

Mehrfachtestungen bei hochbegabten Problemkindern

Mehrfachtests bei gut leistenden und gut gedeihenden Kindern können für die Forschung interessant sein, etwa unter der Frage, wie stabil die Messwerte im Entwicklungsverlauf sind oder welche Faktoren die Gültigkeit der Messwerte beeinflussen. In der Praxis aber werden sich Diagnostiker einig sein, dass Tests, die keiner Problemlösung dienen, zu vermeiden sind. Das Testbusiness, das sich zum Nachweis eines Mindestwertes für die Teilnahme an Hochbegabtenkursen oder Förderprojekten, etabliert hat, lasse ich im Folgenden außer Betracht. In diesem Beitrag interessieren die Kinder, die im Risiko stehen, die trotz oder wegen einer hohen Begabung in ihrer Bildungslaufbahn auf Hindernisse stoßen oder womöglich in eine krisenhafte Persönlichkeitsentwicklung geraten. Was bedeuten Mehrfachtestungen für diese Zielgruppe?

Für die weiteren Überlegungen ist die Unterscheidung von genetisch angelegter und durch Förderung entwickelte Leistung von Bedeutung. Die fluide Intelligenz umfasst das physiologische Funktionieren der Kognition, auf der Basis von genetischen Bedingungen und epigenetischen und weiteren neurobiologischen Festlegungen. Die kristalline Intelligenz umfasst dagegen die Wirkungen, die aus individuellen Schwerpunkten, Übung und Förderung entstehen; sie gibt also überwiegend Information über die Leistung. Die Aufmerksamkeit der Hochbegabtenforschung ist zu einem wesentlichen Teil in diese Richtung gewandert. Damit wird aber das Thema Hochbegabung unklar, denn für den Aufbau von „Expertisierung“ (Hochleistung in einem Schwerpunkt) ist zwar eine gute Basis-Intelligenz wichtig, das Gelingen setzt aber neben konsequenter Förderung und Anstrengungsbereitschaft weitere nicht kognitive Begabungen und Persönlichkeitseigenschaften sowie eine unterstützende soziale Umgebung voraus, so dass die Annahme, das Leistungsniveau steige parallel zum Intelligenzniveau keine Grundlage hat. Karl Landscheidt, ein Schulpsychologe, zitiert einen amerikanischen Autor mit deutlichem Sarkasmus: „It is important to have enough of it (intelligence), but having lots and lots does not buy you that much. My regrets to Mensa, but that is the way things are“ (Hunt 1995, p. 7, zitiert bei Landscheidt, o.J., S. 2).

Die Testentwicklung hat neben dem Kaufman ABC, der beide Bereiche schon immer trennt, mit dem HAWIVA und dem HAWIK IV eine Hinwendung zur Erfassung der fluiden Intelligenz vollzogen.

Bei dem ersten Test eines Kindes geht es also im guten Fall um Diagnostik bei bestehender Problemlage – meist schwierige Erziehungs- und Leistungssituati-

onen. Die Intelligenz wird eingeschätzt, um gesicherte Grundlagen für pädagogisches Handeln in der Schule, der Familie oder im therapeutischen Kontext zu gewinnen. Die Einschätzung wird durch ausführliche Anamnese und ergänzende Persönlichkeitsdiagnostik unterstützt, damit die Ergebnisse der Intelligenzschätzung in einen Kontext gestellt werden können. Die Kinder haben noch keine „Testbiografie“.

Bei jedem weiteren Test geht es von Seiten des Initiators um Kontrolle des ersten und von Seiten des Kindes um Verteidigung seines Status oder um dessen Verbesserung. Denn Tests bringen oft Berechtigungen, es geht um was. Vergessen wird oft, dass mit der Wiederholungstestung Rückwirkungen auf das Denken und Empfinden der Kinder selbst und ihrer sozialen Umgebung angestoßen werden: sowohl in der Familie als auch in Kindergarten, Schule oder weiteren Einrichtungen, in denen sich das Kind bewegt. Bei den Eltern mag der dringende Wunsch nach guten Ergebnissen eine Rolle spielen, weil davon erwünschte Maßnahmen abhängig gemacht werden; bei dem Kind der Wunsch, die Eltern nicht zu enttäuschen, in der Schule mögen Zweifel entstanden sein, ob die zuvor dokumentierte Hochbegabung auch tatsächlich vorliegt, denn das Kind verhält sich doch unerwünscht; wäre es intelligent, dann würde es doch einsehen können ... Eltern und Kind kennen diese Zweifel und sind von den Entscheidungen der Lehrkräfte abhängig. Deren Einstellung zu Hochbegabung ist in der Regel eben so wie die vieler Eltern mit nicht hochbegabten Kindern ambivalent, in Kollegien polarisiert, in Elterngremien fast immer aversiv. Mittelbar, in den Erwartungen der Eltern des Testkindes, spielen diese Risiken der sozialen Rückmeldung immer eine Rolle.

Sie wirken in verschiedenster Weise auf die neue Testsituation ein und verändern damit die Rahmenbedingungen der Testleistung. Die unbefangene Kooperationsbereitschaft des Kindes, seine Freude an der Anstrengung, werden in unbekannter Richtung beeinflusst, ohne dass dies im Testergebnis verrechnet werden kann.

Schulen, Schulämter, Schulpsychologische Dienste arbeiten überwiegend zugewandt und einzelfallbezogen und geben den Familien die Hilfe, die über Krisen und Unsicherheiten hinweg trägt. Es liegt sicher an der besonderen Perspektive, die sich mir bietet, dass ich von dieser positiven Statistik sehr viele negative Abweichungen sehe. Die Schule für verhaltensschwierige Kinder und Jugendliche mit hoher Begabung, die ich in Offenbach mit aufgebaut habe, hat Kinder (aus der ganzen Bundesrepublik von den Jugendämtern zugewiesen), die zwischen zwei und fünf Wiederholungstests durchgemacht haben. In der Ombudsstelle für hochbegabte Problemkinder in Frankfurt erlebe ich Fälle, in denen das Wohlwollen der amtlichen Stellen nicht (mehr) vorhanden ist, so

dass Familien mit ihren opponierenden oder resignierten Kindern in Gegnerschaft zu den Behörden geraten. Die Fälle, in denen Hilfebedarf in Hilflosigkeit und Schädigung führt, sind vermutlich prozentual wenige, absolut gesehen aber viele, mindestens zu viele. In solchen Situationen sollten Wiederholungstests nicht in der Hand einer amtlichen Stelle liegen, die selbst Konfliktpartei ist. Der Test ist dann nicht valide; er entspricht nicht den Handbuchbestimmungen. Dazu ist es nicht erforderlich, die Redlichkeit des Testleiters in Frage zu stellen, es genügt, die seelische Situation des Kindes einzuschätzen. Defensivtests sind unzulässig, und das Kind ist in einer Defensivsituation (dazu weiter unten).

Mein Anliegen ist, Argumentationslinien zu zeigen, um den Betroffenen, den Eltern wie auch den Lehrerinnen und Lehrern und weiteren Zuständigen fundierte Argumente an die Hand zu geben, um unnötige, vor allem aber ungültige und damit schädliche Mehrfachtestungen zu vermeiden bzw. den Betroffenen das Geleit zu geben, das hinausführt.

Mit dem ersten Wiederholungstest entsteht eine Testbiografie. Sie kann sich etwa so lesen (Konstruiert nach Typik, nicht nach konkretem Fall):

In einer dritten Klasse will ein Kind, Arne, nicht mehr am Unterricht teilnehmen; die anderen Kinder ärgern ihn, er reagiert aggressiv. Auf Anraten der Lehrerin wird Arne einer niedergelassenen Psychiaterin vorgestellt, die mit Messergebnis 135 eine Hochbegabung feststellt und ein halbes Jahr mit dem Jungen therapeutisch arbeitet. Der Schule von Arne empfiehlt sie, Unterforderung zu vermeiden, ihm weitere oder andere Aufgaben zu geben, zusätzliche Herausforderungen zu bieten. In der vierten Klasse geht es um die Schullaufbahnentscheidung; Arne hat sein Verhalten nicht geändert, die Leistungen schwanken im mittleren Bereich, er soll der Realschule zugewiesen werden.

Die Mutter erhebt Einspruch. Arne wird von einer Schulpsychologin getestet mit Ergebnis 115. Er geht auf die Realschule. Der Kontakt zu der Psychiaterin wird abgebrochen. Offenbar hat sie falsch getestet. Einige Monate später in der Realschule ist er bald wieder Zielscheibe für die anderen Kinder, es kommt zu ersten Impulsdurchbrüchen; die familiäre Situation spitzt sich zu. Die Mutter verlangt entgegen der augenfälligen Plausibilität die Umsetzung in ein Gymnasium. Arne wird erneut, dieses Mal von einer Mitarbeiterin in einer kommunalen Erziehungsberatungsstelle, getestet; Ergebnis 125. Das Ergebnis wird im Schulamt vorgelegt, die Schulentscheidung wird aber nicht revidiert; man erkenne das Ergebnis nicht an, denn es komme nicht von einem Schulpsychologen. Ein Jahr später, in der 6. Jgst., verweigert Arne den Schulbesuch und wird in einer psychiatrischen Klinik vorgestellt. Hier wird er routinemäßig ein weiteres Mal

getestet, Ergebnis 105. Er wechselt in die Hauptschule. Ein Gutachten zur Feststellung des Sonderpädagogischen Förderbedarfs wird angeregt.

Man kann die Testbiografie bis zu diesem Zeitpunkt so lesen: 135, 115, 125, 105. Die Nachtestungen zeigen keine klare Linie.

Mit Arnes Augen kann man die Testbiografie so lesen:

In der dritten Klasse war meine Mutter unzufrieden mit mir, weil ich in der Schule nicht mehr so gut war. Sie hatte auch Mitleid mit mir, weil die Anderen mich immer geärgert haben. Die Psychiaterin hat mich getestet und mir erklärt, weshalb ich Probleme in der Schule habe; ich wäre schneller als die Anderen und müsse versuchen, mehr Geduld zu haben. Meine Mutter war danach sehr lieb zu mir. Bei dem zweiten Test war meine Mutter enttäuscht, weil ich schlechter als beim ersten Mal war. Die Schulpsychologin hat ihr gesagt, dass die Erklärung der Psychiaterin nicht richtig war. Meine Mutter hat mir aber gesagt, dass die Schulpsychologin nicht Recht hätte, wir könnten aber nichts dagegen machen. Obwohl ich große Angst hatte, noch schlechter zu werden, habe ich bei dem dritten Test mitgemacht; ich war auch wieder besser, aber meine Mutter war dennoch wieder traurig danach. Als sie das Ergebnis der Schulpsychologin geschickt hat, hat die zu ihr gesagt, dass der Test nicht stimmen könnte, die Frau in der Erziehungsberatung hätte auch nicht richtig getestet. Meine Mutter hat sich sehr aufgeregt. Ich bin dann in der Schule sehr schlecht geworden und war richtig wütend auf die anderen Kinder, weil sie mich fertig machen, und auf meine Lehrer, weil sie uns nicht verstehen wollen. Meine Mutter hat dann nicht verhindert, dass ich in die Psychiatrie musste und in die Hauptschule runtergestuft wurde. Sie schreit oft oder weint, weil ich so ein Versager bin.

Dieser Verlauf steht in Gegensatz zu der Erwartung, dass dieser Junge von der Mehrfachtestung eigentlich hätte profitieren müssen (FEGGER, 1988, 110). Albert Ziegler (2004), ein Hochbegabtenforscher aus der Schule von Kurt Heller, vergleicht die Testerfahrung mit der beim Einfädeln eines Fadens – am Anfang mühsam, dann geht es besser – und führt dann aus: „Wie dramatisch diese Übungseffekte sein können, zeigen beispielsweise die Ergebnisse von Längsschnittstudien, in denen der durchschnittliche Intelligenzquotient von der ersten zur letzten Testung nicht selten um mehr als 15 Punkte anstieg. Von der „normalen“ Intelligenz unterscheidet man deshalb auch eine so genannte „Testintelligenz“. - Kurzum: Die Stabilität einer mittels eines Intelligenztests gemessenen Hochbegabung wird in späteren Testungen künstlich überschätzt, weil sich durch jede Intelligenztestung die Testintelligenz erhöht. Durch die hinzukommende Testintelligenz wird man seinen IQ-Vorsprung oder seinen Leis-

tungsvorsprung leichter verteidigen können (sic! Eckerle), oft genug kann man dadurch sogar noch besser abschneiden“ (S. 7).

Ziegler beschreibt also eine Testbiografie, die vom Lernen im Fach Testen bestimmt ist. - Arnes Testbiografie dagegen ist vom Lernen im Fach „Bedingungen für die Zuneigung der Mutter“ bestimmt.

Die Wiederholungstests in unseren Akten steigen nicht im Ergebnis, sondern fallen. Und zwar nicht nur um 15 Punkte, sondern um bis zu 45 Punkte, und das oft bei Testleitern, deren Erfahrung außer Zweifel steht.

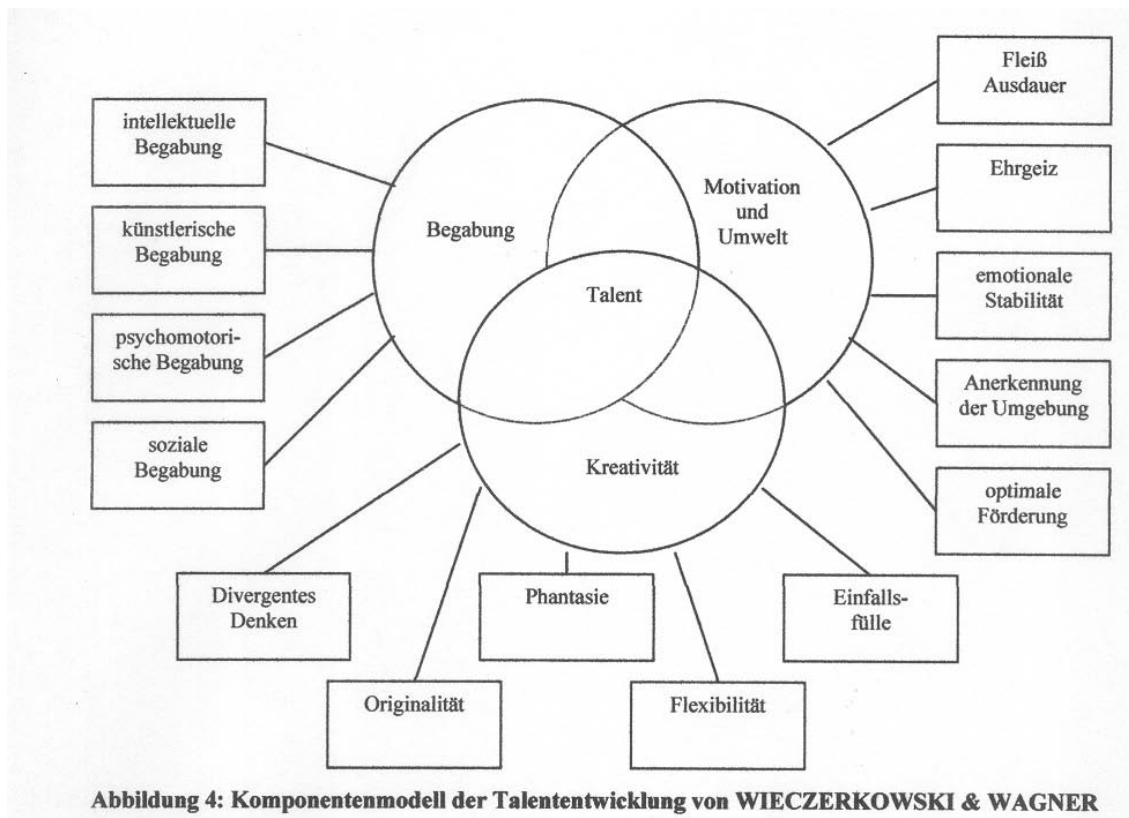
Warum wird nachgetestet?

1. Hypothese: Testergebnisse sind nicht stabil. Wenn ein Test länger zurückliegt, sollte nachgetestet werden.

Vorweg muss klargestellt werden, dass es bei Fragen der Teststabilität nicht um das Stagnieren der individuellen Intelligenz geht, diese entwickelt sich vielmehr im Verlauf; statt dessen geht es um die Stabilität der relativen Position der Probanden in der Population, die dem verwendeten Tests zugrunde liegt. Wenn Messergebnisse stabil sind, heißt das also, dass die Rangplätze der individuellen Entwicklungen ähnlich bleiben.

Greift man in den Fundus von Ergebnissen, die aus der Wissenschaft berichtet werden, dann gerät man bei der Frage nach der Teststabilität in Verwirrung.

Für Ziegler ist Intelligenz die Disposition eines Kindes, später in einem Gebiet herausragende Leistungen zu erbringen. Diese sind nicht allein abhängig von kognitivem Funktionieren, sondern von einem Zusammenspiel fördernder Faktoren. Es geht also um das, was man mit einem Kind entwickelt, nicht um das, was es genetisch hat: „Begabungen sind kein Schicksal, das uns in die Wiege gelegt wird in Form goldener Chromosomen. Begabungen können und müssen aufgebaut werden“ (2004, S. 8).



Aus Kemmer, 2006, S. 24

Aus dieser Sicht müssen Testungen etwas Anderes messen als nur die fluide Intelligenz, sie müssen die Leistung erfassen. Und diese nimmt im guten Fall zu, bei abbrechender Förderung ab; in dem Ergebnis drücken sich mittelbar die biografischen Bedingungen des Aufwachsens in ganzer Breite aus.

Um solchen Grundproblemen aus dem Weg zu gehen, haben sich etwa der Hawiva (Vorschultest) und der HAWIK IV verstärkt auf die genetisch angelegte und von Förderung weniger abhängige sogenannte fluide Intelligenz konzentriert, der Kaufmann ABC erfasst beides, die kristalline und die fluide Intelligenz, getrennt. Mit der Konzentration auf die fluide Intelligenz nimmt die Vorhersage-Genauigkeit der Tests für den IQ im späteren Schulalter und der Adoleszenz zu.

In der Forschung gut gesicherte Ergebnisse weisen darauf hin, dass unter Bedingungen einer stabilen familiären Umgebung die Einflüsse von außen auf die genetisch gegebene Intelligenz relativ gering sind, deren Stabilität also in Abhängigkeit von der sozialen (nicht der Anregungs-) Situation relativ hoch ist, bei ernstesten Deprivationen aber erhebliche Einflüsse auf die gemessene Intelligenz eintreten können, die Stabilität des Messergebnisses also relativ niedrig ist (z.B.

Greenberg, Coie, Lengua und Pinderhughes (1999) in Janke S. 71). Im früheren HAWIK III drückte sich die fluide Intelligenz vor allem im Handlungsteil aus. Die Diskrepanz des (niedrigen)Handlungsteils gegenüber dem (sehr hohen) Verbalteil, der besonders die kristalline Intelligenz gemessen hat, war geradezu ein Indikator für Kinder mit hoher Intelligenz und Störungsentwicklung (oft fälschlich mit Underachievement gleichgesetzt). Die fluide Intelligenz ist die aktuelle Leistung des kognitiven Handelns, nicht Wiedergabe von früher Gelerntem; daher ist sie von situativen Faktoren des Kindes besonders abhängig.

Die Testergebnisse der fluiden Intelligenz werden also durch Stress beeinflusst. Einen Literaturüberblick zu dem Zusammenhang von Bindungsqualität und kognitiver Entwicklung gibt Gärter, 2004, S. 39 ff. - Auf neurobiologischer Ebene wird über den Zusammenhang von Stress und (u.s.) kognitiver Entwicklung in Magdeburg, vor allem von Prof. Braun, geforscht. (vgl. dazu den Aufsatz „Neurobiologische Forschungsergebnisse über den Zusammenhang zwischen Hochbegabung und psychischen Störungen (z.B. ADS) in der Adoleszenz“, unter www.igl-net.de/aktuelles).

Bedenkenswert ist die Beobachtung der Testautoren des HAWIK IV, wonach alle Hochbegabten im Vergleich zu ihren übrigen Ergebnissen in den Untertests Arbeitsgedächtnis und Verarbeitungsgeschwindigkeit die schwächsten Leistungen aufweisen. Diese Beobachtung scheint erwartungswidrig. Die Hochbegabten sollten in diesen zentralen Fähigkeiten sehr gute Werte haben (R.G. Schmid u.a., 2008). Das Ergebnis zeigt, dass die Gruppe der Hochbegabten eine Schwäche teilt, die die Testergebnisse statistisch mindert. Obwohl eine entsprechende Deutung zur Zeit noch voreilig wäre, weise ich darauf hin, dass diese Beobachtung plausibel wird, wenn die Forschungen von Shaw, National Institute of Mental Health, US, hinzugezogen werden (a.a.O.).

Ich gebe gleich einen Überblick über Beobachtungen zur Teststabilität. Zuvor der Hinweis, dass aufgrund der verschiedenen Intelligenz- bzw. Begabungskonzepte und aufgrund der Unkenntnis darüber, was eigentlich in mitgeteilten Nachtestungen getestet wurde, solche Überblicke eigentlich nur geeignet sind zu erkennen, dass das Forschungsfeld unaufgeräumt ist. Eine Meta-Analyse, die die berichteten Ergebnisse vor dem Hintergrund der Untersuchungsmethoden vergleichen würde, wäre sehr erwünscht. Die folgenden Zahlen sind zwar nicht unmittelbar miteinander vergleichbar, sie geben aber doch eine Auskunft darüber, dass die Vorhersagefähigkeit von IQ-Tests ab dem Schulalter nennenswert ist und sich mit zunehmendem Alter stabilisiert. Wenn im Grundschulalter untersucht wird - später natürlich auch, aber vor allem hier - sollten daher immer zugleich eine gründliche Anamnese und begleitend eine Exploration der Persönlichkeits- und Bindungsentwicklung vorgenommen werden, um die Aussagefähigkeit eines Intelligenztests zu schätzen und für eine Beratung in der auslösenden Problemsituation verlässliche Grundlagen zu haben.

Aussagen über die Teststabilität

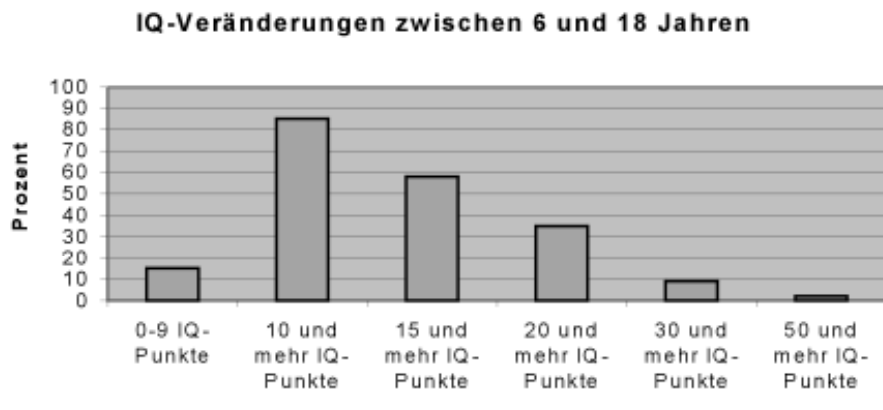
Bei Testwiederholungen stabilisieren sich nach Humphreys & Davey, 1988, und nach Wilson, 1987, (zit. bei Eric Turkheimer and Irving I. Gottesman. 1991., S. 20) die Korrelationen zwischen den Tests; dies ist konsistent mit der Beobachtung, dass mit zunehmendem Alter die Teststabilität steigt.

Tabelle 3.1: Übersicht - Langzeitstudien zur Intelligenzstabilität

Quelle	Alter bei Untersuchung (t_1 , t_2)	Zeitspanne zwischen den Messzeitpunkten	r
Linver, Brooks-Gunn & Kohen (2002)	drei Jahre, fünf Jahre	zwei Jahre	.77
Heller, Baker, Henker & Hinshaw (1996)	letztes Kindergartenjahr, 1.Klasse	ein Jahr	.64
Wilson (1983)	ab fünf Jahren jährlich	ein Jahr	.70
(1986)	ab Jugendalter jährlich	ein Jahr	.90
Mortensen, Andresen, Kruuse, Sanders & Reinisch (2003)	Kindesalter, junges Erwachsenenalter	14 Jahre	.85
Deary, Whiteman, Starr, Whalley & Fox (2004)	11 Jahre, 80 Jahre	69 Jahre	.66
Plomin, Pedersen, Lichtenstein & McClearn (1994)	64 Jahre, 67 Jahre	3 Jahre	.93

Zusammenstellung von Janke, 2008, S. 32

In dieser Zusammenstellung sehen wir die erwähnten Zahlen von Wilson, die die Wiederholungstests betreffen, neben der Angabe von Mortensen et. al., die eine Spanne von 14 Jahren betrifft. Wenn wir annehmen, dass bei Mortensen et al. die Ersttestung mit 8 Jahren stattfand und die spätere mit 22 Jahren, dann kann eine Nähe zu den Angaben von Wilson erkannt werden. Bei der Zeitspanne von 69 Jahren sollten Bedenken bestehen, ob die Messinstrumente in dieser Distanz das Gleiche gemessen haben. Die Angaben von Linver et. al. liegen etwas optimistischer als von Wilson.



Honzik-Studie, berichtet von Ziegler, 2004, S. 7

Ziegler zitiert eine frühe Untersuchung, die Honzik-Studie, und empfindet als Argument für geringe Teststabilität, dass danach die Messwerte von 58 % der Probanden zwischen 6 und 18 Jahren um 15 IQ-Punkte schwanken (2004, S. 7). Meines Erachtens liegt dieser Befund eher in einer ermutigenden Höhe, insbesondere wenn man sich in der gleichen Grafik die Angabe klar macht, dass bei etwa 85 % eine Verschiebung um nur 10 Punkte im gleichen Zeitraum stattfindet. Das hieße nichts Anderes, als dass eine Testung mit 6 Jahren um kaum mehr als im statistischen Fehlerbereich ohnehin anzugeben, variierte. Wichtig ist auch, dass bei diesen Angaben ja nicht eine Richtung enthalten ist, sondern in den Einzelfällen der zuerst gefundene Wert mal über-, mal unterschritten wird.

2. Hypothese: Testwerte sind nicht aussagekräftig (objektiv/valide/reliabel)

Jetzt geht es nicht mehr um das Verhältnis von Testwerten zu verschiedenen Zeitpunkten, sondern um die Aussagekraft eines einzelnen Testergebnisses.

Preusche (2006) stößt in ihrer Untersuchung zur Testfairness an Wiener Grundschulen unerwartet auf einen allgemeinen Effekt, den sie in ihrer Vorgehensweise nicht berücksichtigt hat: das allgemeine Leistungsniveau der einzelnen Schule, in der ein Schüler arbeitet. Dieser Faktor wirke sich auf nahezu alle Fragen der Untersuchung aus. Sehr plausibel scheint ihre Überlegung, dass sich in diesem Faktor Effekte der sozialen Schicht ausdrücken könnten (Stadtteil, S. 157). Kurzberichte über verschiedene Untersuchungen zu diesem Thema gibt auch Janke, S. 71.

Die Einflüsse des Bildungsniveaus der Familie und von sprachlicher Benachteiligung aufgrund eines Migrationshintergrundes sind bekannt und hoch plausibel.

Ich gehe hier nicht näher darauf ein (zum Thema Migrationshintergrund vgl. z.B. Daseking et al., 2008).

Statt dessen gehe ich auf die Bedingungen ein, die dem common sense etwas ferner liegen. Eine Vorverständigung hierzu in der Formulierung von Janke, S. 70, die mit äußerster Kürze die Logik des Misslingens benennt, die für hochbegabte Problemkinder – und natürlich für alle anderen auch – bestehen kann :

„Sie (Masten et al., 2006)gehen davon aus, dass bestimmte Ereignisse im Leben eines Individuums die erfolgreiche Bewältigung der, für das Alter oder den jeweiligen Lebensabschnitt spezifischen, Entwicklungsaufgaben beeinträchtigen können. Hierzu zählen

1) die Person direkt betreffende Einflüsse wie Dysfunktionen des Organismus, körperliche oder psychische Erkrankungen beziehungsweise deren Symptome,

2) kritische Lebensereignisse und

3) soziale Beeinträchtigungen durch andere Personen.

Alle Einflussfaktoren, die die Handlungsmöglichkeiten im Umgang mit den Entwicklungsaufgaben einschränken, können vor allem dann zu Schwierigkeiten führen, wenn die Aufgaben das Erlernen von für die Zukunft entscheidenden Fähigkeiten, Wissen und Ressourcen widerspiegeln.

Das Scheitern bei der Bewältigung der Entwicklungsaufgaben kann auf unterschiedliche Weise zu Psychopathologien beitragen. Es kann zunächst direkt zu Emotionen wie Trauer, Verzweiflung oder Wut führen. Darüber hinaus kann es das Risiko erhöhen, negative Erfahrungen mit Peers, Lehrern oder Eltern zu machen. Dies könnte negative Anpassungen verstärken oder zu einem Kontextwechsel beitragen, der beispielsweise Stigmatisierungen zur Folge hat .“

Ein Beispiel dazu:

Edgar Friederichs und Arnfried Heine, 2000

Das Profil ein und desselben Kindes zu verschiedenen Zeiten:		
	HAWIVA (5 Jahre)	HAWIK-R (7 Jahre)
Verbal-IQ	127	111
Handl.-IQ	96	87

Dieses Kind zeigt eine physiologische Einschränkung bei der visuellen Wahrnehmung. Es hat die langfristige und nicht überwindbare Diskrepanz zwischen Potential und Erbringung als frustrierend erlebt und resigniert. Unter diesen Bedingungen ist die Frage nach der Stabilität des IQ sinnlos, stattdessen die Suche nach Artefakten und Störfaktoren geboten. Zunächst misst der HAWIK-R noch stärker leistungsabhängig als der HAWIVA, daher sind bei Störungsent-

wicklungen Veränderungen nach unten zu erwarten. In dem zu beiden Zeitpunkten niedrigen Handlungs-IQ werden die physiologische Einschränkung, die fehlende Aufgabenzuwendung und mit hoher Wahrscheinlichkeit eine geringe Motivation mit gemessen.

Testleiter halten in solchen Fällen im Beobachtungsprotokoll fest, welche Validitätseinschränkungen sie festgestellt haben. Gegebenenfalls sprechen sie eine Empfehlung für die Schätzung des „wahren“ Wertes aus oder empfehlen, den Test als nicht aussagekräftig einzustufen. Dissertationen bei Kubinger, Holling und Mandl haben sich in den letzten Jahren mit Testfairness, Risiken der Stabilität von IQ-Werten und Persönlichkeitsmerkmalen von Hochbegabten befasst und ergänzen mit einer Vielzahl von Details die bereits langfristig bekannten Einflüsse.

z.B. Testprobanden mit psychischen Störungen

Edgar Friederichs und Arnfried Heine merken zu ihrem Probanden an, dass solche Konstellationen sich vielfach bei hochbegabten Problemkindern fänden. Beides wirke sich auf die Arbeitsgeschwindigkeit und das Arbeitsgedächtnis aus. - Die Autoren des HAWIK IV weisen darauf hin, dass bei ADHS, aber auch bei Depression die Arbeitsgeschwindigkeit und das Arbeitsgedächtnis schwächer seien (R.G. Schmid u.a., 2008). Zu gleichen Ergebnissen kommen Rost und Schermer: Wenn es nicht gelinge, Potential in Performanz umzusetzen, könne Leistungsängstlichkeit entstehen (2001, S. 411). Misserfolge werden auf persönliches Unvermögen zurückgeführt, wohingegen Erfolg dem Zufall zugeschrieben wird. Mit sinkendem Selbstwertgefühl sinkt auch die Erfolgswahrscheinlichkeit, was wiederum zu nicht adäquater Leistung führt. Diese Ausführung spitzt sich zu in der Feststellung, dass „Hochhängstliche in fast allen Schulfächern weniger leisten als emotional stabile Schüler“ (ebda, S. 410) - und natürlich auch im Text.

Calhoun und Mayes gehen davon aus, dass der Index Verarbeitungsgeschwindigkeit des HAWIK-IV durch die strukturellen Veränderungen deutlich sensitiver gegenüber Aufmerksamkeitsstörungen ist als der entsprechende Index des HAWIK-III. Ähnliches wird für den Index Arbeitsgedächtnis des HAWIK-IV im Vergleich zum Index Unablenkbarkeit des HAWIK-III angenommen. Die Autoren des HAWIK IV führen näher aus: „Bei der Interpretation der Ergebnisse ist auch zu berücksichtigen, dass die Aufmerksamkeitsstörung an sich die Testdurchführung und die Testleistung beeinflussen kann. So kann beispielsweise der impulsive Arbeitsstil zu einer höheren Fehlerzahl führen. Andererseits können sich

aber auch eine unstrukturierte Testsituation oder Umgebung oder die fehlende Routine in der Testdurchführung negativ auf das Arbeitsverhalten des Kindes auswirken. Daher stellen Verhaltensbeobachtung und Dokumentation der Testsituation wichtige Momente in der Interpretation des Ergebnisses dar“ (R.G. Schmid1 u.a. 2008).

z.B. Defensive Einstellung des Testprobanden in der Testsituation

Oben (S. 7) habe ich die Beobachtungen der Testautoren des HAWIK IV berichtet, wonach hochbegabte Kinder statistisch in den Skalen Arbeitsgeschwindigkeit und Arbeitsgedächtnis die – relativ zu ihren Stärken – schwächsten Ergebnisse erzielen. Diese Beobachtung schreibt der Zielgruppe also bereits ein Ausgangsrisiko zu, ohne dass eine besondere Ursache zugeordnet wird.

Bei jeder Testwiederholung entstehen Erwartungen der Kindeseltern und des Kindes, die den Test beeinflussen. Davon war in dem Eingangsbeispiel die Rede. Der erneute Test entscheidet vielleicht über die Schullaufbahn, im Erleben des Kindes über die Zufriedenheit der Mutter und über die Gültigkeit der Erklärung, weshalb ihm das Verhalten in der Schule oft nicht gelingt (Identität). Der Test entscheidet auch darüber, ob z.B. die Klassenlehrerin Recht hatte, als sie der Mutter sagte, dass sie keineswegs Anzeichen für Hochbegabung feststellen könne; im Gegenteil, das Kind zeige schlechte Leistungen. Oder ob der Schulpsychologe, der der Mutter vorgehalten hat, dass ihr Beharren auf Hochbegabung Schuld daran habe, dass es ihrem Kind in der Gruppe und bei den Lehrkräften so schlecht gehe. In der Erwartung der Mutter entsteht ein erheblicher Druck, der mittelbar auf das Kind einwirkt, das eben diesem Schulpsychologen im Test gegenüber sitzt.

Der Test wird alle diese Risiken als Resultante im Erleben des Kindes mit messen, ohne sie auszuweisen und ohne dass ein noch so redlicher Testleiter sie einschätzen könnte. Das ist der Grund dafür, dass Dörner Tests unter Defensivbedingungen für unzulässig erklärt (2004) und die Handbucharweisungen der Tests sie sinngemäß ausschließen. Im Handbuch zum Beispiel des HAWIK IV heißt es in den Anweisungen zur Testdurchführung: „Es (das Kind) darf außerdem nicht den Eindruck gewinnen, dass es unter Druck gesetzt wird.“ In der ein oder anderen Formulierung steht diese Anweisung in jedem Testhandbuch.

Diese im Handbuch gestellte Bedingung ist bei entwickelter Störung nicht mit einem Standardverfahren und den zugehörigen Verhaltensanweisungen, und nicht bei einer Vermengung von Gegnerschaft/Macht und Testdurchführung einzulösen; dies um so mehr, wenn man zwei weitere Forschungsergebnisse

hinzu denkt, die immer wieder berichtet werden und aus der Erfahrung mit den Schülern unserer Schule für verhaltensgestörte Hochbegabte geradezu augenfällig sind (beide berichtet bei Kemmer, S. 15).

Terman & Oden fassen in einer Langzeitstudie zur Hochbegabung, bei der es u.a. auch um Versagen bei der Identifikation von Höchstbegabten ging, vier Merkmale für (von ihm so genannte) Underachiever zusammen:

- Fehlende Ausdauer
- Mangelnde Zielorientierung
- Unterlegenheitsempfinden
- Fehlendes Selbstvertrauen

Mandel & Marcus finden in einer Reihe von Untersuchungen ebenfalls ein negatives Selbstkonzept (1988, S. 16-18). Darüber hinaus hätten die Kinder eine negative Einstellung zu Lehrern und seien innerhalb der Familie häufiger mit Konflikten belastet.

Testwerte für Kinder in dieser Situation sind daher nicht valide. Es wird nicht gemessen, was zu messen vorgegeben wird. In solchen Fällen sollte der erste Test geprüft werden. Ist er am Anfang der Störungsentwicklung und nach Beginn des Schulalters von einem erfahrenen Psychologen durchgeführt, stellt er die beste Schätzung dar, die erreichbar ist, auch wenn die oben angeführten Einschränkungen der Teststabilität hier zu berücksichtigen sind. Mindestens aber sollte ein Wiederholungstest außerhalb des Machtbereichs der Schulbehörden durchgeführt werden, um wenigstens den gravierendsten Defensivfaktor heraus zu nehmen.

Mit einer anderen Logik ist auch die Regel zu erstellen, dass das höchste Testergebnis heranzuziehen sei. Da ja das Potential und nicht die Performanz geschätzt werden soll, kann keine Überschätzung in einem ordnungsgemäß durchgeführten Test erfolgen. Nachfolgende niedrigere Testergebnisse sind daher als Artefakte der Testsituation aufzufassen. - Da in der Regel der erste Test auch der höchste ist und das Problem daraus entsteht, dass das Ergebnis nach Beginn der Störungsentwicklung bezweifelt wird, fallen beide Argumentationen in der Regel in der Entscheidung für den gleichen Test zusammen.

Testleitereffekte

Die Diplomarbeit von Ingrid Preusche (2009) wird auf der Homepage des Institut für Entwicklungspsychologie und Psychologische Diagnostik (Klaus Kubinger) zusammengefasst; in dem Text heißt es: Erst durch die besondere

Beobachtungsmöglichkeit bzw. durch eine Diagnostik der Interaktion der Testperson mit dem Testleiter können relevante Informationen erfasst werden. Allerdings birgt die intensive Interaktion von Tp und Tl die Gefahr systematischer Testleitereffekte. Und diese werden in der Fachliteratur nicht genauer erforscht, wohl aber kritisch reklamiert (o.J.).

Frau Preusche geht der Frage im Rahmen der übergreifenden Themenstellung Testfairness genauer nach (Vergleich des AID mit dem HAWIK III). Allerdings führten in ihrer Studie eigens geschulte Testleiter die Tests durch, deren Ergebnisse sie dann untersuchte. Die Testleitereffekte ergeben sich aus Merkmalen und Verhaltensweisen wie Alter, Geschlecht, Sprachkultur, Dominanzstreben Bewusste Fälschungen schloss sie aus. In ihrer Auswertung erwies sich insbesondere das Merkmal Geschlecht als breit wirksam, ohne dass insgesamt eine klare Richtung anzugeben wäre. Bei den eher leistungsstarken Kindern fand sie, dass weibliche Tl bei ihren Tpn bessere Ergebnisse in den Untertests Rechnerisches Denken bzw. Angewandtes Rechnen erzielten. Zusammenfassend kommt sie zu der Aussage: Es „zeigte sich, dass lediglich bei eher leistungsstärkeren Kindern Testleitereffekte auffielen. Ein Beobachterbias seitens der Testleiter gegenüber diesen Kindern scheint zumindest möglich. – Generell scheint es keine Untertests zu geben, die überdauernde Testleitereffekte aufweisen“ (170).

In den Studien zum HAWIK IV ist allerdings in ganz anderer Weise von Fehlern der Testleiter die Rede.

Lipsius kommt in ihrer Dissertation (bei Petermann) 2008 zu der Feststellung, dass bei umfangreichen Testbatterien wie dem HAWIK III und IV Fehler des Testleiters in der Durchführung und Auswertung des Tests bisweilen zu gravierenden Abweichungen zwischen Testergebnis und wahren Wert führten.

Um dem mangelhaften Umgang mit Testverfahren entgegenzuwirken, haben in der Vergangenheit einige Organisationen Testleitlinien und -standards entwickelt. (Hacker, Leutner & Amelang, 1998; Moosbrugger & Hofling, 2006); die Standards für pädagogisches und psychologisches Testen, die von der AERA (American Educational Research Association), der APA (American Psychological Association) und der NCME (National Council on Measurement in Education) entwickelt wurden (Standards for educational and psychological testing, AERA, APA & NCME, 1999); die Richtlinien der International Test Commission (International guidelines for test use, ITC, 2000), in deutscher Übersetzung vom BDP (Berufsverband Deutscher Psychologinnen und Psychologen, 2001).

„Auch die Interpretationsobjektivität ist bei Intelligenztests nicht immer gegeben. Sie betrifft den Grad der Eindeutigkeit, mit der gleichen Werten (also in

diesem Fall Testergebnissen) auch gleiche Merkmalsausprägungen (also hier kognitive Fähigkeiten) zugeordnet werden (Fisseni, 2004, zit. bei Lipsius, S. 52).

Eindeutige Kritik wird an der ungenügenden Praxis geübt, dass kompetente Testleiter die Testungen an nicht kompetente delegieren: „Hall et al. (2005) stellen klar, dass ein qualifizierter Testleiter, der die Tests von unausgebildeten Helfern durchführen lässt, damit die Testsicherheit gefährdet, und fordern, dass diese Testleiter die Testpersonen über diese Vorgehensweise informieren. Sie gehen davon aus, dass die Qualität der Informationen, die man von einem solchen Test bekommt, sehr gering ist“ (zit. bei Lipsius, Petermann und Daseking, 2008, S. 108). Gerade aber die Wechsler-Skalen scheinen besonders häufig von untrainierten Testleitern durchgeführt zu werden (ebda.).

Die Durchführungs-Checkliste des HAWIK umfasst mehrere Hundert Verhaltensweisen oder Fähigkeiten, die benötigt werden, um den Test erfolgreich durchführen und auswerten zu können (Sattler, 2001, zit. ebda.). Die von Kubinger genannten Neben- oder Metabeobachtungen während des Tests sind von wenig routinierten und in der Sache nicht urteilsfähigen Testleitern nicht zu erwarten. Neben den Anforderungen des Tests bleibt keine Aufmerksamkeit für die so wichtige zusätzliche Diagnostik.

Die Herausgeber des HAWIK IV haben durch redaktionelle Maßnahmen, z.B. verkürzte Anweisungen auf den je relevanten Protokollbögen, versucht, Hilfen zu geben. Insgesamt belegen weitere Forschungen aber eine sehr hohe Fehlerrate, wobei die sprachlichen Untertests besonders anfällig waren. Bei mehr als der Hälfte der Bögen führte die Korrektur der Fehler zu einer Veränderung des Gesamt-IQ. Bei 17% führten die Korrekturen sogar zu Gesamt-IQ-Änderungen, die eine abweichende Einordnung in der Klassifikation der kognitiven Funktionen zur Folge hatten. In einer Studie von Slate et al. (1992), an der qualifizierte Testanwender teilnahmen, führten die Fehler zum Teil zu Änderungen im Gesamt-IQ, die Diagnose- und Platzierungsentscheidungen beeinflusst hätten.

Was heißt das alles in der Praxis?

- Testleiter müssen zum Testen qualifiziert sein. Um in einen komplexen Test zu gehen, sind Erfahrung und fortlaufende Selbstkontrolle erforderlich. Die oft anzutreffende Gewohnheit, die Tests von unausgebildeten Helfern – Sprechstundenhilfen, Studenten, jungen PsychologInnen im Berufseinstieg – durchführen zu lassen, gefährdet die Testsicherheit.

- Die verwendeten Tests müssen den neuesten Testnormen entsprechen.
- Dem Test sollte eine mindestens halbstündige Begegnung von Testleiter und Testprobanden vorausgehen, die vom Testleiter dazu genutzt wird, um das Vertrauen des Kindes und der Mutter/der Eltern zu gewinnen.
- Der Testleiter sollte in der Lage sein, sich - trotz der hohen Anforderung - freundlich und gelassen auf die Kommunikation mit dem Kind zu konzentrieren. Während des Tests sollte er über das Testgeschehen hinaus das Verhalten des Kindes analysieren können, um weitere diagnostische Beobachtungen vorzunehmen.
- Wer an einer Konfliktsituation teilhat, die im Vorfeld des Tests entstanden ist, sollte nicht testen. Entscheidungsbefugnis (vor allem über schulische Maßnahmen in Abhängigkeit vom Testergebnis) und Testleitung sollten immer in einer Weise getrennt bleiben, die Defensivtests verhindert.

Nach unseren Akten und Erfahrungen in der Ombuds-Stelle und der Schule bestehen für verhaltensschwierige Kinder und Jugendliche mit hoher Begabung mehrere Risiken, von denen ich zwei herausgreifen will:

1. Die in Psychiatrien durchgeführten Tests haben häufig den Charakter von Defensivsituationen; die Kinder befinden sich in einer inferioren Position, in der sie ihre Möglichkeiten oft nicht nutzen können. Zu der Ausgangsproblematik, die sie in die Klinik geführt hat, tritt die Belastung der Testsituation hinzu. Es handelt sich dabei um Strukturen der Situation, die meist nicht dadurch kompensiert werden können, dass der Testleiter/die Testleiterin besonders sensibel reagiert. Wenn es zu einer Unterschätzung des Potentials gekommen ist, wiegt das im Verlauf schwer, weil die noch immer weithin empfundene ärztliche Autorität sich gegenüber fachlichen Zweifeln oft durchsetzen kann. Das gravierendste Fehlurteil mit anschließender Fehlleitung des Kindes, das ich in meiner Erfahrung erlebt habe, wurde mit einem Wiederholungstest einer kinder- und jugendpsychiatrischen Klinik herbeigeführt, der um 40 IQ-Punkte von dem Test eines anderen Universitätsinstituts abwich und Grundlage für die weitere Versorgung des Kindes wurde. Beide Einrichtungen haben zweifellos das Testmaterial ordnungsgemäß geboten, aber sie haben die Effekte der Testsituation nicht angemessen berücksichtigt.

2. Die von Schulpsychologen durchgeführten Tests stehen tendenziell in ähnlichem Risiko, wenn es aus der Sicht des Kindes um erwünschte Entscheidungen geht, insbesondere wenn zwischen Schule und Eltern oder Kind und Schule ein

konflikthaftes Verhältnis besteht. In unseren Akten sind die Differenzen bei schulpsychologischen Tests zwar nicht ganz so auffallend wie die aus psychiatrischer Umgebung, bei Kindern in Störungsentwicklung liegen die Testergebnisse aber in der Regel niedriger als bei Testungen von niedergelassenen Therapeuten, Ärzten und Psychologen.

SchulpsychologInnen sind Vertreter der Schulbehörde und bereiten schulfachliches Handeln der Schulaufsichtsbehörde und der Schulen vor. In vielen Bundesländern, nicht in Hessen, wird auf die amtliche Zuständigkeit der SchulpsychologInnen verwiesen und im gleichen Schritt für ihre Testdurchführung eine erhöhte Validität und Objektivität in Anspruch genommen. Dieser Anspruch steht fachlich in Gegensatz zu der Forderung, Defensivtests zu vermeiden. – Darüber hinaus schreiben sich die schulpsychologischen Dienste selbst Interessenfreiheit zu, die vermeintlich zu gültigen Messergebnissen beitrage. Mit unserer Erfahrung ist statt dessen vereinbar, dass es unter den SchulpsychologInnen wie unter allen anderen Gruppen von Menschen viele zugewandte und nach bestem Wissen arbeitende gibt, und dann die Anderen, die ihre Zuständigkeit ausgrenzend handhaben, anstatt zu Gunsten des Kindes in eine Helferumgebung hinein zu gehen und mit freien Trägern und niedergelassenen Professionellen zusammen zu arbeiten.

Im Kontext von Mehrfachtestungen immer schädlich ist die Forderung, dass Tests von nicht verbeamteten PsychologInnen einer Nachprüfung bedürften. Nicht nur ist ethisch nicht zu rechtfertigen, die amtliche Macht zur Diskriminierung von Kolleginnen und Kollegen zu missbrauchen; auch ganz konkret schadet diese Anmaßung eben denen, zu deren Gunsten sie scheinbar vorgetragen wird: den Schülerinnen und Schülern, deren Krise überwunden werden soll. Oft kommt es zu Reaktion und Gegenreaktion in mehrfachen Episoden, manchmal artet die Auseinandersetzung in regelrechte Kleinkriege aus. Eltern, die das Pech haben, mit Problemkindern in einem solchen Bezirk zu wohnen, können einsam werden.

Ein aktuelles Beispiel dazu: Vor wenigen Wochen wurde mir ein Fall in der Ombuds-Stelle vorgestellt, in dem innerhalb eines Jahres vier Nachtestungen durchgeführt wurden, drei davon innerhalb eines halben Jahres - im Wechsel von einer Pädagogischen Praxis, Schulbehörde, Sozialpädagogischem Beratungszentrum, Schulbehörde. Der Testproband beschwerte sich darüber, dass es immer die gleichen Tests gewesen seien.

In Fortbildungen der Begabungsdiagnostischen Beratungsstelle in Marburg wird SchulleiterInnen ausdrücklich zu Nachttestungen geraten, wenn nicht eine amtliche Stelle getestet hat. Das Hessische Kultusministerium dagegen lehnt diese Praxis ab. In den meisten anderen Bundesländern werden Nachttestungen gefordert, wenn auch nicht immer durchgesetzt.

Zusammenfassend:

Die Risiken von Mehrfachtestungen liegen neben der Testkompetenz der Testleiter vor allem in der Testsituation, und zwar immer dann, wenn die in den Testhandbüchern geforderten Bedingungen für die Testperson nicht vorliegen: Wohlbefinden, Vertrauen und Frei sein von Druck. Die Risiken werden dann erheblich gesteigert, wenn in der Helferumgebung Frontlinien bestehen, die die Helfer davon abhalten, das Wohl des Kindes über alle anderen (Neben-) Ziele zu stellen.

Liegen mehrere Testergebnisse vor, dann sollte der höchste Test akzeptiert werden. Diese Regel ergibt sich aus der Logik des Testens, dass das Potential, nicht die situationsabhängige Performanz gemessen werden soll. Demnach kann ein Kind, korrekte Testdarbietung vorausgesetzt, nicht zu hoch bewertet werden; es kann lediglich dazu kommen, dass es unter optimierten Bedingungen Leistungen erreicht, die es in seiner schulischen Umgebung nicht wiederholen kann. Gerade das aber ist meistens Anlass für einen Test, weil ja zu entscheiden ist, ob das Kind und die ihm gebotene Leistungsumgebung zueinander passen. Ein unerwartet hoher Wert lenkt dann die Aufmerksamkeit auf die Entwicklung der Lernbedingungen.

Die prognostische Validität der Testung im frühen Grundschulalter reicht aus, um die „Liga, in der ein Kind spielt“ für die Schullaufbahn vorherzusagen. In Abhängigkeit von der Biografie des Kindes kann es zu Problemkonstellationen kommen, die eine erneute Testung rechtfertigen; geht es um eine Entscheidung über schulische Maßnahmen, muss eine Defensivsituation ausgeschlossen werden; daher ist es empfehlenswert, dass nicht ein Mitglied der Schulbehörde den Wiederholungstest durchführt, sondern eine andere Einrichtung gewählt wird.

Eine Häufung von Wiederholungstests bei Kindern zeigt immer ein fachliches Versagen der anfordernden Stellen an.

Literatur

Calhoun, S. L. & Mayes, S. D., 2005. Processing speed in children with clinical disorders. *Psychology in the Schools*, 42, 333-343

- Daseking, M., Lipsius M., Petermann, F., Waldmann, H-Ch. (2008) Differenzen im Intelligenzprofil bei Kindern mit Migrationshintergrund: Befunde zum HAWIK-IV. In: PsyJOURNALS, Volume 17, Number 2 / 2008, 76 - 89
- Dörner, Dietrich, 2004. Der kognitive Kern der Neurowissenschaften. In Neuropsychologie psychischer Störungen, Springer
- Feger, Barbara (1980) „Identifikation von Hochbegabten“. In: Klauer, Karl Josef/Konradt, Hans-Joachim (Hg.) (1980): Jahrbuch für empirische Erziehungswissenschaft 1980. Düsseldorf: Pädagogischer Verlag Schwann, 87-112
- Fisseni, H.-J. (2004). Lehrbuch der psychologischen Diagnostik (3., überarbeitete u. erweiterte Aufl.). Göttingen: Hogrefe
- Friederichs, Edgar; Heine, Arnfried (2000). Intelligenz, Aufmerksamkeitsstörungen mit und ohne Hyperaktivität und Selbstregulationsdefiziten. In: Labyrinth 66, 2000
- Gaertner, Jonna (2004). Die Zusammenhänge zwischen Bindungstyp, Explorationsverhalten und ausgewählten Persönlichkeitsaspekten hochbegabter Kinder. Dissertation Ludwig-Maximilians-Universität München, Fakultät für Psychologie und Pädagogik. URN (NBN): urn:nbn:de:bvb:19-23050. erhältlich unter <http://edoc.ub.uni-muenchen.de/2305/>
- Hacker, H., Leutner, D. & Amelang, M. (Hrsg.). (1998). Standards für pädagogisches und psychologisches Testen. Bern: Huber
- Humphreys, L. G., & Davey, T. C. (1988). Continuity in intellectual growth from 12 months to 9 years. *Intelligence*, 12, 183-197
- Hunt, E. (1995). The role of intelligence in modern society. *American Scientist*.
<http://www.sigmaxi.org/amsci/articles/Hunt-full.html>
- Greenberg, M. T., Coie, J. D., Lengua, L. J. & Pinderhughes, E. E. (1999). Predicting developmental outcomes at school entry using a multiple-risk model: Four American communities. *Developmental Psychology*, 35, 403-417
- International Test Commission, Commission Internationale des Tests & Berufsverband Deutscher Psychologinnen und Psychologen. (2001). Internationale Richtlinien Literaturverzeichnis 207 für die Testanwendung. Version 2000. Deutsche Fassung [On-line]. Verfügbar unter:
http://ftp.zpid.de/pub/tests/itc_richtlinien.pdf [19.12.2008
- Janke, Nina (2008) Risikofaktoren der Höhe und Stabilität von kindlicher Intelligenz im Übergang zur Schule - eine entwicklungspsychologische Sichtweise.
<http://webdoc.sub.gwdg.de/ebook/dissts/Bremen/Janke2008.pdf>
- Kemmer, Michael (2006). Hochbegabungsdagnostik - Messmethoden und Informationsquellen unter besonderer Berücksichtigung der Identifikation von hochbegabten Unerachievten
http://books.google.de/books?id=u3tKhJX89LIC&pg=PA48&lpg=PA48&dq=Michael+Kemmer,+Hochbegabungsdagnostik&source=bl&ots=fD72IKcb0T&sig=bUUe863rVTCryWySRLn_rcqFHA8&hl=de&ei=tse9TbvMcPEswbhobCMBg&sa=X&oi=book_result&ct=result&resnum=6&ved=0CEgQ6AEwBQ#v=onepage&q&f=false

- Kubinger, Klaus (2009). Zusammenfassung zur Diplomarbeit Preusche.
<http://psychologie.univie.ac.at/eppd/studium/psychologische-diagnostik/diplomarbeiten/diplomarbeiten-seit-1998/kubinger/preusche/>
- Landscheidt, Karl, o.J., Begabung und Hochbegabung.
http://bsrwn.lsr-noe.gv.at/www_bsr/Web/Service/www.learn-line.nrw.de/angebote/schulberatung/main/foyer/down/landscheidt_begabung.pdf
- Lipsius, M., Petermann, F. & Daseking, M. (2008). Wie beeinflussen Testleiter die HAWIK-IV-Befunde? *Kindheit und Entwicklung*, 17(2), 107-117
- Mandel, Harvey P./Marcus, Sander I. (1988): *The Psychology of Underachievement. Differential Diagnosis and Differential Treatment*. New York: John Wiley & Sons, Inc.
- Masten, A. S., Burt, K. B. & Coatsworth, J. D. (2006). Competence and psychopathology in development. In D. Cicchetti & D. Cohen (Eds.), *Developmental Psychopathology* (2nd ed., Vol. 3, pp. 696-738). New Jersey: Wiley
- Moosbrugger, H. & Hofling, V. (2006). Testdurchführung und -auswertung. In F. Petermann & M. Eid (Hrsg.), *Handbuch der psychologischen Diagnostik* (S. 449-456). Göttingen: Hogrefe
- Preusche Ingrid (2007) *Fairness von Intelligenztests bei Kindern und Jugendlichen*. Saarbrücken: VDM Verlag
- Rost, Detlef H./Schermer, Franz J. (2001): „Leistungsängstlichkeit“. In: Rost, Detlef H. (Hg.) (2001): *Handwörterbuch Pädagogische Psychologie*. 2., überarb. und erw. Aufl. Weinheim: Psychologie Verlags Union, Verlagsgruppe Beltz
- Sattler, J.M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego: Author
- Schmid, R.G., Petermann, U., Petermann, F., Weiß, C., Daseking, M. (2008). Leistungen des HAWIK-IV in der Intelligenz-diagnostik im Schulalter. In: *Monatsschr Kinderheilkd* 2009 · 57:587–594 / <http://www.springerlink.com/content/c85028ql23657884/fulltext.pdf>
- Slate, J. R., Jones, C. H., Coulter, C. & Covert, T. L. (1992). Practitioners_ administration and scoring of the WISC-R: Evidence that we do err. *Journal of School Psychology*, 30, 77 – 82
- Terman, L.M. & Oden, M.H. (1959). *The gifted group at mid-life: 35 years' follow-up of the superior child*. Stanford, CA; Stanford University Press
- Turkheimer, Eric; Gottesman, Irving I. (1991). Individual Differences and the Canalization of Human Behavior. In: *Developmental Psychology* 1991, Vol. 27, No. 1, 18 - 228-22
- http://studyforquals.pbworks.com/f/Turkheimer%20Canalization_development.pdf
- Wilson, R. S. (1987). A new method of representing mental growth. *Acta Geneticae Medicae et Gemellologiae*, 36, 121-134
- Ziegler, Albert (2004). Stabilität von Intelligenz und Hochbegabung im Vorschulalter. In: *Landesverband Hochbegabung aktuell*, Nummer 13, S. 3 – 8 (http://www.lvh-bw.de/fileadmin/LVH/pdf/LVH-13_s3-8.pdf)