dieser für die Praxis geeignet, weil schlimmstenfalls gleich auffällt, dass das betreffende Kind so nicht testbar ist – für Kinder, die die sprachfreie Instruktion verstehen, ist dieser wie die übrigen Tests psychometrisch äquivalent. Was allerdings an sprachfreien Instruktionen interessierte Praktiker betrifft, sind diese mit den vorliegenden Ergebnissen besser zu informieren, dass eben in einem der Tests des AID 2 die gegebene Instruktion sehr oft nicht die Intention erfüllt. Wagt man also einen Ausblick auf die Einsetzbarkeit der sprachfreien Instruktion bei Kindern mit fremder Muttersprache (oder Beeinträchtigungen des Hörens), so kann davon ausgegangen werden, dass durch die Umgehung der Sprache Kinder fair zu testen sind.

Trotz der weitgehenden hier festgestellten Äquivalenz einer sprachfreien Instruktionversion mit der originalen sprachlichen Version sollte es mit den vorliegenden Ergebnissen gelungen sein, zu zeigen, dass Äquivalenz nicht schlicht angenommen werden kann, sondern empirisch belegt sein muss, bevor dem Praktiker eine solche Version angeboten wird.

Literatur

- Bühner, M.; Schmidt-Atzert, L. (2004): Überprüfung der Äquivalenz einer Test d2-Version für ältere Probanden. *Zeitschrift für Neuropsychologie* 15: 7–13.
- Kindsvater, S.; Sturm, W. (2003): Computer- vs. Papier-Bleistiftvorgabe: Äquivalenzstudie zum nonverbalen Lerntest (NVL): *Zeitschrift für Neuropsychologie* 14: 13–21.

- Klinck, D. (2002): Computergestützte Diagnostik. Göttingen: Hogrefe.
- Kubinger, K. D. (2003): Testtheorie, Probabilistische. In: Kubinger, K. D.; Jäger, R. S. (Hg.): Schlüsselbegriffe der Psychologischen Diagnostik. Weinheim: Beltz PVU, S. 415–423.
- Kubinger, K. D.; Wurst, E. (2000): Adaptives Intelligenz Diagnostikum – Version 2.1 (AID 2). Göttingen: Hogrefe.
- Preusche, I.; Leiss, U. (2003): HAWIK-III – AID 2 – K-ABC im Vergleich – ein Überblick. Report Psychologie 28 (1): 12–26.
- Rasch, D.; Kubinger, K. D. (2006): Statistik für das Psychologiestudium – Mit Softwareunterstützung zur Planung und Auswertung von Untersuchungen sowie zu sequentiellen Verfahren. Heidelberg: Spectrum.
- Wagner-Menghin, M. (2003): Äquivalenzprüfung. In: Kubinger, K. D.; Jäger, R. S. (Hg.): Schlüsselbegriffe der Psychologischen Diagnostik. Weinheim: Beltz PVU, S. 32–34.

Korrespondenzadressen: Mag. Dr. Ingrid Preusche, Arbeitsbereich Psychologische Diagnostik, Fakultät für Psychologie, Universität Wien, Liebiggasse 5, A-1010 Wien;
E-Mail: ingrid.preusche@univie.ac.at
Univ.-Prof. Mag. Dr. Klaus D. Kubinger, Arbeitsbereich Psychologische Diagnostik samt Test- und Beratungsstelle, Fakultät für Psychologie, Universität Wien, Liebiggasse 5, A-1010 Wien;
E-Mail: klaus.kubinger@univie.ac.at