

Ernst Fay (Hrsg.)

1999

## TESTS UNTER DER LUPE II

Kersting, M. (1999). Intelligenz-Struktur-Test 2000 – Testrezension. In E. Fay (Hrsg.), *Tests unter der Lupe - Band 2* (S. 86-113). Lengerich: Pabst Science Publishers.



PABST SCIENCE PUBLISHERS  
Lengerich, Berlin, Düsseldorf, Leipzig,  
Riga, Scottsdale (USA), Wien, Zagreb

# INTELLIGENZ-STRUKTUR-TEST 2000 (IST 2000)<sup>1</sup>

Martin Kersting

## 1 TESTART

Intelligenztest

## 2 TESTMATERIAL

Die vollständige Mappe für den I-S-T 2000 Test umfasst die Handanweisung sowie je ein Testheft des Grund- und des Erweiterungsmoduls und je fünf Antwortbögen für die Formen A und B. Außerdem gehören fünf Rohwertprotokolle für das Erweiterungsmodul (getrennt nach Form A und B) sowie fünf Rohwertprotokolle für kristallisierte und fluide Intelligenz, fünf Vordrucke für Ergebnisprofile und je zwei Auswertungsschablonen für die beiden Formen des Grundmoduls zur Komplettausstattung des Tests. Alle Materialien können einzeln nachbestellt werden. Für den Test wird ein Auswertungsprogramm angeboten. Dieses Programm lag zum Zeitpunkt der Manuskripterstellung noch nicht vor und bleibt in der Rezension unberücksichtigt.

Bis auf wenige im folgenden spezifizierte Details weisen alle Materialien eine hohe Qualität und eine ansprechende Gestaltung auf, die Testhefte sind stabil geklammert. Lediglich bei der Aufgabengruppe 09 finden sich fünf Items mit "Matrizen", die teilweise sehr dünne und schlecht erkennbare Zeichnungen aufweisen. Dies stellt sicherlich einen wirksamen Kopierschutz dar, könnte einigen Testanden<sup>2</sup> aber leistungsbeeinträchtigende Wahrnehmungsprobleme

<sup>1</sup> Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000*. Göttingen: Hogrefe

<sup>2</sup> Personen und Funktionsbezeichnungen gelten jeweils in weiblicher und männlicher Form.

# INTELLIGENZ-STRUKTUR-TEST 2000 (IST 2000)<sup>1</sup>

Martin Kersting

## 1 TESTART

Intelligenztest

## 2 TESTMATERIAL

Die vollständige Mappe für den I-S-T 2000 Test umfasst die Handanweisung sowie je ein Testheft des Grund- und des Erweiterungsmoduls und je fünf Antwortbogen für die Formen A und B. Außerdem gehören fünf Rohwertprotokolle für das Erweiterungsmodul (getrennt nach Form A und B) sowie fünf Rohwertprotokolle für kristallisierte und fluide Intelligenz, fünf Vordrucke für Ergebnisprofile und je zwei Auswertungsschablonen für die beiden Formen des Grundmoduls zur Komplettausstattung des Tests. Alle Materialien können einzeln nachbestellt werden. Für den Test wird ein Auswertungsprogramm angeboten. Dieses Programm lag zum Zeitpunkt der Manuskripterstellung noch nicht vor und bleibt in der Rezension unberücksichtigt.

Bis auf wenige im folgenden spezifizierte Details weisen alle Materialien eine hohe Qualität und eine ansprechende Gestaltung auf, die Testhefte sind stabil geklammert. Lediglich bei der Aufgabengruppe 09 finden sich fünf Items mit "Matrizen", die teilweise sehr dünne und schlecht erkennbare Zeichnungen aufweisen. Dies stellt sicherlich einen wirksamen Kopierschutz dar, könnte einigen Testanden<sup>2</sup> aber leistungsbeeinträchtigende Wahrnehmungsprobleme

<sup>1</sup> Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000*. Göttingen: Hogrefe

<sup>2</sup> Personen und Funktionsbezeichnungen gelten jeweils in weiblicher und männlicher Form.

bereiten. Demgegenüber kann sich ein gutes Sehvermögen bei der Merkfähigkeitsaufgabe "Figurenpaare" auszahlen, da das im Aufgabenheft zwischen den einzuprägenden Figurenpaaren und der Abfrageseite geheftete Zwischenblatt trotz der aufwendigen Farbgestaltung keinen perfekten Sichtschutz gewährt und somit die einzuprägenden Figurenpaare bei der Abfrage "durchschimmern". Solange hier nicht verlagsseitig Abhilfe geschaffen wird, müssen Anwender unbedingt durch das zusätzliche Aufkleben eines weiteren Blatt Papiers das Funktionieren der Aufgabe gewährleisten.

Als ein unglückliches Detail der Gestaltung der Testmaterialien muss die im Testheft des Grundmoduls gewählte Anrede der Diagnostikenden als "Versuchsteilnehmerinnen" und "Versuchsteilnehmer" sowie die Bezeichnung "Versuchsleiter" gewertet werden, da der I-S-T in der Vergangenheit im diagnostischen Ernstfall angewendet wurde, wo gerade der Eindruck von "Experimenten" vermieden werden sollte.

## 3 TESTGLIEDERUNG

Der I-S-T 2000 beinhaltet zwei Module, mit denen sich drei Anwendungsformen realisieren lassen. Das Grundmodul umfasst je drei Aufgabengruppen mit jeweils 20 Items zur verbalen, numerischen und figuralen Intelligenz, die auf einer höheren Generalitätsebene als Indikator für schlussfolgerndes Denken ("Reasoning") berücksichtigt werden können. Ebenfalls im Grundmodul berücksichtigt sind zwei Aufgabengruppen zur verbalen (10 Items) und figuralen (13 Items) Merkfähigkeit, deren Einsatz zusätzlich zu den zuvor genannten neun Aufgabengruppen erfolgen kann. Das Erweiterungsmodul besteht aus einem Test mit 73 Fragen zur Erfassung des erworbenen Wissens (als wesentlicher Indikator für kristallisierte Intelligenz). Dabei geht es um Kenntnisse aus den Bereichen "Geographie/Geschichte", "Kunst/Kultur", "Mathematik", "Naturwissenschaften" und "Alltags-Symbole", wobei die Kenntnisfragen zu jeder Wissensdomäne mit verbalem, numerischem und figuralem Aufgabenmaterial dargeboten werden. Der Wissenstest dient aber ausschließlich der Operationalisierung der kristallisierten Intelligenz, eine separate Auswertung des Wissenstests und/oder der Subskalen des Wissenstests ist nicht vorgesehen.

Sofern das Erweiterungsmodul eingesetzt wird, können die Testleistungen insgesamt über die Bestimmung von Faktorwerten nach fluider und kristallisierter Intelligenz differenziert werden.

Die Bezeichnung der Aufgabengruppen und ihre Zuordnung zu den intellektuellen Fähigkeiten ist in Abbildung 1 aufgeführt. Bei den Aufgabengruppen

"Rechenzeichen", "Matrizen" und "Figurenpaare" handelt es sich – ebenso wie bei den Wissensfragen – um vollständige Neukonstruktionen. Die Aufgaben-Gruppe "Rechenaufgabe" wird – anders als im I-S-T 70 – ohne sprachliche Einkleidung vorgegeben. Die übrigen Aufgabengruppen wurden in geringfügig überarbeiteter Form aus dem I-S-T 70 übernommen, die gegenüber dem I-S-T 70 vorgenommenen Modifikationen sind in der Handanweisung ausführlich dokumentiert.

#### 4 GRUNDKONZEPT / THEORETISCHE GRUNDLAGEN

Mit der Revision und Neukonstruktion des I-S-T 2000 soll den Mängeln des I-S-T 70 abgeholfen werden. Neben psychometrischen Defiziten des I-S-T 70 (siehe z.B. Schaarschmidt, 1997; Schmidt-Atzert, Hommers und Heß, 1995) wurde insbesondere beklagt, dass es der Konstruktion des I-S-T 70 einer ausgewiesenen theoretischen Grundlage ermangelte (Brocke, Beauducel und Tasche, 1998). Dies führte zu Einbußen in der Konstruktvalidität des I-S-T 70. So ließ sich die von Amthauer postulierte Intelligenzstruktur anhand der mit dem I-S-T 70 erhobenen Daten zumeist nicht aufzeigen. Entsprechende Analysen deuteten eher in Richtung einer Zwei- bis Dreifaktorenstruktur des I-S-T 70 (z.B. Brocke et al., 1998; Heyde, 1996; Greif, 1972; Schmidt-Atzert et al., 1995).

Demgegenüber sollte der I-S-T 2000 auf der Basis einer systematischen theoretischen Fundierung und unter Berücksichtigung vorliegender Intelligenzmodelle konstruiert werden.

Als theoretischer Ausgangspunkt der Testkonstruktion wird ein "*hierarchisches Rahmen- bzw. Protomodell der Intelligenzstrukturforschung (HPI)*" formuliert. Dieses Modell wird von den Autoren aus den invarianten Merkmalen verschiedener Intelligenztheorien abgeleitet. Angeführt werden hier sowohl strukturelle Konvergenzen wie die Annahme unterschiedlicher Hierarchie- oder Generalitätsebenen als auch inhaltliche Konvergenzen auf der Ebene der Primärfaktoren. Deutliche inhaltliche Konvergenzen verschiedener Intelligenzmodelle können nach Ansicht der Autoren festgestellt werden "*für die Fähigkeit zum schlussfolgernden Denken (Reasoning), verbale, numerische und räumlich-figurale Fähigkeiten, Ideenflüssigkeit bzw. Kreativität und, mit Einschränkungen, für die Merkfähigkeit und wahrnehmungsbezogene Fähigkeiten*" (Amthauer, Brocke, Liepmann & Beauducel, 1999, S. 8).

Diese Komponenten werden von den Autoren entsprechend als modellübergreifend bedeutsame Intelligenzfaktoren gewertet und bilden offensichtlich

den Ausgangspunkt der Testkonstruktion. Mit dem I-S-T 2000 wird die Messung von fünf der sieben genannten Fähigkeiten avisiert (siehe Abbildung 1).

Aus Gründen der eingeschränkten Auswertungsobjektivität und -ökonomie wird auf die Diagnose der Wortflüssigkeit bzw. des Einfallsreichtums verzichtet. Außerdem werden – gänzlich ohne Begründung – mit dem I-S-T 2000 keine wahrnehmungsbezogenen Fähigkeiten erfasst.

Neben dieser inhaltlichen Festlegung folgt die Testkonstruktion der Strukturanahme der *Multi-Trait-Determination* von Intelligenztestleistungen, derzufolge sich in den Leistungen bei der Bearbeitung einzelner Aufgaben verschiedene Fähigkeiten zugleich manifestieren können. Eine Fähigkeit, die bei der Manifestation verschiedener anderer Fähigkeiten Einfluss nimmt, kann durch die Berücksichtigung einer höheren Generalitätsebene der Auswertung derselben Aufgaben diagnostiziert werden. Dieser Aspekt findet im I-S-T 2000 zweifach Berücksichtigung. Im Grundmodul wird auf einer übergeordneten Generalitätsebene das schlussfolgernde Denken ("*Reasoning*") als Summenscore der verbalen, numerischen und figuralen Intelligenz ausgewertet, und im Erweiterungsmodul werden auf übergeordneter Ebene die Generalfaktoren fluide und kristallisierte Intelligenz gebildet. Diese auf Cattell und Horn zurückgehende Differenzierung stellt nach Ansicht der Autoren eine spezielle Version des "HPI" dar. Hervorzuheben ist, dass bei der Konzeptualisierung und Operationalisierung der fluiden und kristallisierten Intelligenz versucht wurde, den hierarchischen Ansatz Cattells (1987) mit dem Facettenansatz von Guttman (1965) bzw. dem Bimodalitätsansatz von Jäger (1982) zu verbinden. Bei Guttman findet sich bereits die – bislang nicht aufgegriffene – Idee, innerhalb des Fähigkeitsbereichs "application", zu dem auch Wissen gezählt werden kann, entsprechende Inhaltskomponenten (verbal, numerisch und figural) zu differenzieren. Anders als in vielen bisherigen Ansätzen, bei denen fluide Intelligenz häufig lediglich über figurale Aufgabeninhalte (z.B. Advanced Progressive Matrices) und kristallisierte Intelligenz über verbale Aufgaben (z.B. Mehrfachwahl-Wortschatz-Intelligenztest) operationalisiert wurde, wird die fluide und kristallisierte Intelligenz sensu Cattell im I-S-T 2000 vollständig mit der Inhaltsfacette (verbal, numerisch, figural) kreuzklassifiziert, wodurch eine Kontamination dieser Fähigkeiten mit Inhaltsfaktoren vermieden wird. Als Markiervariable für kristallisierte Intelligenz wurde für das Erweiterungsmodul des I-S-T 2000 ein neuer Wissenstests konstruiert, der eine möglichst große Breite unterschiedlicher Kenntnisbereiche ansprechen soll.

Etwas unklar am Grundkonzept bleibt, ob das schlussfolgernde Denken einerseits und die fluide und kristallisierte Intelligenz andererseits auf gleicher Ge-

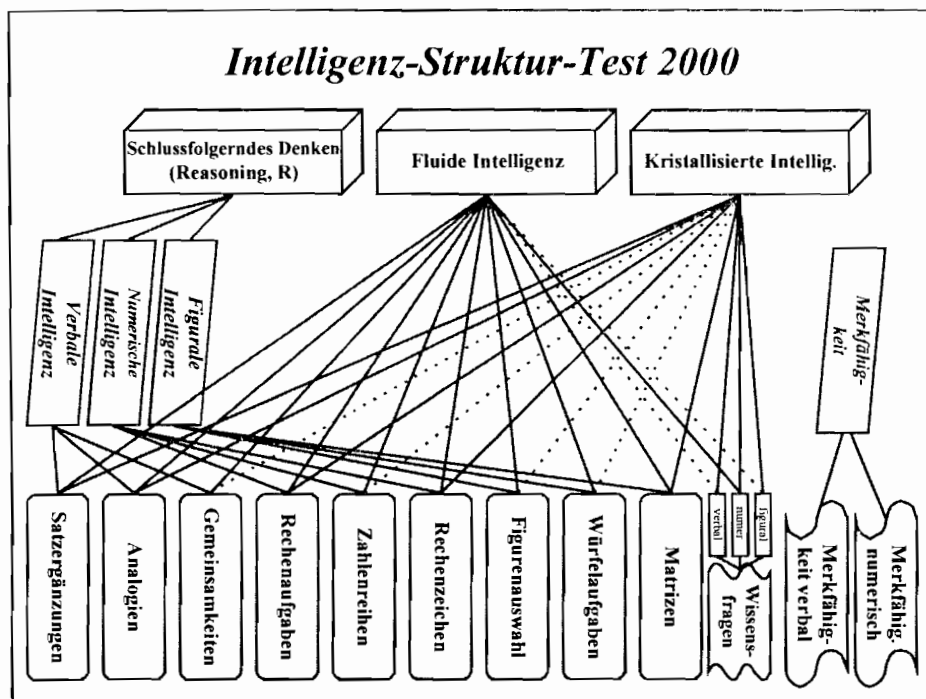


Abb. 1

neralitätssebene angesiedelt sind (wie in Abbildung 1 vereinfachend dargestellt), oder ob der fluiden und kristallisierten Intelligenz ein höheres Generalitätsniveau zukommt. Den Ausführungen in der Handanweisung zufolge bildet die fluide Intelligenz vor allem eine von Akkulturationseinflüssen weitgehend bereinigte Fähigkeit zum schlussfolgernden Denken ab. Diese Bereinigung oder Ausparialisierung wird technisch realisiert, indem bestimmte Wissensanteile mit negativen Gewichten in den Faktorwert für fluide Intelligenz eingehen (gestrichelte Linien in Abbildung 1). Aufgrund der größeren Anzahl von berücksichtigten Variablen könnte man also davon ausgehen, dass der Score für fluide Intelligenz ein höheres Aggregations- bzw. Generalitätsniveau besitzt als der "Reasoning"-Score. Demgegenüber führt die negative Verrechnung der berücksichtigten Variablen aber nicht zu einer Anhebung des Generalitätsniveaus, die Varianzen werden dadurch vielmehr spezifischer, d.h. "reiner". Hinsichtlich der Generalität bleibt die Relation von fluider Intelligenz einerseits

und "Reasoning" andererseits somit zunächst unbeantwortet. Dies stellt aber nicht nur eine offene Frage im Kontext des I-S-T 2000 dar, sondern verweist darüber hinaus auf grundsätzlichen Forschungsbedarf. Anwendern, die sich weniger für den Generalitätsvergleich von Konstrukten interessieren, genügt es jedoch zu wissen, dass schlussfolgerndes Denken mit dem I-S-T 2000 zum einen in einer relativ stärker kulturgebundenen Form ("Reasoning") und zum anderen in einer relativ weniger kulturgebundenen Form ("fluide Intelligenz") erhoben werden kann.

Um die Grundkonzeption des I-S-T 2000 nachvollziehen zu können, sollten die Anwender Teil I und II sowie Teil V der Handanweisung vollständig studieren. Lediglich die Absätze 10.2 und 11.2 in Teil V können übersprungen werden. Die auf den Seiten 3 bis 5 ausgesprochenen Empfehlungen für einen "selektiven Zugriff auf die Handanweisung" sind im Detail unstimmtig. Würde man diesen Empfehlungen Folge leisten, würde man beispielsweise die nur schwer entbehrlichen Darstellungen der mit dem Grundmodul erfassten intellektuellen Fähigkeiten und die Erläuterung ihres Generalitätsniveaus (Abschnitt 2.2) ebenso versäumen wie die Informationen zur Validität des Grundmoduls (Abschnitt 5).

## 5 DURCHFÜHRUNG

### 5.1 Alter

Für den I-S-T 2000 liegen Standardwert-Normen und Angaben von Prozenträngen für Personen im Alter von 15 bis 25 Jahren vor. Auf das Alter bezogene Empfehlungen zum Testeinsatz fehlen.

### 5.2 Formen

Wie bereits erwähnt, lassen sich mit den beiden Modulen des I-S-T 2000 drei Anwendungsformen realisieren. Mit Ausnahme von Forschungsuntersuchungen ist die Anwendung des Grundmoduls obligatorisch, die Merkfähigkeitsaufgaben und / oder der Wissenstest können zusätzlich appliziert werden.

Der Test wird in zwei pseudoparallelen Formen "A" und "B" angeboten, d.h. die Aufgabensubstanz ist in beiden Formen identisch, lediglich die Darbietungsfolge der Items und ggf. der Distraktoren wurde variiert. Aufgrund der identischen Aufgabensubstanz kann mit einiger Wahrscheinlichkeit von der

gleichen Schwierigkeit der Formen ausgegangen werden – ein diesbezüglicher empirischer Beleg ist gleichwohl in Zukunft nachzutragen, da theoretisch die unterschiedliche Itemdarbietung auch Effekte auf die Aufgabengruppen-schwierigkeit zeitigen kann (siehe z.B. Sax & Crommack, 1966).

### 5.3 Durchführung

Die Durchführung der Paper-Pencil-Version des I-S-T 2000 erfordert einen Anweiser. Der Anweiser nimmt die allgemeine Einführung vor, erläutert die einzelnen Aufgaben und achtet auf die Einhaltung der in der Handanweisung genannten Laufzeiten. Die Testanden erhalten pro Testmodul ein Aufgabenheft und einen Antwortbogen, auf dem sie ihre Lösungsversuche notieren.

Auf Seite 72 f. der Handanweisung finden sich als Checkliste gekennzeichnete Durchführungshinweise. Dieser Checkliste zufolge sind 10-20 Personen für eine Testung vorzusehen. Allerdings erfordern – so der weitere Text – mehr Testpersonen offensichtlich lediglich *"entsprechende Räumlichkeiten"*, so dass letztendlich keine Angaben zur Limitierung der gleichzeitig zu testenden Personenzahl getroffen werden. Diese Unterlassung der Angabe einer maximalen Anzahl gleichzeitig zu testender Personen ist zwar üblich, gleichwohl aber bedauerlich. Gerade die Durchführung von "Massentestungen" stellte einen in der Praxis häufigen, schwerwiegenden Verstoß gegen Durchführungsstandards dar, dem durch entsprechende Warnhinweise begegnet werden sollte. Auch zur Frage der Anzahl und Qualifikation der Testleiter finden sich keine Angaben. Die in der Handanweisung abgedruckten allgemeinen und spezifischen Durchführungshinweise sind letztendlich in Stichwortform abgefasst und teilweise nur von erfahrenen Testanweisern nachzuvollziehen. So dürfte sich beispielsweise der Sinn der folgenden Aussage nur dem erschließen, der schon über einige Testerfahrungen verfügt: *"Wenn man in einer Aufgabengruppe wenig Aufgaben bearbeitet hat, bedeutet dies nichts für die Bearbeitung der nächsten Aufgabengruppe"* (S. 73). Ausformulierte Instruktionen wären hier aus Sicht des Rezensenten einer standardisierten Testdurchführung dienlicher. Vorschläge / Richtlinien für den standardisierten Umgang mit (Nach-)Fragen der Diagnostikanden werden ebenfalls nicht gegeben. Ungeachtet dieser Schönheitsfehler im Detail sind durch die Erläuterung der Aufgabengruppen im Testheft, durch die instruktiven Beispielaufgaben und durch die präzisen Vorgaben zur Bearbeitungszeit die Voraussetzungen für eine hohe Durchführungsobjektivität des Verfahrens eindeutig gegeben.

Beim Wissenstest werden die Testanden instruiert, in jedem Fall – auch wenn sie eine Lösung nicht wissen – eine der dargebotenen Lösungen anzukreuzen. Dies ist eine ungewöhnliche Maßnahme zur Vermeidung von "Rateeffekten". Es bleiben die Erfahrungen aus der Praxis abzuwarten, inwieweit die Testanden dieser Aufforderung Folge leisten. Da keine zusätzliche Ratekorrektur vorgenommen wird, wäre eine diesbezüglich mangelnde Compliance der Diagnostikanden problematisch. Die Untersuchungsleitung sollte auf jeden Fall besonderes Augenmerk darauf legen, dass dieser Anweisung Folge geleistet wird. In diesem Kontext muss bereits auf einen Aspekt der Auswertung hingewiesen werden: Hinsichtlich der Auswertung des Erweiterungsmoduls ist leider nicht geregelt, wie mit – instruktionswidrigen – Missing Values bei den Wissensfragen umzugehen ist. Da hier keine gesonderten Regeln greifen, sind Missing Values wie falsche Antworten zu werten (null Punkte).

### 5.4 Zeit

Für die Durchführung der neun Aufgaben des Grundmoduls (Instruktionen und Bearbeitungszeit) werden insgesamt 92 Minuten veranschlagt. Der zusätzliche Einsatz des Merkfähigkeitstests (für Instruktion und Bearbeitung ist ein Zeitbedarf von ca. 12 Minuten einzukalkulieren) und /oder des Erweiterungsmoduls (weitere 37 Minuten für Instruktion und Bearbeitung) ist den Anwendern frei gestellt. Sowohl im Anschluss an die Applikation des Grundmoduls als auch nach den Merkfähigkeitsaufgaben soll eine Pause im Umfang von zehn bzw. (zwischen Merkfähigkeitsaufgaben und Wissenstest) von fünf Minuten eingeschoben werden. Somit ergibt sich eine Gesamtlaufzeit von ca. 92 Minuten für das Grundmodul, von ca. 114 Minuten für die Kombination Grundmodul und Merkfähigkeitsaufgaben sowie von ca. 156 Minuten für die Kombination Grundmodul, Merkfähigkeitsaufgaben und Wissenstest. Angemerkt sei, dass sich die Notwendigkeit der zweiten Pause (zwischen Merkfähigkeitsaufgaben und Wissenstest) nicht unmittelbar erschließt, da die Bearbeitung der Merkfähigkeitsaufgaben lediglich zwölf Minuten dauert. Wenn den Teilnehmern das Grundmodul im Umfang von 92 Minuten in einem pausenlosen Block zugemutet werden kann, wieso muss dann die nur 49 Minuten dauernde Bearbeitung der Merkfähigkeitsaufgaben und des Erweiterungsmoduls durch eine Pause unterbrochen werden? Nachvollziehbar oder nicht: Da die Normen sich auf die genannten Durchführungsbedingungen beziehen, muss die Pause zwischen der Bearbeitung der Merkfähigkeitsaufgaben und dem Erweiterungsmodul nun natürlich eingehalten werden.

## 6 AUSWERTUNG

### 6.1 Modus

Für eine möglichst objektive Auswertung des Grundmoduls und der Merkfähigkeitsaufgaben stehen je Testform Folien zur Verfügung. Diese Auswertungsschablonen sind qualitativ hochwertig in Farbdruck gestaltet. Der helle Blauton der Folienschrift erleichtert die Auswertung, da sich somit die Auswertungsinformationen vom Antwortbogen farblich abheben. Die Folien erlauben die Bestimmung des Rohwerts pro Aufgabengruppe. Durch Addition der Rohwerte der jeweils drei Aufgabengruppen zur verbalen, numerischen und figuralen Intelligenz erhält man die Rohwertsummen der jeweiligen Skalenwerte. Die Summe aller neun Aufgaben des Grundmoduls bildet die Rohwertsumme für den Gesamtwert ("Reasoning"). Auch für die beiden Merkfähigkeitsaufgaben ist die Rohwertsumme zu bilden. Anschließend werden die bereichsspezifischen Rohwertsummen anhand von Tabellen in Standardwerte umgelesen, wobei die angebotenen Tabellen eine Differenzierung nach Alter und / oder Schulbildung der Testanden ermöglichen. Neben den Standardwerten sehen die Tabellen auch Informationen zum jeweiligen Prozentrang vor.

Die Auswertung des Erweiterungsmoduls basiert auf zwei so genannten Rohwertprotokollen. Im Rohwertprotokoll für den Wissenstest ist für jede Subskala dieses Tests eine Spalte vorgesehen, die Zeilen sind für die Einzelitems gedacht. Der Auswerter schaut auf Itemebene anhand der in einer Spalte verzeichneten Lösung, ob die Ankreuzung des Testanden richtig ist oder nicht. Für richtige Lösungen trägt er einen Punkt in die Spalte der zugehörigen Subskala ein. Da die jeweils zugehörige Skala im Rohwertprotokoll weiß, die übrigen Felder aber grau hinterlegt gedruckt sind, ist die Zuordnung zu der jeweiligen Subskala einfach und sicher zu gewährleisten. Die für die sechs Subskalen ermittelten Rohwertsummen werden schließlich auf das andere Rohwertprotokoll (Protokoll für kristallisierte und fluide Intelligenz) übertragen, hier erfolgt auch noch einmal eine Eintragung der Rohwerte der neun Aufgabengruppen des Grundmoduls. Die fünfzehn Rohwerte (neun Aufgabengruppen des Grundmoduls und sechs Wissens-Skalen) werden dann anhand von Tabellen einmal in Punktwerte für fluide und einmal in Punktwerte für kristallisierte Intelligenz umgelesen. Bei diesen Punktwerten handelt es sich um die Werte, die sich bei einer Multiplikation der Rohwerte mit den B-Gewichten zur Berechnung der Faktorwerte ergeben würden. Um die Faktorwerte zu bestimmen, werden die Punktschichten für fluide und kristallisierte Intelligenz abschließend anhand von Tabellen in Standardwerte umgelesen.

Alle ermittelten Werte können auf dem Vordruck "Ergebnisprofil" eingetragen werden.

Die Auswertung ist insgesamt hochgradig standardisiert. Das gewählte Vorgehen zeigt außerdem, dass auch bei Auswertungen "per Hand" das Arbeiten mit Faktorwerten – die eine diskretere bzw. den intendierten Konstrukten angemessenere Abbildung von Fähigkeiten erlauben – möglich ist. Gleichwohl wird das demnächst erscheinende Auswertungsprogramm sicherlich insbesondere als Erleichterung bei der Bestimmung der Faktorwerte willkommen sein. Die nachfolgend aufgelisteten kritischen Anmerkungen zur Auswertung beziehen sich auf Details. Diese Details stellen das jetzige Niveau der Auswertungshilfen nicht in Frage, machen aber auf Verbesserungsmöglichkeiten – die der Anwender teilweise sofort selbst realisieren kann – aufmerksam. So sehen die Folien beispielsweise leider keine "Ausparungen" vor, die es ermöglichen, den ermittelten Rohwert ohne Abheben der Schablone auf den Antwortbogen zu notieren. Hier kann sich der Anwender selbst mit einer Schere Auswertungserleichterung verschaffen.

Problematischer ist es, dass für das Erweiterungsmodul überhaupt keine Auswertungsschablonen vorgesehen sind. Die Auswertung über die Rohwertprotokolle ist ungleich aufwendiger und fehleranfälliger als das übliche Schablonieren. Auch hier werden Anwender, die den I-S-T 2000 häufig bei großen Gruppen einsetzen, sich sinnvollerweise selbst eine Lösungsschablone erstellen.

Ein anderes Detail bezieht sich auf den Antwortbogen des Grundmoduls. Hier ist der Platz für die demographischen Angaben auf dem *Deckblatt* des Antwortbogens vorgesehen. In Folge dessen sind auf den Antwortbogen mit den Lösungen selbst keine demographischen Angaben verzeichnet, was ein unnötiges Umblättern bei der Zuordnung der Daten zu den Personen bedingt.

Als unpraktisch erweist sich schließlich das angebotene Ergebnisprofil. Für die graphische Darstellung der Standardwerte sind gerade einmal vier Zentimeter Platz vorgesehen. Pro Standardwert stehen somit knapp 0,7 Millimeter zur Verfügung. Die Orientierung in der vorgedruckten Graphik wird zusätzlich erschwert, indem Hilfslinien in Schritten von je 2,5 Standardwertpunkten eingezeichnet sind. Da ausschließlich Standardwerte ohne Nachkommastellen vorkommen, sind Hilfslinien an definitionsgemäß nicht vorkommender Position eher verwirrend als hilfreich. Nachteilig ist auch, dass das Ergebnisprofil keine Eintragung des Alters des Testanden vorsieht. Das Feld zum Eintragen des Geburtsdatums allein reicht ohne die (aus verschiedenen Gründen benötigte) auf dem Ergebnisprofil nicht vorgesehene Information zum Untersu-

chungsdatum nicht aus. Nützlich wäre auch eine Kennzeichnung der jeweils zugrunde gelegten Normenbasis.

## 6.2 Zeit

Die Handanweisung sagt nichts darüber aus, wie viel Zeit für die Auswertung schätzungsweise zu veranschlagen ist.

## 7 GÜTEKRITERIEN

### 7.1 Objektivität

Sofern der Testeinsatz regelgerecht erfolgt, ist durch die hohe Standardisierung der Durchführung und Auswertung die (Durchführungs- und Auswertungs-) Objektivität des Verfahrens gewährleistet. Lediglich hinsichtlich der Auswertungsobjektivität der Merkfähigkeitsaufgaben sind aufgrund einiger kleiner (im Abschnitt 7.4 [Normen] angeführten) Fehler zur Sicherung der Auswertungsobjektivität einige Zusatzmaßnahmen zu treffen.

### 7.2 Reliabilität

Zur Reliabilität des Tests werden Maße der inneren Konsistenz und Halbierungskoeffizienten berichtet. Die so ermittelten Reliabilitäten der Skalen und des Gesamtwertes können mit Werten über .80 bis .97 als gut bis sehr gut bezeichnet werden. Diese Reliabilitätsschätzungen basieren auf dem Pool der jeweiligen Einzelaufgaben. Die Reliabilitätsschätzungen für die einzelnen Aufgabengruppen sind ebenfalls berichtet und fallen teilweise deutlich geringer aus. So ergab sich für Gymnasiasten beispielsweise ein Cronbach's Alpha von .51 für die Aufgabengruppe "Satzergänzungen" und von .54 für die Aufgabengruppe "Matrizen". Diese unbefriedigenden Werte sind kein Problem, solange der Test ordnungsgemäß / der Handanweisung entsprechend eingesetzt wird. Ein separater Einsatz und / oder eine separate Interpretation der Aufgabengruppen oder Subskalen – wie es beim I-S-T 70 sowohl in der Praxis als auch in der Forschung erlitten wurde – wird von den Autoren (S. 82) explizit als gravierende Verletzung diagnostischer Standards und als Missbrauch gekennzeichnet.

Die innere Konsistenz des Wissenstests – dessen gesonderter Einsatz allerdings ebenfalls nicht vorgesehen ist – beträgt .90.

Für die bei der Anwendung des Erweiterungsmoduls berechneten Faktorwerte für fluide und kristallisierte Intelligenz werden ebenfalls Reliabilitätsschätzungen benötigt. Da bislang weder eine Wiederholungsstudie mit dem I-S-T 2000 durchgeführt wurde, noch ein Paralleltest vorliegt, konnten hier keine Standardverfahren zur Reliabilitätsschätzung für die Faktorwerte eingesetzt werden. Bezüglich dieser Problematik präsentieren die Autoren einen methodisch interessanten Lösungsvorschlag, der auch unabhängig von dem hier rezensierten Verfahren Beachtung finden sollte. Grundidee dieser neuen Bestimmungsmethode ist es, die Faktorwerte zweimal – aufgrund jeweils halbiertes Variablen – zu schätzen und dann die so gebildeten Faktorwerte miteinander zu korrelieren. Für zwei unterschiedliche Datensätze werden entsprechende, Spearman-Brown korrigierte, "split-half-Reliabilitäten" in Höhe von  $r=.86$  und  $r=.88$  für den Faktorwert kristallisierte Intelligenz und  $r=.90$  und  $r=.93$  für den Faktorwert fluide Intelligenz berichtet.

Die zur Zeit vorliegenden Reliabilitätsdaten sollten in nächster Zeit um Ergebnisse anderer Reliabilitätsschätzverfahren ergänzt werden. Dies gilt insbesondere für die Faktorwerte, da sich die mittels des innovativen Bestimmungsverfahrens geschätzten Reliabilitäten aufgrund mangelnder Erfahrung mit dieser Methode nicht so gut bewerten lassen. Alternativen für die Zukunft bilden die Bestimmung der Paralleltest-Reliabilität sowie Angaben zu Stabilität (Retest-Reliabilität). Die Bestimmung der Reliabilität über die Paralleltestmethode setzt natürlich die Existenz eines "echten" Paralleltests, mit anderem Aufgabenmaterial, voraus. Bezüglich der Retest-Reliabilität müsste vorab erörtert werden, ob hinsichtlich der kristallisierten Intelligenz die Wiederholungsmethode überhaupt ein angemessenes Schätzverfahren darstellt. Angemessen wäre die Retest-Schätzmethode nämlich nur dann, wenn man von einer stabilen Wissensleistung ausgeht. Zu diskutieren wäre aber, ob die kristallisierte Intelligenz insgesamt oder für spezifische Gruppen (z.B. jüngere Diagnostikanden bei einem Messwiederholungsabstand von einem Jahr) ein vergleichsweise weniger stabiles Merkmal ist als die fluide Intelligenz.

Zusätzlich zur Thematik der Reliabilität soll an dieser Stelle der Rezension auch kurz auf die berichteten Trennschärfekoeffizienten eingegangen werden. Dem Testkonzept entsprechend sind hier die Gütekriterien der interpretierbaren Skalen maßgeblich für die Testbewertung. Vor diesem Hintergrund erstaunt es etwas, dass dem Bericht der Gütekriterien der Aufgabengruppen recht viel Raum in der Handanweisung eingeräumt wird. Begründet wird dies damit, dass sich aus diesen Kennwerten Hinweise zur Pflege und Entwicklung

des Instruments ableiten lassen (S. 34). Die Werte sind somit interessant für die kleine Gruppe der Testkonstruktoren. Ungeachtet dessen wäre es aber für die ungleich größere Gruppe der Testanwender interessant gewesen, mehr über die dem Testkonzept entsprechenden Kennwerte zu erfahren. So hätte man beispielsweise die Trennschärfen als Korrelation der Items mit den höher aggregierten Skalen (verbale, numerische und figurale Intelligenz sowie "Reasoning" und Merkfähigkeit) berechnen und berichten können. Dies gilt umso mehr, als die berichteten Trennschärfen für die Aufgabengruppen bei den Tests "Matrizen" und "Satzergänzungen" Fragen aufwerfen. Bezogen auf die Gesamtgruppe ist nur ein Item der Aufgabengruppe "Matrizen" hinsichtlich seiner Trennschärfe, einer Klassifikation von Ebel (1972) folgend, "brauchbar", 30% der übrigen Items müssen hinsichtlich ihrer Trennschärfen als "wenig brauchbar" (0.20 bis 0.29) und 65% als "unbrauchbar" (Trennschärfe unter .20) angesehen werden. Für die Aufgabengruppe "Satzergänzung" ist lediglich der Range der Trennschärfen angegeben, der allerdings mit einem Maximum von .31 nichts Gutes verspricht. Es ist davon auszugehen, dass die interpretierbaren Trennschärfen (Item-Gesamtskala) günstiger ausfallen, der entsprechende Bericht fehlt aber in der Handanweisung.

### 7.3 Validität

In der Handanweisung werden die Analysen zur Validität des Grundmoduls und des Erweiterungsmoduls getrennt dargestellt.

Zur Konstruktvalidität des Grundmoduls werden Dimensionsanalysen der faktoriellen Struktur angeführt. Die Analysen mit den neun Aufgaben zur verbalen, numerischen und figuralen Intelligenz liefern eine überzeugende Bestätigung der diesbezüglich dreifaktoriellen Struktur, die sich auch als hinreichend invariant gegenüber den untersuchten vier Altersgruppen erweist. Erste Hinweise auf die konvergente und diskriminante Validität des Grundmoduls können aus den Korrelationen der Skalen "verbal", "numerisch", "figural" sowie der Gesamtskala "Reasoning" einerseits und vier verschiedenen Intelligenzmaßen andererseits abgeleitet werden. Die anhand der Daten von 170 bis 180 Schülern errechneten Daten zeigen verhältnismäßig geringe Zusammenhänge mit dem als Indikator für Speed eingesetzten "d2"-Test sowie substantielle Zusammenhänge mit dem "CFT-20" und mit dem Wissensanteil des "HAWIE-R". Die Interkorrelationen der zur konvergenten und diskriminanten Validität herangezogenen Intelligenzmaße untereinander sind nicht dargestellt. Eine interessante Ergänzung der Befunde wären Strukturvergleiche zwischen den mit dem I-S-T 2000 erhobenen Daten und solchen Daten, die mit anderen

gängigen Strukturtests (z.B. dem BIS-4 Test von Jäger, Süß und Beauducel, 1997) erhoben wurden.

Als externes Kriterium stehen Schulnoten in fünf Fächern von 151 bis 202 Schülern zur Verfügung. Der Gesamtwert ("Reasoning") kovariiert insbesondere ( $r=.45$ ) mit der Schulleistung im Fach "Mathematik" sowie – in nominell geringerem Ausmaß – mit den naturwissenschaftlichen Fächern. Die Schulleistungen in den sprachlichen Fächern korrelieren in nominell deutlich geringerem Maße mit den sprachlichen intellektuellen Fähigkeiten ( $r=.20$  bis  $r=.25$ ). Für die Interpretation der Befunde wäre es hilfreich zu erfahren, welcher Schultyp, welche Altersgruppe und welches Unterrichtsniveau (z.B. Grundkurs oder Leistungskurs) den Schulnoten zugrunde liegt. Auch Informationen zur psychometrischen Qualität der als Kriterien verwendeten Schulnoten (z.B. deren Reliabilität) wären hilfreich gewesen. Zumindest die Interkorrelation der Schulnoten untereinander hätte berichtet werden können.

Aus nicht genannten und nicht nachvollziehbaren Gründen bleiben die Merkfähigkeitsaufgaben bei den Analysen zur Validität des Grundmoduls vollständig ausgespart. Obwohl auf Seite 15 der Handanweisung explizit angekündigt wird, dass der I-S-T 2000 die Erfassung von **vier** grundlegenden Intelligenzfaktoren ermöglicht, werden keine entsprechenden Analysen berichtet. Auch bei dem Bericht des Zusammenhangs der Skalen mit anderen Tests und mit verschiedenen Schulnoten bleibt die Merkfähigkeit außen vor, so dass über die Validität dieser Skala nichts ausgesagt werden kann.

Die mit dem Wissenstest des Erweiterungsmoduls gewonnenen Daten wurden mit Hilfe der Multidimensionalen Ähnlichkeitsstruktur-Analyse umfassend untersucht. Analysiert wurden hypothesengeleitet voraggregierte Wissensfragen. Für den im Erweiterungsmodul eingesetzten Wissenstest konnten sowohl die intendierte wissensthematische Facette also auch die Kodierungsfacette (verbal, numerisch, figural) nachgewiesen werden. Die Dimensionsanalysen rechtfertigen die Bildung von mehreren Wissensskalen, deren gemeinsame Analysen mit den neun Skalen des Grundmoduls (ausgenommen der Merkfähigkeit) den Ausgangspunkt für die Dimensionsanalysen zur fluiden und kristallisierten Intelligenz bilden. Die Daten stimmen mit der intendierten zweifaktoriellen Struktur gut überein. Für die Dimensionsanalysen zur fluiden und kristallisierten Intelligenz wurden verschiedene Verfahren (exploratorische und konfirmatorische Faktorenanalysen, Multidimensionale Ähnlichkeitsstruktur-Analyse) eingesetzt, so dass die Ergebnisse als methodeninvariant gelten können. Den Analysen zufolge ist es mit dem Erweiterungsmodul gelungen, einen nach Wissensdomänen und Kodierungen (verbal, numerisch, figural) binnendifferenzierten Kenntnistest zu konstruieren. Der gemeinsame Einsatz des Grund-



und des Erweiterungsmoduls des I-S-T 2000 erlaubt eine Diagnose der fluiden und kristallisierten Intelligenz, die weitgehend unabhängig ist von der Kodierung des Aufgabenmaterials. Bei der Interpretation der Daten ist die Faktorinterkorrelation in Höhe von .42 zu berücksichtigen.

Anhand der Daten von 173 nicht näher beschriebenen Schülern konnte die konvergente Validität des Maßes für kristallisierte Intelligenz anhand der substantiellen Korrelationen zu dem Wissensanteil des "HAWIE-R" sowie zum "Mehrfachwahl-Wortschatz" nachgewiesen werden. Als Indikator der konvergenten Validität der fluiden Intelligenz gilt den Autoren der substantielle Zusammenhang mit den Matrizenaufgaben des CFT 20. Auch der Nachweis der diskriminanten Validität konnte anhand der Korrelationen der Faktorwerte mit dem "d2" sowie mit den bereits genannten Intelligenzmaßen geführt werden.

Schließlich werden Zusammenhänge zwischen den Indikatoren der fluiden und kristallisierten Intelligenz einerseits und ausgewählten Schulnoten andererseits berichtet. Das berichtete Korrelationsmuster unterstreicht nach Ansicht der Autoren die Differenzierung in fluide und kristallisierte Intelligenz.

Die Validierung der Messungen der fluiden und kristallisierten Intelligenz an Schulnoten ist interessant und problematisch zugleich. Voraussetzung für den von den Autoren realisierten Validierungsansatz ist, dass einige Schulnoten eher ein Indikator für fluide, andere Schulnoten hingegen eher ein Indikator für kristallisierte Intelligenz darstellen. So hat nach Ansicht der Autoren z.B. bei der Schulleistung im Fach "Geschichte" die kristallisierte Intelligenz eine größere Bedeutung als bei der Schulleistung im Fach "Physik". Die dieser Ansicht zugrundeliegenden Annahmen werden von den Autoren aber nicht expliziert. Sicherlich kann man sich vorstellen, dass naturwissenschaftliche Fächer beispielsweise von außerschulischen Akkulturationsbedingungen – z.B. dem Sozialstatus, Bildungsstatus der Erziehungsberechtigten usw. – weniger berührt sind als traditionelle Bildungsfächer wie "Geschichte" oder Sprachen. Andererseits dürften Schulleistungen aber in toto, also *unabhängig* vom jeweiligen Fach, ein Indikator für die Einflüsse der Akkulturation und der Bildung sein. Eine dritte Argumentationslinie ergibt sich, wenn man berücksichtigt, dass Schulleistungen nicht unwesentlich den Wissenserwerb betreffen. Gerade die erste Phase des Wissenserwerbs wird häufig in Zusammenhang mit der fluiden Intelligenz gebracht. Es ist ja gerade das namensgebende Element der Investmenttheorie, dass fluide Intelligenz *investiert* werden muss, um Wissen zu erwerben. Es könnte dementsprechend auch vermutet werden, dass der Zusammenhang der Maße für fluide und kristallisierte Intelligenz mit den Schulleistungen in Abhängigkeit davon variiert, in welcher Wissenserwerbsphase sich die Schüler gerade befinden. Welche Kriterien man immer zur Validierung

der fluiden und kristallisierten Intelligenz heranzieht – die zugrunde liegenden Annahmen wären in jedem Fall zu explizieren. Eine konsensfähigere Validierung der Skalen für fluide und kristallisierte Intelligenz könnte sich beispielsweise an vielfach replizierten Befunden zur altersspezifischen Intelligenzentwicklung orientieren und/oder den Ansatz kulturvergleichender Studien nutzen. Die in der Handanweisung verzeichneten Hinweise zur Kriteriumsvalidität der fluiden und kristallisierten Intelligenz sind dürftig. Dies ist für ein neues Testverfahren nicht anders zu erwarten. Hilfreich wären aber Hinweise auf Literatur über Validitäten für Indikatoren fluider und kristallisierter Intelligenz aus anderen Operationalisierungen.

#### 7.4 Normen

Die Skalen zur verbalen, numerischen und figuralen Intelligenz sowie der diesbezügliche Summenwert wurden anhand der Daten von insgesamt 1285 Personen im Alter von 15 bis 25 Jahren normiert. Die Gesamtgruppe wird nach Schulbildung (Gymnasiasten versus Nicht-Gymnasiasten) differenziert. Innerhalb der 925 Personen umfassenden Gymnasialgruppe werden vier Altersgruppen, innerhalb der anderen Gruppe werden zwei Altersgruppen unterschieden. Die Aufteilung der Gruppe der Nicht-Gymnasiasten in zwei altersspezifische Gruppen von unbekannt bleibender Größe strapaziert einerseits die mit 360 Personen für eine Normierung ohnehin knapp kalkulierte Gruppengröße, andererseits bleiben beim Sammelbegriff "Nicht-Gymnasiasten" Differenzierungswünsche offen. Zumindest weitergehende Informationen über den Bildungsstand dieser Gruppe wären dringend erforderlich. Dies gilt auch hinsichtlich der nicht genannten Verteilung der Geschlechter.

Die Berechnung der Standardwerte für die genannten Skalen erfolgt über die Rohwertsummen der jeweiligen Aufgabengruppen. Auf eine vor der Aggregation vorgenommene Korrektur (Standardisierung) der pro Aufgabengruppe unterschiedlichen Streuungen wird aus pragmatischen Gründen verzichtet. Auf Seite 39 der Handanweisung heißt es diesbezüglich, dass die auf der Basis standardisierter Streuungen und die auf Basis einfacher Summenscores berechneten Skalen zu  $r=.97$  bzw. zu  $r=.96$  (Gesamtwerte) und höher miteinander korrelieren. Die Einschätzung der Autoren, dass Korrelationen in dieser Höhe eine Standardisierung der aufgabengruppenspezifischen Streuungen entbehrlich machen, wird vom Rezensenten nicht geteilt. Während eine Übereinstimmung zwischen den auf unterschiedliche Art und Weise berechneten Summenscores in der berichteten Höhe für gruppenorientierte Auswertungskonzepte ausreichen mag, kann eine Abweichung in fast 6% der

Fälle bei einer auf Einzelfallentscheidungen ausgerichteten Diagnostik sehr wohl problematisch sein.

Für den Gesamtwert ("Reasoning") werden zusätzlich zu den genannten sechs Gruppen spezifischer Alters- und Bildungskombinationen drei weitere Normgruppen angeboten. Möglich ist zusätzlich eine Eichung aufgrund der Gruppe (1) aller Gymnasiasten und (2) aller Nicht-Gymnasiasten (jeweils ohne Altersdifferenzierung) sowie (3) aufgrund einer stratifizierten Gesamtstichprobe (N=625). Bei dieser zuletzt genannten Gruppe wurde der Anteil der Gymnasiasten an der gebildeten Gesamtgruppe auf ca. 40% beschränkt. Während bei allen bisher angeführten Normen der Rohwert / die Rohwertsumme in eine Ausprägung der Z-Skala (MW=100; SD=10) transformiert wird, wird für die zuletzt genannte Vergleichsbasis (Normierung des Gesamtwertes ["Reasoning"] auf Basis einer stratifizierten Gesamtgruppe) zusätzlich auch die IQ-Skala (MW=100; SD=15) angeboten.

Den für die Merkfähigkeitsaufgaben gebildeten Normen liegt mit 435 Gymnasiasten und 331 Nicht-Gymnasiasten eine weniger umfangreiche Normbasis zugrunde. Zusätzlich zu diesen beiden Gruppen wird auch hier eine stratifizierte Gesamtgruppe (40% Gymnasiasten, N=534) als Referenz angeboten. Innerhalb der untersuchten Gruppe zeigten sich keine altersspezifischen Ausprägungen der Merkfähigkeit, so dass diesbezüglich auf weitergehende Differenzierungen in Altersgruppen verzichtet wurde. Während bei den übrigen Aufgaben des Grundmoduls die Berechnungen aufgrund der Rohwertsummen erfolgen, werden in Tabelle A10 bei den Merkfähigkeitsaufgaben für die stratifizierte Gesamtgruppe (und nur für diese) sowohl Standardwerte für die Rohwertsummen als auch Standardwerte für standardisierte Summen angeboten. In der Handanweisung heißt es dazu auf Seite 39, dass diese Standardisierung aufgrund der unterschiedlich großen Aufgabenzahl in den verbalen und figuralen Aufgaben notwendig sei. Im nächsten Satz wird dann aber unter Verweis auf die hohe Korrelation ( $r \geq .95$ ) der auf unterschiedliche Art und Weise gebildeten Skalen gleichwohl geraten, auf diese zusätzliche Information in den Normtabellen zu verzichten – eine Empfehlung, die angesichts der berichteten Streuungsunterschiede doch überrascht. Wie auch immer: Da die unterschiedlichen Berechnungsweisen im Einzelfall zu (geringfügig) abweichenden Ergebnissen führen können, sollten Anwender diesem Umstand bei den Merkfähigkeitsaufgaben Aufmerksamkeit schenken. So reicht es beispielsweise nicht aus, für die Auswertung herangezogenen Assistenzkräften nur das Kapitel 11 zur Auswertung (in dem dieser Aspekt nicht thematisiert wird) und die Normtabellen zu erläutern. Um Einbußen in der Auswertungs-

objektivität zu vermindern, muss hinsichtlich der Tabelle A10 eine – der Seite 39 der Handanweisung entsprechende – eindeutige Festlegung getroffen werden.

Bei dem eigentlich nicht benötigten Abschnitt der Tabelle A10, den Standardwerten für die standardisierten Rohwertsummen, stellt sich noch weiterer Regelungsbedarf ein. Ungeklärt ist nämlich, was mit einem standardisierten Summenwert geschieht, der in ein Intervall fällt. So wird beispielsweise in Tabelle A 10 lediglich festgelegt, dass ein Testand für einen standardisierten Summenwert in Höhe von "178" den Standardwert "87" erhält. Die nächste Information bezieht sich dann bereits auf einen standardisierten Summenwert von "181", dem der Standardwert "89" zugewiesen wird. Wie mit Werten in den nicht-definierten Bereichen umzugehen ist, wurde nicht geregelt. Auswerter können versucht sein, einen standardisierten Summenwert von z.B. "179" entweder dem für "178" oder dem für "181" definierten Wert zuzuordnen oder aber eigenständig in der Tabelle zu extrapolieren. In den Tabellen A10 bis A12 wurde außerdem gelegentlich vergessen, den Rohwertausprägungen "Null" sowie teilweise auch dem Rohwert "Eins" einen eindeutigen Standardwert zuzuordnen. Auch hinsichtlich dieser Fälle sollte der Anwender durch eigene Festlegungen vorab regeln, wie sich die Auswertungskräfte bei denjenigen Rohwerten bzw. bei denjenigen standardisierten Summenwerten verhalten sollen, welche sich nicht eindeutig zu Standardwerten zuordnen lassen. Aus den berichteten Prozenträgen lässt sich bezüglich der Rohwertausprägungen die Empfehlung ableiten, den bezüglich der Standardwerte nicht definierten Rohwerten den Standardwert zuzuweisen, der für den kleinsten definierten Rohwert aufgeführt ist.

Während in der Handanweisung für das Grundmodul ein eigener Abschnitt "Normen" vorgesehen ist, fehlt ein entsprechender Gliederungspunkt für das Erweiterungsmodul. Die Beschreibung der diesbezüglich relevanten Normgruppe läuft daher Gefahr, im Text etwas unterzugehen. Die entsprechenden Informationen finden sich auf Seite 68 der Handanweisung. Die Normbasis für die Faktorwerte zur fluiden und kristallisierten Intelligenz umfasst 248 Schüler im Alter von 15 bis 25 Jahren. Für die Faktorwerte für kristallisierte und fluide Intelligenz wird jeweils nur eine Normtabelle angeboten, d.h. diese Auswertung kann nicht nach Alter und Schulbildung differenziert erfolgen.

Insgesamt ist zu kritisieren, dass für die unterschiedlichen Anwendungsformen des I-S-T 2000 (nur) unterschiedliche Normgruppen angeboten werden. Bei einer vollständigen Nutzung aller Testmodule kann dies dazu führen, dass die auf dem Ergebnisprofil verzeichneten Werte auf dem Vergleich mit einer jeweils anderen Referenzgruppe basieren.

Zusätzlich zu den Tabellen B3 und B4 mit den Standardwerten für die Faktorwerte sind in der Handanweisung vermutlich irrtümlicherweise noch zwei weitere Tabellen (B5 und B6) mit Standardwerten "nur für Aufgaben mit positiven Punktwerten" abgedruckt, die zur Auswertung nicht benötigt werden und die in der Handanweisung nicht erläutert sind.

In der Handanweisung finden sich keine Angaben zu geschlechtsspezifischen Leistungsausprägungen.

Eine letzte kritische Anmerkung zu den Normtabellen betrifft die mit den abgebildeten Normtabellen einhergehende Möglichkeit zum Testmissbrauch. Es ist sehr lobenswert, dass in der Handanweisung zum I-S-T 2000 explizit Anwendungen genannt werden, die als Missbrauch und Verletzung diagnostischer Standards zu werten sind. Als diskussionsbedürftig muss in diesem Kontext aber gewertet werden, dass im Rahmen der Normtabellen nicht nur Standardwerte für die Rohwertsummen, sondern auch Standardwerte für die einzelnen Aufgabengruppen angeboten werden – obgleich die Standardwerte pro Aufgabengruppe der Testintention entsprechend nicht benötigt werden. Dieses "Angebot" könnte geradezu als Einladung missverstanden werden, entgegen der expliziten und gut begründeten Missbrauchswarnung auf Seite 82 der Handanweisung die Aufgabengruppen eben doch isoliert einzusetzen und zu interpretieren. Begründet wird das Angebot der Normwerte für Aufgabengruppen auf Seite 79 der Handanweisung mit deren Verwendbarkeit im Forschungskontext. Auch für den Vergleich von Testwerten bei den relativ wenig modifizierten Aufgabengruppen des I-S-T 70 mit denen im I-S-T 2000 kann der Rückgriff auf Normen für die Aufgabengruppen nach Ansicht der Autoren sinnvoll sein. Ein besserer Schutz gegen Testmissbrauch hätte bestanden, wenn man Anwendern mit solchen Spezialinteressen (Forschung, Vergleich mit I-S-T 70 Werten) für den Fall der Kontaktaufnahme mit den Autoren Unterstützung zugesagt und dafür die Angabe von Normen für nichtinterpretierbare Aufgabengruppen unterlassen hätte.

## 8 KRITIK

Mit dem I-S-T 2000 wurde gegenüber dem I-S-T 70 ein deutlicher diagnostischer Fortschritt erzielt, der zahlreichen Anwendern und Diagnostikern zugute kommt. Fortschritte zeigen sich zunächst auf der psychometrischen Ebene. Schwerwiegende Probleme des I-S-T 70, wie z.B. die veraltete Normbasis, die niedrigen Reliabilitäten für den Subtest Wortauswahl und die Mittelwertsunterschiede zwischen Form A und Form B bei den Aufgabengruppen

"Wortauswahl" und "Analogien" wurden überwunden. Es ist zu wünschen, dass der mangelhafte I-S-T 70 nun endlich seinen verdienten Ruhestand antritt und nicht weiter eingesetzt wird. Die entscheidenden Fortschritte wurden auf konzeptioneller und theoretischer Ebene erzielt. Durch die konsequente Ausrichtung der Testkonstruktion auf eine explizite theoretische Grundlage konnte ein konstruktvalides Instrument geschaffen werden. Während für den I-S-T 70 die intendierte Struktur letztendlich nicht nachgewiesen werden konnte, kann die dem I-S-T 2000 zugrundegelegte Struktur – mit Ausnahme der in den Strukturanalysen leider ausgespart bleibenden Merkfähigkeit – prägnant aufgezeigt werden. Auch wenn die die Merkfähigkeit berücksichtigenden Strukturanalysen noch nachgetragen werden müssen, kann jetzt bereits aufgrund der Operationalisierung auch für die Diagnose der Merkfähigkeit ein Fortschritt gegenüber dem I-S-T 70 konstatiert werden. Durch die Berücksichtigung verbaler und figuraler Materialien bei den Merkfähigkeitsaufgaben wurde die mit dem I-S-T 70 praktizierte, völlig unbegründete Beschränkung der Diagnose der Merkfähigkeit auf verbale Inhalte überwunden. Die theoriegeleitete Konstruktion macht sich vor allem auf der Ebene der aggregierten Werte positiv bemerkbar. Im Gegensatz zum Gesamtwert des I-S-T 70, der aufgrund der unterschiedlichen Anzahl von Aufgabengruppen zu den einzelnen Dimensionen (und der damit einhergehenden Gewichtung dieser Dimensionen) nur bedingt als globales Intelligenzmaß interpretiert werden konnte, erlaubt der I-S-T 2000 eine stringent operationalisierte Erfassung des schlussfolgernden Denkens. Die aus Sicht des Rezensenten maßgebliche Innovation besteht in der Umsetzung der als *Multi-trait-Determination* von Intelligenztestleistungen bezeichneten Strukturannahme, die sich als Diagnose des schlussfolgernden Denkens und als Diagnose der fluiden und kristallisierten Intelligenz manifestiert. Durch das Erweiterungsmodul wird es mit dem I-S-T 2000 erstmalig möglich, im deutschen Sprachraum Diagnosen der fluiden und kristallisierten Intelligenz (junger) Erwachsener weitgehend unabhängig von Inhaltskomponenten zu erstellen und somit die Investmenttheorie von Cattell und Horn deutlich besser zu nutzen. Dabei wurde die häufig praktizierte, aber unbegründet enge Operationalisierung fluider Intelligenz über figurale Reasoningaufgaben (z.B. Matrizen) und kristallisierte Intelligenz über verbale Aufgaben (z.B. Mehrfachwahl-Wortschatztest) überwunden.

Durch die Option zur Berücksichtigung der Generalitätsebene des schlussfolgernden Denkens sowie vor allem mit den Faktorwerten für fluide und kristallisierte Intelligenz geht der I-S-T 2000 über eine bloße technische Revision des I-S-T 70 deutlich hinaus. Dies war auch erforderlich, da das rein technisch-psychometrische Funktionieren eines Messinstruments zwar eine notwendige, keinesfalls aber eine hinreichende Grundlage für die Fähigkeitsdiagnostik dar-

stellt. Ausgangspunkt der Fähigkeitsdiagnostik sind und bleiben Konstruktannahmen. Diese Konstruktannahmen sind u.a. handlungsleitend bei der Auswahl und Kombination der diagnostischen Instrumente, bei der Konzeptualisierung des Kriteriumsverhaltens und bei der Interpretation der diagnostischen Befunde. (Zusätzlich zeitigen klare Konstruktannahmen positive Effekte bei der regelgeleiteten Item- und Testkonstruktion.) Intelligenzstrukturtests basieren notwendigerweise auf Konstruktannahmen zur Intelligenzstruktur. Entsprechende Annahmen sind mit Strukturanalysen zu prüfen. Das Konzept des I-S-T 70 führte zu Daten, in denen die intendierte Intelligenzstruktur nicht aufgezeigt werden konnte. Die Bezeichnung Intelligenzstrukturtest sollte Instrumenten vorbehalten werden, mit denen nicht nur irgendeine, sondern die theoretisch intendierte und erläuterte Intelligenzstruktur aufgezeigt werden kann. In diesem Sinne wird das mit der Produktbezeichnung "I-S-T" verbundene Versprechen eines Intelligenzstrukturtests erst jetzt – fast ein halbes Jahrhundert nach der Präsentation des "I-S-T" im Jahre 1953 – mit dem I-S-T 2000 eingelöst.

Neben dieser positiven Würdigung sind andere Aspekte kritisch zu werten. Wie bei jedem neuen Test stehen anfangs noch einige Daten und /oder Weiterentwicklungen aus. Die Autoren haben sich auf Seite 6 der Handanweisung selbst die Bereitstellung einer weiteren Parallelfarm mit neuen Aufgaben und die Ausdehnung des Normierungsbereiches ins Pflichtenheft geschrieben. An anderer Stelle nennen sie die Ermittlung der Wiederholungs-Reliabilität (Stabilität) als zukünftige Aufgabe. Weiterhin sehen sie kennwertbedingt Modifikationsbedarf bei den Aufgabengruppen "Satzergänzung" und "Matrizen" (S. 34). Ergänzt werden kann der Wunsch nach Strukturvergleichen zwischen dem I-S-T 2000 und anderen Intelligenzstrukturtests sowie der Wunsch nach Zusammenhangsanalysen zwischen dem "alten" I-S-T 70 und dem "neuen" I-S-T 2000. Dringend benötigt werden Daten zur Kriteriumsvalidität, insbesondere zur Kriteriumsvalidität der Faktorwerte für fluide und kristallisierte Intelligenz. Sinnvoll wäre es außerdem, die Normen für die Faktorwerte kompatibel zu den Normen der Skalen des Grundmoduls und der Merkfähigkeitsaufgaben zu gestalten. Auch auf der Ebene der ansonsten Vorbildlich durchgeführten Strukturanalysen könnte noch investiert werden. Unbedingt nachzutragen sind Strukturanalysen, die auch die Merkfähigkeitsaufgaben berücksichtigen. In der gegenwärtigen Fassung der Handanweisung bleibt die Merkfähigkeit völlig isoliert. Zusätzlich sollten auch weitergehende Generalitätsanalysen durchgeführt werden. Die Analyse der Ladungen der Zweifaktoren-Lösung (fluide und kristallisierte Intelligenz) sowie die Interkorrelation der Faktoren lässt für den Fall zusätzlicher Analysen einen weiteren, generelleren Faktor (im Sinne von

Spearman's  $g$  oder Cattells  $g_{(m)}$ ) erwarten, der im I-S-T 2000 bislang nicht spezifiziert wird.

Kritisch ist auf das "theoretische Modell" der Testkonstruktion einzugehen. Die Idee einer integrativen Betrachtung konkurrierender Intelligenzmodelle stellt eine vielversprechende Ausgangsbasis der Testentwicklung dar. Mit der Berücksichtigung häufig etablierter Primärfaktoren einerseits und der Hierarchieannahme andererseits betont das *"hierarchische Rahmen- bzw. Protomodell der Intelligenzstrukturforschung"* (HPI) offensichtlich jene Positionen der Intelligenzstrukturforschung, über die forschungshistorisch betrachtet eine hohe Übereinstimmung herrschte, bevor im Zuge der Fortschreibung von verschiedenen Autoren weitere – interessante, aber wiederum divergierende – Positionen bezogen wurden. Dieser Grundgedanke und das der Testkonstruktion des I-S-T 2000 zugrunde liegende *"hierarchische Rahmen- bzw. Protomodell der Intelligenzstrukturforschung"* (HPI) ist allerdings in der Handanweisung nicht hinreichend präzise und ausführlich dargestellt, zahlreiche Fragen bleiben offen. Welche Intelligenzmodelle wurden konkret auf inhaltliche und formale Konvergenzen geprüft? Während im Text zahlreiche Modelle angerissen werden, sind in Abbildung 2.1 auf Seite 7 der Handanweisung lediglich drei Ansätze (warum diese drei?) herausgegriffen. Welches waren die Kriterien für die Annahme einer Konvergenz? Auch der Schritt vom HPI-Modell zum I-S-T 2000 ist nicht vollständig nachvollziehbar, wenn beispielsweise die zumindest mit Einschränkungen als *"modellübergreifend bedeutsam"* (S. 10) bewerteten *"wahrnehmungsbezogenen Fähigkeiten"* (S. 8) ohne eine Begründung zwar im HPI-Modell, nicht aber im I-S-T 2000 Berücksichtigung finden. Erläuterungsbedürftig erscheint auch, warum beim schlussfolgernden Denken Aufgabengruppen für alle drei Inhaltsklassen (verbal, numerisch und figural) dargeboten werden, während sich die Diagnostik der Merkfähigkeit auf die Inhalte verbal und figural beschränkt.

Es muss insgesamt in Frage gestellt werden, ob es sich bei dem *"hierarchischen Rahmen- bzw. Protomodell der Intelligenzstrukturforschung"* wirklich um ein konsistentes, eigenständiges Intelligenzstrukturmodell im engeren Sinne handelt. Der Modellstatus setzt empirisch fundierte und somit über Assoziationen hinausgehende Fokussierungen der mit dem HPI in Verbindung gebrachten Fähigkeitsbegriffe voraus, wobei sicherlich auch auf das Aufgabenmaterial der im HPI vorgeblich integrierten Ansätze rekuriert werden müsste. Die Geltungsbereiche des Modells wären empirisch auszuloten. Vor dem Hintergrund der in der I-S-T 2000 Handanweisung skizzierten Überlegungen erscheint der Modellbegriff strapaziert. Die Überlegungen stellen gleichwohl einen interessanten und diagnostisch tragfähigen theoretischen Bezugsrahmen dar, der auf

Gemeinsamkeiten existierender Intelligenzmodelle verweist. Dieser Verweis auf andere Intelligenzmodelle, ihre Gemeinsamkeiten und Abweichungen fällt allerdings etwas kryptisch aus, was vor allem den intelligenztheoretisch weniger ausgewiesenen Praktiker benachteiligt. Hinsichtlich der mit dem I-S-T 2000 gemessenen Fähigkeiten finden sich beispielsweise überwiegend nur Kurzverweise auf die den Konvergenzüberlegungen zugrunde gelegten Intelligenzmodelle sowie allgemeine Ausführungen zu Konstrukten, deren Komplexität, Offenheit usw. Bei aller Zustimmung zu diesen wissenschaftstheoretisch orientierten Ausführungen muss festgehalten werden, dass der abstrakt bleibende HPI-Gedanke nur unzureichend praxisverwertbar aufbereitet wurde. Von der Erkenntnis der Konvergenz verschiedener Positionen der Intelligenzstrukturforschung profitiert ohne weitere Erläuterung nur jener Anwender, welcher diese Positionen – die in der Handanweisung nicht ausgeführt werden – kennt. Bezüglich dieser Passagen der Handanweisung gilt: hier schreiben universitäre Spezialisten für universitäre Spezialisten. Der I-S-T wurde bislang aber häufig in der Praxis angewandt, so dass zusätzliche Erläuterungen hier der Sache einen guten Dienst erweisen würden. Für den Fall, dass entsprechende Erläuterungen den Rahmen einer Handanweisung sprengen, könnten diese Ausführungen auch in einer gesonderten Arbeit publiziert werden.

Hinsichtlich der Validität wurden die Defizite der Kriteriumsvalidierung bereits genannt. Mit Ausnahme der nur bedingt informativen Zusammenhänge zu den Schulnoten liegen keine Daten zur Kriteriumsvalidität vor. Dies ist bei einem neuen Test nicht ungewöhnlich. Bei dem hier gegebenen Sonderfall der Überarbeitung und Erweiterung eines der am häufigsten eingesetzten Tests im deutschen Sprachraum ist es aber überraschend, dass die Autoren nicht auf die einige Anwender sicherlich interessierende Frage eingehen, inwieweit Erfahrungen mit dem I-S-T 70 auf den I-S-T 2000 übertragen werden können. Da zahlreiche Aufgabengruppen fast unverändert aus dem I-S-T 70 übernommen wurden, bzw. die Modifikationen lediglich einer Verbesserung der psychometrischen Qualität dienen, könnten – mit zu explizierenden Ausnahmen und Einschränkungen – einige der mit dem I-S-T 70 gewonnenen Validitätsbelege auch bei Überlegungen zum Einsatz und zur Interpretation des I-S-T 2000 hilfreich sein. Die Autoren üben diesbezüglich vollständige Zurückhaltung, was gegebenenfalls zu einer Unterschätzung der Einsatzmöglichkeiten des I-S-T 2000 führen kann. Eine nützliche Grundlage für Überlegungen zur Validitätsgeneralisierung wären empirische Studien zum Zusammenhang des I-S-T 70 und des I-S-T 2000.

## 9 EMPFEHLUNGEN

Der Intelligenz-Strukturtest wurde vor fast einem halben Jahrhundert von Amthauer in seiner Funktion als Leiter der *Deutschen Gesellschaft für Personalwesen* e.V. entwickelt (siehe Althoff, 1990). Der I-S-T stellt vermutlich das im deutschsprachigen Raum am häufigsten benutzte Intelligenztestverfahren dar (Schmidt-Atzert, Hommers und Heß, 1995). Grund für die Beliebtheit des Tests war sicherlich nicht nur die Gunst der ersten Stunde, derzufolge es sich beim I-S-T um einen der ersten deutschsprachigen Gruppenintelligenztests nach dem Krieg handelte. Der Test konnte vielmehr mit einem attraktiven Aufgabenpool, akzeptablen Gütekriterien und einem auf den ersten Blick plausibel erscheinenden Konzept aufwarten. Insbesondere die Profile für bestimmte Berufsgruppen erfreuten sich einer hohen Face-Validity. Die Einsatzhäufigkeit des I-S-T wurde u.a. für die Erfassung einer umfassenden Eichstichprobe und für Studien zur Kriteriumsvalidität genutzt (z.B. Schmidt-Atzert & Deter, 1993a und 1993b). Wesentliche psychometrische Mängel des Verfahrens (siehe z.B. Schaarschmidt, 1997; Schmidt-Atzert et al., 1995) und vor allem die weitgehend theoriendiskonforme Konstruktion (siehe z.B. Brocke et al., 1998) wurden lange Zeit toleriert. Mit dem I-S-T 2000 von Amthauer, Brocke, Liepmann und Beauducel (1999) wurde jetzt endlich eine Version des I-S-T vorgestellt, die vorhandene Vorzüge wie z.B. das attraktive und zahlreich erprobte Aufgabenmaterial beibehält, psychometrische und theoretische Defizite des I-S-T 70 ausräumt und darüber hinaus innovative diagnostische Möglichkeiten wie die Diagnostik der fluiden und kristallisierten Intelligenz eröffnet. Angesichts der teilweise massiven Probleme des I-S-T 70, die bei Funktionsmängeln einzelner Items anfangen, über unterschiedliche Schwierigkeiten der beiden Testformen gehen und bis zu empirisch nicht zu rechtfertigenden Strukturannahmen reichen, muss der wider besseren Wissens erfolgende weitere Einsatz des veralteten I-S-T 70 eindeutig als fachlich begründungsbedürftig gekennzeichnet werden. Dies gilt insbesondere, falls der Gesamtwert des I-S-T 70 als Indikator der Allgemeinen Intelligenz interpretiert wird, obwohl hier die unterschiedlichen Fähigkeitsbereiche mit unterschiedlichen Gewichten einfließen, oder falls psychometrisch nicht zu rechtfertigende Profilauswertungen vorgenommen werden. Mit dem I-S-T 2000 ist allen Anwendern eine Alternative zum I-S-T 70 aufgezeigt, wobei mit der Konstruktion des I-S-T 2000 zielgerichtet die Defizite des bisherigen Verfahrens überwunden wurden. Dabei muss sich mancher Anwender allerdings von vielleicht lieb gewonnenen Gewohnheiten, wie dem isolierten Einsatz einzelner Aufgabengruppen und/oder wie auch immer zusammengesuchter "Kurzskalen" verabschieden. Diese Praktiken waren weder mit dem I-S-T 70, noch sind sie mit dem I-S-T 2000 zu

rechtfertigen. Der modulare Aufbau des I-S-T 2000 kommt aber den unterschiedlichen Anwendungszwecken entgegen.

Der I-S-T 70 war nicht nur eines der am häufigsten verwendeten, sondern auch eines der am besten analysierten Intelligenztestverfahren. Mit dem neuen I-S-T 2000 – insbesondere mit den innovativen Aspekten wie den Faktorwerten für fluide und kristallisierte Intelligenz – wurden zwangsläufig noch nicht so viele Erfahrungen gesammelt. Einige in Zukunft anstehende "Hausaufgaben" der Testautoren wurden benannt (z.B. Erweiterung und Vereinheitlichung der Normbasis, Integration der Merkfähigkeit in die Strukturanalysen, weitere Analyse zu den Gütekriterien, insbesondere zur Kriteriumsvalidität usw.). Aber auch die Anwender sollten zur Verfahrenspflege und Evaluation des I-S-T 2000 beitragen, indem sie ihre Erfahrung den Autoren mitteilen und nach Möglichkeit darüber hinaus auch Daten für Normierungen und Evaluationsstudien zur Verfügung stellen (siehe auch den folgenden Abschnitt).

## 10 EINE VERFAHRENSÜBERGREIFENDE SCHLUSSBEMERKUNG

Die Geschichte des I-S-T verdeutlicht wie kaum ein anderer Test das Schnecken-tempo des Fortschritts auf dem Gebiet psychometrischer Verfahren. Zwischen dem I-S-T, dem I-S-T 70 und dem jetzt vorgestellten I-S-T 2000 liegen jeweils fast 25 Jahre. Da das wesentliche Modifikationselement beim I-S-T 70 eine solidere Normenbasis war, kann man vereinfachend auch formulieren, dass fast fünfzig Jahre vergehen mussten, bevor mit dem I-S-T 2000 eine psychometrisch und theoretisch verbesserte Testfassung präsentiert wurde. Dabei musste der I-S-T 70 den Altersvergleich mit einigen alternativen Verfahren nicht scheuen. Die aktuelle Form des WILDE-Tests von Althoff und Jäger (1994) geht letztendlich auf die 1962 für die *Deutsche Gesellschaft für Personalwesen* konzipierte Fassung dieses Tests zurück, der LPS von Horn wurde im gleichen Jahr veröffentlicht usw. Nun ist "alt" nicht gleichbedeutend mit "schlecht" und "neu" ist nicht mit "gut" zu verwechseln. Wenn man überhaupt verallgemeinern kann, so markieren wohl eher die extremen Pole dieses Kontinuums die Abgründe der psychometrischen Diagnostik. Auf der einen Seite des Kontinuums finden sich "alte", ungeachtet aller anderslautender Erkenntnisse in der schon immer praktizierten Art und Weise angewendete und interpretierte Verfahren. Demgegenüber stehen auf der anderen Extremausprägung neue, vielleicht elektronisch aufwendig gestaltete, "bunte" Verfahren, die häufig schneller modifiziert werden, als man evaluieren kann. Mit Max Planck darf man daran erinnern, dass das Erkennen dem Anwenden vorausgehen muss, und vor Bilderstürmerei warnen. Das mehr oder minder beliebige Beispiel des "alten" I-S-T

zeigt aber auch, dass nachweisliche psychometrische und theoretische Defizite dem Überleben und der Einsatzhäufigkeit eines Verfahrens offenbar keinen Abbruch tun. Im Kontext der Qualitätssicherung und Qualitätsoptimierung der psychometrischen Diagnostik müssen nicht nur die Tests, sondern auch die Vertriebs- und Kommunikationsmethoden auf den Prüfstand. Ein wesentliches Hindernis auf dem Weg zur Qualitätssteigerung ist der Umstand, dass Testkonstrukteure sowie an Evaluationen interessierte Fachleute einerseits und Anwender andererseits kaum miteinander kommunizieren. Die Diskrepanz zwischen der Häufigkeit der Testeinsätze auf der einen und der für Evaluationen und Normierungen zur Verfügung stehenden dünnen Datenbasis auf der anderen Seite, ist ein erschreckendes Symptom der genannten Funkstille. Diesem Missstand könnte beispielsweise teilweise durch die Einrichtung testbezogener Internetseiten abgeholfen werden. Im Rahmen dieser Internetpräsenz könnten Anwender ihre Erfahrungen zur Durchführung und Auswertung untereinander und mit den Testverantwortlichen austauschen; könnten Anwender testbezogene Fragen stellen, aus der Praxiserfahrung gewonnene Verbesserungsvorschläge unterbreiten, ihre eigenen Testanwendungssituationen und diagnostischen Strategien vorstellen sowie Möglichkeiten der Kooperation mit interessierten Testentwicklern, Forschern usw. diskutieren. Träger dieses Internetservices sollten entweder die Verlage und/oder berufsständische Organisationen sein, beispielsweise das von der Föderation Deutscher Psychologenvereinigungen eingesetzte Testkuratorium. Darüber hinaus sollten die Verlage den Anwendern verstärkt (wie es bereits teilweise geschieht) Anreize (auch pekuniärer Art) für die Rücksendung von normentauglichen Daten anbieten. Nach wie vor sind auch Ideen gesucht, wie Belangen der praktischen Diagnostik mehr Geltung in einschlägigen Fachzeitschriften verschafft werden kann. Dies gilt umso mehr, als der Trend zur Internationalisierung der psychologischen Forschung und die andiskutierte Reduzierung / Umwidmung der deutschsprachigen wissenschaftlichen Fachzeitschriften (siehe Gigerenzer, 1999) die letzten verbleibenden Publikationsräume für teilweise sprach- und kulturgebundene – und somit in international ausgerichteten Zeitschriften kaum publizierbare – Aspekte der angewandten psychometrischen Diagnostik bedroht (siehe Wottawa, 1999). Schließlich muss gewährleistet werden, dass psychodiagnostische Tests zu Forschungszwecken eingesetzt werden können. Hier besteht häufig ein Konflikt zwischen den finanziellen Interessen der Verlage einerseits und der Finanzknappheit der Universitäten und Forschungseinrichtungen andererseits. De facto sehen sich viele Wissenschaftler nicht in der Lage, zu Forschungszwecken eingesetzte psychodiagnostische Instrumente legal (käuflich) zu erwerben, so dass es möglicherweise zu illegalen Verfahreenseinsätzen kommt. Um diesbezüglichen

Vorwürfen zu entgehen, werden dann aber beispielsweise nur Subskalen eingesetzt, Items werden verfremdet, Skalen werden umgetauft usw. Konsequenz ist, dass die für einen kumulativen Erkenntnisfortschritt notwendigen Daten nicht anfallen oder nicht genutzt werden können. Lösungen sind hier nur zu erzielen, wenn das heikle Thema offen angegangen wird. Hier müssen die Interessengruppen gemeinsam über einen Kompromiss verhandeln, moderiert wiederum von neutraler, aber kompetenter Seite. Nicht zuletzt ist aber auch der praktisch arbeitende Anwender aufgerufen, sich Innovationen gegenüber aufgeschlossen zu verhalten. Das zu beklagende Festhalten an überkommenen, defizitären Testpraktiken ist allerdings nicht nur eine Herausforderung für jeden Anwender, sondern auch eine Herausforderung für die Testautoren und Verlage. Für neue Verfahren sollten Informationsveranstaltungen, Schulungen und eben Möglichkeiten zum Erfahrungsaustausch angeboten werden (z.B. internetbasierte Diskussionsforen, jährliche Anwendertreffen usw.).

"Den Schildbürgern wurde eine schöne neue Sonnenuhr geschenkt, und weil sie ihnen so gut gefiel, wollten sie das Geschenk nicht Wind und Wetter aussetzen und errichteten ein Dach über der Uhr." Wünschen wir dem I-S-T 2000 und anderen wohlkonstruierten neuen Verfahren, wünschen wir allen am Fortschritt der psychometrischen Diagnostik interessierten Gruppen, dass die Tests dem Wind und Wetter der diagnostischen Praxis ausgesetzt werden, dass Testkonstrukteure, Forscher und Anwender ihre Erfahrungen miteinander austauschen, und dass die reichlichen Erfahrungen dann zeitnah in nachhaltigen praktischen Verbesserungen münden.

## 11 LITERATUR

- Althoff, K. (1990). Zur Geschichte der Deutschen Gesellschaft für Personalwesen e.V. und ihrer Bedeutung für die Entwicklung der Berufseignungsdiagnostik in Deutschland. *Diagnostica*, 36, 249-263.
- Amthauer, R. (1970). *I-S-T 70. Intelligenz-Struktur-Test*. Göttingen: Hogrefe.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000*. Göttingen: Hogrefe.
- Brocke, B., Beauducel, A. & Tasche, K. (1998). Der Intelligenz-Struktur-Test: Analysen zur theoretischen Grundlage und technischen Güte. *Diagnostica*, 44, 84-99.
- Cattell, R.B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: Elsevier.
- Ebel, R.L. (1972). *Essentials of educational measurement*. New Jersey: Englewood Cliffs.
- Gigerenze, G. (1999). Internationalisierung der psychologischen Forschung in Deutschland, Österreich und der Schweiz: Sieben Empfehlungen. *Psychologische Rundschau*, 50, 101-105.
- Greif, S. (1972). *Gruppenintelligenztests. Untersuchungen am WIT, IST, LPS und AIT*. Frankfurt/M.: Lang.
- Guttman, L. (1965). A faceted definition of intelligence. In R. Eiferman (Ed.), *Studies in Psychology, scripta hierosolymitana* (Vol. 14, pp. 166-181). Jerusalem: The Hebrew University.
- Heyde, G. (1996). Weitere Daten zum I-S-T 70. *Report Psychologie*, 21, 138.
- Jäger, A.O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-226.
- Jäger, A.O. & Althoff, K. (1994). *Der WILDE-Intelligenz-Test (WIT)* (revidierte Auflage). Göttingen: Hogrefe.
- Jäger, A.O., Süß, H.M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. Form 4. Handanweisung*. Göttingen: Hogrefe.
- Sax, G. & Cromack, T. (1966). The effects of various forms of item arrangements on test performance. *Journal of Educational Measurement*, 3, 309-311.
- Schaarschmidt, U. (1997). Testrezenion zum Intelligenz-Struktur-Test (I-S-T 70). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, 106-108.
- Schmidt-Atzert, L., Hommers, W. & Heß, M. (1995). Der I-S-T 70. Eine Analyse und Neubewertung. *Diagnostica*, 41, 108-130.
- Schmidt-Atzert, L., & Deter, B. (1993a). Intelligenz und Ausbildungserfolg: Eine Untersuchung zur prognostischen Validität des I-S-T 70. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 52-63.
- Schmidt-Atzert, L., & Deter, B. (1993b). Die Vorhersage des Ausbildungserfolgs bei verschiedenen Berufsgruppen durch Leistungstests. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 191-196.
- Wottawa, H. (1999). Stellungnahme der Leitung Fachgruppe "Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik" zum Problembereich Internationalisierung der psychologischen Forschung. *Psychologische Rundschau*, 50, 109-111.