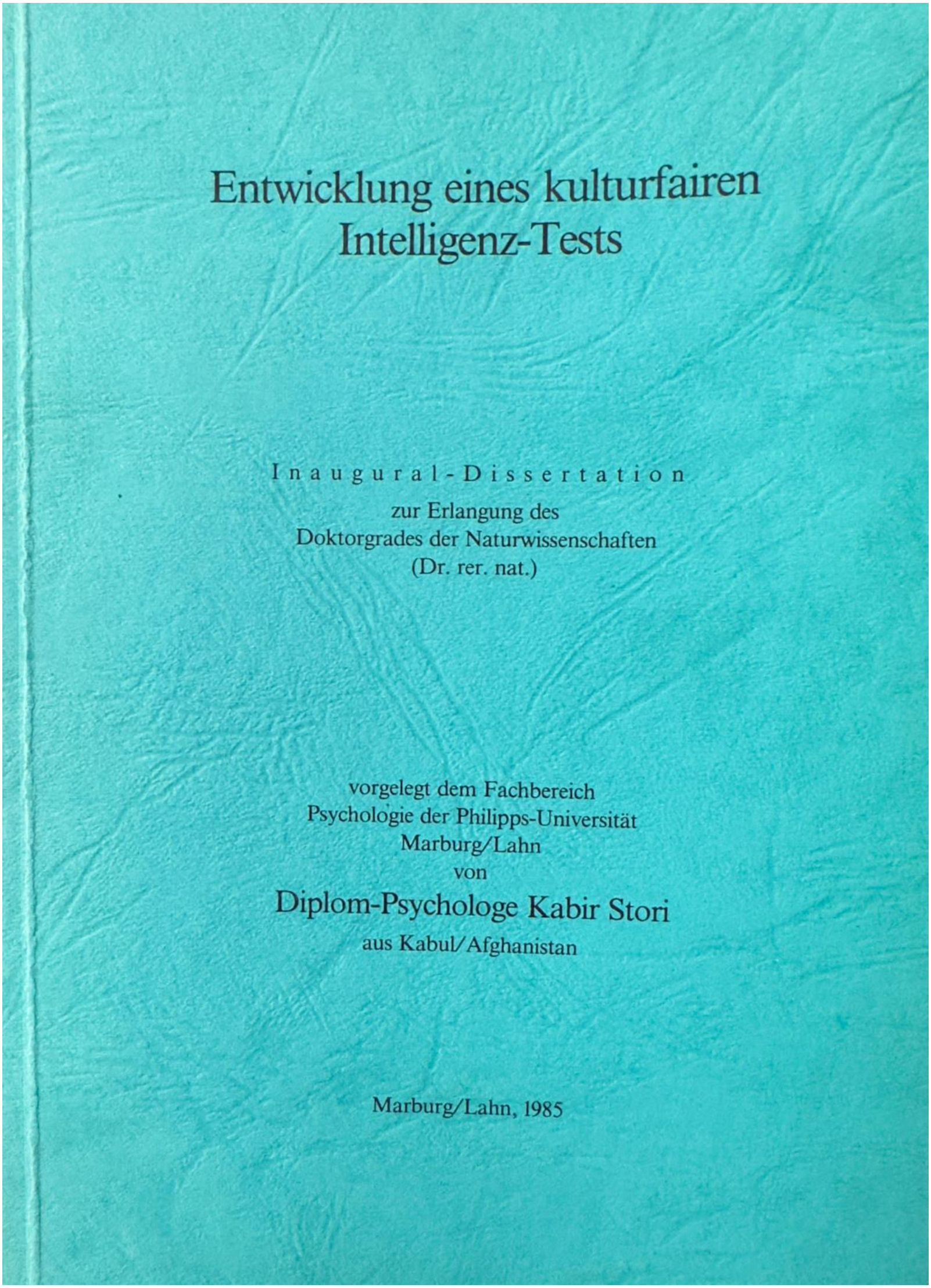
Entwicklung eines kulturfairen Intelligenz-Tests



Vom Fachbereich Psychologie der Philipps-Universität Marburg/Lahn als Dissertation angenommen am: 9.4.1985

Tag der mündlichen Prüfung am:
26.4., 7.5. und 13.5.1985
Berichterstatter:
Prof. Dr. Hartmann Scheiblechner
Mitberichterstatter:
Prof. Dr. Lothar Tent

Entwicklung eines kulturfairen Intelligenz-Tests

Inaugural-Dissertation zur Erlangung des
Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)
vorgelegt dem Fachbereich Psychologie der Philipps-Universität Marburg/Lahn
von

Diplom-Psychologe Kabir Stori
aus Kabul/Afghanistan

Marburg/Lahn, 1985
Deutschland

Inhaltsverzeichnis

	Seite
1. Einleitung	1
1.1. Intelligenzstruktur in verschiedenen Kulturen	1
1.2. Populationsabhängigkeit der Testergebnisse	11
1.2.1. Probabilistische Testmodelle	12
1.2.2. Das Rasch-Modell	14
1.2.3. Parameterschätzung	17
1.2.4. Modelltests	18
1.3. Der kulturfreie Intelligenz-Test	22
1.3.1. Entwicklung des Begriffs kulturfreier Intelligenz-Test	22
1.3.2. Kritik an der Kulturfreiheit der sprachfreien Intelligenz-Tests	23
1.3.3. Versuch einer Definition eines kulturfreien Intelligenz-Tests	31
1.4. Grundlage des Naturformen-Intelligenz-Tests (N-I-T)	35
1.4.1. Materialauswahl	35
1.4.2. Aufbau der Aufgaben des N-I-T	36

1.4.2.1. Naturformen-Intelligenz-Test, Vorform (N-I-Tv) 37
1.4.2.2. Endform des N-I-T 44
1.4.3. Die Beantwortungsmethode 44
2. Fragestellung 46
3. Methode 49
4. Die Untersuchung in Afghanistan 50
4.1. Einführung in die afghanische Gesellschaft 50
4.2. Gegenstand der Untersuchung in Afghanistan 56
4.3. Methode 57
4.4. Ergebnisse 62
4.4.1. Analyse nach der klassischen Testtheorie 62
4.4.2. Die Analyse der Tests anhand des Rasch-Modells nach der Methode der bedingten Wahrscheinlichkeit 65
4.5. Diskussion 68
5. Die Untersuchung in der Bundesrepublik Deutschland 73
5.1. Gegenstand der Untersuchung in der Bundesrepublik Deutschland 75
5.2. Methode 75
5.3. Ergebnisse 81

5.4. Diskussion 85
6. Die Analyse der Aufgaben des N-I-TB auf Anpassung des Rasch-Modells nach der Methode der bedingten Wahrscheinlichkeit anhand der verschiedenen Gruppen der Afghanen und Deutschen. 91
6.1. Fragestellung 91
6.2. Methode 91
6.3. Ergebnisse 92
6.4. Diskussion 97

7. Einige mit dem Rasch-Modell zusammenhängende Frage. 104
7.1. Innere Teilungskriterien 104
7.1.1. Fragestellung 104
7.1.2. Methode 104
7.1.3. Die Ergebnisse 105
7.1.4. Diskussion 105
7.2. Homogenisierung der Trennschärfe 107
7.2.1. Problemstellung 107
7.2.2. Methode 107
7.2.3. Die Ergebnisse 108
7.2.4. Diskussion 108

8. Zusammenfassende Diskussion 111
9. Zusammenfassung 117
10. Kurze Zusammenfassung 125
Literaturverzeichnis
Anhang A
Anhang B
Anhang C
Anhang D

Einleitung

Gegenstand dieser Arbeit ist die Fragestellung, ob es möglich ist, einen "kulturfairen" Intelligenz-Test zu konstruieren.
Diese Fragestellung hängt mit dem Problem der Intelligenzstruktur in verschiedenen Kulturen und der Populationsabhängigkeit von Testergebnissen zusammen. Aus diesem Grunde werden diese Probleme kurz dargestellt und diskutiert.

1.1. Intelligenzstruktur in verschiedenen Kulturen

Für quantitative Vergleiche der Intelligenz ist die qualitative oder strukturelle Ähnlichkeit der Intelligenz bei den zu untersuchenden Probanden oder Kulturgruppen erforderlich.

Wenn "diese Forderung nicht oder in zu geringem Maße erfüllt ist, wird jeder weitere quantifizierende oder andere Vergleichsversuch invalide, da die erhaltenen Daten unterschiedliche und damit unvergleichli-
che" (Schwebcke, Lück, Jandron,
1973, S. 440)
Intelligenzbereiche repräsentieren. In einigen Untersuchungen ist gezeigt worden, "daß die Intelligenzstruktur afrikanischer Stammesneger z.T. erheblich von der weißer Vergleichsgruppen abweicht",
(Schwebcke, Lück, Jandron, 1973,

S. 441). Selbst bei Weißen - und zwar innerhalb einer Kultur - gibt es eine Anzahl auf faktorenanalytischen Untersuchungen basierende Hypothesen, die auf unter-
schiedliche Intelligenzstrukturen bei verschiedenen Alters- und Leistungsgruppen

hinweisen, nämlich die Alters-, die Begabungs- und Leistungsdifferenzierungs-Hypothesen.
Die Altersdifferenzierungs-Hypothese beinhaltet, daß die Intelligenz sich im Laufe der Entwicklung differenziert, die Anzahl der Intelligenzfaktoren zunimmt und die Struktur der Intelligenz komplexer wird.

Die Begabungsdifferenzierungs-Hypothese besagt, daß Personen mit höherer Intelligenz eine differenziertere Intelligenzstruktur aufweisen als Personen mit niedriger Intelligenz.

Nach der Leistungsdifferenzierungs-Hypothese wird die Differenziertheit der Intelligenzstruktur als Funktion der absoluten Höhe der Leistung betrachtet, d.h., je höher die Leistung, desto differenzierter die Intelligenz- struktur.

Den Differenzierungs-Hypothesen läßt sich das Rasch-Modell als Nullhypothese gegenüberstellen, wonach für die Lösung einer Menge Intelligenzaufgaben

nur eine einzige invariante latente Fähigkeit ausschlaggebend ist.
Invariant bedeutet nicht, daß diese Fähigkeit im Laufe der Entwicklung stets konstant bleibt, "sondern, daß sich die Veränderung oder Entwicklung auf einer einzigen Dimension vollzieht" (F i s c h e r , 1974, S. 311). Hierzu ist insbesondere die Untersuchung von S p a d a (1970) zu erwähnen. S p a d a (1970) untersuchte 629 männliche und weibliche Schüler und Absolventen allgemeinbildender Höherer Schulen mit drei Intelligenztests, und zwar dem Matrizentest, dem Reihentest und dem Worttest und analysierte sie mit Hilfe des R a s c h - M o d e l l s i n b e z u g a u f Bedeutungsinvarianz gegenüber dem variablen Alter. Keiner dieser drei Tests war in ursprünglicher Form im Sinne des Rasch-Modells homogen, aber nach Ausschaltung e i n i g e r A u f g a b e n w u r d e i m F a l l e d e s Matrizentests und des Reihentests für die ganze Stichprobe der 11- bis 30jährigen und im Worttest nur für die Teilstichprobe der 14- bis 30jährigen Modellanpassung erzielt.
Das bedeutet, daß man bei dem Matrizen- und Reihen-Test die Leistung von 11jährigen und

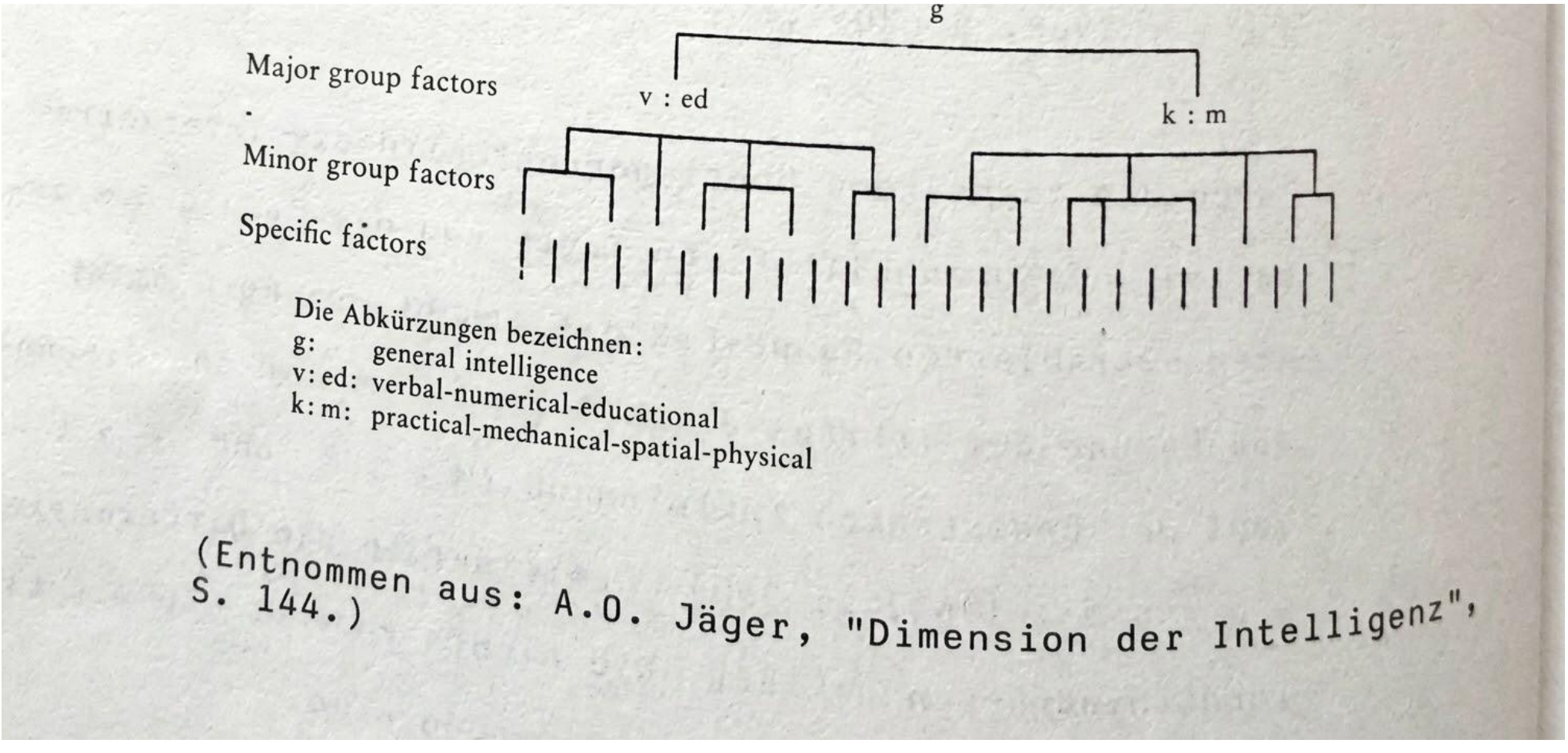
bei dem Worttest die Leistung von 14jährigen mit 20- oder 30jährigen direkt vergleichen kann. Aus der Gegenüberstellung des Rasch-Modells und den auf Faktorenanalysen basierenden Differenzierungs-Hypothesen

ergibt sich, daß unterschiedliche Auffassungen über die qualitative Entwicklung der Intelligenz letzten Endes auch eine Frage der Methode ist. Es kommt eben darauf an, welche Methode man anwenden und bevorzugen will. Darüberhinaus ist der Optimismus, den man in früheren Jahren der faktorenanalytischen Intelligenzforschung entgegengebrachte, inzwischen verebbt. Grundsätzliche Kritik richtet sich gegen die der Faktorenanalyse zugrunde gelegten Korrelationskoeffizienten, deren Höhe einerseits von den Testvariablen und andererseits von der jeweiligen Personengruppe abhängt. Merz und Kalver am (1965) sprechen in diesem Zusammenhang von "simultaner Überlagerung". Diese liegt dann vor, "wenn die Probanden einer Stichprobe nicht homogen in bezug auf Merkmale sind, die ihrerseits mit der

Testleistung korrelieren. Als derartige Überlagerungsvariablen können etwa das Alter, der Entwicklungsstand oder das Übungsniveau von Probanden fungieren" (A m e l a n g, L a n g e r, 1968, S. 205).
Durch die "simultane Überlagerung" wird die Interkorrelation und Kommunalität von Tests und die Ladung des ersten extrahierten Haupt-Faktors erhöht, dagegen nimmt die Ladung der letzten extrahierten Faktoren und die Anzahl der bedeutsamen Faktoren ab (M e r z und K a l - v e r a m, 1965, S. 290). Damit werden die Differenzierungs-Hypothesen geklärt, die nichts anderes als die Ergebnisse der simultanen Überlagerung sind.

Überträgt man diese Ergebnisse auf Resultate der faktorenanalytischen Intelligenzforschung im interkulturellen Bereich, so läßt sich die Feststellung der unterschiedlichen Intelligenzstruktur bei den verschiedenen Völkern auf "simultane Überlagerungs"-Effekte zurückführen.
Über die Intelligenzstruktur selbst gibt es verschiedene Auffassungen, die sich z.T.

widersprechen. So erklärt z.B. S p e a r m a n (1904, 1927) die Intelligenzleistung durch das Zusammenwirken eines allgemeinen (g) Faktors, der in jeder Leistung eine Rolle spielt, und eines spezifischen (s) Faktors, der für die Leistung in speziellen Aufgaben verantwortlich ist. Diese Auffassung ist später von der Englischen Schule, vor allem von B u r t (1949) und V e r n o n (1961) durch Hinzufügung von großen und kleinen Gruppenfaktoren zu einer hierarchischen Faktorentheorie der Intelligenz wie folgt erweitert worden.



(Entnommen aus: A.O. Jäger, "Dimension der Intelligenz", S. 144.)

(Entnommen aus: A.O. Jäger, "Dimension der Intelligenz", S. 144.

Nach der Auffassung von T h u r s t o n e (1938) gibt es eine Vielzahl von Faktoren, die für verschiedene Leistungen verantwortlich sind. Er spricht von neun nebeneinanderstehenden Faktoren, von denen sechs definiert sind, nämlich

a) verbal comprehension (sprachliches Verständnis)

b) verbal fluency (Schnelligkeit der Wortfindung)

c) space (Raumvorstellung oder Visualisation)

d) number (Rechengewandtheit)

e) memory (Gedächtnis)

f) reasoning (Denken = induktives und deduktives Denken).

M e i l i (1965, S. 31-140) spricht von vier Intelligenzfaktoren. M e i l i s Konzept der Intelligenzfaktoren beruht auf einer gestaltpsychologischen Theorie, nach der sämtliche Denkleistungen Prozesse sind, für die die Gestaltgesetze gelten. Er vertritt die Ansicht, daß es möglich sei, die Intelligenzfaktoren so zu definieren, daß sie mit der gestalttheoretischen Auffassung des

Denkens übereinstimmen, und charakterisiert von daher folgende Faktoren:

a) Komplexitätsfaktor

Es handelt sich dabei um eine Fähigkeit, die in komplexen Aufgaben gefordert wird. Dazu erklärt M e i l i (1965, S. 52), "daß es nicht so sehr auf die Anzahl oder Ausdehnung der Elemente ankommt, sondern auf die Art ihres Zusammenwirkens". Die Wirkung einfacher und kompli-

zierter Strukturen erläutert M e i l i am Beispiel eines etwa fünfjährigen Kindes, das fähig ist, Gewichte "zweipolig", d.h. nach schweren und leichten Gewichten, zu unterscheiden. Doch ist es nicht in der Lage, dieselben Gewichte in eine mehrgliedrige Rangordnung zu bringen.

b) Plastizitätsfaktor

Dieser Faktor bezeichnet die Fähigkeit zur Umstrukturierung und steht, wie M e i l i sagt, in engem Kontakt mit der geistigen Entwicklung des Kindes.

c) Ganzheitsfaktor

Der Ganzheitsfaktor wird bestimmt durch die Fähigkeit, getrennt gegebene Inhalte - z.B.

bekannte und neue Tatsachen - zu einem Ganzen zusammenzufügen.

d) Flüssigkeitsfaktor

Dieser Faktor charakterisiert die Fähigkeit des leichten Gleitens von einem Denkinhalt zum anderen.

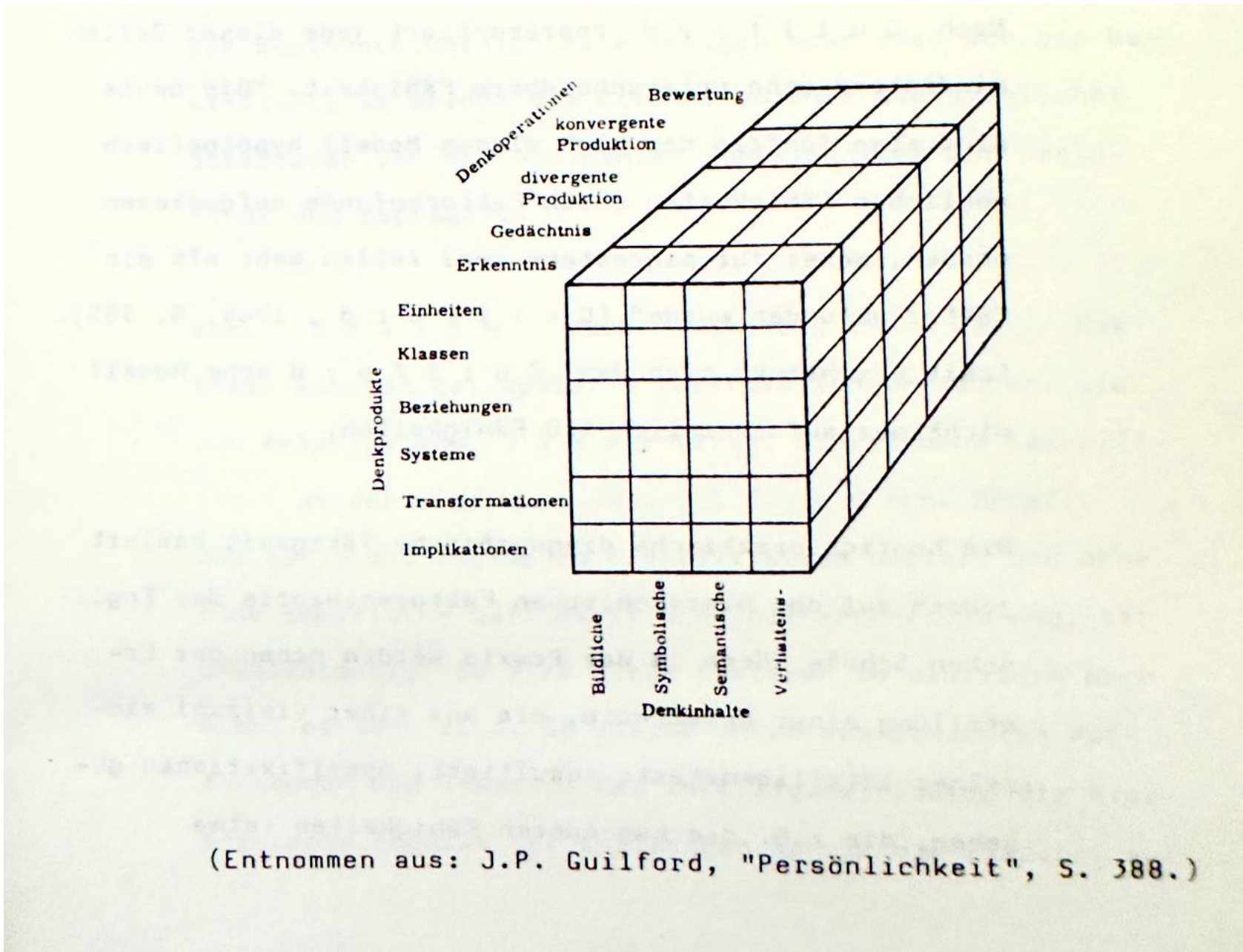
G u i l f o r d (1964, S. 353-397, 1967, S. 60-68) verweist in seinem theoretisch begründeten Intelligenzstruktur-Modell sogar auf 120 und mehr Intelligenzfaktoren. Er klassifiziert die intellektuellen Fähigkeiten nach folgenden drei Gesichtspunkten:

a) nach den Denkoperationen (Bewertung, k o n v e r g e n t e P r o d u k t i o n, d i v e r g e n t e Produktion, Gedächtnis und Erkenntnis),

b) nach den Denkprodukten (Einheiten, K l a s s e n, B e z i e h u n g e n, S y s t e m e, Transformationen und Implikationen),

c) nach den Denkinhalten (bildlich, symbolisch, semantisch, Verhalten).

Diese drei Klassifikationsgesichtspunkte ergeben ein dreidimensionales Schema, das sich graphisch wie folgt darstellen lässt:



(Entnommen aus: J.P. Guilford, "Persönlichkeit", S. 388.)

Seite 9

Dieser Würfel soll die Gesamtstruktur der Intelligenz aufzeigen. In der Quaderdarstellung „ist natürlich jedes einzelne Kästchen dreidimensional vorzustellen, da jedes einen Schnittpunkt von drei Dimensionen

repräsentiert: eine bestimmte Denkoperation ist mit einem bestimmten Denkinhalt und mit einer bestimmten Denkproduktion ausgezeichnet“ (Guilford, 1964, S. 389).

Das dreidimensionale Schema gliedert sich in fünf Arten von Denkoperationen, sechs Arten von Denkprodukten und vier Arten von Denkinhalten, also 5×6×4 Zellen. Derart ergibt sich die Summe von 120 Fähigkeiten.

Nach Guilford repräsentiert jede dieser Zellen mindestens eine unterscheidbare Fähigkeit. „Bis heute sind etwa fünfzig der nach diesem Modell hypothetisch möglichen Fähigkeiten durch Faktorenfunde aufgezeigt worden, wobei für mindestens zwei Zellen mehr als ein Faktor gefunden wurde“ (Guilford, 1964, S. 389).
Somit beschränkt sich das Guilfordsche Modell nicht nur auf lediglich 120 Fähigkeiten.
Die heutige praktische diagnostische Tätigkeit basiert jedoch auf der hierarchischen Faktorentheorie der Englischen Schule, denn in der Praxis werden neben der Erstellung einer Gesamtnote, die aus einer Vielzahl einzelner Intelligenztests resultiert, Spezifikationen

gegeben, die z.B. die besonderen Fähigkeiten (etwa

Abstraktionsfähigkeit, praktisch-rechnerisches Denken etc.) charakterisieren.

Auch Amthauer (1955, S. 5) stützt sich auf die hierarchische Faktorentheorie der Englischen Schule, indem er ablehnt, daß Intelligenzleistungen und die ihnen zugrunde liegenden Fähigkeiten ein bloßes Nebeneinander bilden. Er formulierte seine Arbeitshypothese dahingehend, „daß alle Leistungen und die ihnen zugrunde liegenden, gedachten Fähigkeiten in einem gefügehaften Zusammenhang stehen, und daß sie jeweils eine Struktur aufweisen. Eine Struktur nämlich, in der die Glieder hierarchisch geordnet sind. Wie in einer Hierarchie der dominante Bereich alle übrigen mehr oder weniger beeinflußt, so werden die Glieder solcher hierarchischer Strukturen von den dominanten Schwerpunkten her beeinflußt und geprägt“.

Weder Meilis noch Thurstones Faktoren konnten bei späteren Verifikationsversuchen, wie zum Beispiel von Jäger (1967) hinreichend identifiziert werden. Auch das Guilfordsche Modell, das auf einer logischen Klassifikation basiert und ohne eine empirische Gültigkeit sinnvoll erscheinen mag, ist unvollständig, da alle diese Faktoren im einzelnen noch nicht bekannt sind. So bleibt unter den bekannten Auffassungen und Theorien der Intelligenzstruktur die hierarchische Theorie der Englischen Schule die einzige, auf

der nicht nur die praktische diagnostische Tätigkeit der Intelligenz gründet, sondern auch die theoretische Grundlage der interkulturellen Intelligenzforschung.

1.2. Populationsabhängigkeit der Testergebnisse

Die Populationsabhängigkeit bzw. Stichprobenabhängigkeit ist eine Eigenschaft der klassischen Testtheorie. Selbst die wichtigsten Grundbegriffe dieser Theorie,

nämlich Zuverlässigkeit, Gültigkeit und Homogenität (Gleichheit der Faktorenstruktur bei allen Items) basieren auf Korrelationskoeffizienten, die populationsabhängigen Charakter haben. Dies bedeutet, daß die Interpretation dieser Begriffe nur in bezug auf eine bestimmte Population gerechtfertigt ist. (Fischer, 1974, S. 133, Boesch und Eckensberger, 1969, S. 548–549)

Die Eigenschaft der Populationsabhängigkeit der klassischen Testtheorie ist ein großer Nachteil dieser Theorie, insbesondere im interkulturellen Bereich, da man mit einem Meßinstrument bzw. Intelligenztest, der populationsabhängig ist, die Angehörigen verschiedener Kulturen hinsichtlich des zu messenden Merkmals bzw. der Intelligenz nicht vergleichen kann.

Selbst innerhalb eines Kulturkreises kann man mit den Tests bzw. Intelligenztests, die nach der klassischen Testtheorie entwickelt worden sind, die Intelligenz z.B. …

Site 12

eines 14jährigen mit der eines 20jährigen nicht vergleichen. Ein Vergleich ist nur in bezug auf die jeweilige Altersgruppe möglich.

Aus diesem Grunde haben in den letzten Jahren die probabilistischen Testmodelle die Aufmerksamkeit der Psychologen auf sich gelenkt.

(Anm.: Bei der Erläuterung der probabilistischen Testmodelle wird aus drucktechnischen Gründen e anstelle von "epsilon" und x anstelle von "Xi" verwendet.)

1.2.1. Probabilistische Testmodelle: Der Grundgedanke probabilistischer Testmodelle ist, daß die beobachtbaren (manifesten) Variablen durch Wahrscheinlichkeitsrelationen mit der latenten Dimension verbunden sind, wobei die manifesten Variablen als Indikatoren oder Symptome der latenten Dimension angesehen werden und nicht mit der latenten Dimension gleichzusetzen sind. (Fischer, 1968, S. 114; 1974, S. 148–149)

Die allgemeine mathematische Form eines dichotomen probabilistischen Testmodells für

einen psychologischen bzw. für einen Intelligenz-Test sieht folgendermaßen aus:

$$P(+/ i, x) = f_i(x)$$

Site 13

Das bedeutet: jede Person mit der Fähigkeit x hat dieselbe Chance $f_i(x)$ die Aufgabe i richtig zu lösen. Die Funktion $f_i(x)$ heißt Aufgabencharakteristik. Es wird zwischen monotonen und nicht monotonen Aufgabencharakteristiken unterschieden. Bei Intelligenztests findet man die meisten Aufgaben mit monotonen Charakteristiken. Durch Spezifikation der Funktion $f_i(x)$ entsteht die spezielle Struktur eines probabilistischen Modells. Dazu werden Zusatzannahmen gemacht. Eine wichtige Zusatzannahme ist die "lokale stochastische Unabhängigkeit" der Antworten. Diese Annahme läßt sich durch folgende Gleichung ausdrücken:

$$P(+ \text{ bei i und } + \text{ bei j}/x) = P(+ \text{ bei i}/x)\, P(+ \text{ bei j}/x)$$

Das bedeutet, daß die Wahrscheinlichkeit, mit der eine Vp mit der Fähigkeit x die Aufgabe i

und j richtig löst, sich aus dem Produkt der beiden einzelnen Wahrscheinlichkeiten zusammensetzt.

Die Annahme der stochastischen Unabhängigkeit beinhaltet, daß der statistische Zusammenhang zwischen Aufgaben, d.h. die Korrelationen untereinander, nur auf ihre gemeinsame Bindung an die latente Dimension x zurückzuführen ist und auf keine andere Beziehung. Würde diese latente Dimension herauspartialisiert, so müßten die Korrelationen zwischen den Aufgaben verschwinden

(Fischer, 1968, S. 82; 1974, S. 157). Hier ist die Parallele zur Faktorenanalyse zu entdecken, die auf der ähnlichen Annahme beruht, daß nach Extraktion aller Faktoren die Korrelationen zwischen den Aufgaben nur noch zufällig um Null variieren.

Auch durch die Annahme lokaler stochastischer Unabhängigkeit wird die Struktur eines probabilistischen Modells noch

nicht bestimmt. Es bedarf weiterer zusätzlicher Annahmen. Nach Art dieser zusätzlichen Annahmen unterscheiden sich verschiedene Modelle. Die wichtigsten dieser probabilistischen Modelle sind von Lazarfeld (1950), Birnbaum (1958) und Rasch (1960) entwickelt worden. In der psychologischen Praxis hat sich jedoch bisher das probabilistische Testmodell von Rasch (1960, 1961, 1966) am besten bewährt (s. dazu Spada, 1970, Fischer und Spada, 1970, Stapf, 1970, Wendeler, 1968) und "als neue Methode der Testkonstruktion in verschiedenen Gebieten der Psychologie weite Verbreitung gefunden" (Stelzl, 1972, S. 298).

1.2.2. Das Rasch-Modell: Das Rasch-Modell "ist ein indirektes Verfahren der eindimensionalen Indikatorskalierung, das zwei Reaktionskategorien zuläßt (ja/nein), richtig/ falsch oder gelöst/nicht gelöst)" (Wakenhut, 1974, S. 34).

Site 15

Der dänische Mathematiker Rasch (1960) ging von der Gegenüberstellung zweier latenter

Parameter aus, nämlich der Fähigkeit x der Personen und Leichtigkeit e der Aufgaben. Die Fähigkeit x wird durch die Menge der Aufgaben repräsentiert. Die Antworten der Personen und die Aufgaben dienen als Indikatoren der beiden latenten Parameter. Diese Gegenüberstellung verweist auf die symmetrische Eigenschaft des Modells, in der die Rollen von Aufgaben und Personen prinzipiell austauschbar sind (Wakenhut, 1974, S. 34).

Der Angelpunkt des Rasch-Modells besteht in der Bestimmung der Funktion $f_i(x)$ durch das Postulat der spezifischen Objektivität, die sein Modell charakterisiert.

$$P\ (a_{vi}/x_v, e_i) = (x_v e_i)^{a_{vi}} / (1 + x_v e_i)$$

Die Wahrscheinlichkeit, daß ein Proband v mit der Fähigkeit x_v die Aufgabe i mit der Leichtigkeit e_i positiv ($a_{vi} = 1$) bzw. negativ ($a_{vi} = 0$) beantwortet, ist die angegebene Funktion von x_v und e_i. Sie steigt monoton mit x für $a_{vi} = 1$. Der Begriff der "spezifischen Objektivität" basiert auf der Überlegung, daß die Wissenschaft nicht nur aus der Überprüfung von Tatsachenfeststellungen besteht, sondern

auch aus objektiven Vergleichen. Die Begriffe Objektivität und Vergleiche sollen

Site 16

dadurch geklärt werden, daß sie einen eindeutigen und spezifischen Sinn erhalten (Fischer, 1968, S. 109–110; 1974, S. 407–408). Es soll daher eine endliche oder unendliche Menge von Objekten und eine endliche oder unendliche Menge von Agentien angenommen werden. In der Psychologie können die Objekte Vpn und die Agentien Aufgaben sein. Vergleichen heißt, das Objekt mit einem Agens, z.B. in der Psychologie eine Vp mit einer Aufgabe in Kontakt zu bringen. Jede Vp, die mit einer Aufgabe in Kontakt gebracht wird, erzeugt ein bestimmtes Resultat, nämlich: gelöst oder nicht gelöst. Bringt man 2 Vpn mit einer Anzahl von Aufgaben in Kontakt, so läßt sich aus den beobachteten Resultaten eine Aussage über das Verhältnis der beiden Personen machen. Diese Aussage bezieht sich nicht auf die beobachteten Resultate, sondern auf die Eigenschaft der Vpn, nämlich die Fähigkeiten. (Fischer, 1968, S. 110; 1974, S. 409) Wenn die Aussage nicht davon abhängt, welche und

wieviele Aufgaben aus einem homogenen Intelligenztest ausgewählt wurden, dann heißt die Aussage spezifisch objektiv (Fischer, 1968, S. 111; 1974, S. 409). Das heißt mit anderen Worten, daß das Verhältnis zwischen zwei Fähigkeiten bei jeder dieser Aufgaben gleich oder konstant bleiben soll. Umgekehrt gilt dies wegen der symmetrischen Eigenschaft des Modells auch für das Verhältnis der Schwierigkeit bzw. Leichtigkeit zweier Aufgaben, die von der gewählten Stichprobe bzw. Refe-

renzgruppe unabhängig sein soll und daher auch populationsunabhängigen Charakter haben soll (Fischer und Scheiblechner, 1970, S. 24; Fischer, 1968, S. 111; 1974, S. 409).

Das Postulat der spezifischen Objektivität, nämlich der aufgabenunabhängige Probanden-Vergleich und probandenunabhängige Aufgaben-Vergleich weist auf die Möglichkeit hin, daß die beiden Parameter “Fähigkeit der Probanden” und “Leichtigkeit der Aufgaben” unabhängig voneinander bestimmt werden

können. Die Aufgabenschwierigkeiten sollen nicht in die Bestimmung der Personenparameter eingehen. Die Voraussetzung dafür ist, daß im Rasch-Modell nur erschöpfende Statistiken verwendet werden, bei denen die Aufgaben trotz der unterschiedlichen Schwierigkeiten gleich gewichtet werden und die Anzahl der gelösten Aufgaben die gesamte Information über die Personenparameter liefert (Fischer, 1974, S. 184–200; 1968, S. 94–105). Das führt dazu, daß alle Probanden mit der gleichen Anzahl gelöster Aufgaben zu einer Rohwertgruppe zusammengefaßt werden und den gleichen Parameterwert erhalten.

1.2.3. Parameterschätzung: Im praktischen Testfall sind die genauen Wahrscheinlichkeiten nicht bekannt, deshalb müssen diese und damit auch die Parameter x_v und e_i mit
Site 18

Hilfe der empirischen Lösungshäufigkeiten geschätzt werden. Ausgangspunkt der Parameterschätzung ist eine Antwortmatrix, deren Spalte nach der Lösungshäufigkeit der verschiedenen Aufgaben in der untersuchten

Stichprobe und die Zeilen nach den Probandenrohwertgruppen angeordnet sind (Kuhl, 1977, S. 63; Wakenhut, 1974, S. 42). Dabei bilden die Probanden mit der gleichen Anzahl richtig gelöster Aufgaben eine Rohwertgruppe. Alle Probanden einer Rohwertgruppe erhalten den gleichen Wert x_v, denn die Anzahl der gelösten Aufgaben bildet die erschöpfende Statistik. Personen, die keine oder alle k-Aufgaben richtig gelöst haben, können nicht in die Antwortmatrix aufgenommen werden, da über sie keine quantitative Leistungsangabe möglich ist. Die Zahl der richtigen Antworten variiert deshalb zwischen 1 und k-1 (Fischer/Scheiblechner, 1970, S. 27; Wakenhut, 1974).

Für die Parameterschätzung gibt es verschiedene Methoden. Die theoretisch befriedigendste ist jedoch die bedingte Maximum-Likelihood-Methode zur Schätzung unbekannter Verteilungsparameter (Fischer, 1974, S. 230), in der die Personenparameter x_v und Itemsparameter e_i unabhängig voneinander geschätzt werden.

1.2.4. Modelltests: Spezifische Objektivität ist eine grundlegende Eigenschaft und ein großer Vorteil des

Site 19

Rasch-Modells. “Spezifische Objektivität ist in einem Datensatz dann und nur dann gegeben, wenn die Items homogen sind, also das Postulat der lokalen stochastischen Unabhängigkeit erfüllt ist, und gleichzeitig aber auch die Items logistische Itemcharakteristikfunktionen aufweisen” (Wakenhut, 1974, S. 44). Diese beiden Bedingungen werden simultan überprüft. Zur Überprüfung der Modellvoraussetzungen gibt es verschiedene Verfahren, die sich in parameterfreie und parametrische unterteilen lassen. Die parameterfreien Verfahren scheitern bis jetzt an ungelösten rechentechnischen Problemen (s. Stene, 1968). Aus diesem Grunde arbeitet man praktisch mit dem parametrischen Modelltest. Am meisten angewandt werden die graphische Modellkontrolle, der Modelltest nach Fischer und Scheiblechner (1970) und der Likelihood Quotienten-Test von Andersen (1973), also ein globaler Modelltest. Trotz der

Kritik, die Van den Wollenberg (1979, S. 31–34) an dem Fischer-Scheiblechner-Modelltest übt, schreibt er: “The Fischer-Scheiblechner approach has an important advantage over the likelihood ratio test. In the present approach, contributions to the statistic of individual items are known; this is very valuable when sources of model violation are sought for” (Van den Wollenberg, 1979, S. 32). Aus diesem Grund verliert der Fischer-Scheiblechner-Modelltest trotz seiner Mängel in der Testtheorie nicht seinen Wert. Er kann zur Konstruktion spezifisch objektiver

Tests verwendet werden, denn er ist gegenüber der graphischen Modellkontrolle exakt und gegenüber dem globalen Modelltest von Andersen (1973) hat er den Vorteil, daß er Hinweise auf störende Items liefert, durch deren Selektion Modellanpassung erzielt werden kann.

Der wichtigste Einwand gegen die t-Statistik zur Anpassungsprüfung eines Einzelitems und die X^2-Statistik zur Prüfung der globalen

Modellgeltung von Fischer und Scheiblechner besteht darin, daß sie normierungsabhängig sind. So kann z.B. die Wahl unterschiedlicher Einheitsitems die Ergebnisse ändern. Wesentliche Verzerrungen sind jedoch bei Produktnormierung in der Regel nicht zu erwarten. Deshalb können die Methoden von Fischer und Scheiblechner mit gebotener Vorsicht weiter verwendet werden. Das derzeit beste Vorgehen besteht in der Verwendung des Likelihood-Quotententests von Andersen zur globalen Modelltestung und der Q_1 und Q_2-Statistik von van den Wollenberg zur Itemselektion, da die entsprechenden Programme jetzt zur Verfügung stehen.

Fischer und Scheiblechner benutzten die Informationsfunktion zur Konstruktion einer Testgröße für die Kontrolle der Modellgeltung (Fischer und Scheiblechner, 1970, S. 32–34,

Fischer, 1974, S. 296–297).

Teilt man wie in der graphischen Modellkontrolle die Personenstichprobe nach

einem geeigneten Kriterium (z. B. Testrohwert, Alter, Geschlecht, Nationalität usw.) in zwei Teilgruppen und ermittelt unabhängig voneinander denselben Itemparameter, so ergeben sich die Schätzungen $e_i^{(1)}$ und $e_i^{(2)}$. Da die der Parameterschätzung zugrundeliegenden Maximum-Likelihood-Schätzungen asymptotisch normalverteilt sind mit dem Erwartungswert gleich dem Parameterwert und der Varianz v gleich dem Kehrwert der Informationsfunktion $I(e_i)$, lassen sich die Differenzen zweier Parameterschätzungen in eine asymptotisch standardnormalverteilte Variable Z_i verwandeln

$$Z_i = (e_i^{(1)} - e_i^{(2)}) / \sqrt{(v_1 + v_2)} \qquad v_1 = 1 / I^{(1)}(e_i)$$
$$v_2 = 1 / I^{(2)}(e_i)$$

Fischer und Scheiblechner haben die Z_i-Werte quadriert und über alle Items i=1,...,k aufsummiert, um eine annähernd Chi-Quadrat-verteilte Prüfgröße für den ganzen Test zu erhalten.

$$X^2 = \Sigma_{i=1}^{k} Z_i^2 \qquad \text{mit } df = k-1$$

Der Z-Test kann bei der Konstruktion homogener Tests

zur Aufgabenselektion benutzt werden. Dabei wird die Aufgabe mit dem größten Z-Wert ausgeschlossen, jedoch jedes Mal nur eine einzige Aufgabe, da die Parameterschätzung und der Z-Test nicht ganz unabhängig voneinander sind. An der X^2-Prüfgröße wird überprüft, ob der Gesamttest dem Modell angepaßt ist.

1.3. Der kulturfreie Intelligenz-Test

1.3.1. Entwicklung des Begriffs kulturfreier Intelligenz-Test

Kurz vor dem 1. Weltkrieg begann man in den Vereinigten Staaten von Amerika zur Auslese von Rekruten (unterschiedlicher nationaler Herkunft, Sprache, Schulausbildung) einen Intelligenz-Test zu entwickeln. Diese Bemühungen wurden verstärkt durch das zunehmende Interesse an der interkulturellen Forschung.

So wurde eine sprachfreie Version des “Army-Alpha-Test” unter dem Motto ‘Chancengleichheit für alle’ entwickelt, der unter dem Namen “Army-Beta-Test” bekannt wurde (Groffman, 1971, S. 181, Drenth, 1969, S. 23), der erstmals während des 1. Weltkrieges Verwendung fand. Seitdem wurden eine große Anzahl sprachfreier Intelligenz-Tests entwickelt. Dazu zählen der “Progressive Matrices Test” des englischen Psychologen Raven (1938) und der “Figure Reasoning Test” von Daniels (1949).

Mit der Zeit verstärkte sich die Meinung, daß diese sprachfreien Intelligenz-Tests kulturunabhängig seien. So nannte z.B. Cattel (1961) seinen Intelligenz-Test (CFT) infolge der Sprachfreiheit “culture free”.

1.3.2. Kritik an der Kulturfreiheit der sprachfreien Intelligenz-Tests: Zwischenzeitlich hat die Kritik bezüglich der Kulturunabhängigkeit sprachfreier Intelligenz-Tests zugenommen. So stellte z.B. Merz (1968) fest, daß die Leistung bei einem Nicht-

Verbalen-Intelligenz-Test sich erheblich steigern kann, wenn man die Versuchspersonen (Vpn) auffordert, alles, was sie sehen und denken, laut auszusprechen bzw. zu verbalisieren. Dagegen hat die Verbalisierung bei einem verbalen Intelligenz-Test geringen oder fast gar keinen Einfluß. Diese Tatsache wird dahingehend interpretiert, "daß gerade bei solchen anscheinend sprachfreien Tests die Sprache und damit die kulturelle Umwelt doch einen erheblichen Einfluß hat.

Einmal mag die kulturelle Umwelt Verbalisieren bei Denkaufgaben nahelegen oder unterdrücken: In unserer westlichen Zivilisation wird etwa – im Gegensatz zu früheren Zeiten – lautes Lesen und Denken systematisch aberzogen. Zum anderen mag die kulturelle Umwelt Wörter für Kategorien bereithalten, die geeignet sind, die in den Testaufgaben relevanten Gesichtspunkte zu verbalisieren; in einer anderen Kultur mögen die entsprechenden Kategorien

fehlen oder ungebräuchlich sein, wodurch Verbalisieren ausgeschlossen wird. Die Kennzeichnung solcher Tests als 'kulturfair' wird damit zweifelhaft." (Merz, 1968, S. 132)

Dies verdeutlicht, daß die sprachfreien Intelligenz-Tests entgegen ihrer Benennung doch nicht sprachfrei sind.

Drenth und van der Flier (1976, S. 140) schreiben: "A great many non-verbal tasks are just as dependent on cultural factors as verbal tasks are". Ebenso stellt sich in neuerer Zeit heraus – und das immer deutlicher –, daß sprachfreie Tests u.U. sogar eher kulturellen Einflüssen unterliegen als sprachliche Tests (Schwebcke, u.a., S. 438, Merz/Stelzl, 1977, S. 80–83).

Damit ist die Ansicht, daß verbale Intelligenz-Tests nicht kulturfrei gemacht werden können, in das Gegenteil verkehrt. Das bedeutet, daß verbale Tests u.U. eher geeignet sind, kulturfair zu werden.

Jedoch liegt bei einigen Ländern die Schwierigkeit darin, daß noch nicht alle

Sprachen, die gesprochen werden, zur Schriftsprache geworden sind (z.B. in Afghanistan die turkmenische und nuristanische Sprache, u.a.).

In einer früheren Untersuchung, die ich (während meines Studiums im Rahmen einer Jahresarbeit 1974 an der Universität Frankfurt am Main) in Afghanistan und in der Bundesrepublik Deutschland durchgeführt habe, wurde festgestellt, daß bei einer Umstellung der geometrischen Anordnung der Teile eines Items des Figure Reasoning Tests im Vergleich zum originalen Figure Reasoning Test die Ergebnisse sich änderten.

Vor Beginn der Untersuchung wurden die Aufgaben des Figure Reasoning Tests sowie das ganze Figure Reasoning Testheft und auch die Antwortbögen, die von links nach rechts angeordnet sind, von rechts nach links umgestellt. Dann wurden die Untersuchungen an zwei Gruppen (A und B) von jeweils 100 Schülern im Alter von vierzehn bis siebzehn Jahren in Afghanistan sowie in der

Bundesrepublik Deutschland durchgeführt. Die Gruppe A bekam den Figure Reasoning Test (FRT), die Gruppe B erhielt den umgestellten Figure Reasoning Test (UFRT). Nach etwa einem Monat wurde unter Inversion der Bedingungen die Untersuchung in Afghanistan sowie in der Bundesrepublik Deutschland wiederholt. Die Zeit zur Beantwortung der Aufgaben wurde bei beiden Gruppen sowohl in Afghanistan als auch in der Bundesrepublik Deutschland auf dreißig Minuten beschränkt.

Es ergab sich, daß die afghanischen Schüler beim umgestellten Figure Reasoning Test (UFRT), der ihrer Schreib-

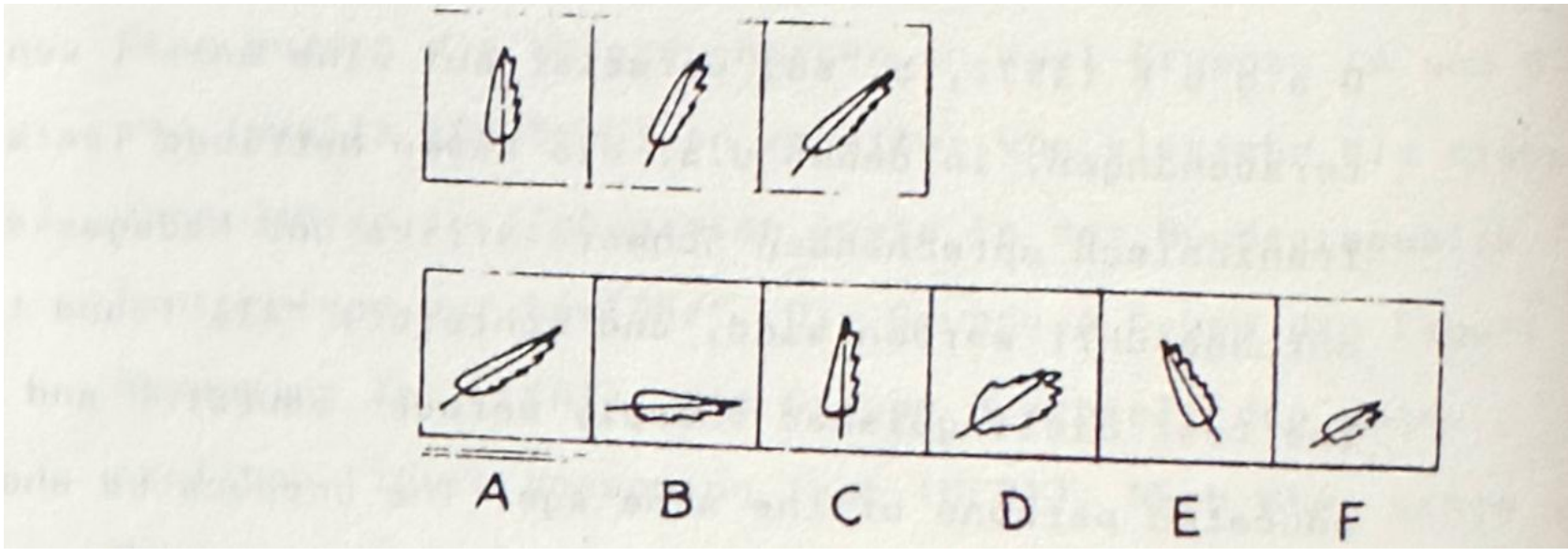
weise angepaßt war, im Durchschnitt signifikant mehr Aufgaben richtig lösten als beim Figure Reasoning Test (FRT), bei dem sie gegen ihre Denk- und Schreibgewohnheiten vorgehen mußten, wohingegen bei den deutschen Schülern das umgekehrte Ergebnis eintrat.

Daraus kann gefolgert werden, daß das Denken durch Erziehung bzw. Sozialisation in bestimmte Richtungen geprägt wird, und daß außer der Sprache noch weitere Faktoren bei figuralen Intelligenz-Tests, nämlich die Richtung der Aufgabenbeantwortung eine Rolle spielen, die kulturgebundenen Charakter haben.

Dague (1972, S. 68) verweist auf eine Anzahl von Untersuchungen, in denen u.a. die Raven Matrices Tests im französisch sprechenden Schwarz-Afrika und Madagaskar durchgeführt worden sind, und schreibt: "All found that the test distinguished sharply between educated and uneducated persons of the same age. The uneducated showed certain characteristic bad responses. Scores were generally proportional to the length of schooling and inversely proportional to the time elapsed since leaving school. Illiterate adults did not get beyond the 5- or 6-year level". Er schreibt weiter, daß mit dem "Culture free" Intelligenz-Test von Cattel dieselben Ergebnisse erzielt worden sind (Dague, 1972, S. 68). Der Test darf und soll im Ausprägungsgrad der Dimension differieren, aber nicht in der Faktorenstruktur. Diese Tests

haben aber zwischen Ausgebildeten und Analphabeten so stark unterschieden, daß man all diese erwachsenen Analphabeten, die nicht einmal ein Intelligenzniveau von Fünf- oder Sechsjährigen erreichten, als schwachsinnig bezeichnen müßte, was gegen die allgemeine, alltägliche Erfahrung, die man mit Analphabeten gemacht hat, spricht.

Ebenso analysierte Liungman (1973) einige Aufgaben von sprachfreien Intelligenztests (Raven Matrices und culture free-Test von Cattel) hinsichtlich der Kulturunabhängigkeit bzw. -abhängigkeit. Er erläuterte dieses u.a. am Beispiel folgender Aufgabe des culture free-Tests von Cattel:



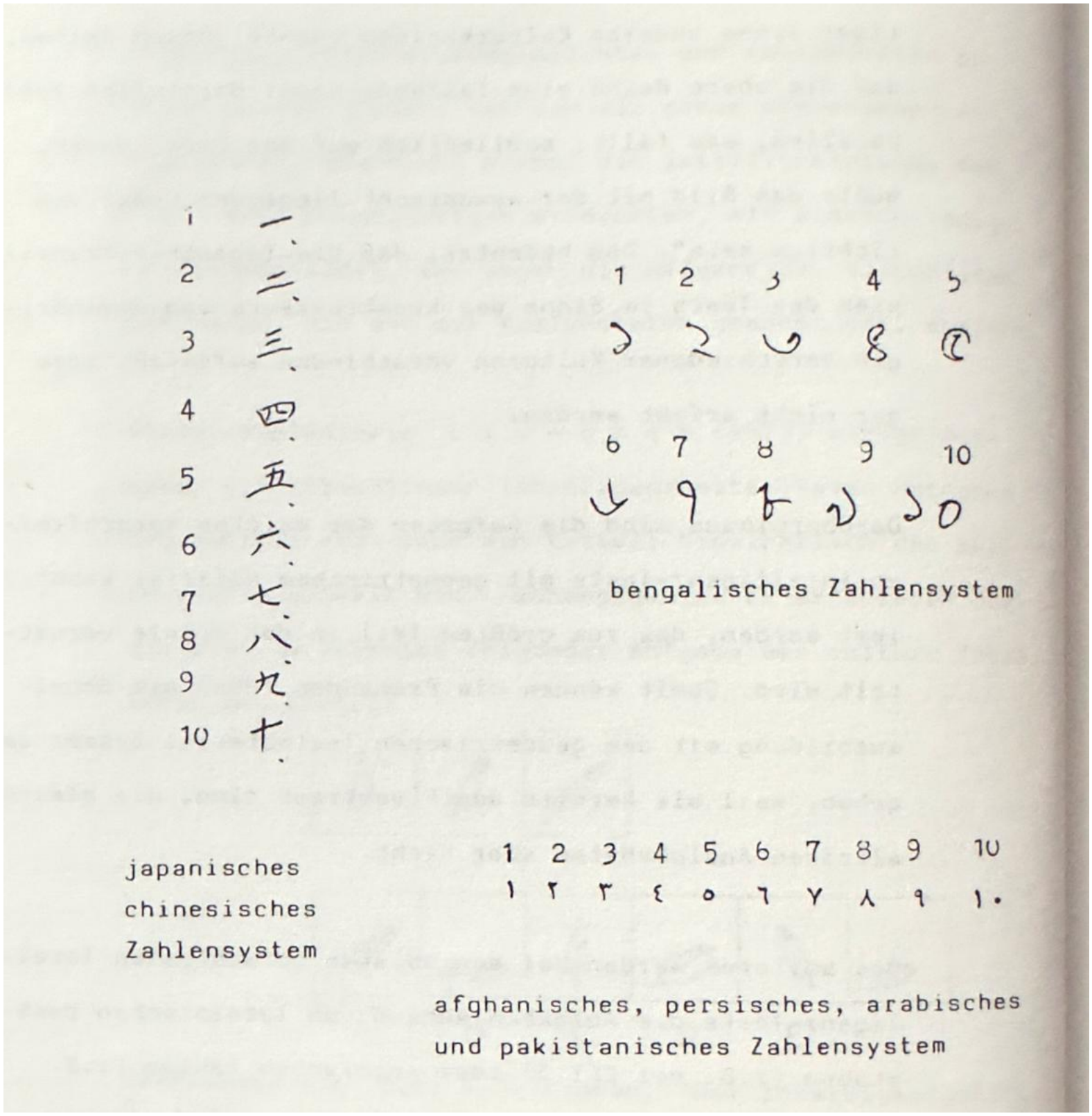
(Entnommen aus: Carl G. Liungman, “Der Intelligenzkult”, S. 126)

Liungman (1973, S. 124–125) schreibt: “Die Bilder der oberen Reihe sollen eine gleichförmige Bewegung (beispielsweise die Bewegung des Uhrzeigers) darstellen, die in zeitlich gleichen Abständen abgebildet wurden. Die Lösung soll das Bild sein, das diese Bewegung in einem weiteren, gleichen Intervall darstellt.*) *) Gemeint ist Antwort A. Ein Angehö-

riger eines anderen Kulturkreises könnte jedoch meinen, daß die obere Reihe eine fallende Feder darstellen soll. Da alles, was fällt, schließlich auf der Erde landet, müßte das Bild mit der waagerecht liegenden Feder das richtige sein”. Das bedeutet, daß die logischen Prinzipien des Tests im Sinne des Konstrukteurs von Angehörigen verschiedener Kulturen verschieden aufgefaßt oder gar nicht erfaßt werden.

Darüberhinaus sind die Aufgaben der meisten sprachfreien Intelligenz-Tests mit geometrischem Material konstruiert worden, das zum größten Teil in der Schule vermittelt wird. Somit können die Probanden (Pbn) mit Schulausbildung mit dem geometrischen Testmaterial besser umgehen, weil sie bereits damit vertraut sind, die gleichaltrigen Analphabeten aber nicht.

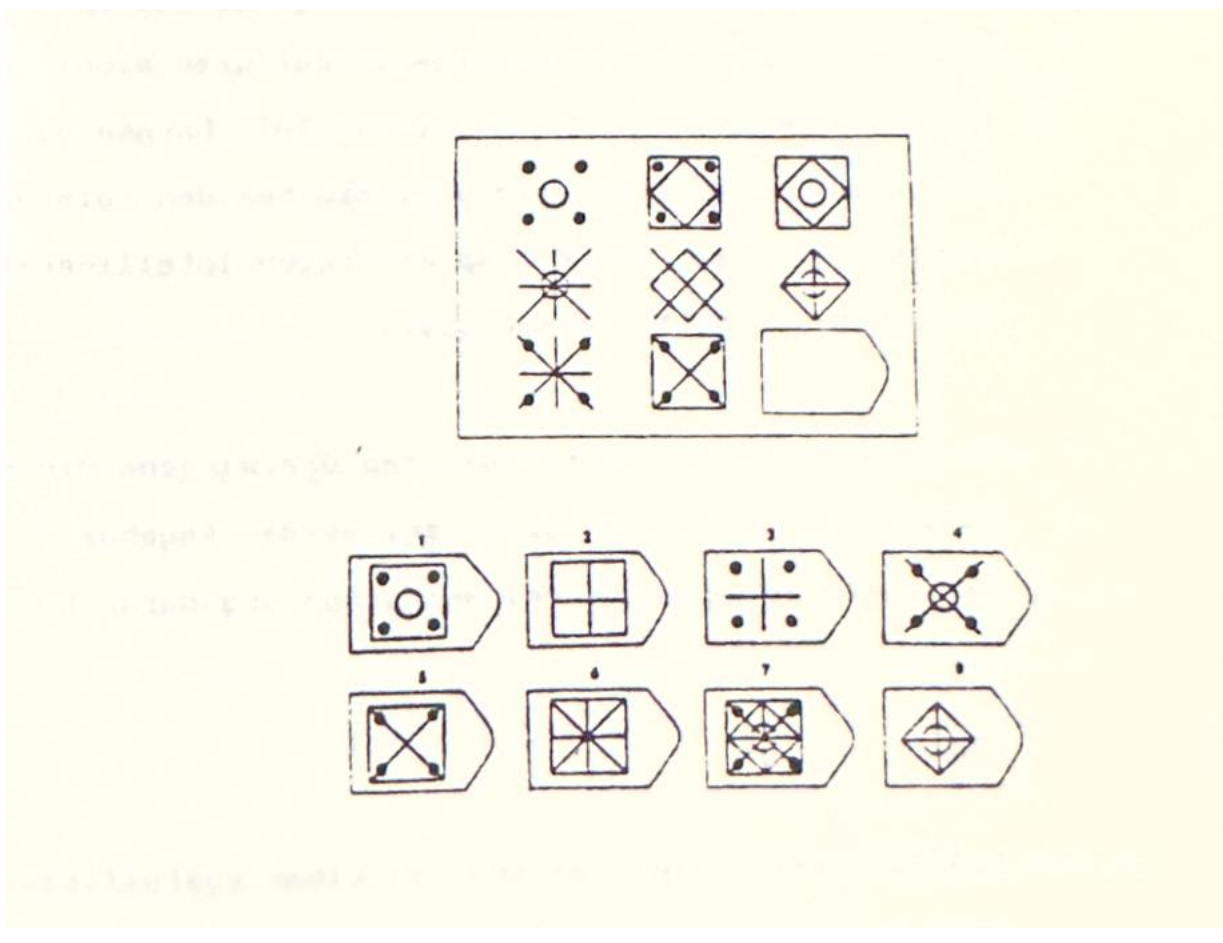
Des weiteren werden bei den meisten sprachfreien Intelligenz-Tests die Aufgaben anhand von lateinischen Buchstaben (z.B. bei CFT 3) oder arabischen Zahlen (z.B. bei FRT, PMT) von den Pbn beantwortet. Da in einigen Kulturen, wie z.B. in China, Afghanistan oder Bangladesh, statt des arabischen Zahlensystems ein anderes Zahlensystem benutzt wird, ist es unangemessen, daß die Testaufgaben ausschließlich mit dem arabischen Zahlensystem bzw. dem lateinischen Alphabet beantwortet werden, weil die Angehörigen verschieder Kulturen damit unterschiedlich vertraut sind.



[Auch bei einigen Aufgaben der sprachfreien Intelligenz-Tests scheint die Methode der A u f g a b e n b e a n t w o r t u n g schulausbildungsgebundenen Charakter zu haben. Z.B. läßt sich folgende Aufgabe des “Standard Progressive Matrices”-Tests auf zwei verschiedene Weisen beantworten: einmal über die Zeile oder Spalte durch die Addition der

beiden ersten komplexen Figuren, wobei die gleichen

Teile beider sich aufheben und die ungleichen Teile die letzte Figur ergeben. Also ist sie eine Art Plusminus-Aufgabe, die in der Schule gelernt wird. Zum anderen kann die Aufgabe beantwortet werden über das Gesamtfeld durch die Zerlegung der “komplexen Figuren” in einfache geometrische Figuren, die man aus der Schule kennt, z.B. Pluszeichen, Malzeichen,



Quadrat usw. Jede Art dieser einfachen geometrischen Figuren ist im Gesamtfeld

viermal vertreten. Aus den fehlenden Figuren ergibt sich die Antwort.

(Entnommen aus dem “Standard Progressive Matrices”-Test)

Außerdem sind alle sprachfreien Intelligenz-Tests bis jetzt nach der klassischen Testtheorie entwickelt worden

und deshalb populationsabhängig. Ein wesentlicher Nachteil der klassischen Testtheorie ist, daß die Testparameter von den Personenparametern abhängig sind (Boesch und Eckensberger, 1969, S. 548). Dies macht einen Vergleich der Ergebnisse von Angehörigen verschiedener Kulturen von vornherein unmöglich.

Auch die Gütekriterien der klassischen Testtheorie, nämlich die Gültigkeit und Zuverlässigkeit der sprachfreien Tests, sind bei verschiedenen Kulturen nicht zufriedenstellend

(Dague, 1972, S. 70). Ebenso bleibt die strenge Zeiteinschränkung, die bei den Leistungstests, darunter auch den sprachfreien Intelligenztests üblich sind, nicht ohne Kritik.

Da es in den westlichen Kulturen überwiegend die Regel ist, nach der Zeit zu arbeiten, werden Angehörige anderer Kulturen wegen mangelnder Gewöhnung daran benachteiligt.

1.3.3. Versuch einer Definition eines kulturfreien Intelligenz-Tests: Die Definition eines kulturfreien Intelligenz-Tests läßt sich am besten aus der Kritik an der Kulturfreiheit der sprachfreien Intelligenz-Tests ableiten, denn sie negiert nicht nur, daß der sprachfreie Intelligenz-Test kulturfrei ist, sondern liefert

Transkription der Seiten 32–51

– 32 –

auch Ansätze und zeigt Forderungen auf, die zur Definition und Entwicklung eines kulturfairen Intelligenz-Tests beitragen können. Aus der Kritik an der Kulturfreiheit der

sprachfreien Intelligenz-Tests lassen sich folgende Forderungen ableiten:

a) Testmaterial
Mit dem Testmaterial sollen die inter- und intrakulturellen Gruppen (wie z.B. Analphabeten und Ausgebildete, männlich und weiblich usw.) in relativ gleichem Maße vertraut sein.

b) Richtungsunabhängigkeit
Die Beantwortung der Aufgaben soll nicht von einer bestimmten Richtung abhängig sein (z.B. von rechts nach links, von oben nach unten oder umgekehrt).

c) Beantwortungsmethode
Mit der Methode der Aufgabenbeantwortung sollen die inter- und intrakulturellen Gruppen in relativ gleichem Maße vertraut sein, insbesondere die Analphabeten und Ausgebildeten. D.h. sie soll nicht schulgebundenen Charakter haben.

d) Zeiteinschränkung
Der Test soll nicht unter strenger Zeiteinschränkung durchgeführt werden.

e) Modellkonformität

Der Test soll für inter- und intrakulturelle Gruppen dem Rasch-Modell konform sein. Diese Forderung schließt die Forderungen der Bedeutungsinvarianz und

der Populationsunabhängigkeit ein.

1. **Bedeutungsinvarianz: Der Testinhalt bzw. die logischen Prinzipien der Aufgaben sollen für inter- und intrakulturelle Gruppen (wie z.B. Analphabeten und Ausgebildete, männlich und weiblich usw.) keine unterschiedliche Bedeutung haben.**
2. **Populationsunabhängigkeit: Der Test soll nicht populations- bzw. stichprobenabhängig sein. D.h. die Parameterschätzung soll nicht systematisch verändert werden, wenn man bei der Schätzung von einer Stichprobe zur anderen übergeht.**

f) Gütekriterien

Der Test soll den testtheoretischen Anforderungen (Gültigkeit und Zuverlässigkeit) genügen.

g) Verbalisierung
Durch Verbalisieren der Testaufgaben soll bei den Probanden verschiedener Kulturen keine Leistungssteigerung erzielt werden.

Sind die obengenannten Forderungen bei einem Intelligenz-Test erfüllt, so läßt sich dieser als kulturfair bezeichnen, denn durch Erfüllung dieser Forderungen werden bei dem Test alle bisher bekannten ausbildungs- und kulturspezifischen Faktoren eliminiert oder neutralisiert.

Ein kulturfairer Test ist nicht auf dem gemeinsamen Gut aller Kulturen aufgebaut, sondern auf dem identischen

Gut der einzelnen Kulturen, denn bei ihm sind alle kulturspezifischen Faktoren, wie z.B. Richtung bei der Aufgabenbeantwortung nicht eliminiert, sondern ausgeglichen bzw.

neutralisiert. (Die Links-Rechts-Methode bei der Aufgabenbeantwortung der Europäer entspricht der Rechts-Links-Methode der Afghanen, Perser und Araber.)

Ein solcher Intelligenztest ist nicht frei von kulturellen Einflüssen, wohl aber frei von unterschiedlich starken Einflüssen verschiedener Kulturen. Er erfaßt einen höheren Grad angeborener Intelligenz, weil er den relativ gleichstarken Einflüssen der einzelnen Kulturen unterliegt, d.h. der spezielle Einfluß der einzelnen Kulturen ist ausgeglichen. Anders ausgedrückt: Dieser Test stellt relativ ähnliche Voraussetzungen bzw. Bedingungen für die Angehörigen der einzelnen Kulturen her.

Bei einem solchen kulturfairen Test werden die interkulturellen und intrakulturellen Gruppen (Analphabeten und Ausgebildete) durch ausbildungs- oder kulturspezifische Faktoren weder bevorzugt noch benachteiligt, da diese eliminiert oder neutralisiert worden sind. Man kann ihn deshalb im weiteren Sinne als "kulturfrei" bezeichnen.

Dieser Konstruktion liegt das allgemein anerkannte Prinzip der Objektivität zugrunde, nach dem der Beobachtungsbefund unabhängig vom Beobachter sein muß. Auch Objektivität ist – wie der Begriff "kulturfreier" Intelligenz-

Test – im strengen Sinne unmöglich zu erreichen. Hierüber wurde insoweit Einigung erzielt, daß Objektivität vorliegt, wenn intersubjektive Gleichheit hinsichtlich desselben Sachverhalts erreicht ist. So ist ein Intelligenz-Test auch dann als "kulturunabhängig" oder "kulturfrei" zu bezeichnen, wenn dieser unter gleich starken Einflüssen der verschiedenen Kulturen steht.

1.4. Grundlage des Naturformen-Intelligenz-Tests (N-I-T)

Der Naturformen-Intelligenz-Test (N-I-T) ist unter Berücksichtigung der Kritik der Kulturabhängigkeit der sprachfreien Intelligenz-Tests entwickelt worden.

1.4.1. Materialauswahl: Bei der Materialauswahl für den Naturformen-Intelligenz-Test war der Gegensatz von Kultur und Natur maßgebend. Dabei soll alles das, was durch Menschen geschaffen worden ist, und derjenige Teil der Natur, der durch Menschen verändert worden ist, unter Kultur verstanden werden.

Aus Naturgegenständen wurden diejenigen ausgewählt, die feste Formen haben, jedem Menschen bekannt und universal sind, wie zum Beispiel Halbmonde, Sterne, Teile des menschlichen Körpers, Hände, Füße usw.

Aus technischen Gründen wurden bei dem Naturformen-Intelligenz-Test nur Halbmonde und Sterne verwendet, mit Ausnahme einer Gedächtnisaufgabe, bei der auch die Formen von Händen und Füßen gezeichnet sind (s. Anhang C). Mit diesen Figuren sind im Gegensatz zu den geometrischen Figuren, die bei den meisten sprachfreien Intelligenz-Tests verwendet worden sind, in unterschiedlichen Kulturen sowohl die Analphabeten als auch die ausgebildeten Menschen relativ gleich gut

vertraut, weil sie Sterne und Halbmonde nicht nur als Erscheinungen am Himmel kennen, sondern auch aus gebräuchlichen Abbildungen, z.B. der Nationalflaggen einiger Länder, auf denen Sterne allein oder zusammen mit Halbmonden abgebildet sind. (35 Länder der Welt führen als Emblem in der Nationalflagge Sterne, auf etwa sechs Flaggen sind Halbmonde und Sterne zusammen abgebildet. Auch auf einigen Militäruniformen sind Sterne als Rangabzeichen angebracht.)

1.4.2. Aufbau der Aufgaben des N-I-T

Mit den Halbmonden und Sternen wurden die Aufgaben so konstruiert, daß einerseits die Beantwortung der Aufgaben nicht richtungsabhängig und andererseits einige der logischen Prinzipien von traditionellen Intelligenz-Tests (wie FRT, PMT und Lps) auf sie übertragen sind.

1.4.2.1. Naturformen-Intelligenz-Test, Vorform (N-I-T_V): Es wurden 120 Aufgaben (und eine Gedächtnisaufgabe) konstruiert (s. Anhang C).

Die Aufgaben wurden nach der Beantwortung, der Anordnung und der Anzahl der Prämissenfiguren (Kreise oder Ringe) in acht Untertests aufgeteilt.

Die Untertests 1, 3, 5 und 7 haben je 20 Aufgaben, die Untertests 2, 4, 6 und 8 je zehn. Jeder Untertest enthält zusätzlich vier Übungsaufgaben (s. N-I-T_v in Anhang C).

Die acht Untertests (U) des N-I-T_v lassen sich je nach den vier Hauptlösungsmethoden der Aufgaben zu vier Teiltests (a, b, c, d) folgendermaßen zusammenfassen:

Teiltest a = U_1 + U_2 Additionsmethode
Teiltest b = U_3 + U_4 Gemeinsamkeitsmethode
Teiltest c = U_5 + U_6 Gleiche Anzahl oder Häufigkeitsmethode
Teiltest d = U_7 + U_8 Abzieh-Methode

Zur Beantwortung einiger Aufgaben bei jedem Untertest sind zusätzliche Methoden notwendig, nämlich im Gesamtfeld die Berücksichtigung der Lage oder Anzahl der Halbmonde, die nach bestimmten Regeln angeordnet sind (s. u.a. Aufgaben 7 und 10 des

Untertests 1 der Vorform des N-I-T in Anhang C).

Ebenso lassen sich je nach der Anordnung und Anzahl der Prämissenfiguren die acht Untertests zu zwei Teiltests (A und B) zusammenfassen:

$$\text{Teiltest A} = U_1 + U_3 + U_5 + U_7$$
$$\text{Teiltest B} = U_2 + U_4 + U_6 + U_8$$

Die Aufgaben des Teiltests A bestehen aus fünf bis acht Prämissenfiguren, während die Aufgaben des Teiltests B sich aus neun Prämissenfiguren zusammensetzen.

Untertest 2 und 8 haben jeweils eine Grundaufgabe, wobei die Grundaufgabe des Untertests 8 sich aus der Grundaufgabe des Untertests 2 ableitet. Die Schwierigkeiten der übrigen Aufgaben der einzelnen Untertests im Vergleich zu deren Grundaufgaben kann man logisch abstufen.

Grundaufgabe des Untertests 2

[Abbildung: neun Kreise mit Sternen, u.a. “5-Sterne”]

– 39 –

Die Richtung ist bei dieser Aufgabe (“Magisches Quadrat”) ausgeschaltet. In jeder Spalte, Zeile oder Diagonale befinden sich 15 Sterne.

Die Grundaufgabe des Untertests 2 vereinfacht durch Zahlen dargestellt:

2 9 4
7 5 3
6 1 8

Wenn zu jeder Prämisse der Grundaufgabe die gleiche Zahl von Sternen addiert oder subtrahiert wird, ändert sich an dem Prinzip (oder der Gleichung) nichts. Zu jeder Prämisse werden z.B. zwei Sterne hinzugefügt.

4 11 6
9 7 5
8 3 10

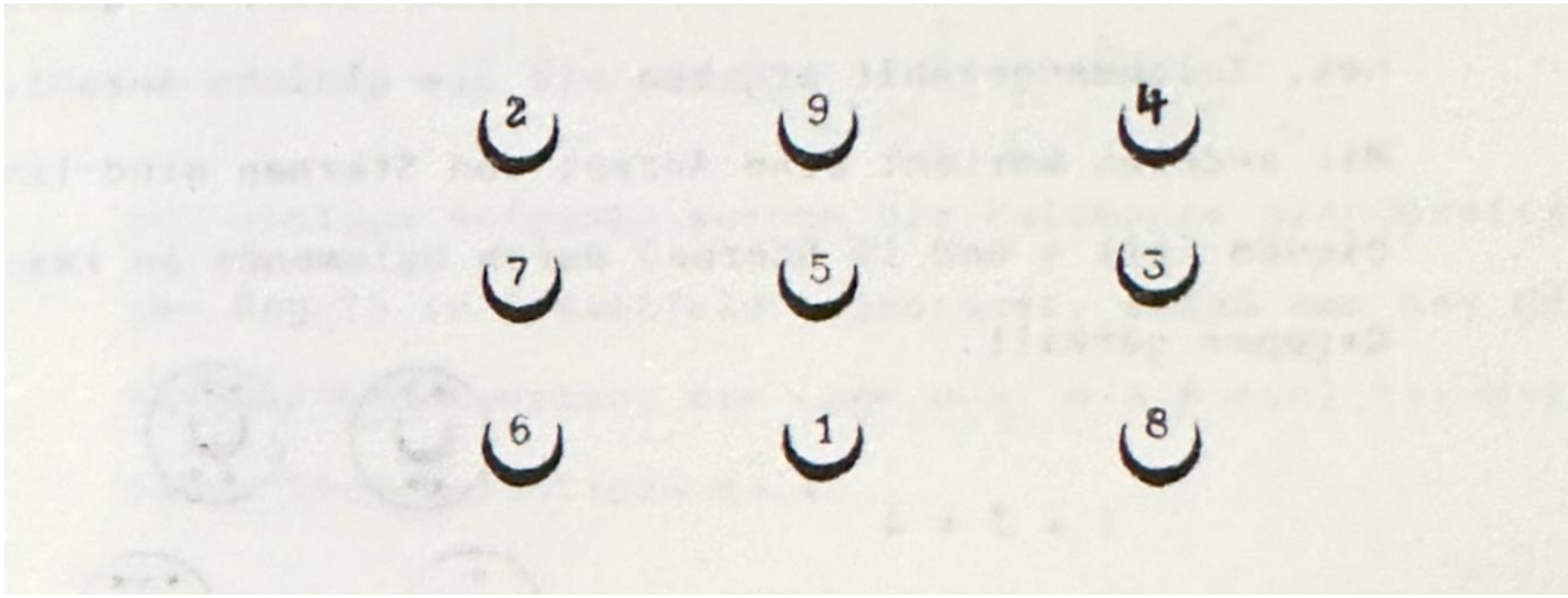
An dem Prinzip hat sich nichts geändert. Es ergibt sich bei jeder Spalte, Zeile oder Diagonale die Zahl 21 bzw. 21 Sterne. Diese Aufgabe ist schwerer als die Grundaufgabe, denn man muß dabei 21 Sterne zählen. Bei der Grundaufgabe sind es nur 15 Sterne.

An dem Prinzip ändert sich ebenfalls nichts, wenn die gegenüberliegenden Reihen, Spalten oder Spalten mit

Reihen der Grundaufgabe des Untertests 2 ausgetauscht werden. Damit kann man eine Anzahl von Aufgaben konstruieren, die den gleichen Schwierigkeitsgrad haben.(s. Anhang B)

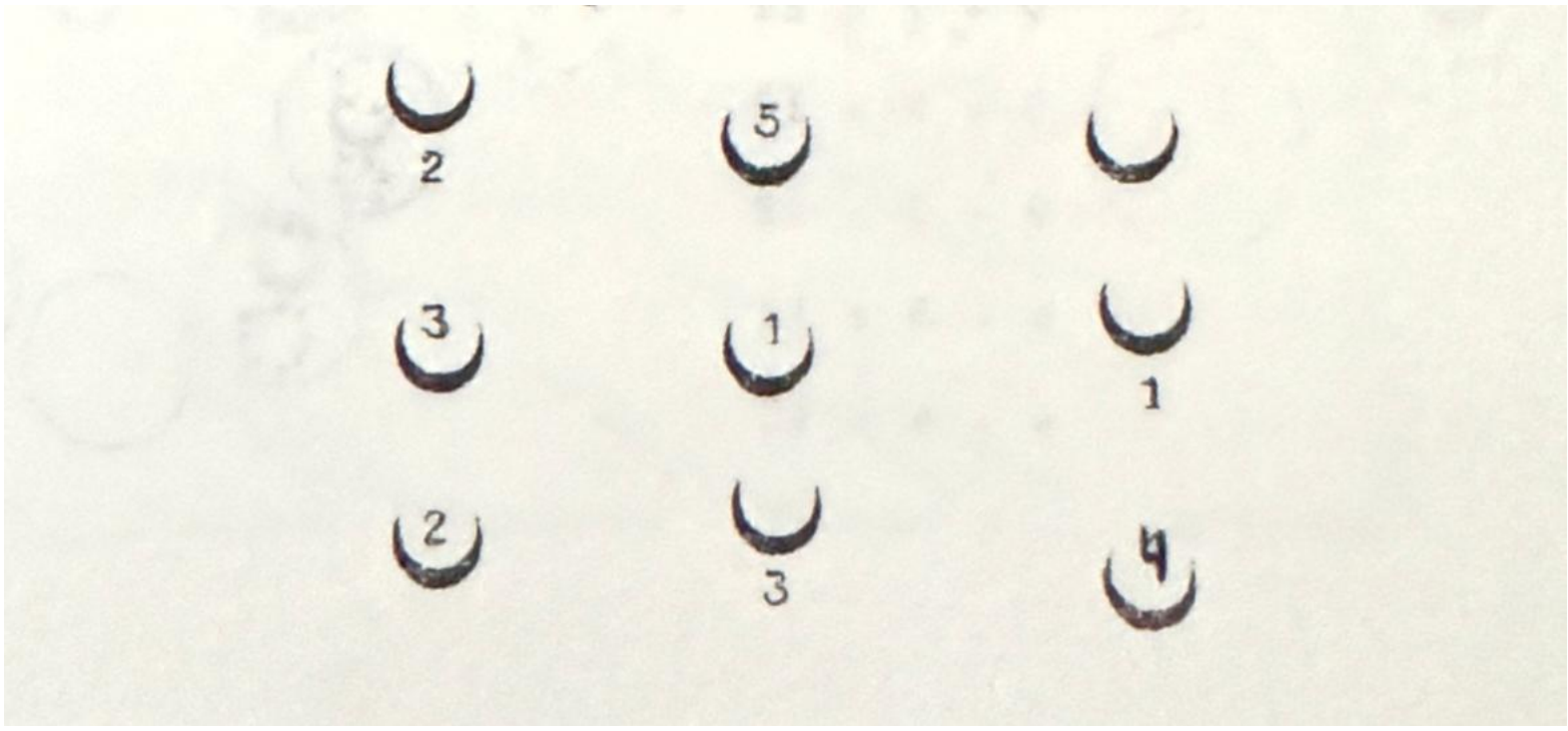
Ableitung der Grundaufgabe des Untertests 8 von der Grundaufgabe des Untertests 2:

Ein Halbmond wurde jeder Prämisse der Grundaufgabe hinzugefügt, somit ergibt sich folgende Aufgabe:



Dann wurden von jeder Prämisse vier (Sterne) abgezogen. Die Minuszahlen wurden unter die Halbmonde, die Pluszahlen über die Halbmonde geschrieben und damit die Grundaufgabe des Untertests 8 konstruiert.

Grundaufgabe des Untertests

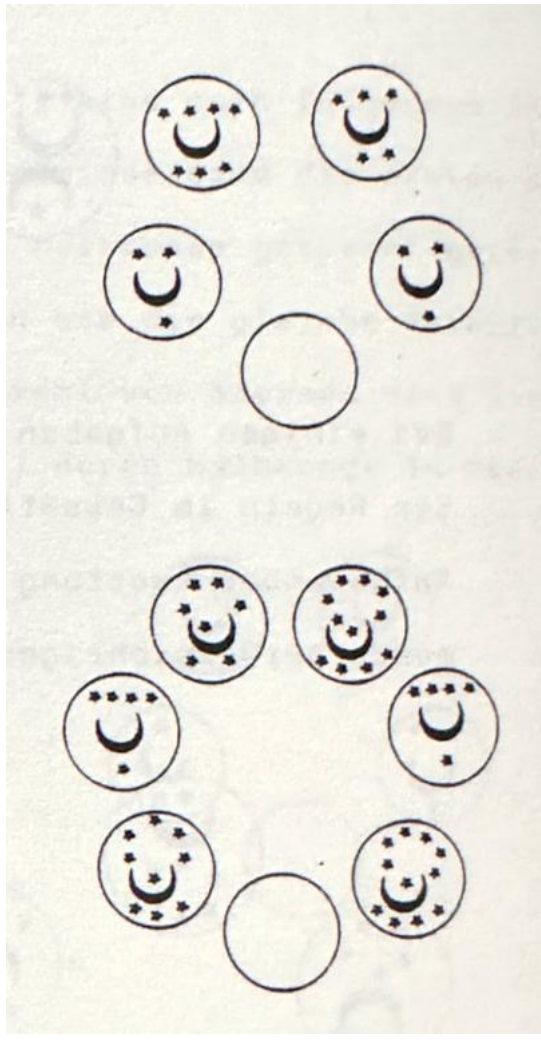


Wenn bei jeder Spalte oder Reihe oder Diagonale die Zahlen (Sterne) über den Halbmonden und die Zahlen (Sterne) unter den Halbmonden getrennt zusammengezählt und voneinander abgezogen werden, erhält man stets das gleiche Ergebnis. Im Falle der Grundaufgabe des Untertests 8 drei (Sterne).

Die Aufgaben des Untertests 1 sind nach folgendem Schema konstruiert: Bei jeder Prämisse sind die ersten beiden Zahlen als Sterne durch Halbmonde getrennt gezeichnet. Zusammengezählt ergeben sie die gleiche Anzahl. Mit anderen Worten: Eine Anzahl von Sternen sind (in diesem Fall 4 und 12 Sterne) durch Halbmonde in zwei Gruppen geteilt.

1 + 3 = 4
2 + 2 = 4
2 + 2 = 4
3 + 1 = 4

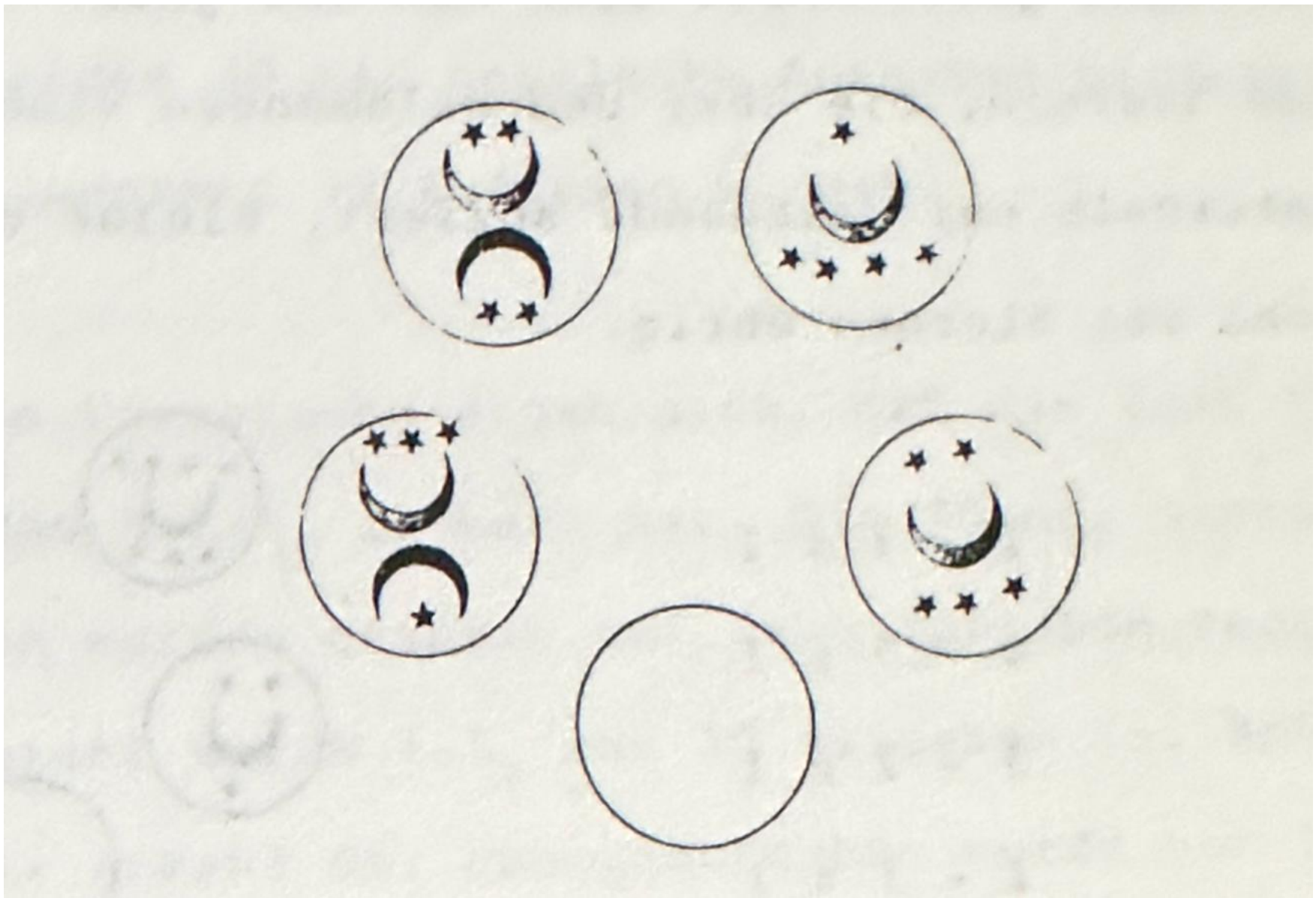
3 + 9 = 12
6 + 6 = 12
5 + 7 = 12
3 + 9 = 12
9 + 3 = 12



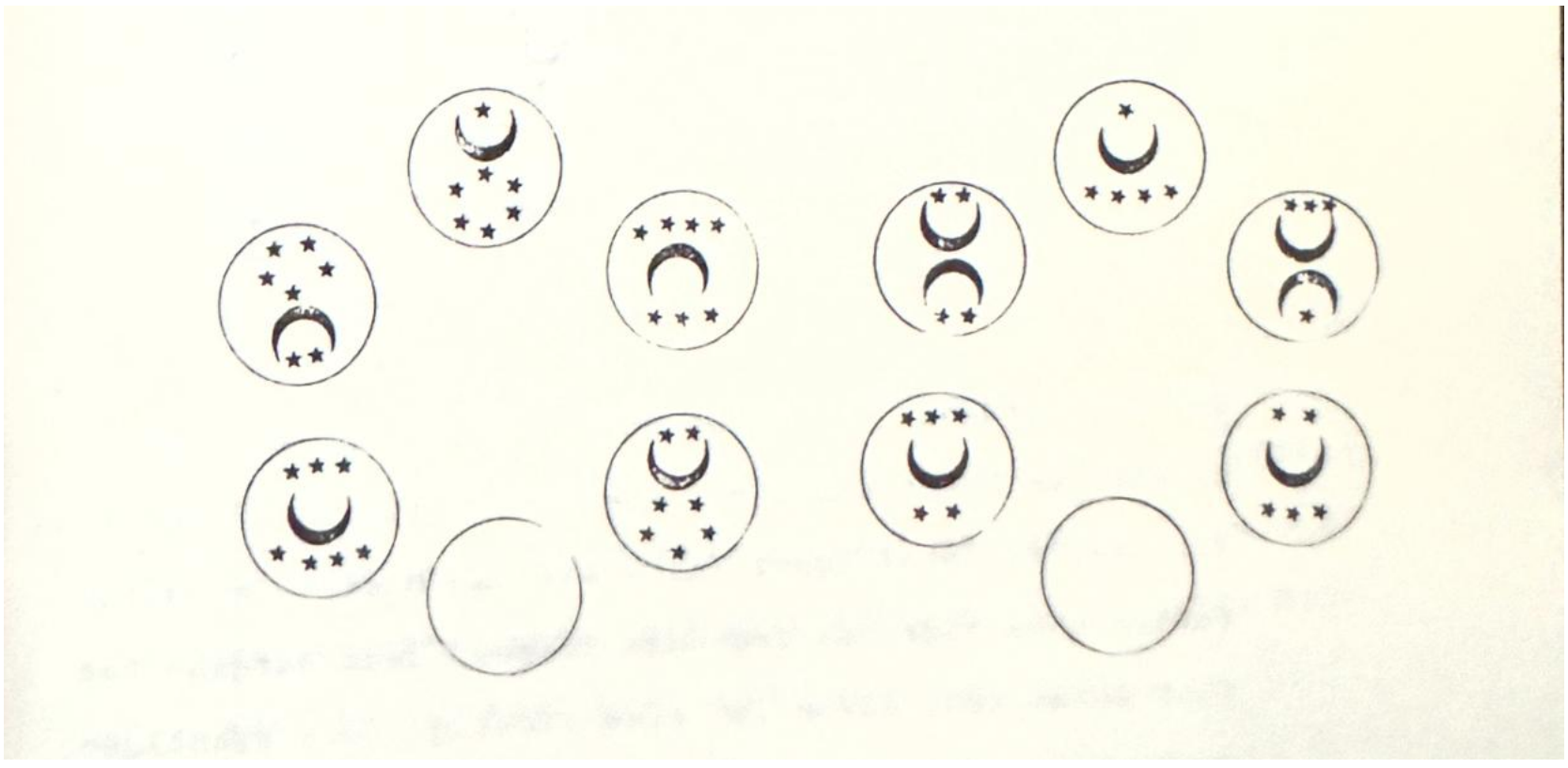
4 + 8 = 12
6 + 6 = 12

– 42 –

Bei manchen Aufgaben wurde bei einigen Prämissen anstelle eines Sternes ein Halbmond hinzugefügt. In diesem Falle ergeben Sterne und Halbmonde zusammen bei jeder Prämisse die gleiche Anzahl von Figuren.



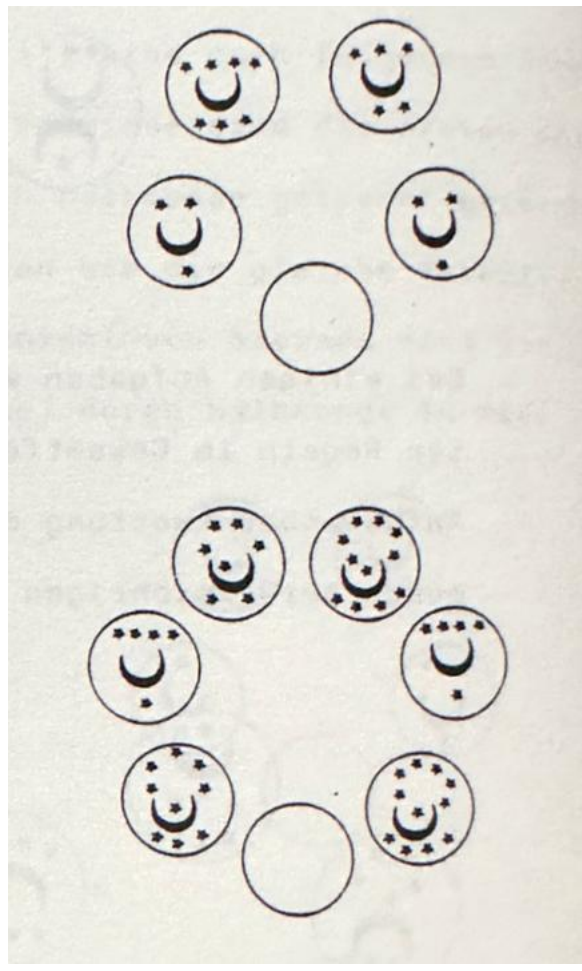
Bei einigen Aufgaben wurden die Halbmonde nach bestimmten Regeln im Gesamtfeld angeordnet, sodaß man bei der Aufgabenbeantwortung die Lage oder die Anzahl der Halbmonde berücksichtigen soll.



Als Grundlage für die Aufgaben des Untertests 7 dient folgendes Schema: Bei jeder Prämisse ist die erste Zahl als Gruppe von Sternen über einem Halbmond und die zweite unter den Halbmond gezeichnet. Wenn man bei jeder Prämisse von den Sternen, die über den Halbmonden stehen, die Sterne unterhalb der Halbmonde abzieht, bleibt die gleiche Anzahl von Sternen übrig.

2 – 1 = 1
4 – 3 = 1
3 – 2 = 1
2 – 1 = 1

6 – 3 = 3
4 – 1 = 3
5 – 2 = 3
7 – 4 = 3
4 – 1 = 3
7 – 4 = 3



Bei den Aufgaben der Untertests 3, 4, 5 und 6 handelt es sich um Übertragung der logischen Prinzipien der traditionellen Intelligenz-Tests auf das Muster der Naturformen ohne zugrundeliegendes Schema. Jede Aufgabe hat fünf Antworten, davon ist eine richtig. (Die richtigen Antworten der Aufgaben der einzelnen Untertests und der Übungsaufgaben des N-I-T sind im Anhang A, Tabelle 1 dargestellt.)

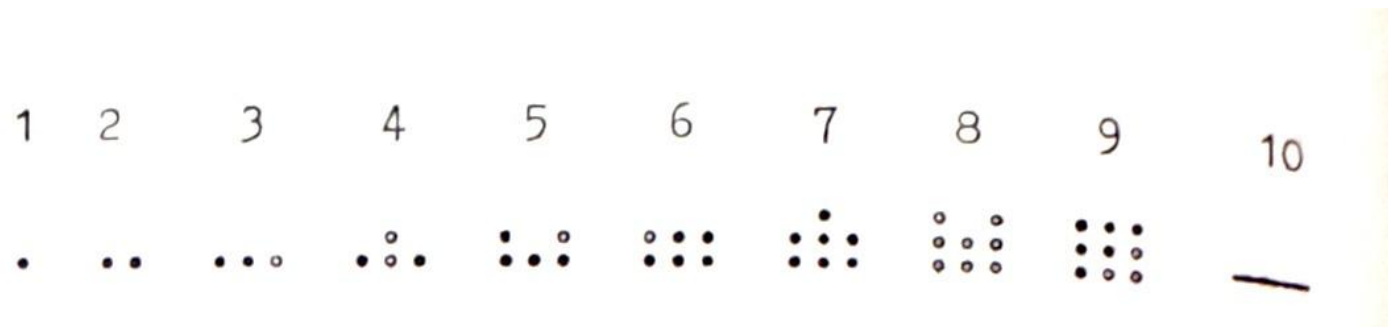
1.4.2.2. Endform des N-I-T: Die Endform des N-I-T besteht aus zwei Tests, dem N-I-T_1 und dem N-I-T_2. Bei der Konstruktion des N-I-T_1 wurden die je zehn Aufgaben der Untertests 2, 4, 6 und 8 des Teiltests B gemischt und außerdem 30 neu gebildete Aufgaben hinzugefügt, so daß es zusammen 70 Aufgaben wurden.

In einem Vorversuch ergab sich, daß die Zeit für 70 Aufgaben des N-I-T_1 zu kurz war. Die 30 neu hinzugefügten Aufgaben wurden deshalb auf zehn Aufgaben reduziert. Somit besteht der N-I-T_1 aus 50 Aufgaben (s. Anhang C). Auch die Anzahl der Übungsaufgaben wurde von 16 auf

acht Aufgaben reduziert (s. Mischungsschema Anhang A, Tabelle 2).

Der N-I-T_2 wurde aus 50 Aufgaben (s. Anhang C) der Untertests 1, 3, 5 und 7 (jeder Untertest hat 20 Aufgaben) des Teiltests A, die miteinander gemischt wurden, konstruiert (s. Mischungsschema Anhang A, Tabelle 3).

1.4.3. Die Beantwortungsmethode: Beim N-I-T wird mit dem folgenden von verschiedenen Sprachen oder Ziffern unabhängigen "Punktzahlensystem" gearbeitet, von dem man annehmen kann, daß alle Kulturen in gleichem Maße mit ihm vertraut bzw. nicht vertraut sind.



1 2 3 4 5 6 7 8 9 10

Bei obigem Punktzahlensystem geht jede Reihe bis drei. Die Punkte der unvollständigen Reihe werden so eingesetzt, daß die Symmetrie erhalten bleibt. Bei zehn wird ein Strich gesetzt.

Dies ist eine Vereinfachung, die der Beschleunigung und Übersichtlichkeit der Bearbeitung dient. (Die Instruktionen sind in Anhang C zusammengestellt.)

2. Fragestellung

Es soll untersucht werden, ob der Naturformen-Intelligenz-Test (N-I-T)* kulturfair ist. Unter kulturfairem Intelligenz-Test verstehen wir einen Test, bei dem die Forderungen, die in der Einleitung aus der Kritik an der Kulturunabhängigkeit sprachfreier Intelligenz-Tests abgeleitet wurden, erfüllt sind, nämlich die Forderungen: a) Testmaterial, b) Richtungsunabhängigkeit, c) Beantwortungsmethode, d) Zeiteinschränkung, e) Modellkonformität, f) Gütekriterien, g) Verbalisierung.

Die Forderungen “Testmaterial”, “Richtungsunabhängigkeit” und “Beantwortungsmethode” sind bei dem Naturformen-Intelligenz-Test bereits erfüllt, weil

das Material bzw. die Figuren, mit denen der Test konstruiert ist, jedem Menschen bekannt und universal sind (Halbmonde, Sterne und Teile des menschlichen Körpers). Mit diesen Figuren sind im Gegensatz zu den geometrischen Figuren sowohl Analphabeten als auch ausgebildete Menschen in unterschiedlichen Kulturen relativ gleich gut vertraut.

Man kann hier den Einwand erheben, daß es sich bei dem

- Die Abkürzung N-I-T (Naturformen-Intelligenz-Test) wird benutzt, wenn allgemein von allen Formen des Tests geredet wird. Im Falle einer bestimmten Form des Naturformen-Intelligenz-Tests bekommt die Abkürzung zusätzlich einen Index, z.B. N-I-T_1, N-I-T_2.

N-I-T, wenn auch die Figuren universal sind, um Rechenaufgaben handelt. Dieser Einwand verliert jedoch seine Gültigkeit, wenn man zwischen einfachem und komplexem Rechnen

unterscheidet. Auch der Analphabet ist mit kleinen Rechenmengen im Alltag vertraut.

Die Aufgaben des Naturformen-Intelligenz-Tests sind so konstruiert, daß die Methode der Aufgabenbeantwortung keine Schulausbildung voraussetzt und die Beantwortung der Aufgaben nicht richtungsgebunden ist.

Darüberhinaus wird bei der Beantwortung des N-I-T mit einem von verschiedenen Sprachen oder Ziffern unabhängigen "Punktzahlensystem" gearbeitet, von dem man annehmen kann, daß die Angehörigen verschiedener Kulturen in gleichem Maße damit vertraut bzw. nicht vertraut sind. Auch der Analphabet kann mit diesem "Punktzahlensystem" gut umgehen. Es wird dafür keine Schulausbildung vorausgesetzt.

Zu der Forderung "Zeiteinschränkung" ist zu sagen, daß die Zeit bei den einzelnen Tests bzw. Untertests festgelegt, aber ausreichend war. Ein großer Teil der Probanden konnte daher genügend Aufgaben lösen, so daß diese Forderung z.T. als erfüllt angesehen werden kann.

Die Forderungen “Modellkonformität” und “Gütekriterien” sollen empirisch überprüft werden.

– 48 –

Die Forderung “Verbalisierung” wird in dieser Arbeit nicht berücksichtigt. Will man sie überprüfen, so soll es durch die gleiche Methode wie bei Merz (1968) geschehen, in der die Probanden aufgefordert werden, alles, was sie sehen und denken, zu verbalisieren.

– 49 –

3. Methode

Die Methode besteht darin, daß der N-I-T zusammen mit einer Anzahl der traditionellen Intelligenztests, die als Kontrolle dienen sollen, in verschiedenen Kulturen an vergleichbaren Gruppen durchgeführt wird. Es soll eine Gruppe von Analphabeten getestet und anschließend nach der probabilistischen Testtheorie bzw. nach dem Rasch-Modell und der klassischen Testtheorie analysiert werden.

Damit lassen sich die Forderungen “Modellkonformität” und “Gütekriterien” überprüfen.

Bei dieser Arbeit wird jedoch die Untersuchung auf zwei extrem unterschiedliche Kulturen, nämlich Afghanistan und die Bundesrepublik Deutschland, eingeschränkt.

4. Die Untersuchung in Afghanistan

Zum besseren Verständnis dieser Untersuchung ist zunächst eine kurze Einführung in die afghanische Gesellschaft angebracht.

4.1. Einführung in die afghanische Gesellschaft

Im Gegensatz zur Industriegesellschaft in der Bundesrepublik Deutschland ist Afghanistan eine Agrargesellschaft, mit einer breiten Unterschicht, einer schmalen Mittelschicht und einer dünnen Oberschicht.
“Die große Masse der Kleinbauern, Pächter und Landarbeiter, der Wanderhirten mit geringem

Viehbesitz, der Gelegenheitsarbeiter, der mehr oder weniger abhängigen Handwerker und Einzelhändler und der ungelernten Staatsbediensteten, bildet die Unterschicht, die fast 95 % der Bevölkerung ausmacht. Die Mittelschicht bilden auf dem Lande die Familien mit größerem Grundbesitz und in den Städten die unabhängigen Handwerker." (Krause, 1972, S. 204) Zur Oberschicht gehören die Stammesfürsten, die Großhändler und die Beamten mit höherem Rang.

Vergleicht man die dreistufige Schichtung der afghanischen Gesellschaft mit der der Bundesrepublik Deutsch-

land, so wird man feststellen, daß die Unterschicht nicht nur breiter und die Mittelschicht nicht nur schmaler ist als die der Bundesrepublik Deutschland, sondern daß ein zur Mittelschicht gehörender Afghane einem zur Unterschicht gehörenden Deutschen entspricht, d.h. daß eine Verschiebung der Gesamtschichtverteilung vorliegt.

Die Bevölkerung Afghanistans gliedert sich in etwa 30 verschiedene ethnische Gruppen mit unterschiedlichen Sprachen, Sitten und Gebräuchen. Dies ist auf zahlreichen Völkerwanderungen und Eroberungszügen nach Indien über Afghanistan zurückzuführen. (Hayatullah, 1967, S. 48)

Die wichtigsten ethnischen Gruppen sind die Paschtunen, Balutschen, Uzbeken, Tadschiken, Turkmenen und Hazara. Die Paschtunen sind die eigentlichen Afghanen, nach denen das Land Afghanistan benannt ist.

“Der iranische Westen nennt sie Afghanen, der indische Osten Pathanen oder ebenfalls Afghanen. Sie selbst bezeichnen sich als Paschtunen oder in den nordöstlichen Siedlungsgebieten, wo die härtere Aussprache überwiegt, als Pachtunen.” (Klimburg, 1966, S. 284)

Obwohl bis jetzt keine Volkszählung stattgefunden hat, schätzt man die Zahl der Paschtunen auf über 30 Millionen. Die Hälfte davon wohnt in Pakistan. Daneben gibt

es zwei bis drei Millionen Paschtunen, die Nomaden sind und keinen festen Wohnsitz haben. Sie wandern je nach Jahreszeit zwischen Afghanistan und Pakistan. Über den Anteil der seßhaften Paschtunen in Afghanistan liegen in der Literatur verschiedene Schätzungen vor. Jedoch ist allen Schätzungen gemeinsam, daß die seßhaften Paschtunen über 65 % der Gesamtbevölkerung ausmachen sollen.

“Die Paschtunen oder ‘echten’ Afghanen (zum Unterschied von den ‘unechten’, das sind die nichtafghanischen Staatsangehörigen Afghanistans), die verwirrenderweise häufig auch als Pachtunen oder Pathanen Erwähnung finden, bilden das stärkste ethnische Element im Land.” (Klimburg, 1966, S. 110–111) Die Paschtunen leben sowohl in Afghanistan als auch in Pakistan nach den Gesetzen des Paschtunwali, das die ungeschriebene Verfassung der Paschtunen ist. Es enthält nicht nur eine Reihe von Ehren-Eigenschaften, die alle Paschtunen, ob arm oder reich, haben sollen, sondern auch eine bestimmte Anzahl

von Regeln, die von Generation zu Generation weitergegeben werden. Die höchste Instanz des Paschtunwali ist die Ratsversammlung (Dschirga), die politische, rechtliche und soziale Beschlüsse faßt und durchführt. Eine Angelegenheit wird solange diskutiert, bis Einstimmigkeit erreicht ist.

"Der demokratischen Einstellung der Freien gemäß und

zur Vermeidung von Eifersüchteleien hat die Dschirga weder eine Sitzordnung noch einen Vorsitzenden. Man sitzt im Kreis auf herrenlosem, folglich neutralem Boden und jeder kann aufstehen und das Wort ergreifen, kommen und gehen, wann immer er will. Für die Beschlüsse einer Dschirga ist nicht die Stimmenmajorität – es wird nie abgestimmt –, sondern das Fehlen einer offenen Opposition maßgebend. Die Befolgung der Beschlüsse gilt als wichtige soziale Pflicht der Männer. Zuwiderhandelnde laufen Gefahr, geächtet zu werden. Weigert sich zum Beispiel im Falle eines 'Strafprozesses' der Angeklagte, die

durch Dschirga-Beschluß festgelegten Wehrgeldzahlungen an den Kläger zu zahlen, hat letzterer das Recht" (Klimburg, 1966, S. 122), sich das ihm zustehende durch Unterstützung der Öffentlichkeit mit Gewalt von dem Pflichtvergessenen zu holen.

Die Ratsversammlung ist keine ständige Institution, sondern nur eine provisorische Versammlung für einen bestimmten Konflikt. Sie hat keinen zentralisierten Charakter, deshalb ist eine Ratsversammlung (Dschirga) von der anderen unabhängig und damit auch ihre Beschlüsse. Die Beschlüsse einer anderen Ratsversammlung über ähnliche Fälle können jedoch als Grundlage bei Entscheidungen dienen. Beschlüsse werden dann zur Regel erhoben, wenn von mehreren Ratsversammlungen ähnlich gelagerte Fälle ähnlich beurteilt wurden.

Die Beschlüsse einer Ratsversammlung (Dschirga) gelten lediglich im Einflußbereich der Versammlungsteilnehmer.

Jeder Paschtune hat das Recht, an einer Ratsversammlung (Dschirga) teilzunehmen, denn nach den Ehren-Eigenschaften sind alle Paschtunen gleich. Die jungen Teilnehmer einer Ratsversammlung haben meistens eine Zuhörerrolle, obwohl sie ebenfalls das Recht besitzen, sich zu Wort zu melden. Meistens ergreifen die alten Leute in einer Dschirga das Wort und spielen die aktive und dominante Rolle. Daraus ergibt sich, daß die Richter des Gerichtshofes des Paschtunwali die alten Leute sind, die über die Rechte und Pflichten sowie über die Sitten und Normen der Paschtunen wachen.

Die Dschirga-Größe und der Umfang betragen je nach Wichtigkeit der Konflikte von zwei bis .. Millionen Paschtunen. Die Dschirga kann durch den Angeklagten oder Beklagten oder einen Dritten einberufen werden. In manchen Fällen kommt die Dschirga auch spontan durch die Wichtigkeit eines Problems zusammen.

Das Paschtunwali wird jedem Paschtunen von Kindestagen an eingeprägt. Die Lernstätten des Paschtunwali sind die Familie, die Gastzimmer

(Hujra) und der Sommergarten für die Gäste (Dira).

Die paschtunische Familie gehört nicht zum Typ der

Kleinfamilie, wie wir sie aus der westlichen Welt kennen. Man kann die paschtunische Familie als “erweiterte” oder “Verbandsfamilie” bezeichnen, in der mehrere Brüder mit ihren Frauen und Kindern, mit ihren verheirateten Söhnen und Töchtern und deren Kindern zusammenleben. Eine solche Familie ist wie eine Gruppe.

Jedes Mitglied der Familie hat eine festgelegte Stellung. Einige haben untergeordnete, andere haben übergeordnete Stellungen; hierbei hat das Alter den Vorrang. Eine Ausnahme macht das weibliche Geschlecht, das der Autorität der Männer unterworfen ist. Jedoch besitzen die älteren Frauen eine gewisse Autorität (Mutter, Großmutter, Urgroßmutter, Tante). Der älteste Mann ist das Familienoberhaupt, dessen

Anordnungen und Einfluß sich alle Familienmitglieder zu fügen haben.

Die niedrige Stellung der Frauen ist nicht nur auf die Rechtsgesetze des Paschtunwali zurückzuführen, sondern auch darauf, daß die Frauen aus Gründen des Arbeitsmangels fast gar keine Verdienstmöglichkeiten haben. Sie leben deshalb von den Männern in Abhängigkeit. Andererseits gibt es in Afghanistan und Pakistan keine Altersversicherung, und so stellen die männlichen Kinder wegen der Arbeits- und Verdienstmöglichkeiten für die Eltern eine Versicherung dar. Aus diesem Grund werden beide Geschlechter von Geburt an unterschiedlich behandelt.

Die Stellung und Rolle der einzelnen Familienmitglieder hängt nach außen oder in der Gesellschaft von der Stellung und Rolle ihrer Familie in der Gesellschaft ab. Die Rolle und Stellung der Familie hängt wiederum von der sozialen und Stammeszugehörigkeit ab. Z.B. wird eine Familie mit vielen Männern oder

Stammesmitgliedern, die zur unteren Schicht gehört, dem Familienstatus und der Rolle der Familie entsprechen, die zur mittleren Schicht gehört und weniger Männer und Stammesmitglieder hat.

Faßt man das Gesagte noch einmal zusammen, so ergibt sich, daß das Alter nur innerhalb der Familie Vorrang hat, während in der Gesellschaft nicht das Alter, sondern die Familie als Gruppe (als Ganzes) die Stellung und den Status bestimmt.

4.2. Gegenstand der Untersuchung in Afghanistan

Der Gegenstand der Untersuchung in Afghanistan beschränkt sich auf folgende Fragen:

1. ob die Vorform des Naturformen-Intelligenz-Tests (N-I-T_v) den klassischen Testkriterien genügt,
2. ob die Aufgaben der Vorform des Naturformen-Intelligenz-Tests (N-I-T_v) für verschiedene Intelligenzgruppen in Afghanistan, Ausgebildete und Analphabeten,

männliche und weibliche Probanden dem Rasch-Modell konform sind.

– 57 –

4.3. Methode

Die Vorform des Naturformen-Intelligenz-Tests (N-I-T_v) wurde zusammen mit folgenden traditionellen Tests, die als Kontrolle dienten, in der Provinz Kuner (Ostafghanistan) und Kabul im Jahre 1976 mit Schülern und Analphabeten durchgeführt: *

1. Figure-Reasoning-Test (FRT)
2. CFT_3-Cattel-Weiß-Form A 1. Teil
3. Leistungsprüfungssystem Lps von Horn, Aufgabengruppe 9 und 10, Form B
4. Progressive matrices von J.C.Raven (PMT), Sets 1 und 2.

Für die Zusammenstellung der Tests waren folgende Gesichtspunkte maßgebend:
a) der Test soll in Originalform in verschiedenen Ländern anwendbar sein,
b) er soll ein Gruppentest sein.

Bei den Analphabeten handelt es sich um Personen, die weder schreiben noch lesen können.

Es wurden 300 Paschtunen (Afghanen) im Alter von 17 bis 20 Jahren (bei dem angegebenen Alter handelt es sich um Schätzwerte, da den meisten Probanden das genaue Geburtsdatum unbekannt war) getestet, davon waren 262 Schüler,

***) Weiteres siehe auch Anhang C**

die übrigen 38 waren Analphabeten. Der Anteil der weiblichen Probanden betrug 36, der Anteil der männlichen 264. Wegen des großen Umfanges der Untersuchung wurde diese auf drei Tage verteilt. Damit wurde einerseits ein Ermüdungseffekt bei den Probanden vermieden, andererseits konnte der Wunsch der Schuldirektoren erfüllt werden, den Unterricht nicht den ganzen Tag ausfallen zu lassen.

In der Schule wurden 20 bis 30 Probanden auf einmal getestet. Dabei wurde die Reihenfolge so eingeteilt, daß jeder Test ungefähr gleich häufig in jeder Position auftrat. Die Untersuchung mit den Analphabeten wurde mit dem N-I-T zu fünf bis zehn Probanden durchgeführt, während bei den klassischen Tests einzeln untersucht wurde. Einige Probanden konnte am zweiten oder dritten Tag an der Untersuchung nicht teilnehmen.

Tabelle I: Verteilung nach weiblichen und männlichen Probanden sowie nach Analphabeten und Schülern nach den einzelnen Tests anhand der afghanischen Stichprobe

Test	männl./weibl.	Analphabeten	Schüler	Insgesamt
N-I-T	195 / 32	35	142	227
FRT	225 / 35	33	227	260
PMT	187 / 20	27	180	207
CFT	178 / 11	0	189	189
Lps 9B	190 /33	30	193	223
Lps 10B	190 / 33	30	193	223

Tabelle II: Verteilung der getesteten afghanischen Analphabeten nach Alter und Geschlecht

Alter	männlich	weiblich	Insgesamt
17	2.	3	5
18	1	0	1
19	0	0	0
20	26	6	32

Tabelle III: Verteilung der getesteten afghanischen Analphabeten nach Berufen

Beruf	Anzahl der Analphabeten
Bauern	19
Hausfrauen.	9
Schuldiener	5
Kleinhändler	1
Lohnarbeiter.	1
Friseure	1
Fährmann	1
Viehzüchter.	1

Tabelle IV: Verteilung der getesteten afghanischen Schüler nach Alter und Geschlecht

Alter	weibl.	männl.	Insgesamt
17	14	98	112
18	11	58	69
19	2	43	45
20.	0	36	36

Tabelle V: Verteilung der getesteten weiblichen und männlichen afghanischen Schüler in den einzelnen Schulen

Schule	männl.	weibl.	Insgesamt
Omerra Khan Lycée (Gymnasium)	90	–	90
De Mazar Lycée (Gymnasium)	114	–	114
De Rahmann Baba Lycée (Gymnasium)	31	–	31
De Asad Abad de Naswan Lycée (Mädchen-Gymnasium der Asad Abad)	–	27	27

Tabelle VI: Verteilung der getesteten weiblichen und männlichen afghanischen Schüler nach der Schulklasse

Klasse	männl.	weibl.	Insgesamt
10	51	7	58
11	63	11	74
12	121	9	130

Tabelle VII: Verteilung der getesteten afghanischen Schüler nach dem Beruf des Vaters

Beruf des Vaters	Anzahl der getesteten Schüler
Bauern	214
Beamte mittleren Ranges	13
Kleinhändler	7
Angestellte	5
Armeeoffiziere mittleren Ranges	5
Khan (Großgrundbesitzer)	3
Handwerker	3
Lehrer	2
Mullah (islamische Geistliche)	2
Viehzüchter	2

Volksbürgermeister 1
Kleinlandbesitzer 1
Abgeordnete 1
Tagelohnarbeiter 1
Volksbürgermeister-Bote 1
Kraftfahrer 1

Die in Afghanistan durchgeführten Tests wurden anhand der erhaltenen Daten nach der klassischen Testtheorie analysiert und nach den Kriterien der Rohwertgruppen, Ausbildung und Geschlecht, auf Konformität des Rasch-Modells geprüft.

4.4. Ergebnisse

4.4.1. Analyse nach der klassischen Testtheorie: Die einzelnen Tests (FRT, CFT_3, Lps 9 B, Lps 10 B, PMT und N-I-T_V) wurden anhand der afghanischen Probanden nach der klassischen Testtheorie analysiert. (Die Formeln für innere Konsistenz und Reliabilität wurden nach Kuder und Richardson berechnet.)

Tabelle VIII: Ergebnisse der Analyse der einzelnen Intelligenztests nach der klassischen Testtheorie anhand der afghanischen Probanden

Test	Anzahl der Aufgaben der Tests	Mittelwerte	Standardabweichungen	innere Konsistenz (Reliabilität)
FRT	45	14,7538	5,6828	0,7842
CFT_3	50	12,518	3,724	0,515
Lps 9 B	40	11,6996	5,4464	0,7920
Lps 10 B	40	13,2915	5,1157	0,7815
PMT	36	8,0435	4,6514	0,7707
N-I-T_V	120	64,5374	17,9881	0,9301

Tabelle IX: Ergebnisse der Analyse der Untertests des CFT_3-Cattel-Weiß-Form A, 1. Teil nach der klassischen Testtheorie anhand der afghanischen Probanden

Untertest	Anzahl d. Aufgaben	Mittelwerte	Standardabweichung	Innere Konsistenz (Reliabilität)
CFT_3				
1	13	4,1481	1,9489	0,4933
2	14	2,8783	1,6170	0,4633
3	13	2,6614	1,5402	0,1878

4 10 2,8307 1,2141 0,1013

Gesamttest 50 12,518 3,724 0,515

Tabelle X: Ergebnisse der Analyse der Untertests der Vorform des Naturformen-Intelligenz-Tests (N-I-T_v) nach der klassischen Testtheorie anhand der afghanischen Probanden.

Untertest	Anzahl d. Aufgaben	Mittelwerte	Standardabweichung	Innere Konsistenz (Reliabilität)
1	20	13,6123	4,3167	0,8501
2	10	5,563	2,3128	0,6378
3	20	11,9427	4,8166	0,8571
4	10	4,6872	2,4358	0,7248
5	20	9,0088	3,4809	0,6907
6	10	3,3700	1,8066	0,4001
7	20	12,0749	5,2964	0,8933
8	10	4,2952	3,1140	0,8380
Gesamttest	120	64,5374	17,9881	0,9301

Tabelle XI: Ergebnisse der Analyse des N-I-T_v anhand der afghanische Probanden nach

Eliminierung* von neun nicht trennscharfen oder mit dem eigenen Untertest negativ korrelierenden Aufgaben des N-I-T_v nach der klassischen Testtheorie

Untertest	Anzahl d. Aufgaben	Mittelwerte	Standardabweichung	Innere Konsistenz (Reliabilität)
1	18	12,6960	4,3014	0,8624
2	10	5,5463	2,3128	0,6378
3	20	11,9427	4,8166	0,8571
4	8	4,4493	2,4207	0,7809
5	18	8,1586	3,4909	0,7178
6	8	3,0925	1,7372	0,4424
7	19	11,9339	5,4206	0,9073
8	10	4,2952	3,11400,8386	
Gesamttest	111	62,1145	18,0963	0,9347

*) Bei den eliminierten Aufgaben handelt es sich um solche, die wir als unbrauchbar (wie z.B. nicht trennscharf, zu große Z-Werte usw.) aus der Analyse herausgenommen haben, während es sich (Fortsetzung S. 64)

Tabelle XII: Die Analyse der vier Teiltests (a, b, c, d) des N-I-T_v (111 Aufgaben) nach der klassischen Testtheorie anhand der afghanischen Probanden

Teiltest	Anzahl d. Aufgaben	Mittelwerte	Standardabweichung	Innere Konsistenz (Reliabilität)
a	28	18,2423	5,7464	0,8577
b	28	16,3921	6,6406	0,8920
c	26	11,2511	4,3900	0,7284
d	29	16,2291	7,5973	0,9190
Gesamttest	111	62,1145	18,0963	0,9347

(Zur Zusammensetzung der N-I-T-Teiltests a, b, c, d siehe S. 37 dieser Arbeit.)

Tabelle XIII: Die Reliabilitätskoeffizienten des N-I-T_v anhand der afghanischen Stichproben mittels der Spearman-Brown-Formel auf vergleichbare Länge mit dem FRT, Lps 9 B, Lps 10 B und PMT

Anzahl der Aufgaben des Tests	innere Konsistenz (Reliabilität)	Errechnete Reliabilität des N-I-T
FRT 45	0,7842	0,8530

Lps 9 B	40	0,7920	0,8376
Lps 10B	40	0,7815	0,8376
PMT	36	0,7707	0,8227

Die Aufgabenanalyse nach der klassischen Testtheorie anhand der afghanischen Probanden des Figure Reasoning Tests (FRT) sind in Tabelle 4, des CFT_3-Cattel-Weiß Form A, erster Teil in Tabelle 5, des Leistungsprüfsystems Form B,

(Fortsetzung Fußnote von S. 63)
bei den ausgeschiedenen um Aufgaben handelt, die bei der einen oder anderen Gruppe von weniger als 3% der Probanden nicht richtig beantwortet oder von mehr als 97% der Probanden richtig beantwortet worden sind. Analog gilt dies wegen der Symmetrie des Modells (Probanden und Aufgaben) für die ausgeschiedenen Versuchspersonen.

Aufgabengruppe 9 in Tabelle 6 und Aufgabengruppe 10 in Tabelle 7 sowie des Progressive Matrices Tests, Sets 1 und 2 in Tabelle 8 und der Vorform des Naturformen-

Intelligenz-Tests (N-I-T_v) in Tabelle 9, Anhang A dargestellt.

4.4.2. Die Analyse der Tests anhand des Rasch-Modells nach der Methode der bedingten Wahrscheinlichkeit

Ausgenommen wurde der CFT_3-Cattel-Weiß, Form B, 1. Teil, der einen geringeren Reliabilitätskoeffizienten (r=0,515) zeigte und deshalb seine weitere Analyse nicht lohnte. Alle übrigen Tests (FRT, Lps 9 B, Lps 10 B, PMT und N-I-T_v) wurden nach den Kriterien der Rohwertgruppen, Ausbildung und Geschlecht anhand der afghanischen Probanden auf Konformität mit dem Rasch-Modell analysiert.

Tabelle XIV: Ergebnisse der Rasch-Analyse des FRT, Lps 9 B, Lps 10 B, PMT, N-I-T-Teiltest A und N-I-T-Teiltest B anhand der afghanischen Probanden in bezug auf die Rohwert-, Ausbildungs- und Geschlechtsgruppen.

VpnRohwertgr.	Ausbildung	Geschlecht	
Gruppe 1	131 33 (afgh. Analphabeten)	3	5 (weibl. Afghanen)

Gruppe 2 129 227 (afgh. Schüler) 225 (männl. Afghanen)
Gesamt 260 260 260
Ausgeschiedene Versuchspersonen 0 0 0
FRT: Chi-Quadrat(X^2) 138,0119 139,2651 76,0736
Freiheitsgrade(df) 44 44 40

Fortsetzung Tabelle XIV:

VpnRohwertgr. Ausbildung Geschlecht
Information(I) 17,29,6738 1742,0705 1633,0306
Signifikanzniveau(S) P=0.1%(s.) P=0.1%(s.) P=0.1%(s.)
Gruppe 1 : 120 30 (afgh. Analphabeten) 35(weibl. Afghanen)
Gruppe 2 103 193 (afgh. Analphabeten) 190(männl. Afghanen)
Gesamt: 223 223 223
Ausgeschiedene Versuchspersonen 0 0 0
Lps9B: Chi-Quadrat(X^2) 80,8210 115,5182 51,8000
Freiheitsgrade(df) 37 39 37
Information(I) 1285,5492 1313,5891 1282,1162

Signifikanzniveau(S)	P=0.1%(s.)	P=0.1%(s.)	P=5%(n.s.)
Gruppe 1	116	30 (afgh. Analphabeten)	33(weibl. Afghanen)
Gruppe 2	107	193 (afgh. Schüler)	190 (männl. Afgh.)
Gesamt	223	223	223
Ausgeschiedene Versuchspersonen	0	0	0
Lps10B: Chi-Quadrat(X^2)	73,7883	196,1403	111,3370
Freiheitsgrade(df)	31	33	33
Information(I)	1072,5907	1099,7792	1108,4905
Signifikanzniveau(S)	P=0.1%(s.)	P=0.1%(s.)	P=0.1%(s.)

Gruppe 1	96	27 (afgh. Analphabeten)	1 9 (weibl. Afgh.)
Gruppe 2	105	174 (afgh. Schüler)	181 (männl. Afgh.)
Gesamt	207	207	207
Ausgeschiedene Versuchspersonen	6	6	7
PMT: Chi-Quadrat(X^2)	132,4684	70,7794	53,1643
Freiheitsgrade(df)	34	35	29
Information(I)	934,9035	969,5416	858,4455
Signifikanzniveau(S)	P=0.1%(s.)	P=0.1%(s.)	P=1%(s.)

Fortsetzung Tabelle XIV:

VpnRohwertgr.	Ausbildung	Geschlecht
Gruppe 1	115 35 (afgh. Analphabeten)	3 2 (weibl. Afgh.)
Gruppe 2	112 192 (afgh. Schüler)	195 (männl. Afgh.)
Gesamt	227 227	227
Ausgeschiedene Versuchspersonen	0 0	0
N-I-T_a: Chi-Quadrat(X^2)	255,9132 452,4322	144,4417
Freiheitsgrade(df)	73 71	73
Information(I)	2968,6807 2879,1343	3001,7804
Signifikanzniveau(S)	P=0.1%(s.) P=0.1%(s.)	P=0.1%(s.)
Gruppe 1	113 35 (afgh. Analphabeten)	32 (weibl Afgh.)
Gruppe 2	114 192 (afgh. Schüler)	195 (männl. Afgh.)
Gesamt	227 227	227
Ausgeschiedene Versuchspersonen	0 0	0
N-I-T_B: Chi-Quadrat(X^2)	98,2686 63,5164	143,8476
Freiheitsgrade(df)	35 35	35
Information(I)	1601,7934 1608,9886	1589,9147

Signifikanzniveau(S) P=0.1%(s.) P=1%(s.) P=0.1%(s.)

(Zur Zusammensetzung der N-I-T_a und N-I-T_B siehe S. 38 dieser Arbeit.)

Tabelle XV: Die X^2-Werte der Raschanalyse der einzelnen Tests in bezug auf die afghanischen Rohwert-, Ausbildungs- und Geschlechtsgruppen dividiert durch entsprechende Freiheitsgrade (df).

Rohwertgruppen X^2/df	Ausbildung X^2/df	Geschlecht X^2/df
FRT 3.136634	3.1651159	1.90184
Lps9B 2.1843513	2.9620051	1.4
Lps10B 2.3802677	5.9436454	3.3738484
PMT 3.8961294	2.0222685	1.8332517
N-I-T_a 3.5056602	6.3736929	1.9786534
N-I-T_B 2.8076742	1.8147542	4.1099314

Die Z-Werte der Rasch-Analyse anhand der afghanischen Probanden der Aufgaben des FRT sind in Tabelle 10, des Lps9B in Tabelle 11, des Lps10B in Tabelle 12, des PMT in Tabelle

13, des N-I-T-Teiltest A in Tabelle 14 und Teiltest B in Tabelle 15, Anhang A dargestellt.

4.5. Diskussion

Aus der Tabelle VIII geht hervor, daß – abgesehen vom CFT_3, der einen geringeren Reliabilitätskoeffizienten aufweist – alle anderen Tests, nämlich FRT, Lps 9 B, Lps 10 B, PMT und N-I-T, zufriedenstellende Reliabilitätskoeffizienten bzw. innere Konsistenz zeigen. Darunter steht der N-I-T mit einem Reliabilitätskoeffizienten von r=0,9301 an erster Stelle. Dieser höhere Reliabilitätskoeffizient des N-I-T_V ist nicht auf die Länge des N-I-T_V zurückzuführen. Die Reliabilitätskoeffizienten des N-I-T_V, die mittels der Spearman-Brown-Formel auf vergleichbare Länge mit den traditionellen Tests (FRT, Lps9B, Lps10B und PMT) errechnet worden sind, sind deutlich höher als die der traditionellen Tests (vgl. Tabelle XIII).

Damit ist ein Kriterium der klassischen Testtheorie, nämlich die Reliabilität beim N-I-T_V als erfüllt anzusehen, also wird die Frage 1 zum Teil positiv beantwortet.

Die Aufgaben der traditionellen Tests (FRT, Lps 9 B, Lps 10 B, PMT und die der vier Untertests des CFT_3) sind nicht im Sinne der Konstrukteure ansteigend schwieriger (s. P-Werte bei Tabellen 4, 5, 6, 7, 8 in Anhang A). Auch beim N-I-T_v sind die Aufgaben der einzelnen Untertests nicht ansteigend schwieriger, aber die meisten Aufgaben haben einen mittleren Schwierigkeitsgrad (s. P-Werte in Tabelle 9, Anhang A).

Wie bei den traditionellen Tests korrelieren auch beim N-I-T_v einige Aufgaben mit dem eigenen Untertest zu Null (sind also nicht trennscharf) oder negativ (s. Tabellen 4, 5, 6, 7, 8 und 9, Anhang A). Das bedeutet, daß diese Aufgaben unbrauchbar sind. Jedoch ist nach Eliminierung von sieben nicht trennscharfen und zwei mit dem Untertest negativ korrelierten Aufgaben der Reliabilitätskoeffizient des N-I-T nun gering von r=0,9301 auf r=0,9347 erhöht (s. Tabelle VIII, X und XI).

Aus Tabelle XIV ist ersichtlich, daß in bezug auf die Intelligenzgruppen (Rohwertgruppen) alle durchgeführten Tests (FRT, Lps 9 B, Lps 10 B, PMT, N-I-T_a und N-I-T_B) modellinkonform sind.

Auch in bezug auf Ausbildung sind, abgesehen vom N-I-T_B, der auf dem 1%-Niveau signifikant ist, alle übrigen Tests auf dem 0,1%-Niveau signifikant, also dem Rasch-Modell nicht konform.

Hinsichtlich des Geschlechts ist Lps 9 B nicht signifikant (P=5%, s. Tabelle XIV), also dem Rasch-Modell praktisch hinreichend konform, während PMT (bei P=1%) und alle übrigen durchgeführten Tests signifikant (P=0,1%) sind, d.h. dem Rasch-Modell nicht konform sind.

In bezug auf Rohwert- und Geschlechtsgruppen der Afghanen läßt sich bei einigen dieser Tests nach Ausschaltung einiger modellinkonformer Aufgaben die Modellkonformität erreichen, während in bezug auf die Ausbildung die Modellkonformität kaum zu erreichen ist. Jedoch kann der N-I-T_B nach

Ausschaltung von ein oder zwei modellinkonformen Aufgaben auch in bezug auf die Ausbildungsgruppen der Afghanen auf Modellkonformität gebracht werden (vgl. Tabellen XIV, XV und 10, 11, 12, 13, 14 und 15, Anhang A).

Die einzelnen Tests (FRT, Lps 9 B, Lps 10 B, PMT, N-I-T_a, N-I-T_B) haben in bezug auf die Rohwert-, Ausbildungs- und Geschlechtsgruppen in Afghanistan unterschiedliche X^2-Werte (vgl. Tabelle XV).

Diese Tatsache bestätigt die Annahme, daß die Größe der Prüfgröße (X^2) und damit die Modellkonformität eines Tests von den gewählten Analyse-Kriterien wie z.B. Intelligenzgruppen, Ausbildung, Geschlecht usw. abhängt.

Wenn man von Modellkonformität spricht, so ist diese nur qualitativ zu verstehen, nicht quantitativ. Wenn ein Test in bezug auf irgendwelche Kriterien, z.B. Geschlecht, dem Rasch-Modell konform ist, so heißt dies nicht,

daß beide Geschlechter (weibliche und männliche Probanden) bei diesem Test im Durchschnitt gleich viele Aufgaben lösen, sondern daß dieser Test für die beiden Geschlechter bedeutungsinvariant ist, d.h. dieselbe Bedeutung hat.

Ein Beispiel dafür ist in unserer Untersuchung in Afghanistan der Lps 9 B, der in bezug auf das Geschlecht dem Rasch-Modell konform ist, obwohl bei ihm die weiblichen Probanden im Durchschnitt 10,909 und die männlichen Probanden im Durchschnitt 11,837 Aufgaben richtig beantwortet haben, also die weiblichen Probanden weniger als die männlichen Probanden. Dasselbe gilt beim N-I-T_B in Afghanistan, der nach Ausschaltung von zwei Aufgaben in bezug auf die Ausbildung (Analphabeten und Schüler) dem Rasch-Modell konform wird. Auch bei diesem Test haben die Analphabeten im Durchschnitt 14,600 Aufgaben und die Schüler im Durchschnitt 18,500 Aufgaben richtig beantwortet, also die Analphabeten weniger als die Schüler.

Des weiteren kann man aus den Ergebnissen der Analyse

Transkription der Seiten 72–90

– 72 –

der einzelnen Tests nach dem Rasch-Modell entnehmen, daß bei den traditionellen Tests (FRT, Lps 9 B, Lps 10 B und PMT) in bezug auf Rohwert-, Ausbildungs- oder Geschlechtsgruppen einige Aufgaben ausgeschieden sind (vgl. Tabelle 10, 11, 12, 13), während beim N-I-T_B weder in bezug auf die Rohwertgruppen und Ausbildung noch in bezug auf das Geschlecht Aufgaben ausgeschieden sind (s. Tabelle 15).

Bei den meisten der ausgeschiedenen Aufgaben beim PMT ist dies darauf zurückzuführen, daß sie in der einen oder anderen Gruppe von 3% der Probanden nicht richtig beantwortet worden sind, während bei den übrigen traditionellen Tests die meisten der ausgeschiedenen Aufgaben darauf zurückzuführen sind, daß sie mehr als 97% der Probanden richtig beantwortet haben.

Teilt man die durchgeführten Tests in abstrakte und nichtabstrakte, so sind N-I-T_B und PMT abstrakter als N-I-T_A, FRT, Lps 9 B und Lps 10 B. Die abstrakten Tests liefern gemessen an den X^2-Werten bessere Ergebnisse in bezug auf die Ausbildung als die nicht abstrakten, so daß man den Eindruck bekommt, daß die abstrakten Tests für Schüler und Analphabeten bedeutungsinvarianter sind als die nichtabstrakten (vgl. Tabelle XV).

5. Die Untersuchung in der Bundesrepublik Deutschland

Für die empirische Testuntersuchung an den Schulen in der Bundesrepublik Deutschland war wie in Afghanistan die Erlaubnis des Kultusministeriums notwendig. Während sie mir in Afghanistan vom damaligen Kultusminister Omer Wardok durch ein Gespräch ohne irgendwelche Bedingungen (auch die Wahl der Schulen und Zeit der Untersuchung war mir überlassen) erteilt wurde (siehe Anhang D), wurde sie in der Bundesrepublik Deutschland beim

Schulkollegium, beim Regierungspräsidenten in Düsseldorf am 21.5.1979 beantragt.

Das Schulkollegium zeigte sich entsprechend der Richtlinien für die Genehmigung von empirischen Untersuchungen und Befragungen in Schulen und dem Runderlaß des Kultusministers vom 17.2.77-III C 1.36-75/ O-426/77 (siehe Anhang D) unter anderem unter folgenden Bedingungen mit der Untersuchung einverstanden.

a) Es sollte ein Verzeichnis der Schulen, an denen die Untersuchung geplant war, Datum der Untersuchung, Einverständnis der Schulleiter, der betroffenen Lehrer und der Schüler vorgelegt werden.

b) Die Dauer der Untersuchung sollte drei Schulstunden nicht überschreiten.

c) Die Ergebnisse sollten dem Schulkollegium schriftlich mitgeteilt werden.

Dafür war notwendig, zunächst mit dem Schulleiter und den Lehrern Kontakt aufzunehmen; so habe ich mich mit dem Schulleiter und den Lehrern in Verbindung gesetzt und ihnen den Grund meiner Untersuchung erläutert. Sie haben, nachdem sie mit ihren Schüler gesprochen hatten, ob sie an einer solchen Untersuchung teilnehmen wollten, sich mit der Untersuchung unter den Bedingungen einverstanden erklärt, daß die Ergebnisse den Schülern mitgeteilt werden und daß die Untersuchung nicht mit allen drei Stunden an einem Tag durchgeführt wurde, sondern an zwei verschiedenen Tagen mit genügend Abstand, und zwar einmal zwei Stunden und einmal eine Stunde in derselben Klasse. Da die zwei Schulstunden für die Durchführung des gesamten N-I-T nicht ausreichend waren, wurde er in zwei Teiltests (Teiltest A und Teiltest B) zerlegt (s. dazu S. 38). Für die Untersuchung kam N-I-T-Teiltest B in Betracht, denn dieser könnte in bezug auf die Ausbildung (Analphabeten und Schüler) nach Ausschaltung von zwei modellinkonformen Aufgaben dem Rasch-Modell konform werden.

Daß der N-I-T-Teiltest B aus vier Untertests (Untertest 2, 4, 6 und 8) des N-I-T bestand und daß für jeden Untertest eine extra Instruktion erforderlich war, hätte störend wirken können. Deshalb war es angebracht, zwei Mischformen des N-I-T zu konstruieren, nämlich N-I-T_1 und N-I-T_2 (s. dazu S. 38 und 44).

Die gemischten Tests N-I-T_1 und N-I-T_2 (s. S.38 und 44) haben den Vorteil, daß jeweils nur eine Instruktion notwendig ist und eine Störung während der Testdurchführung durch eine neue Instruktion vermieden wird. Als Vergleich wurde der Figure-Reasoning-Test herangezogen, weil eine Schulstunde Zeit für die Durchführung dieses Tests ausreichend war.

5.1. Gegenstand der Untersuchung in der Bundesrepublik Deutschland

Der Gegenstand der Untersuchung in der Bundesrepublik Deutschland beschränkt sich auf die Frage, ob die Aufgaben der N-I-T_B, N-I-T_1 und N-I-T_2 für die verschiedenen

Intelligenzgruppen, männliche und weibliche Probanden dem Rasch-Modell konform sind.

5.2. Methode

Der N-I-T_B, N-I-T_1 und der N-I-T_2 wurden zusammen mit dem Figure-Reasoning-Test (FRT), der als Kontrolle diente, in Köln und Umgebung an Schulen durchgeführt und anschließend in bezug auf Rohwertgruppen und Geschlecht auf Anpassung des Rasch-Modells analysiert.

An der Untersuchung nahmen insgesamt 509 Schüler teil, davon waren 332 weibliche und 177 männliche Schüler im Alter von 16-19 Jahren.

Tabelle XVI: Das Datum der Untersuchung an den einzelnen Schulen in der Bundesrepublik Deutschland

Name und Adresse der Schule	Datum der Untersuchung

Irmgardis-Schule, Schillerstraße, 5000 Köln 51 14.8. – 12.8.1979
Albertus-Magnus-Gymnasium, Ottostraße 87, 5000 Köln 30 27.8. – 31.8.1979
Norbert-Gymnasium, 4047 Dormagen 6 10.9. – 18.9.1979
Hürth-Gymnasium, Bonnstraße 64/66, 5030 Hürth 19.9. – 16.11.1979
Gymnasium Mülheim, Genovevastraße 58-62, 5000 Köln-Mülheim 7.11. – 23.11.1979
Hansa-Gymnasium, Hansaring 56-58, 5000 Köln 1 5.12. – 19.12.1979

Tabelle XVII: Die Anzahl der weiblichen und männlichen getesteten deutschen Schüler in den einzelnen Schulen

Schule	weiblich	männlich	Gesamt
Irmgardis-Schule Köln	9	0	9
Albertus-Magnus-Gymnasium Köln	25	28	53
Norbert-Gymnasium Dormagen	0	18	18
Hürth-Gymnasium	98	79	177
Gymnasium Köln-Mülheim	173	30	203

Fortsetzung Tabelle XVII:

Schule	weiblich	männlich	Gesamt
Hansa-Gymnasium Köln	27	22	49
Summe	332	177	509

Tabelle XVIII: Verteilung der weiblichen und männlichen getesteten deutschen Schüler nach dem Alter

Alter der Schüler	weiblich	männlich	Gesamt
16	108	65	173
17	104	47	151
18	94	46	140
19	26	19	45

Die Kontaktaufnahme mit den einzelnen Schuldirektoren und Lehrern konnte nur nacheinander und nicht gleichzeitig erfolgen. Bei den Gesprächen ergab sich, daß in einigen Klassen – meistens in pädagogischen Kursen – gerade das Thema Intelligenz-Test behandelt wurde. Deshalb bestand Interesse an der gleichzeitigen Durchführung eines Intelligenztests, der eine praktische Ergänzung des theoretischen Unterrichtsstoffes darstellte.

Die Schüler zeigten Interesse an der Untersuchung, und sie freuten sich auf die Ergebnisse, die ihnen eine bessere Selbsteinschätzung ermöglichten. Auf Wahrung der Anonymität wurde bei Verteilung der Ergebnisse anhand des FRT geachtet.

Jeder Schüler erhielt nach dem Test in einem geschlossenen Umschlag den IQ-Wert und Prozentrang in bezug auf die gesamtdeutsche Bevölkerung zwischen zehn und 60 Jahren mitgeteilt.

Die Schüler zeigten nach den Ergebnissen des FRT im Durchschnitt einen IQ von 112,35. Für die Mädchen ergab sich im Durchschnitt ein IQ von 111,69 und für die Jungen ein IQ von 113,59.

Tabelle XIX: Die Verteilung der weiblichen und männlichen getesteten deutschen Schüler nach dem Intelligenzquotienten (IQ-Werte) anhand des FRT

IQ	weiblich	männlich	Gesamt
89	0	1	1
91	1	0	1
93	1	1	2
95	3	1	4
97	2	1	3
98	3.	3	6
100	9	5	14
102	9	6	15
103	17	2	19
105	19	9	28
107	18	13	31
109	29	12	41
111	40	14	54
112	38	11	49
114	39	10	49
116	28	19	47
118	26	16	42
119	23	16	39
121	14	13	27
123	6	13	19
125	2	6	8
126	3	4	7
128	1	1	2

Tabelle XX: Die Verteilung der weiblichen und männlichen sowie der gesamten anhand des Figure-Reasoning-Tests getesteten deutschen Schüler nach der Termanschen Klassifikation der Intelligenz: Unterdurchschnittliche Intelligenz (IQ<90), durchschnittliche Intelligenz (IQ=90-110), gute Intelligenz (IQ=110-120) und sehr gute Intelligenz (IQ>120).

IQ-Bereich	Anzahl der Mädchen	in %
Unterdurchschnittliche Intelligenz (IQ<90)	0	0,0000
Durchschnittliche Intelligenz (IQ=90-110)	112	33,734939
Gute Intelligenz (IQ=110-120)	194	58,433734
Sehr gute Intelligenz (IQ>120)	26	7,8313253

IQ-Bereich	Anzahl der Jungen	in %
Unterdurchschnittliche Intelligenz (IQ<90)	1	0,5649717
Durchschnittliche Intelligenz(IQ=90-110)	53	29,943502
Gute Intelligenz (IQ=110-120)	86	48,587570
Sehr gute Intelligenz (IQ>120)	37	20,903054

IQ-Bereich	Anzahl der Schüler Gesamt	in %

Unterdurchschnittliche Intelligenz (IQ<90) 1 0,1964636
Durchschnittliche Intelligenz(IQ=90-110) 165 32,416502
Gute Intelligenz (IQ=110-120) 280 55,009823
Sehr gute Intelligenz (IQ>120) 63 12,377210

Da der FRT als Kontrolle diente, wurde er an denselben Schülern wie N-I-T_B, N-I-T_1 und N-I-T_2 durchgeführt. Wegen der Vergleichbarkeit mit N-I-T_B, N-I-T_1 und N-I-T_2 wurde er auch der Stichprobe entsprechend FRT_B, FRT_1 und FRT_2 genannt. Die Anzahl der Schüler, die am FRT_2 insgesamt teilgenommen haben, setzt sich aus der Anzahl der Schüler, die an FRT_B, FRT_1 und FRT_2 teilgenommen haben, zusammen.

Tabelle XXI: Die Verteilung der weiblichen (w) und männlichen (m) deutschen Schüler nach den einzelnen Tests sowie die Mittelwerte und Standardabweichungen der einzelnen Tests

Test	Vpn		Mittelwerte	Standard-abweichung
$N\text{-}I\text{-}T_B$	w	96	31,396	6,623
	m	44	31,523	7,242
$N\text{-}I\text{-}T_1$	w	101	20,574	6,579
	m	102	21,794	7,564
$N\text{-}I\text{-}T_2$	w	111	30,793	9,171
	m	40	30,750	9,559
FRT_B	w	114	35,360	3,947
	m	41	36,976	3,883
FRT_1	w	107	35,850	3,395
	m	96	36,521	4,065
FRT_2	w	111	34,694	3,702
	m	40	34,525	5,862
FRT	w	332	35,295	3,713
	m	177	36,175	4,560

Obwohl N-I-T_B und FRT_B wie auch $N\text{-}I\text{-}T_1$ und FRT_1 in den gleichen Schulklassen durchgeführt worden sind, ist die Anzahl der Versuchspersonen beim N-I-T_B nicht die gleiche wie beim FRT_B und beim $N\text{-}I\text{-}T_1$ nicht die gleiche wie

beim FRT_1. Der Grund dafür ist, daß diese Tests an verschiedenen Tagen durchgeführt wurden

und einige Schüler bei der ersten bzw. zweiten Testdurchführung fehlten.

5.3. Die Ergebnisse

Die 40 Aufgaben des N-I-T_B, 50 Aufgaben des N-I-T_1, 50 Aufgaben des N-I-T_2, die 45 Aufgaben des FRT_B, FRT_1, FRT_2 und FRT wurden in bezug auf die Rohwertgruppen und Geschlecht nach der Methode der bedingten Wahrscheinlichkeit anhand der deutschen Probanden auf Anpassung des Rasch-Modells geprüft.

Tabelle XXII: Die Ergebnisse der Rasch-Analyse des N-I-T_B, N-I-T_1, N-I-T_2, FRT_B, FRT_1 und FRT_2 und FRT in bezug auf die Rohwert- und Geschlechtsgruppen der deutschen Schüler

VpnRohwertgr.	Geschlecht	
Gruppe 1	60	90 weibl. deutsche Schüler
Gruppe 2	71	41 männl. deutsche Schüler
Gesamt	131	131
Ausgeschiedene Versuchspersonen	0	0
N-I-T_B: Chi-Quadrat(X^2)	34,5064	48,1057
Freiheitsgrade(df)	35	39
Information(I)	660,0808	696,6531

Signifikanzniveau(S) P=95%(n.s.) P=95%(n.s.)

Fortsetzung Tabelle XXII:

VpnRohwertgr.	Geschlecht
Gruppe 1 98	101 weibl. deutsche Schüler
Gruppe 2 105	102 männl. deutsche Schüler
Gesamt 203	203
Ausgeschiedene Versuchspersonen 0	0
N-I-T_1: Chi-Quadrat(X^2) 80,1969	62,9955
Freiheitsgrade(df) 40	46
Information(I) 1442,2651	1535,2858
Signifikanzniveau(S) P=0,1%(s.)	P=10%(n.s.)

VpnRohwertgr.	Geschlecht
Gruppe 1 76	111 weibl. deutsche Schüler
Gruppe 2 74	39 männl. deutsche Schüler
Gesamt 150	150
Ausgeschiedene Versuchspersonen 1	1
N-I-T_2: Chi-Quadrat(X^2) 141,3630	50,0079
Freiheitsgrade(df) 47	47
Information(I) 1167,6872	1198,0359
Signifikanzniveau P=0,1%(s.)	P=95%(n.s.)

VpnRohwertgr. Geschlecht

Gruppe 1 79 114 weibl. deutsche Schüler
Gruppe 2 75 41 männl. deutsche Schüler
Gesamt 154 155
Ausgeschiedene Versuchspersonen 1 0
FRT_B: Chi-Quadrat(X^2) 62,1845 26,0819
Freiheitsgrade(df) 21 26
Information(I) 492,5611 550,9697
Signifikanzniveau(S) P=0,1%(s.) P=95%(n.s.)

VpnRohwertgr. Geschlecht
Gruppe 1 108 105 weibl. deutsche Schüler
Gruppe 2 91 96 männl. deutsche Schüler
Gesamt 199 199
Ausgeschiedene Versuchspersonen 4 4
FRT_1: Chi-Quadrat(X^2) 84,1293 41,2391
Freiheitsgrade(df) 20 28
Information(I) 642,0905 752,7233
Signifikanzniveau(S) P=0,1%(s.) P=10%(n.s.)

Fortsetzung Tabelle XXII:

VpnRohwertgr. Geschlecht
Gruppe 1 73 111 weibl. deutsche Schüler
Gruppe 2 78 40 männl. deutsche Schüler
Gesamt 151 151

Ausgeschiedene Versuchspersonen	0	0
FRT$_2$: Chi-Quadrat(X^2)	32,2659	43,9598
Freiheitsgrade(df)	21	30
Information(I)	501,9341	592,0746
Signifikanzniveau(S)	P=10%(n.s.)	P=5%(n.s.)

Vpn	Rohwertgr.	Geschlecht	
Gruppe 1	266	332	weibl. deutsche Schüler
Gruppe 2	238	177	männl. deutsche Schüler
Gesamt	504	509	
Ausgeschiedene Versuchspersonen	5	0	
FRT: Chi-Quadrat(X^2)	131,1342	63,6758	
Freiheitsgrade(df)	22	30	
Information(I)	1717,4103	1927,6419	
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)	

Tabelle XXIII: Ergebnisse der Raschanalyse des N-I-T$_1$ anhand von 40 Aufgaben und nach Eliminierung von drei modellinkonformen Aufgaben anhand von 37 Aufgaben des N-I-T$_1$ und nach Eliminierung von 9 modellinkonformen Aufgaben anhand von 41 Aufgaben des N-I-T$_2$ in bezug auf die Rohwert- und Geschlechtsgruppen der deutschen Schüler

Vpn	Rohwertgr.	Geschlecht

Gruppe 1 102 101 weibl. deutsche Schüler
Gruppe 2 101 102 männl. deutsche Schüler
Gesamt 203 203
Ausgeschiedene Versuchspersonen 0 0
N-I-T_1 (40 Aufgaben): Chi-Quadrat(X^2) 63,3868
53,3854
Freiheitsgrade(df) 38 38
Information(I) 1391,9418 1393,3040
Signifikanzniveau P=1%(s.) P=95%(n.s.)

Fortsetzung Tabelle XXIII:

VpnRohwertgr. Geschlecht
Gruppe 1 95 101 weibl. deutsche Schüler
Gruppe 2 108 102 männl. deutsche Schüler
Gesamt 203 203
Ausgeschiedene Versuchspersonen 0 0
N-I-T_1 (37 Aufgaben): Chi-Quadrat(X^2) 54,0759
47,6910
Freiheitsgrade(df) 35 35
Information(I) 1258,3643 1258,8450
Signifikanzniveau(S) P=5%(n.s.) P=10%(n.s.)

VpnRohwertgr. Geschlecht
Gruppe 1 72 39 weibl. deutsche Schüler

Gruppe 2	77	110 männl. deutsche Schüler
Gesamt	149	149
Ausgeschiedene Versuchspersonen	2	2
N-I-T_2 (41 Aufgaben): Chi-Quadrat(X^2)	43,0287	38,4716
Freiheitsgrade(df)	38	38
Information(I)	942,5047	944,9481
Signifikanzniveau(S)	P=95%(n.s.)	P=95%(n.s.)

Tabelle XXIV: Chi-Quadratgröße (X^2) der Rasch-Analyse der einzelnen Tests (N-I-T_B, N-I-T_1, N-I-T_2, FRT_B, FRT_1, FRT_2 und FRT) in bezug auf die Rohwerte und Geschlechtsgruppen der deutschen Probanden dividiert durch die entsprechenden Freiheitsgrade (df)

Test	Rohwertgruppen X^2/df	Geschlecht X^2/df
N-I-T_B	0,9858971	1,2334794
N-I-T_1	2,0049225	1,3694673
N-I-T_2	3,0077234	1,0639978
FRT_B	2,9611666	1,00315
FRT_1	4,206465	1,472826
FRT_2	1,5364714	1,4653266
FRT	5,4606454	2,1225266

Tabelle XXV: Die Anzahl der ausgeschiedenen Aufgaben (Lösungswahrscheinlichkeit kleiner 3% oder größer 97%) bei den einzelnen Tests in einer Raschanalyse anhand der deutschen Probanden

Tests	Anzahl der ausgeschiedenen Aufgaben	
	Rohwertgruppen	Geschlechtsgruppen
N-I-T_B	4	0
$N\text{-}I\text{-}T_1$	1	1
$N\text{-}I\text{-}T_2$	1	1
FRT_B	23	18
FRT_1	24	16
FRT_2	23	14

Die Z-Werte der Rasch-Analyse der Aufgaben anhand der deutschen Probanden des N-I-T_B sind in Tabelle 16, des $N\text{-}I\text{-}T_1$-50-Aufgaben in Tabelle 18, $N\text{-}I\text{-}T_1$-40-Aufgaben nach Eliminierung von 3 Aufgaben in Tabelle 19, $N\text{-}I\text{-}T_2$-50-Aufgaben in Tabelle 20, $N\text{-}I\text{-}T_2$-50-Aufgaben nach Eliminierung von neun Aufgaben in Tabelle 21, FRT_B in Tabelle 22, FRT_1 in Tabelle 23, FRT_2 in Tabelle 24 und FRT in Tabelle 25, Anhang A zusammen dargestellt.

5.4. Diskussion

Aus den Ergebnissen der Untersuchung in der Bundesrepublik geht hervor, daß die X^2-Werte des N-I-T_B sowohl in bezug auf Rohwertgruppen (Intelligenzgruppen) als auch in bezug auf das Geschlecht (bei P=95%) nicht signifikant sind. D.h., der N-I-T_B ist dem Rasch-Modell bezüg-

lich der deutschen Probanden konform und besitzt damit spezifische Objektivität für die Deutschen (vgl. Tabelle XXII).

Die X^2-Werte des N-I-T_1 (50 Aufgaben) sind in bezug auf Rohwertgruppen (bei P=0,1%) signifikant und in bezug auf das Geschlecht (bei P=10%) nicht signifikant. Das bedeutet, daß dieser Test hinsichtlich der Rohwertgruppen mit der Rasch-Modell unverträglich ist, aber in bezug auf das Geschlecht konform ist.

Beim N-I-T_1 (40 Aufgaben) handelt es sich um eine Mischform des N-I-T_B. Während der N-I-T_B, wie schon gesagt, sowohl in bezug auf die Rohwertgruppen als auch hinsichtlich des

Geschlechts (bei P=95%) nicht signifikant und damit dem Rasch-Modell angepaßt war (s. Tabelle XXII); ist die Chi-Quadratgröße des N-I-T_1 (40 Aufgaben) in bezug auf die Rohwertgruppen (bei P=1%) signifikant und in bezug auf das Geschlecht (bei P=5%) nicht signifikant. Das bedeutet, daß dieser Test in bezug auf die Rohwertgruppen dem Rasch-Modell nicht konform ist, aber hinsichtlich des Geschlechts dem Rasch-Modell konform ist. Diese unterschiedlichen Ergebnisse sind unter anderem auf die Mischung der je 10 Aufgaben der vier Untertests des N-I-T_B zurückzuführen, aus der N-I-T_1 (40 Aufgaben) resultierte. Mit anderen Worten: Je mehr Testvariable in einem Test vorhanden sind, desto größere X^2-Werte und schlechtere Anpassung an das Rasch-Modell liegt vor.

Nach Eliminierung von drei modellinkonformen Aufgaben (Aufgaben 6, 12 und 25) könnte N-I-T_1 (40 Aufgaben) auch in bezug auf die Rohwertgruppen (bei P=5%) und in bezug auf das Geschlecht (bei P=10%) nicht signifikant

sein, d.h. dem Rasch-Modell angepaßt werden (s. Tabelle XXIII).

Die X^2-Werte des N-I-T_2 sind in bezug auf die Rohwertgruppen (bei P=0,1%) signifikant und in bezug auf das Geschlecht (bei P=95%) nicht signifikant. Das heißt, der N-I-T_2 ist in bezug auf die Rohwertgruppen dem Modell von Rasch nicht konform, aber in bezug auf das Geschlecht dem Rasch-Modell konform (s. Tabelle XXII). Nach Ausschaltung von neun modellinkonformen Aufgaben (Aufgabe 11, 15, 17, 35, 37, 39, 40, 44 und 46) könnte auch dieser Test sowohl in bezug auf die Rohwertgruppen als auch in bezug auf das Geschlecht (bei P=95% Nicht-Signifikanz) dem Rasch-Modell angepaßt werden (vgl. Tabelle XXIII).

Vergleicht man die X^2-Werte des Figure-Reasoning-Tests (FRT_B) mit dem N-I-T_B, der an denselben Versuchspersonen durchgeführt wurde, so ergibt sich, daß der FRT_B in bezug auf die Rohwertgruppen (bei P=0,1%) signifikant und in bezug auf das Geschlecht (bei P=95%) nicht signifikant ist. Das heißt, daß der FRT_B in bezug auf die Rohwertgruppen der deutschen Probanden dem Rasch-Modell

nicht konform ist, aber in bezug auf das Geschlecht dem Rasch-Modell konform ist (vgl. Tabelle XXII), während

N-I-T_B sowohl in bezug auf die Rohwertgruppen als auch in bezug auf das Geschlecht bei den deutschen Probanden (bei P=95% nicht signifikant) dem Rasch-Modell angepaßt ist.

Aus dem Vergleich der X^2-Werte des Figure-Reasoning-Tests(FRT_1) mit dem $N\text{-}I\text{-}T_1$ (50 Aufgaben), die an der gleichen Stichprobe durchgeführt wurden, ergibt sich, daß FRT_1 und $N\text{-}I\text{-}T_1$ beide in bezug auf die Rohwertgruppen (bei P=0,1%) signifikant und in bezug auf das Geschlecht (bei P=10%) nicht signifikant sind. Sie sind in bezug auf die Rohwertgruppen dem Rasch-Modell nicht angepaßt, aber in bezug auf das Geschlecht angepaßt. Nach Ausschaltung einiger Aufgaben könnte der $N\text{-}I\text{-}T_1$ jedoch auch in bezug auf die Rohwerte dem Modell angepaßt werden (vgl. Tabelle XXIII).

Vergleicht man die Ergebnisse des Figure-Reasoning-Tests (FRT_2) mit dem N-I-T_2, der an derselben Stichprobe durchgeführt wurde wie FRT_2, so sind die X^2-Werte des FRT_2 in bezug auf die Rohwertgruppen (bei P=10%) und in bezug auf das Geschlecht (bei P=5%) nicht signifikant, d.h. dem Modell von Rasch angepaßt. Dagegen sind die X^2-Werte des N-I-T_2 in bezug auf die Rohwertgruppen (bei P=0,1%) signifikant und in bezug auf das Geschlecht (bei P=95%) nicht signifikant, also hinsichtlich der Rohwertgruppen mit dem Rasch-Modell nicht konform, aber in bezug auf das Geschlecht konform. Jedoch könnte nach Eliminie-

rung von neun Aufgaben (Aufgaben 11, 15, 17, 35, 37, 39, 40, 44 und 46) die X^2-Werte des N-I-T_2 in bezug auf die Rohwertgruppen (bei P=95% nicht signifikanz) auch dem Rasch-Modell angepaßt werden (s. Tabelle XXIII).

Aus der Tabelle XXIV ist ersichtlich, daß die X^2-Werte der einzelnen Tests in bezug auf die Rohwertgruppen und Geschlechtsgruppen der

deutschen Probanden unterschiedliche Größen aufweisen. Also sind die Ergebnisse der Rasch-Analyse von den gewählten Außenkriterien abhängig. Ist ein Test in bezug auf ein Außenkriterium dem Rasch-Modell angepaßt, so bedeutet das nicht, daß er in bezug auf andere Kriterien auch dem Rasch-Modell angepaßt sein muß.

Außerdem geht aus der Tabelle XXV hervor, daß bei der Rasch-Analyse des N-I-T_B, N-I-T_1 und N-I-T_2 anhand der deutschen Probanden weniger Aufgaben ausgeschieden sind (vgl. auch Tabelle 16, 17, 20, 22, 23 und 24, Anhang A) als bei dem als Vergleich dienenden FRT (FRT_B, FRT_1, FRT_2).

Die Ausscheidung so vieler Aufgaben vermindert die Bedeutung der übriggebliebenen Aufgaben des FRT_B als Test. Bei den meisten der ausgeschiedenen Aufgaben ist dies darauf zurückzuführen, daß sie in einer der Rohwertgruppen und Geschlechtsgruppen von mehr als 97% der Probanden richtig gelöst worden sind.

Vergleicht man die Ergebnisse des FRT_B, FRT_1, FRT_2 und FRT untereinander, so läßt sich feststellen, daß FRT_B und FRT_1 in bezug auf die Rohwertgruppen dem Modell von Rasch nicht konform sind, aber in bezug auf das Geschlecht dem Rasch-Modell konform sind. Der FRT_2 ist sowohl in bezug auf die Rohwertgruppen als auch in bezug auf das Geschlecht dem Rasch-Modell konform. Der FRT, bei dem alle drei Stichproben zusammengefaßt worden sind, ist hinsichtlich der beiden Kriterien (Rohwertgruppen und Geschlecht) dem Rasch-Modell nicht konform (vgl. Tabelle XXII).

Transkription der Seiten 91–110

– 91 –

6. Die Analyse der Aufgaben des N-I-T_B auf Anpassung des Rasch-Modells nach der Methode der bedingten Wahrscheinlichkeit anhand der verschiedenen Gruppen der Afghanen und Deutschen

6.1. Fragestellung

Es ergibt sich die Fragestellung, ob die Aufgaben des N-I-T_B für Afghanen und Deutsche dem Rasch-Modell konform sind.

6.2. Methode

Die Aufgaben des N-I-T_B wurden anhand der erhaltenen Daten in Afghanistan und der Bundesrepublik Deutschland auf Konformität mit dem Rasch-Modell analysiert. Als Vergleich dient der Figure Reasoning Test (FRT) von Daniels.

6.3. Ergebnisse

Tabelle XXVI: Ergebnisse der Rasch-Analyse des N-I-T_B und des FRT in bezug auf Afghanen und Deutsche, afghanische und deutsche Schüler, weibliche afghanische und weibliche deutsche Schüler, männliche afghanische und männliche deutsche Schüler

Rohwertgruppen	Nationalitätsgruppen
Gruppe 1 180	227(Afghanen)

Gruppe 2	178	131(Deutsche)
Gesamt	358	358
Ausgeschiedene Versuchspersonen	9	9
N-I-T_B: Chi-Quadrat(X^2)	170,9656	337,1715
Freiheitsgrade(df)	39	39
Information(I)	2486,4692	2452,2121
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)

Rohwertgruppen	Nationalitätsgruppen	
Gruppe 1	160	192(afghan. Schüler)
Gruppe 2	163	131(deutsche Schüler)
Gesamt	323	323
Ausgeschiedene Versuchspersonen	9	9
Chi-Quadrat(X^2)	123,8733	307,8697
Freiheitsgrade(df)	39	39
Information(I)	2226,8955	2188,9226
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)

Rohwertgruppen	Nationalitätsgruppen	
Gruppe 1	54	23(weibl.afgh. Schüler)
Gruppe 2	59	90(weibl.deutsche Schüler)
Gesamt	113	113
Ausgeschiedene Versuchspersonen	6	6
Chi-Quadrat(X^2)	44,7422	150,4124
Freiheitsgrade(df)	37	37
Information(I)	653,0645	608,2220
Signifikanzniveau(S)	P=10%(n.s.)	P=0,1%(s.)

Fortsetzung Tabelle XXVI:

	Rohwertgruppen	Nationalitätsgruppen
N-I-T_B Gruppe 1	108	169(männl.afgh. Schüler)
Gruppe 2	102	41(männl.deutsche Schüler)
Gesamt	210	210
Ausgeschiedene Versuchspersonen	0	0
Chi-Quadrat(X^2)	202,0757	174,7352
Freiheitsgrade(df)	39	39
Information(I)	1474,4133	1481,4022
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)

	Rohwertgruppen	Nationalitätsgruppen
Gruppe 1	375	258(Afghanen)
Gruppe 2	388	509(Deutsche)
Gesamt	763	767
Ausgeschiedene Versuchspersonen	6	2
FRT: Chi-Quadrat(X^2)	371,7382	1025,4964
Freiheitsgrade(df)	27	33
Information(I)	3026,9998	3331,0679
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)

	Rohwertgruppen	Nationalitätsgruppen
Gruppe 1	379	225(afgh. Schüler)

Gruppe 2	351	509(deutsche Schüler)
Gesamt	730	734
Ausgeschiedene Versuchspersonen	6	2
Chi-Quadrat(X^2)	233,5913	931,5090
Freiheitsgrade(df)	25	33
Information(I)	2719,2615	3062,0984
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)

	Rohwertgruppen	Nationalitätsgruppen
Gruppe 1	173	26(weibl.afgh. Schüler)
Gruppe 2	182	332(weibl.deutsche Schüler)
Gesamt	355	358
Ausgeschiedene Versuchspersonen	3	0
Chi-Quadrat(X^2)	68,3517	216,4440
Freiheitsgrade(df)	23	26
Information(I)	1289,8876	1151,5340
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)

Fortsetzung Tabelle XXVI:

	Rohwertgruppen	Nationalitätsgruppen
FRT Gruppe 1	186	197(männl.afgh. Schüler)
Gruppe 2	190	177(männl.deutsche Schüler)
Gesamt	376	374
Ausgeschiedene Versuchspersonen	2	4

Chi-Quadrat(X^2)	381,429	440,7595	
Freiheitsgrade(df)	32	31	
Information(I)	1579,9691	1533,9733	
Signifikanzniveau(S)	P=0,1%(s.)	P=0,1%(s.)	

Tabelle XXVII: Chi-Quadratgröße (X^2) der Rasch-Analyse des Naturformen-Intelligenz-Tests (N-I-T)-Teiltest B und des Figure-Reasoning-Tests (FRT) in bezug auf die Rohwert- und Nationalitätsgruppen (Afghanen und Deutsche) dividiert durch die entsprechenden Freiheitsgrade (df)

Test	Rohwertgruppen X^2/df	Nationalitätsgruppen X^2/df
N-I-T_B	4.3837333	Afghanen/Deutsche 8.645423
	3.1762384	afgh./deutsche Schüler 7.8940948
	1.2092486	weibl.afgh.Schüler/weibl.deutsche Schüler 4.0652
	5.1814282	männl.afgh.Schüler/männl.deutsche Schüler 4.4803897
FRT	13.768081	Afghanen/Deutsche 31.075648
	9.343652	afgh./deutsche Schüler 28.227545
	2.971813	weibl.afgh.Schüler/weibl.deutsche Schüler 8.3247692
	11.919656	männl.afgh.Schüler/männl.deutsche Schüler 14.218048

Tabelle XXIII: Die Anzahl (N), die Mittelwerte (X) und die Streuungen (S) der afghanischen und deutschen Probanden beim Naturformen-Intelligenz-Test-Teiltest B (N-I-T_B) und Figure-Reasoning-Test (FRT)

		N	$\bar{X}$	S
N-I-T_B (40 Aufgaben)	Afghanistan	227	17,90	6,63
	Deutschland	140	31,44	6,82
FRT (45 Aufgaben)	Afghanistan	260	14,75	5,68
	Deutschland	509	35,61	4,02

Tabelle XXIX: Die Anzahl der weiblichen (w) und männlichen (m) afghanischen Probanden, die Anzahl der afghanischen Analphabeten (A) und Schüler (Sch) bei den einzelnen Tests, die Mittelwerte ($\bar{X}$), Streuungen (S) dieser Gruppen beim Figure-Reasoning-Test (FRT), beim Leistungsprüfsystem Aufgabengruppe 9, Form B (Lps9B) und Aufgabengruppe 10 B (Lps10B), beim Progressive Matrices Test (PMT), beim Naturformen-Intelligenz-Test-Teiltest A (N-I-T_A) und Teiltest B (N-I-T_B)

Test		Anzahl	$\bar{X}$	S
FRT	w	35	16,200	4,303
	m	225	14,529	5,956
Lps 9 B	w	33	10,909	5,095
	m	190	11,837	5,520
Lps 10 B	w	33	18,970	4,838
	m	190	15,395	4,893
PMT	w	20	6,300	3,948
	m	187	8,230	4,704
N-I-T_A	w	32	44,438	7,658
	m	195	47,000	13,410
N-I-T_B	w	32	17,969	5,070
	m	195	17,887	6,887

Fortsetzung Tabelle XXIX:

Test		Anzahl	$\bar{X}$	S
FRT	A	33	15,455	5,574
	Sch	227	14,652	5,716
Lps 9 B	A	30	13,133	4,848
	Sch	193	11,477	5,562
Lps 10 B	A	30	15,600	6,072
	Sch	193	15,974	4,874
PMT	A	27	9,630	2,789

	Sch	180	7,896	4,843
N-I-T_A	A	35	39,029	9,532
	Sch	192	48,026	12,822
N-I-T_B	A	35	14,500	5,595
	Sch	192	18,500	6,665

Die Z-Werte der Rasch-Analyse der Aufgaben des N-I-T_B in bezug auf Afghanen und Deutsche sind in Tabelle 26, in bezug auf afghanische und deutsche Schüler in Tabelle 26, in bezug auf weibliche afghanische und weibliche deutsche Schüler in Tabelle 28, in bezug auf männliche afghanische und männliche deutsche Schüler in Tabelle 29 und des FRT in bezug auf Afghanen und Deutsche in Tabelle 30, in bezug auf afghanische und deutsche Schüler in Tabelle 31, in bezug auf weibliche afghanische und weibliche deutsche Schüler in Tabelle 32, in bezug auf männliche afghanische und männliche deutsche Schüler in Tabelle 33, Anhang A dargestellt.

6.4. Diskussion

Die X^2-Werte des N-I-T_B sind sowohl nach den Kriterien der Rohwertgruppen als auch nach der Nationalität (Afghanen und Deutsche) signifikant (P=0,1%). Das heißt, daß dieser Test sowohl in bezug auf die Rohwertgruppen als auch in bezug auf die Nationalität (Afghanen und Deutsche) dem Rasch-Modell nicht konform ist, aber, wie aus den Z-Werten ersichtlich ist, gibt es dabei einige Aufgaben, die in bezug auf diese beiden Kriterien (Rohwertgruppen und Nationalität) sich mit dem Rasch-Modell vertragen (vgl. Tabelle 26, Anhang A).

Nach Eliminierung der Analphabeten bei den Stichproben der Afghanen wurden die X^2-Werte des N-I-T_B sowohl in bezug auf die Rohwertgruppen als auch in bezug auf die Nationalität (afghanische und deutsche Schüler) etwas kleiner als die X^2-Werte des N-I-T_B bei der Gesamtstichprobe mit Analphabeten, aber sie bleiben (bei P=0,1%) signifikant (vgl. Tabelle XXVI und XXVII).

Die Analyse des N-I-T_B nach der Ausschaltung der Analphabeten und der männlichen Schüler, also nur anhand der weiblichen afghanischen

und deutschen Schüler, ergab, daß die X^2-Werte des N-I-T_B, sowohl in bezug auf die Rohwertgruppen als auch in bezug auf die Nationalität (weibliche afghanische und weibliche deutsche Schüler),

viel kleiner wurde, und zwar hinsichtlich der Rohwertgruppen (bei P=10%) nicht signifikant, allerdings in bezug auf die Nationalität immer noch (bei P=0,1%) signifikant sind (vgl. Tabelle XXVI und XXVII). Auch die Anzahl der mit dem Modell verträglichen Aufgaben wurde größer (vgl. Tabelle XXVI und XXVII).

Bei der Analyse des N-I-T_B nach Ausschaltung der Analphabeten und der weiblichen Schüler, also anhand der männlichen afghanischen und männlichen deutschen Schüler, wurden die X^2-Werte des N-I-T_B bei gleichen Freiheitsgraden in bezug auf die Rohwertgruppen größer und in bezug auf die Nationalität (männliche afghanische und männliche deutsche Schüler) kleiner als die X^2-Werte des N-I-T_B anhand der Analyse der Gesamtstichproben. Die X^2-Werte sind jedoch in beiden Fällen in bezug auf die

Rohwertgruppen und Nationalitätsgruppen (bei P=0,1%) signifikant (vgl. Tabelle XXVI).

Die X^2-Werte des Figure-Reasoning-Tests (FRT), der als Vergleich dient, sind sowohl in bezug auf die Rohwertgruppen als auch in bezug auf die Afghanen und Deutschen (bei P=0,1%) signifikant (vgl. Tabelle XXVI). Außerdem geht aus den Z-Werten hervor, daß bei diesem Test kaum oder nur sehr wenige Aufgaben existieren, die sich mit dem Modell von Rasch vertragen (vgl. Tabelle 30, 31, 32 und 33, Anhang A).

Nach Ausschaltung der Analphabeten bei den afghanischen Probandengruppen wurden die X^2-Werte des FRT sowohl in bezug auf die Rohwertgruppen als auch in bezug auf die Nationalität (afghanische und deutsche Schüler) etwas kleiner als die X^2-Werte des FRT anhand der Gesamtstichprobe mit Analphabeten, aber sie blieben (bei P=0,1%) signifikant (vgl. Tabelle XXVI).

Die X^2-Werte der Rasch-Analyse des FRT anhand der weiblichen afghanischen und weiblichen deutschen Schüler sind sowohl in bezug auf die Rohwertgruppen als auch in bezug auf die Nationalität (weibliche afghanische und weibliche deutsche Schüler) kleiner als die X^2-Werte des FRT anhand der Gesamtstichprobe. Dagegen sind die X^2-Werte des FRT anhand der Analyse der männlichen afghanischen und männlichen deutschen Schüler in bezug auf die Rohwertgruppen größer und in bezug auf die Nationalitäten (männliche afghanische und männliche deutsche Schüler) kleiner als die X^2-Werte des FRT anhand der Analyse der Gesamtstichprobe (vgl. Tabelle XXVII). Aber die X^2-Werte des FRT bleiben in bezug auf die Rohwertgruppen und in bezug auf weibliche afghanische und weibliche deutsche Schüler sowie männliche afghanische und männliche deutsche Schüler (bei P=0,1%) signifikant (vgl. Tabelle XXVI).

auch in bezug auf die Nationalitätsgruppen erheblich kleiner sind als die des FRT. Das bedeutet, daß der N-I-T_B gemessen an der X^2-

Größe in bezug auf die Afghanen und Deutschen hinsichtlich des Rasch-Modells erheblich besser ist als der FRT.

Die X^2-Größe des FRT in bezug auf die Rohwertgruppen und die Nationalitätsgruppen läßt keine Hoffnung auf die Modell-Verträglichkeit einiger Aufgaben dieses Tests; während diese Hoffnung beim N-I-T_B nicht aufzugeben ist (vgl. Tabelle XXVI und XXVII). Die X^2-Werte des N-I-T_B sind schon anhand der Analyse der weiblichen afghanischen und weiblichen deutschen Schüler in bezug auf die Rohwertgruppen (bei P=10%) nicht signifikant, d.h. dem Rasch-Modell konform. Die X^2-Werte des N-I-T_B sind in bezug auf die Nationalität (weibliche afghanische und weibliche deutsche Schüler) auch nicht besonders hoch, so daß der N-I-T_B nach Ausschaltung einiger Aufgaben dem Modell angepaßt werden kann (vgl. Tabelle XXVI und 28, Anhang A). Außerdem ist beim FRT fast 1/3 der Aufgaben bei all diesen Analysen sowohl bei den Rohwertgruppen als auch bei den Nationalitätsgruppen ausgeschieden, während beim N-I-T_B nur bei der Analyse anhand der weiblichen afghanischen und weiblichen

deutschen Schüler bei Rohwertgruppen und Nationalität jeweils zwei Aufgaben ausgeschieden sind (vgl. Tabelle 28, 30, 31, 32 und 33, Anhang A).

Ferner beträgt der durchschnittliche Unterschied zwischen den afghanischen und den deutschen Probanden beim N-I-T_B 13,54 Aufgaben (vgl. Tabelle XXIII). Das bedeutet, daß die Deutschen beim N-I-T_B im Durchschnitt 13,54 Aufgaben mehr richtig beantwortet haben als die Afghanen. Rechnet man die Anzahl der 40 Aufgaben des N-I-T_B auf die vergleichbare Größe von 45 Aufgaben um, so würde dieser Unterschied 15,2325 Aufgaben betragen, während er beim FRT im Durchschnitt 20,86 Aufgaben beträgt (vgl. Tabelle XXIII). Also ist der durchschnittliche Unterschied zwischen deutschen und afghanischen Probanden beim N-I-T_B kleiner als beim FRT.

Außerdem haben die Analphabeten in Afghanistan, abgesehen vom Lps 10 B, bei den traditionellen Tests (FRT, Lps 9 B, PMT) im Durchschnitt mehr Aufgaben richtig

beantwortet als die afghanischen Schüler, während dies beim N-I-T_B umgekehrt der Fall ist. Die Analphabeten haben im Durchschnitt 14,500 Aufgaben und die Schüler im Durchschnitt 18,500 Aufgaben richtig beantwortet (vgl. Tabelle XXIX). Die Ergebnisse des N-I-T_B entsprechen der Erwartung, daß die Analphabeten wegen ihrer ungünstigeren Lebensbedingungen etwas weniger Aufgaben richtig beantworten können als die Schüler.

Ansonsten zeigen die Ergebnisse des N-I-T_B und des FRT einen ähnlichen Trend, nämlich daß die X^2-Werte dieser

beiden Tests bei der Analyse nach Ausschaltung von Analphabeten anhand der afghanischen und deutschen Schüler in bezug auf die beiden Kriterien (Rohwertgruppen und Nationalität) kleiner geworden sind als die bei der Analyse dieser Tests anhand der Gesamtstichprobe (Afghanen und Deutsche), und nach Ausschaltung der männlichen Probanden wurden sie noch kleiner (vgl. Tabelle XXVII). Die Unterschiede zwischen

diesen X^2-Werten sind erheblich groß, so daß es nicht gerechtfertigt ist, wenn man sie auf einen statistischen Effekt (Stichproben-Umfang) zurückführt. Vielmehr liegt die Vermutung nahe, daß, je mehr Kriterienvariablen (Ausbildung, Geschlecht, Alter, Sozialstatus, Schultypen usw.) in einer Stichprobe vorhanden sind, desto größer die X^2-Werte sind und sich somit modellinkonforme Ergebnisse des Tests ergeben, der anhand dieser Stichprobe analysiert wurde.

Aus dem Vergleich der X^2-Werte des N-I-T_B, der anhand der weiblichen afghanischen und weiblichen deutschen Schüler analysiert wurde, mit den X^2-Werten des N-I-T_B, der anhand der männlichen afghanischen und männlichen deutschen Schüler analysiert wurde, ergibt sich, daß die X^2-Werte des N-I-T_B nach den beiden Kriterien (Rohwertgruppen und Nationalität) bei den weiblichen Schülern kleiner sind als bei den männlichen Schülern. Dasselbe gilt auch für die X^2-Werte des FRT (vgl. Tabelle XXVII). Dies läßt sich u.a. dadurch erklären, daß die

Untersuchung an den männlichen Schülern in Kabul und der Provinz Kunar (Ostafghanistan) durchgeführt wurde, während die Untersuchung an den weiblichen Schülern auf eine Schule in der Provinz Kunar beschränkt war. D.h., daß die Anzahl der Kriterienvariablen in der männlichen Stichprobe größer ist als in der weiblichen Stichprobe und sich deshalb für die männlichen Probanden größere X^2-Werte ergeben als für die weiblichen Probanden.

7. Einige mit dem Rasch-Modell zusammenhängende Fragen

7.1. Innere Teilungskriterien

Nachdem, wie in einigen anderen Untersuchungen (s. Fischer, 1974, S. 320), auch in unserer Untersuchung gezeigt wurde, daß die Ergebnisse hinsichtlich der Modellkonformität eines Tests von den Außenkriterien (Geschlecht, Ausbildung, Alter usw.) der Stichproben abhängen, wurde unser Interesse auf die Frage der Abhängigkeit des

Ergebnisses von den inneren Teilungskriterien der Stichprobe (Rohwertgruppen) gelenkt.

7.1.1. Fragestellung

Hängt das Ergebnis hinsichtlich der Modellkonformität eines Tests von den Teilungskriterien der Rohwertgruppen ab, d.h. wird ein anderes Ergebnis erzielt, wenn man die Probanden nach den Rohwerten des Gesamttests oder nach den Rohwerten von Teiltests in Gruppen teilt?

7.1.2. Methode

Der in Afghanistan durchgeführte Figure-Reasoning-Test (FRT) soll durch schrittweise Eliminierung der modellin-

konformen Aufgaben hinsichtlich der Rohwertgruppen (Intelligenzgruppen) dem Rasch-Modell angepaßt werden. Die konformen Aufgaben des FRT sollen dann nach geraden und ungeraden Aufgaben-Nummern in zwei Hälften (FRT Teiltest A und FRT Teiltest B)

geteilt werden. Dann soll einmal nach den Rohwerten der gesamten modellangepaßten Aufgaben des FRT mit Hilfe der klassischen Testanalyse und zum anderen nach den eigenen Rohwerten der Aufgaben des FRT-Teiltest A und des FRT-Teiltest B die Gruppen gebildet und auf Modellkonformität analysiert werden.

7.1.3. Die Ergebnisse

Die Ergebnisse sind in den Tabellen 34, 35 und 36 im Anhang A dargestellt.

7.1.4. Diskussion

Aus den Ergebnissen (s. Tabelle 34, Anhang A) geht hervor, daß die Aufgaben des FRT nach Eliminierung von 11 modellinkonformen Aufgaben (Aufgaben 6, 8, 13, 14, 16, 17, 19, 24, 27, 29 und 41) in bezug auf die afghanischen Rohwertgruppen dem Rasch-Modell (bei P=95% nicht signifikant) angepaßt wurden. Die modellkonformen Aufgaben wurden nach den Aufgaben-Nummern zu je 16 Auf-

gaben in zwei Teiltests (FRT-Teiltest_A und FRT-Teiltest_B) geteilt.

Die 16 Aufgaben des FRT-Teiltest_A, die nach den eigenen Rohwertgruppen und nach den Rohwertgruppen der 32 modellangepaßten Aufgaben des FRT analysiert worden sind, zeigen zwar unterschiedliche Z-Werte, aber das Signifikanzniveau der Modellkonformität der beiden Analysen bleibt das gleiche (vgl. Tabelle 35, Anhang A). Dasselbe gilt auch für die Analysen des FRT-Teiltest_B. Jedoch sind beim FRT-Teiltest_A in bezug auf die eigenen Rohwertgruppen eine Aufgabe und eine Versuchsperson ausgeschieden, während beim FRT-Teiltest_B zwei Aufgaben und vier Versuchspersonen ausgeschieden sind. Der Trennscore ist bei beiden Teiltests in bezug auf die eigenen Rohwertgruppen vier richtig gelöste Aufgaben und in bezug auf die 32 modellangepaßten Aufgaben des FRT zehn richtig gelöste Aufgaben.

Zusammenfassend läßt sich in bezug auf unsere Fragestellung sagen, daß die Ergebnisse der Analyse im Gegensatz zu den

Außenkriterien der Gruppen von den inneren Teilungskriterien der Gruppen nicht abhängen. Ist ein Test einmal dem Rasch-Modell in bezug auf die inneren Kriterien (Rohwertgruppen) konform, so ändert sich am Ender-gebnis nichts, wenn man die Probanden nach den Rohwerten des Gesamttests oder nach den Rohwerten von Teil-

tests in Gruppen teilt.

7.2. Homogenisierung der Trennschärfe

7.2.1. Problemstellung

Es wurde in einigen Arbeiten (Wendler, 1968, Spada, 1970) darauf hingewiesen, daß das Rasch-Modell zur Homogenisierung der Trennschärfe führt. Dieser Hinweis führt im Rahmen der Frage, warum bestimmte Aufgaben nicht dem Rasch-Modell konform sind, zu folgender Annahme:

Hypothese: Beim Rasch-Modell werden die Aufgaben, deren Trennschärfe über oder unter

dem Variationsbereich liegt, in dem sich die Mehrzahl der Trennschärfen der Aufgaben befindet, eliminiert.

7.2.2. Methode

Die Z-Werte des Figure-Reasoning-Tests (FRT), des Progressive-Matrices-Tests (PMT), des Leistungsprüfsystems (Lps), Aufgabengruppe 9 und 10, Form B, die anhand der erhobenen Daten in Afghanistan in bezug auf die Rohwertgruppen, Ausbildung (Analphabeten und Schüler) und Geschlecht auf Konformität mit dem Rasch-Modell analysiert

worden sind, sollen der Trennschärfe der Aufgaben dieser Tests gegenübergestellt werden. Die Z-Werte und die Trennschärfe sollen hinsichtlich ihres Zusammenhangs miteinander verglichen werden.

7.2.3. Die Ergebnisse

Die Ergebnisse sind in den Tabellen 37, 38, 39 und 40, Anhang A dargestellt.

7.2.4. Diskussion

Will man die Trennschärfen der Aufgaben des Figure-Reasoning-Tests (FRT) in bezug auf die Z-Werte miteinander vergleichen, so läßt sich feststellen, daß in bezug auf die Rohwertgruppen 11 Aufgaben (Aufgaben 6, 8, 13, 14, 16, 17, 19, 24, 27, 29 und 41) einen Z-Wert von größer als zwei haben. Davon liegen die Trennschärfen von sieben Aufgaben (Aufgaben 6, 8, 16, 17, 19, 24 und 27) oberhalb des Variationsbereichs und von drei Aufgaben (Aufgaben 13, 14, 41) unterhalb des Variationsbereichs, in dem sich die Mehrzahl der Trennschärfen der Aufgaben befindet.

Dagegen läßt sich bei den Z-Werten des FRT in bezug auf Ausbildung und Geschlecht eine solche Beziehung zwischen

Z-Werten, die größer als zwei sind, und der Trennschärfe der Aufgaben des FRT nicht feststellen. So haben zum Beispiel die Aufgaben 13 und 14 des FRT in bezug auf die

Rohwertgruppen höhere Z-Werte und kleinere Trennschärfen. Sie sind jedoch in bezug auf Ausbildung und Geschlecht modellkonform (vgl. Tabelle 37, Anhang A).

Beim Progressive-Matrices-Test (PMT) haben auch in bezug auf die Rohwertgruppen 11 Aufgaben (Aufgaben 1, 2, 3, 6, 11, 13, 16, 17, 23, 27 und 29) einen Z-Wert, der höher ist als zwei. Davon haben sieben Aufgaben (Aufgaben 1, 2, 3, 6, 11, 16, 17) eine Trennschärfe über dem Variationsbereich und die übrigen vier Aufgaben (Aufgaben 13, 23, 27, 29) unter dem Variationsbereich, in dem sich die Mehrzahl der Trennschärfen befindet. Es läßt sich jedoch bei den Aufgaben des PMT in bezug auf Ausbildung und Geschlecht eine solche Beziehung zwischen den Z-Werten über zwei und den Trennschärfen nicht erkennen (vgl. Tabelle 38, Anhang A).

Auch beim Leistungsprüfsystem (Lps), Aufgabengruppe 9 und 10, Form B ließ sich feststellen, daß die meisten Aufgaben, die in bezug auf die Rohwertgruppen einen Z-Wert über zwei haben, entweder eine bessere oder schlechtere Trennschärfe aufweisen als die

Mehrzahl der Aufgaben. Aber zwischen den Z-Werten über zwei in bezug auf Ausbildung und Geschlecht und den Trennschärfen der Aufgaben ließ sich eine solche Beziehung nicht erkennen.

– 110 –

Zusammenfassend läßt sich mit diesen Ergebnissen die Hypothese bestätigen, daß beim Rasch-Modell die Aufgaben mit einer Trennschärfe, die über oder unter dem Variationsbereich liegt, in der sich die Mehrzahl der Trennschärfen der Aufgaben befindet, eliminiert werden müssen. D.h. das Rasch-Modell erfordert Aufgaben mit relativ gleichen Trennschärfen. Die Beziehung zwischen den Trennschärfen und der Modellkonformität der Items tritt jedoch nur bei Modellkontrollen nach den Rohwertkriterien und nicht bei anderen Kriterien, wie z.B. Ausbildung und Geschlecht in Erscheinung.

Hier ist der transkribierte Text der Seiten (111–125 sowie Literaturverzeichnis):

8. Zusammenfassende Diskussion

Will man unter einem kulturfairen Intelligenz-Test im Sinne unserer Definition einen Test verstehen, bei dem kulturspezifische Faktoren eliminiert oder neutralisiert sind, so müssen die Forderungen "Testmaterial", "Richtungsunabhängigkeit", "Beantwortungsmethode", "Zeiteinschränkung", "Modellkonformität", "Gütekriterien" und "Verbalisierung" erfüllt werden (s. S. 32f.).

Eine klare Antwort auf unsere Frage, ob bei dem Natur-Formen-Intelligenz-Test (N-I-T) alle diese Forderungen erfüllt sind und er damit als kulturfair betrachtet werden kann, ist nicht leicht.

In unserer Untersuchung wurde gezeigt, daß das Ergebnis der Rasch-Analyse zwar nicht von den inneren Teilungskriterien der Rohwertgruppen abhängt, aber dafür von den Außenkriterien Intelligenz, Ausbildung und Geschlecht. Ein und derselbe Test kann in

bezug auf ein Kriterium mit dem Modell verträglich und in bezug auf andere Kriterien mit dem Modell nicht oder weniger verträglich sein. D.h. die Modellverträglichkeit eines Tests in bezug auf ein Kriterium bedeutet nicht die Modellverträglichkeit des Tests in bezug auf andere Kriterien. Diese Tatsache demonstriert die Relativität des Untersuchungsergebnisses. Man kann also nicht allgemein sagen, daß

ein Test dem Rasch-Modell konform bzw. nicht konform ist, sondern lediglich in bezug auf bestimmte Kriterien, für die er sich als konform oder nicht konform erwiesen hat.

Ein anderes Problem ist die Frage, warum einige Aufgaben eines Tests in bezug auf bestimmte Kriterien mit dem Rasch-Modell unverträglich sind. Diese Frage ist in der Literatur nicht genügend diskutiert, jedoch wird unter anderem auf folgende Gründe hingewiesen:

1. Abweichung von der logistischen Form (monoton steigend) der Item-Charakteristiken

2. Fehlende stochastische Unabhängigkeit
3. Raten der Probanden bei Aufgabenbeantwortung
4. Abweichende Trennschärfe von der Trennschärfe der meisten Aufgaben
5. nicht bedeutungsinvariante Inhalte der Aufgaben.

Es ist nicht in jedem Einzelfall möglich, eindeutig festzustellen, worauf die Modellunverträglichkeit einer Aufgabe zurückzuführen ist. Es wurde jedoch in unserer Untersuchung bei der Gegenüberstellung der Trennschärfe und der Z-Werte der Aufgaben des FRT, PMT, Lps 9 B, Lps 10 B in bezug auf die Rohwertgruppen, Ausbildung und Geschlecht die Beobachtung gemacht, daß die Modellunverträglichkeit einer Aufgabe nur in bezug auf die Rohwertgruppen auf die inhomogene Trennschärfe dieser Aufgabe

zurückzuführen ist, aber nicht in bezug auf die Ausbildung und das Geschlecht. Bei externen Teilungskriterien ist nicht die Trennschärfe des Items bezüglich des Rohwerts, sondern die

Trennschärfe des Items bezüglich des externen Teilungskriteriums (Geschlecht, Ausbildung) maßgeblich.

Ferner sind für eine Rasch-Analyse inhaltlich homogene Stichproben, die sich nur in bezug auf eine Kriteriumsvariable unterscheiden, erforderlich. Eine Analyse eines Tests anhand inhaltlich heterogener Stichproben oder Mehrkriterien-Stichproben liefert bei ein und demselben Test, wie in unserer Untersuchung bei dem Figure-Reasoning-Test (FRT) gezeigt wurde (bei einer Stichprobe dem Rasch-Modell konform, bei einer anderen nicht konform), unterschiedliche Ergebnisse.

Unter einer inhaltlich heterogenen oder Mehrkriterien-Stichprobe soll eine Probandengruppe verstanden werden, die aus mehreren Untergruppen zusammengesetzt ist, die sich in bezug auf mehrere Kriterien, wie z.B. Alter, Geschlecht, Intelligenz, Sozialstatus, Schultyp usw., unterscheiden.

Ferner könnte bei einer Mehrkriterien-Stichprobe durch Eliminierung von Probanden

einer Kriteriumsvariablen dieser Stichprobe festgestellt werden, daß die X^2-Werte

Seite 102

beiden Tests bei der Analyse nach Ausschaltung von Analphabeten anhand der afghanischen und deutschen Schüler in bezug auf die beiden Kriterien (Rohwertgruppen und Nationalität) kleiner geworden sind als die bei der Analyse dieser Tests anhand der Gesamtstichprobe (Afghanen und Deutsche), und nach Ausschaltung der männlichen Probanden wurden sie noch kleiner (vgl. Tabelle XXVII). Die Unterschiede zwischen diesen X^2-Werten sind erheblich groß, so daß es nicht gerechtfertigt ist, wenn man sie auf einen statistischen Effekt (Stichproben-Umfang) zurückführt. Vielmehr liegt die Vermutung nahe, daß, je mehr Kriterienvariablen (Ausbildung, Geschlecht, Alter, Sozialstatus, Schultypen usw.) in einer Stichprobe vorhanden sind, desto größer die X^2-Werte sind und sich somit modellinkonforme Ergebnisse des Tests ergeben, der anhand dieser Stichprobe analysiert wurde.

Aus dem Vergleich der X^2-Werte des N-I-T_B, der anhand der weiblichen afghanischen und weiblichen deutschen Schüler analysiert wurde, mit den X^2-Werten des N-I-T_B, der anhand der männlichen afghanischen und männlichen deutschen Schüler analysiert wurde, ergibt sich, daß die X^2-Werte des N-I-T_B nach den beiden Kriterien (Rohwertgruppen und Nationalität) bei den weiblichen Schülern kleiner sind als bei den männlichen Schülern. Dasselbe gilt auch für die X^2-Werte des FRT (vgl. Tabelle XXVII). Dies läßt sich u.a. dadurch erklären, daß die

erheblich kleiner werden, was nicht nur auf die Verkleinerung des Stichprobenumfangs zurückzuführen ist, sondern auch auf die Ausschaltung der Kriteriumsvariablen, so daß die Vermutung naheliegt, daß, je mehr eine Stichprobe inhaltlich heterogen ist, das Gesamtergebnis desto modellunverträglicher ist. Dies läßt sich vielleicht dadurch erklären, daß beim Rasch-Modell erschöpfende Statistiken verwendet werden. D.h., Probanden mit der gleichen Anzahl gelöster Aufgaben

werden gleiche Parameter zugeordnet. Dies ist für inhaltlich homogene Stichproben gerechtfertigt, aber nicht für inhaltlich heterogene Stichproben.

Die Analyse des N-I-T_B in bezug auf die weiblichen afghanischen und weiblichen deutschen Schüler zeigt, daß die meisten Aufgaben des Tests dem Modell verträglich sind, so daß nach Ausschaltung einiger modellunverträglicher Aufgaben Modellkonformität erreicht werden kann (vgl. Tabelle 28, Anhang A). Auch die Modellkonformität des N-I-T_B in bezug auf die männlichen afghanischen und deutschen Schüler ist nach Ausschaltung einiger modellunverträglicher Aufgaben nicht ausgeschlossen, obwohl diese stärker inhaltlich heterogene Stichproben sind (vgl. Tabelle 29, Anhang A) als die weiblichen afghanischen und deutschen Schüler. Hingegen erscheint die Modellkonformität beim Figure-Reasoning-Test (FRT), der als Vergleich dient, weder in bezug auf die weiblichen

afghanischen und deutschen Schüler noch in bezug auf die männlichen afghanischen und deutschen Schüler durch Ausschaltung selbst einer großen Anzahl von Aufgaben erreichbar zu sein.

Darüberhinaus ist der N-I-T_B in Deutschland sowohl in bezug auf die Rohwertgruppen als auch in bezug auf das Geschlecht dem Rasch-Modell angepaßt (vgl. Tabelle XXII). Auch die Mischform des N-I-T_B, der N-I-T_1, ist nach Eliminierung von 3 Aufgaben in bezug auf die Rohwertgruppen (bei P=5% nicht signifikant) und in bezug auf das Geschlecht (bei P=10% nicht signifikant) dem Rasch-Modell angepaßt (vgl. Tabelle XXIII), während der Figure-Reasoning-Test, der als Vergleich dient, nur in bezug auf das Geschlecht dem Rasch-Modell angepaßt ist, aber nicht in bezug auf die Rohwertgruppen (vgl. Tabelle XXII).

Der N-I-T_B zeigte auch in Afghanistan in bezug auf die Ausbildung (Analphabeten und Schüler), die als wichtiges Kriterium eines kulturfairen Intelligenz-Tests gelten muß, ein besseres Ergebnis als die als Vergleich dienenden klassischen Tests (FRT, PMT, Lps 9

B, Lps 10 B), so daß er nach Ausschaltung zweier Aufgaben dem Rasch-Modell angepaßt werden konnte. Jedoch lieferte der PMT im Vergleich zu den übrigen klassischen Tests in bezug auf die Ausbildung recht gute Ergebnisse. Der

N-I-T_B und der PMT unterscheiden sich von den übrigen Intelligenz-Tests unter anderem im Grad der Abstraktion, was zu der Vermutung führen kann, daß abstrakte Aufgaben für Analphabeten und Schüler bedeutungsinvarianter sind als nicht abstrakte.

Der durchschnittliche Unterschied zwischen deutschen und afghanischen Probanden ist beim N-I-T_B kleiner als beim FRT, der als Vergleich dient (vgl. Tabelle XXIII). Vor allem haben beim N-I-T_B entsprechend der Erwartung die Analphabeten in Afghanistan im Durchschnitt etwas weniger Aufgaben richtig beantwortet als die Schüler, während dies, abgesehen vom Lps 10 B, bei den traditionellen Tests (FRT, Lps 9 B, PMT) umgekehrt der Fall

ist, was gegen die alltägliche Erfahrung und allgemeine Erwartung spricht.

Zusammenfassend läßt sich sagen, daß der N-I-T_B nach der klassischen Testtheorie vor allem in der Schwierigkeit viel adäquater ist und größere innere Konsistenz (Reliabilität) hat als die traditionellen Tests. Bei der inter- und intrakulturellen Untersuchung hinsichtlich der Konformität mit dem Rasch-Modell hat der N-I-T_B viel bessere Ergebnisse gezeigt als die traditionellen Tests, die als Vergleich dienen.

9. Zusammenfassung

Die Frage der Entwicklung eines kulturfairen Intelligenz-Tests, die der Gegenstand dieser Arbeit ist, hat nicht nur wissenschaftlich-theoretische Bedeutung, sondern verdient auch ein großes praktisches Interesse, insbesondere bei den Vielvölkerstaaten, weil die Intelligenz ein wichtiger Faktor bei der Vorhersage von Ausbildung und Berufserfolg ist. Die internationale praktische Bedeutung dieser Frage wird auch von Tag zu Tag größer, da die

Bevölkerungskontakte, vor allem auf dem Gebiet der Erziehung und Bildung, zugenommen haben und weiterhin zunehmen werden.

Mit dem Problem kulturfairer Intelligenz-Test befaßte man sich schon während der Zeit des Ersten Weltkrieges in Amerika bei der Auslese der Wehrpflichtigen wegen deren heterogener Merkmale, wie z.B. nationale Herkunft, Sprache und Schulausbildung. So wurde eine sprachfreie Version, und zwar der “Army-Alpha-Test” unter dem Motto ‘Chancengleichheit für alle’ entwickelt, der unter dem Namen “Army-Beta-Test” bekannt wurde. Seitdem wurde eine große Anzahl sprachfreier Intelligenz-Tests entwickelt. Mit der Zeit verstärkte sich die Meinung, daß sprachfreie Intelligenz-Tests kulturunabhängig seien.

Jedoch nahm auch die Kritik bezüglich der Kulturfreiheit der sprachfreien Intelligenz-Tests zu. Diese Kritik widerlegt nicht nur, daß der sprachfreie Intelligenz-Test

kulturfrei ist, sondern sie liefert auch Ansätze und zeigt folgende Forderungen auf, die an einen kulturfairen Intelligenz-Test gestellt werden müssen.

a) Testmaterial
Mit dem Testmaterial sollen die inter- und intrakulturellen Gruppen (wie z.B. Analphabeten und Ausgebildete, männlich und weiblich usw.) in relativ gleichem Maße vertraut sein.

b) Richtungsunabhängigkeit
Die Beantwortung der Aufgaben soll nicht von einer bestimmten Richtung abhängig sein (z.B. von rechts nach links, von oben nach unten oder umgekehrt).

c) Beantwortungsmethode
Mit der Methode der Aufgabenbeantwortung sollen die inter- und intrakulturellen Gruppen in relativ gleichem Maße vertraut sein, insbesondere die Analphabeten und Ausgebildeten. D.h., daß die Beantwortungsmethode nicht schulgebundenen Charakter haben soll.

d) Zeiteinschränkung
Der Test soll nicht unter strenger Zeiteinschränkung durchgeführt werden.

e) Modellkonformität
Der Test soll für inter- und intrakulturelle Gruppen dem

Seite 119

Rasch-Modell konform sein. Diese Forderung schließt die Forderungen der Bedeutungsinvarianz und der Populationsunabhängigkeit ein.

1. Bedeutungsinvarianz: Der Testinhalt bzw. die logischen Prinzipien der Aufgaben sollen für inter- und intrakulturelle Gruppen (wie z.B. Analphabeten und Ausgebildete, männlich und weiblich usw.) keine unterschiedliche Bedeutung haben.

2. Populationsunabhängigkeit: Der Test soll nicht populationsabhängig sein. D.h. die Parameterschätzung soll nicht systematisch verändert werden, wenn man bei der Schätzung von einer Stichprobe zu anderen übergeht.

f) Gütekriterien
Der Test soll den klassisch-testtheoretischen Anforderungen (Gültigkeit und Zuverlässigkeit) genügen.

g) Verbalisierung
Durch Verbalisieren der Testaufgaben soll bei den Probanden verschiedener Kulturen keine Leistungssteigerung erzielt werden.

Da die traditionellen sprachfreien Intelligenz-Tests (PMT, FRT, CFT) diese Forderungen nicht erfüllen, wurde ein eigener Intelligenz-Test unter dem Namen Naturformen-Intelligenz-Test (N-I-T) konstruiert.

Seite 120

Bei der Materialauswahl für den Naturformen-Intelligenz-Test war der Gegensatz von Kultur und Natur maßgebend. Dabei soll alles das, was durch Menschen geschaffen worden ist, und derjenige Teil der Natur, der durch Menschen verändert worden ist, unter Kultur verstanden werden.

Aus Naturgegenständen wurden diejenigen ausgewählt, die feste Formen haben, wie zum Beispiel Halbmonde, Sterne, Teile des menschlichen Körpers, Hände, Füße usw.

Aus technischen Gründen wurden bei dem Naturformen-Intelligenz-Test nur Halbmonde und Sterne verwendet, mit Ausnahme einer Gedächtnisaufgabe, bei der auch die Formen von Händen und Füßen gezeichnet sind.

Wir sehen, daß beim N-I-T die erste Forderung "Testmaterial" bereits erfüllt ist, weil das Material bzw. die Figuren, mit denen der Test konstruiert ist, jedem Menschen bekannt und universal sind. Mit diesen Figuren sind im Gegensatz zu geometrischen Figuren sowohl Analphabeten als auch ausgebildete Menschen in unterschiedlichen Kulturen relativ gleich gut vertraut.

Auch die Forderungen "Richtungsunabhängigkeit" und "Beantwortungsmethode" sind beim N-I-T bereits erfüllt, da die Aufgaben des N-I-T nicht richtungsgebunden sind und die Methode der Aufgabenbeantwortung keine Schulausbil-

dung voraussetzt. Dabei wird mit einem von verschiedenen Sprachen oder Ziffern unabhängigen "Punktzahlensystem" gearbeitet, von dem man annehmen kann, daß die Angehörigen verschiedener Kulturen in gleichem Maße damit vertraut bzw. nicht vertraut sind.

Die Forderung "Zeiteinschränkung" ist beim N-I-T ebenfalls als erfüllt anzusehen, da ausreichend Zeit zur Beantwortung der Aufgaben zur Verfügung gestellt wird, so daß die meisten Probanden genügend Aufgaben beantworten.

Es bleiben noch die Forderungen "Modellkonformität", "Gütekriterien" und "Verbalisierung". Die Forderung "Verbalisierung" wurde in dieser Arbeit nicht überprüft.

Für die Überprüfung der Forderungen "Modellkonformität" und "Gütekriterien" wurde der N-I-T zusammen mit einer Anzahl

traditioneller Tests (PMT, Lps 9 B, Lps 10 B, FRT, CFT) in Afghanistan durchgeführt. In der Bundesrepublik wurde die Untersuchung auf den Naturformen-Intelligenz-Test-Teiltest B (N-I-T_B) und auf den Figure-Reasoning-Test (FRT) beschränkt, weil das Schulkollegium nur für drei Stunden die Erlaubnis erteilt hatte. Es wurden außerdem zwei Mischformen des N-I-T konstruiert, nämlich N-I-T_1 und N-I-T_2, und durchgeführt.

Bei der Analyse nach der klassischen Testtheorie ergab

sich, daß der Naturformen-Intelligenz-Test (N-I-T) in der Schwierigkeit adäquater ist (vgl. Tabelle 4, 5, 6, 7, 8 und 9, Anhang A) und größere innere Konsistenz (Reliabilität) besitzt (vgl. Tabelle XIII) als die traditionellen Intelligenz-Tests (PMT, Lps 9 B, Lps 10 B, FRT und CFT).

Da das Ergebnis der Rasch-Analyse*) von den gewählten Kriterien, nämlich Ausbildung, Geschlecht, Alter usw., abhängt, kann man

nicht allgemein sagen, ob der N-I-T dem Rasch-Modell konform bzw. nicht konform ist.

So ist er nach dem Ergebnis der Rasch-Analyse in bezug auf die Ausbildung, die als wichtiges Merkmal eines kulturfairen Intelligenz-Tests gelten muß, anhand der afghanischen Schüler und Analphabeten den traditionellen Tests (FRT, PMT, Lps 9 B, Lps 10 B, CFT) überlegen. Es ließ sich nach Ausschaltung einiger Aufgaben Modellkonformität erreichen.

In der Bundesrepublik Deutschland weist der N-I-T_B nach dem Modell von Rasch in bezug auf die Rohwert- und Geschlechtsgruppen bessere Ergebnisse auf als der Figure-

***) Die Itemparameter der Rasch-Analyse, die eine Reihe theoretischer Konsistenzprüfungen ermöglichen könnten, konnten nicht angegeben werden, weil diese Unterlagen durch einen Umzug während meiner Abwesenheit verlorengegangen sind.**

Reasoning-Test (FRT), der als Vergleich diente. Er ist sowohl in bezug auf die Rohwertgruppen als auch in bezug auf das Geschlecht dem Rasch-Modell (bei P=95% nicht signifikant) angepaßt, während der FRT in bezug auf das Geschlecht dem Rasch-Modell konform ist, aber nicht in bezug auf die Rohwertgruppen. Auch bei den Mischformen des N-I-T (N-I-T_1 und N-I-T_2) läßt sich in bezug auf diese beiden Kriterien nach Ausschaltung einiger Aufgaben Modellkonformität erreichen. Also ist dieser Test auch für die deutschen Probanden spezifisch objektiv.

Auch der durchschnittliche Unterschied der Afghanen und Deutschen bei der Aufgabenbeantwortung ist beim N-I-T_B kleiner als beim FRT.

Darüber hinaus haben die Analphabeten beim N-I-T_B im Durchschnitt etwas weniger Aufgaben richtig beantwortet als die Schüler, während bei den meisten traditionellen Tests das Gegenteil der Fall ist. Aber es liegt die Vermutung nah, daß dies auf mangelnde Motivation der Schüler zurückgeführt werden kann. Die Untersuchung wurde in einer Zeit (im

Jahre 1976) in Afghanistan durchgeführt, in der die einzelnen politischen Gruppen bzw. Parteien ihren Einfluß auf die Schulen verbreiteten. Die Schüler wurden ideologisiert. Sie stellten das ganze System einschließlich der Schulleistung in Frage und waren negativ dagegen eingestellt.

Seite 124

Diese negative Einstellung erstreckte sich auch auf die traditionellen Intelligenztests, weil diese mit bestimmten Schulleistungen, nämlich der Geometrie, Ähnlichkeit aufwiesen. Beim N-I-T war dies nicht der Fall, weil er nicht mit einer Schulleistung identisch war.

Tatsächlich müßten bei den traditionellen Intelligenztests im Vergleich mit dem N-I-T die Schüler viel besser abschneiden als die Analphabeten.

Darüber hinaus ist zu erwarten, daß der Übungseffekt beim N-I-T während der Testuntersuchung und bei darauffolgenden Wiederholungen viel geringer sein wird als bei traditionellen Intelligenztests.

Ferner zeigt die Rasch-Analyse des N-I-T_B in bezug auf die Nationalitätsgruppen (Afghanen und Deutsche) gemessen an der X^2-Größe dreimal bessere Ergebnisse als der FRT (vgl. Tabelle XXVII). Außerdem ist die Anzahl der ausgeschiedenen Aufgaben beim FRT so groß, daß die Erwartung auf übrigbleibende modellkonforme Aufgaben, die man als Test bezeichnen könnte, sehr gering ist, während diese Erwartung beim N-I-T_B fortbestehen kann.

Somit ist ausgeschlossen, daß der FRT in bezug auf die Nationalität (Afghanen und Deutsche) dem Rasch-Modell konform wird, während dies beim N-I-T_B zu erwarten ist, insbesondere, wenn die Analyse anhand inhaltlich homogener Stichproben durchgeführt wird.

10. Kurze Zusammenfassung

Die Frage der Entwicklung eines kulturfairen Intelligenz-Tests wurde zum Gegenstand der

Untersuchung gemacht. Dabei wurden aus der Kritik der Kulturabhängigkeit der sprachfreien Intelligenz-Tests Forderungen abgeleitet, die ein Intelligenz-Test erfüllen muß, um als kulturfair bezeichnet werden zu können. Unter einem kulturfairen Intelligenz-Test verstehen wir einen Test, der zwar nicht frei von kulturellen Einflüssen ist, aber die Unterschiede ausgleicht. Es wurde mit universalen, jedem Menschen bekannten Naturformen der Naturformen-Intelligenz-Test (N-I-T) konstruiert, auf den die logischen Prinzipien der traditionellen sprachfreien Intelligenz-Tests übertragen wurden.

Der N-I-T wurde zusammen mit traditionellen sprachfreien Intelligenz-Tests (PMT, FRT, Lps9B, Lps10B, CFT_3), die als Vergleich dienten, in Afghanistan und der Bundesrepublik Deutschland durchgeführt und anschließend nach der klassischen Testtheorie und dem Rasch-Modell analysiert.

Es zeigte sich, daß der N-I-T sowohl nach der klassischen Testtheorie als auch nach der Analyse hinsichtlich der Konformität mit dem

Rasch-Modell den traditionellen Tests überlegen ist.

Er erfüllt die meisten Forderungen, die an einen kulturfairen Intelligenz-Test zu stellen sind. Damit ist er kulturfairer als die traditionellen sprachfreien Intelligenz-Tests.

Seite 1
Anhang C
Handanweisung zum Naturformen-Intelligenz-Test-Form 1 ($N\text{-}I\text{-}T_1$)
1. Einführung

Der Naturformen-Intelligenz-Test-Form 1 ($N\text{-}I\text{-}T_1$) ist ein Verfahren zur Bestimmung der Intelligenz. Beim $N\text{-}I\text{-}T_1$ sind die Forderungen (s. 32, 33 dieser Arbeit), die an einen kulturfairen Intelligenz-Test gestellt werden müssen, zum großen Teil erfüllt. Deshalb ist er kulturfairer als die sprachfreien traditionellen Intelligenz-Tests. Damit ist nicht gesagt, daß er frei von kulturellen Einflüssen ist, wohl aber frei von unterschiedlich starken Einflüssen verschiedener Kulturen. Somit schafft er für Angehörige verschiedener Kulturen identische Voraussetzungen.

Der N-I-T_1 besteht aus 50 Aufgaben, jede Aufgabe hat fünf Antwortmöglichkeiten, davon ist jeweils eine richtig.
Der N-I-T_1 ist besonders für die Gruppen-Untersuchung geeignet. Auch Analphabeten können mit ihm in kleinen Gruppen von fünf bis zehn Personen untersucht werden. Er kann aber auch im Einzelversuch angewandt werden. Da die Vorschriften für den Einzelversuch im Prinzip die gleichen sind wie für die Gruppenuntersuchung, wird in der Anweisung nur der allgemeine Fall der Gruppendurchführung behandelt.

Anhang C
Zur Durchführung des N-I-T_1 benötigt man ein Testheft und einen Antwortbogen.
2. Die Testdurchführung
Der Versuchsleiter erklärt am Anfang mit wenigen Worten den Zweck der Untersuchung. Dabei bemüht er sich, eine Atmosphäre lockerer Aufnahmebereitschaft zu schaffen. Dann verteilt er die Antwortbögen und Testhefte mit der Bemerkung: “Bitte, schreiben Sie nur auf dieses Blatt und nicht in das

Büchlein!". Anschließend fordert der Versuchsleiter die Probanden auf, ihre Personalien und das Datum auf dem Antwortbogen zu notieren (er zeigt dabei mit dem Finger oder Bleistift auf die entsprechenden Stellen). Für die Analphabeten trägt der Versuchsleiter die Personalien selbst in den entsprechenden Rubriken ein, nachdem er die Versuchspersonen danach gefragt hat.
Danach erklärt der Versuchsleiter das Punktzahlensystem mit den Worten: "Sie sehen hier lauter Punkte. Diese haben eine bestimmte Bedeutung, die leicht zu verstehen ist. Ein Punkt zählt eins, zwei Punkte zwei und so weiter bis zu neun Punkten, die selbstverständlich neun bedeuten. Für zehn Punkte haben wir einen Strich gemacht. Das ist übersichtlicher und leichter zu zählen als Punkte. Ein Strich gilt als zehn und zwei Striche sind zwanzig Punkte, drei Striche sind dreißig und so weiter,

fünf Striche sind fünfzig. Ein Punkt und ein Strich bedeuten also 11, zwei Punkte und ein Strich sind 12, neun Punkte und ein Strich sind

19, ein Punkt und zwei Striche bedeuten 21, ein Punkt und drei Striche sind 31 usw.
Nun passen Sie bitte gut auf. An beiden Seiten rechts und links dieses Blattes, das wir Antwortbogen nennen, sehen Sie große Punkte bzw. Zahlen und in der Mitte kleine Punkte bzw. Zahlen. Die großen Punkte bzw. Zahlen sind die Seitenzahl dieses Büchleins. Bitte vergleichen Sie, – stimmt! Auf jeder Seite ist eine Aufgabe und darunter die fünf Antwortringe, darüber sind die kleinen Punkte bzw. Zahlen. Dies ist die Antwortnummer der Aufgabe, die in der Mitte Ihres Antwortbogens abgebildet ist. (Bei der Erklärung weist man im Antwortbogen und Testheft auf die entsprechenden Stellen.)
Haben Sie noch irgendwelche Fragen zu diesem Punktzahlen-System?
Instruktion: Blättern Sie bitte die erste Aufgabe auf. Sie sehen lauter Sterne, die nach bestimmten Regeln angeordnet sind. Sie sollen diese Regeln herausfinden. Wenn Sie in jeder Reihe oder Spalte die Sterne zusammenzählen, ergibt sich 15 Sterne (der Versuchsleiter zeigt mit dem Finger oder Bleistift – entsprechend der Schreibweise der Probanden – zunächst auf die erste Reihe, danach auf die erste Spalte,

dann auf die dritte Reihe und die dritte Spalte). Sie sollen nun herausfinden, wieviel

Seite 4
Anhang C
Sterne in den leeren Ring kommen, damit es in dieser Reihe bzw. Spalte ebenfalls 15 Sterne werden. Die richtige Antwort ist: fünf Sterne. Nun suchen wir bei den Antwortringen den Ring mit fünf Sternen. Dieser ist in der Mitte. Darüber steht zufälligerweise auch die Zahl fünf. Wir haben die Zahl fünf in Ihrem Antwortbogen gestrichen. Das bedeutet, daß der Antwortring mit der Zahl fünf richtig ist. Wenn Sie sich nun in Gedanken den Antwortring mit den fünf Sternen anstelle des leeren Ringes vorstellen, werden Sie feststellen, daß sich von allen Seiten – auch diagonal – 15 Sterne ergeben. Bitte, zählen Sie nach und überzeugen Sie sich. Hat jemand eine Frage?
Wir gehen nun zur zweiten Aufgabe über. Hierbei ist zunächst das Gemeinsame aller Ringe herauszufinden. Die Lage der Halbmonde soll mit berücksichtigt werden. Sie sehen, das Gemeinsame aller Ringe sind zwei Sterne. Deren Plazierung ist dabei unwichtig. In jeder

Reihe oder Spalte ist außerdem je ein Halbmond nach oben und nach unten, sowie ein Ring ohne Halbmonde, nur mit zwei Sternen zu sehen. Betrachtet man das Gesamtfeld, so sind drei Halbmonde nach oben, drei Halbmonde nach unten gezeichnet und zwei Ringe mit zwei Sternen. Dies soll dreimal vorkommen. Die richtige Antwort ist also die eins. Bitte, tragen Sie diese in Ihren Antwortbogen ein. Haben sie noch Fragen?

Seite 5
Anhang C
Blättern Sie nun bitte weiter zur nächsten Übungsaufgabe. Hierbei geht es um die gleiche Anzahl der Sterne. Im Gesamtfeld sind dreimal ein Stern, dreimal drei Sterne, aber nur zweimal vier Sterne. Damit das Muster ergänzt wird, sollen dreimal vier Sterne vorkommen, außerdem soll noch die Lage der Halbmonde berücksichtigt werden. In jeder Reihe oder Spalte gibt es zwei Halbmonde nach unten und einen Halbmond nach oben. Die richtige Antwort ist wieder die eins. Vergessen Sie bitte nicht, die eins in Ihrem Antwortbogen zu streichen. Haben Sie noch Fragen?

Nun zur vierten Aufgabe. Wenn in jeder Reihe oder Spalte die über den Halbmonden stehenden Sterne zusammengezählt und davon die unter den Halbmonden befindlichen Sterne abgezogen werden, bleiben sechs Sterne übrig, also immer die gleiche Anzahl von Sternen. Die richtige Antwort ist die drei. Bitte, streichen Sie diese in Ihrem Antwortbogen und zählen Sie nach, ob es stimmt.
Nun gehen wir weiter zur Aufgabe fünf. Das Gemeinsame aller Ringe sind drei Sterne. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt die vier als richtige Antwort.) Bitte vergessen Sie nicht, diese in Ihrem Antwortbogen zu streichen. Haben Sie noch Fragen?
Nun zur sechsten Aufgabe. Wenn in jeder Reihe bzw.

Anhang C
Spalte sowohl die Sterne über den Halbmonden als auch die Sterne unter den Halbmonden getrennt zusammengezählt und voneinander abgezogen werden, bleibt nichts übrig. Hat jemand die richtige Antwort gefunden? (Der

Versuchsleiter bestätigt die drei als richtige Antwort.) Bitte, streichen Sie diese in Ihrem Antwortbogen. Haben Sie noch Fragen?
Bei der siebten Aufgabe wird ein Halbmond wie ein Stern gezählt, ganz gleich, in welcher Lage er sich befindet, ob er mit oder ohne Stern im Kreis steht. In jeder Reihe bzw. Spalte sind 15 Figuren. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt die vier als richtige Antwort.) Haben Sie noch Fragen?
Und nun zur achten Aufgabe. Sie sehen, alles ist dreimal vorhanden. Auch die Lage der Halbmonde soll mit berücksichtigt werden. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt: "Ja, die zwei ist richtig.")
Haben Sie noch Fragen? Wenn nicht, fangen Sie bitte an. Ich bitte Sie, nicht zu sprechen, nicht laut zu denken oder zu zählen. Sollten Sie aus Versehen etwas falsch anstreichen, machen Sie ein Kreuz daraus. Nach dieser Korrektur streichen Sie die richtige Antwort an. Ich mache noch einmal darauf aufmerksam, daß nur auf dem Antwortbogen und nicht ins Testheft geschrieben werden soll.

Anhang C

Wenn Sie bei einer Aufgabe stecken bleiben, verwenden Sie nicht zuviel Zeit darauf. Lassen Sie die entsprechende Antwort leer und gehen Sie zur nächsten Aufgabe weiter. Sie haben eine Stunde Zeit.

Nach 30 Minuten sagt der Versuchsleiter: Die Hälfte der Zeit ist um. Für die restlichen Aufgaben bleibt lhnen noch eine halbe Stunde.

3. Testauswertung

Jede richtig beantwortete Aufgabe wird mit einem Punkt bewertet. Zur Auszählung der richtig beantworteten Aufgaben am besten eine Schablone benutzen. Im folgenden sind die richtigen Antworten der Übungsaufgaben und der Testaufgaben des N-I-T_1 dargestellt

A

N-I-T_1

Übungsaufgabe	Richtige Antwort
1	5
2	1
3	1
4	3
5	4

6	3
7	4
8	2

Anhang C

|Testaufgabe|Richtige Antwort|Testaufgabe|Richtige Antwort|

|---|---|---|---|
1	2	26	4
2	2	27	3
3	5	28	5
4	1	29	4
5	4	30	1
6	5	31	1
7	4	32	3
8	4	33	1
9	5	34	1
10	5	35	1
11	4	36	2
12	3	37	3
13	4	38	4
14	4	39	3
15	3	40	5
16	5	41	1

|17	|4	|42	|5	|
|18	|5	|43	|1	|
|19	|2	|44	|2	|
|20	|2	|45	|5	|
|21	|5	|46	|4	|
|22	|2	|47	|2	|
|23	|2	|48	|4	|
|24	|2	|49	|3	|
|25	|2	|50	|3	|

Hinweis: Die rechte Spalte (26–50) ist auf dem Foto schräg abgebildet; die Zuordnung der Zahlen bitte am Original gegenprüfen.

Anhang C

Naturformen-Intelligenz-Test-Form 2 ($N\text{-}I\text{-}T_2$)

Der $N\text{-}I\text{-}T_2$ besteht aus 50 Aufgaben und 8 Übungsaufgaben.

Bei diesem Test wird auch wie beim $N\text{-}I\text{-}T_1$ zunächst der Untersuchungszweck und das Punktzahlensystem erklärt und die Personalien notiert. Dann wird folgende Instruktion gegeben.

Instruktion: Zur ersten Übungsaufgabe: Sie sehen innerhalb dieser Ringe Sterne und

Halbmonde, die nach bestimmten Regeln angeordnet sind. Sie sollen diese Regeln herausfinden. Bitte, blättern Sie die erste Aufgabe auf. (Je nach Schreibweise der Probanden – für Analphabeten ist dies gleichgültig – zeigt der Versuchsleiter auf den rechten oder linken Ring neben dem leeren Ring). Er sagt:

Anhang C

“Wenn Sie innerhalb dieses Ringes die Sterne unter den Halbmonden und die darüberstehenden Sterne zusammenzählen, ergibt dies sechs Sterne. Dasselbe gilt auch für die anderen Ringe. Bitte kontrollieren Sie nach, ob es stimmt.” (Der Versuchsleiter wartet solange, bis die Probanden nachgezählt haben.) In jedem Ring sind also sechs Sterne und ein Halbmond. Das gleiche übertragen wir in den leeren Ring. Nun suchen wir in der unteren Reihe aus den fünf Möglichkeiten den Ring, in dem sich sechs Sterne und ein

Halbmond befinden. Die Antwort Nr. 4 ist also richtig. Wir haben deshalb im Antwortbogen die vier angestrichen.
Ich möchte Sie noch darauf hinweisen, daß es immer nur eine richtige Antwort gibt.
Nun zur zweiten Aufgabe: In jedem Ring sehen Sie einen alleinstehenden Stern, dessen Stellung dabei keine Rolle spielt. Es geht um das Gemeinsame aller Ringe und dies ist ein Stern. Die richtige Antwort ist die Nummer 1, also der Ring, in dem sich ein Stern befindet. Bitte, streichen Sie die eins in Ihrem Antwortbogen an. Hat jemand noch eine Frage? Blättern Sie weiter zur dritten Aufgabe. Hierbei kommt es auf die gleiche Anzahl von Sternen im Gesamtfeld an. Gehen wir das doch einmal im einzelnen durch. Im Gesamtfeld sind zweimal zwei Sterne, zweimal ein Stern, aber nud [sic] einmal drei Sterne. Diese drei Sterne sollen jedoch

auch zweimal vorkommen, damit das Muster ergänzt wird. Nun ist es leicht, die richtige

Antwort zu finden. Es ist die Nummer 2. Bitte, vergessen Sie nicht, die richtige Antwort in Ihrem Antwortbogen anzustreichen.
Nun zur vierten Aufgabe: Wenn Sie in jedem Ring von den Sternen, die über den Halbmonden stehen, die darunter befindlichen Sterne abziehen, bleibt immer eins übrig. Bei dieser Aufgabe ist die richtige Antwort die Nummer 4. Hat jemand noch Fragen?
Wir gehen nun zur fünften Aufgabe weiter. Hierbei werden alle Halbmonde wie Sterne gezählt, die Anzahl der Halbmonde in den Ringen soll außerdem berücksichtigt werden. Die richtige Antwort ist die drei und nicht die eins. In jedem Ring sind sechs Figuren. Hat jemand noch Fragen?
Wir gehen nun zur sechsten Aufgabe. Das Gemeinsame in allen Ringen sind zwei Sterne. Mitberücksichtigt wird die Anzahl der Halbmonde in den Ringen. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt: Ja, die drei ist richtig.)

7. Aufgabe. Hierbei wird die Lage der Halbmonde mitberücksichtigt. Alles soll dreimal vorkommen. Hat jemand die richtige

Antwort gefunden? (Der Versuchsleiter bestätigt: Ja, die eins ist richtig.)

Seite 12
Anhang C

8. Aufgabe. In jedem Ring bleiben drei übrig. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt: Ja, die vier ist richtig.) Haben Sie noch Fragen? Wenn dies nicht der Fall ist, fangen sie bitte an. Ich möchte Sie bitten, nicht zu sprechen oder laut zu denken. Sollten Sie irrtümlich etwas falsch anstreichen, so kreuzen Sie es aus. Bitte, beachten Sie, daß nur auf dem Antwortbogen und nicht ins Testheft geschrieben werden soll. Wenn Sie bei einer Aufgabe stecken bleiben, verwenden Sie bitte nicht zuviel Zeit darauf. Lassen Sie die entsprechende Antwort leer und gehen Sie zur nächsten Aufgabe über. Für die Beantwortung der Aufgaben steht Ihnen eine Stunde zur Verfügung.

Nach Ablauf von 30 Minuten sagt der Versuchsleiter: Die Hälfte der Zeit ist um. Sie haben noch eine halbe Stunde vor sich.

Seite 13

Anhang C
Naturformen-Intelligenz-Test-Vorform (N-I-T_V)*

Der N-I-T_V besteht aus acht Untertests, jeder Untertest hat vier Übungsaufgaben. Untertest, 1, 3, 5 und 7 haben je 20 Aufgaben, während Untertest 2, 4, 6 und 8 aus je 10 Aufgaben bestehen.

Bei diesem Test wird auch wie beim N-I-T_1 zunächst der Untersuchungszweck und das Punktzahlensystem erklärt und die Personalien notiert. Dann wird folgende Instruktion gegeben.

Instruktion zu Untertest 1: Bitte blättern Sie die erste Übungsaufgabe auf. Der Versuchsleiter zeigt, je nach der Schreibweise der Probanden, mit dem Finger auf den rechten oder linken Ring neben dem leeren Ring – für Analphabeten ist dies gleichgültig. Er sagt: Wenn Sie innerhalb

*) Aus ökonomischen Gründen wurde beim N-I-T_V nur der letzte Teil (Gedächtnisaufgaben) gedruckt.

Anhang C

dieser Ringe die Sterne, die Sie unterhalb der Halbmonde sehen, mit den Sternen, die über den Halbmonden stehen, zusammenzählen, ergeben sich sechs Sterne. Dasselbe gilt für die Ringe. Bitte zählen Sie nach, ob es stimmt. (Der Versuchsleiter wartet solange, bis die Probanden nachgezählt haben.)

In jedem Ring sind also sechs Sterne und ein Halbmond. Damit das Muster ergänzt wird, tragen wir in den leeren Ring ebenfalls sechs Sterne und einen Halbmond ein.

Nun suchen wir auf dem Blatt unten bei den fünf Antworten sechs Sterne und einen Halbmond. Die Antwort vier ist also richtig. Bitte streichen Sie diese in Ihrem Antwortbogen. (Der Versuchsleiter zeigt die richtige Stelle.) Haben Sie noch Fragen?

Nun zur zweiten Übungsaufgabe, die etwas anders ist. Ein Halbmond wird wie ein Stern gezählt, außerdem wird die Anzahl der Halbmonde mit berücksichtigt. In jedem Ring sind sechs Figuren. Die richtige Antwort ist die drei und nicht die eins, weil auch die Anzahl der Halbmonde mit berücksichtigt wird. Haben Sie noch Fragen?

Nun gehen wir zur Übungsaufgabe drei. Hat schon jemand eine Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: Ja, die zwei ist richtig.)

Seite 15
Anhang C
Nun blättern Sie bitte weiter zur vierten Übungsaufgabe. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: Ja, die eins ist richtig.) Haben Sie noch eine Frage? Wenn nicht, fangen Sie bitte an. Ich bitte Sie, nicht zu sprechen, nicht laut zu zählen und nicht laut zu denken. Sie haben zwanzig Minuten Zeit. Wenn Sie bei einer Aufgabe steckenbleiben, dann verwenden Sie nicht zuviel Zeit darauf. Gehen Sie zur nächsten Aufgabe über und lassen Sie die entsprechenden Antworten frei. (Nach 20 Minuten sagt der Versuchsleiter: "Dankeschön, die Zeit ist um.")
Untertest 2: Wir machen zunächst wieder einige Übungsaufgaben. Bitte blättern Sie die erste Übungsaufgabe auf. Wenn Sie die Sterne in jeder Reihe bzw. Spalte zusammenzählen, ergibt sich 15. (Der Versuchsleiter zeigt mit dem Finger oder Bleistift, entsprechend der

Schreibweise der Probanden, zunächst auf die erste Reihe, dann auf die erste Spalte, danach auf die dritte Reihe und dritte Spalte.) Sie sollen herausfinden, wieviele Sterne in den leeren Ring kommen sollen, damit es in dieser Reihe oder Spalte auch 15 Sterne werden. Die richtige Antwort lautet fünf Sterne. Nun suchen wir bei den Antwortringen den Ring mit fünf Sternen. Dieser ist in der Mitte. Darüber steht zufälligerweise auch die Zahl fünf. Wenn Sie sich in Gedanken den Antwortring fünf mit den fünf Sternen anstelle des leeren Rings vorstellen, dann

ergeben sich von allen Seiten 15 Sterne, auch diagonal. Bitte zählen Sie nach und überzeugen Sie sich. Vergessen Sie bitte nicht, die fünf, also die richtige Antwort, in Ihrem Antwortbogen zu streichen.

Bitte blättern Sie jetzt weiter zur nächsten Aufgabe. Bei dieser Aufgabe ist es etwas anders als vorher. Und zwar wird ein Halbmond wie ein Stern gezählt. Außerdem wird noch die Anzahl und Lage der Halbmonde berücksichtigt. In jeder Reihe oder Spalte ist

ein Halbmond nach oben und ein Halbmond nach unten abgebildet, einmal zwei Halbmonde, insgesamt 15 Figuren. Die richtige Antwort ist die zwei. Bitte, zählen Sie nach. Hat jemand noch eine Frage?
Nun gehen wir zur dritten Übungsaufgabe. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt Nr. 1 als richtige Antwort. Es sind diesmal in jeder Reihe oder Spalte 18 Sterne.)
Blättern Sie nun bitte weiter zur vierten Übungsaufgabe. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: ja, die vier ist richtig.) Hat jemand noch eine Frage? Wenn nicht, fangen Sie bitte an. Sie haben 15 Minuten Zeit. (Nach 15 Minuten sagt der Versuchsleiter: “Dankeschön, die Zeit ist um”.)

Seite 17
Anhang C
Untertest 3: Bitte blättern Sie die erste Übungsaufgabe auf. Diesmal brauchen Sie nicht zu zählen, stattdessen sollen Sie das Gemeinsame aller Ringe herausfinden. In jedem Ring gibt es einen alleinstehenden Stern. Die Stellung dieses einzelnen Sternes ist nicht

wichtig. Das Gemeinsame aller Ringe ist also ein Stern. Die richtige Antwort ist der Ring mit einem Stern, die Nummer eins. Bitte streichen Sie die Eins in Ihrem Antwortbogen durch. Hat jemand noch eine Frage?
Blättern Sie bitte weiter zur zweiten Aufgabe. Das Gemeinsame aller Ringe sind diesmal zwei Sterne. Außerdem wird noch die Anzahl der Halbmonde mit berücksichtigt. Die richtige Antwort ist die Drei. Haben Sie noch Fragen?
Nun zur dritten Übungsaufgabe. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: Ja, die Eins ist richtig.)
Jetzt blättern Sie bitte um zur vierten Übungsaufgabe. Hat schon jemand die Antwort? (Der Versuchsleiter bestätigt: Ja, die Eins ist wieder richtig.) Bitte, vergessen Sie nicht, die Eins in Ihrem Antwortbogen durchzustreichen. Haben Sie noch Fragen? Wenn nicht, können Sie anfangen. Sie haben wieder zwanzig Minuten Zeit. (Nach zwanzig Minuten sagt der Versuchsleiter: “Dankeschön, die Zeit ist um”.)

Untertest 4: Bitte, blättern Sie die erste Übungsaufgabe auf. Dieser Teil der Aufgabe ist genauso, wie bei Teil drei. Das Gemeinsame aller Ringe ist ein Stern. Die Stellung des Sterns ist nicht wichtig. Die richtige Antwort ist die Vier. Bitte, streichen Sie diese in Ihrem Antwortbogen durch. Hat dazu jemand eine Frage?

Nun gehen wir zur zweiten Übungsaufgabe. Das Gemeinsame aller Ringe sind diesmal zwei Sterne, außerdem soll noch die Lage der Halbmonde mit berücksichtigt werden. Sie sehen in jeder Reihe bzw. Spalte einen nach oben und einen nach unten gerichteten Halbmond, sowie einen Ring, in dem nur zwei Sterne abgebildet sind. Wenn man das Gesamtfeld betrachtet, sieht man drei Halbmonde, die nach oben und drei Halbmonde, die nach unten gerichtet sind, und zwei Ringe mit zwei Sternen. Dies soll dreimal vorkommen. Die richtige Antwort ist also die Eins.

Nun blättern Sie bitte weiter zur dritten Übungsaufgabe. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: Ja, es ist die Vier.)

Jetzt gehen wir zur vierten Übungsaufgabe. Hat jemand die richtige Antwort? (Der Versuchsleiter bestätigt: Ja, die Eins ist richtig.) Haben Sie dazu noch eine Frage? Wenn nicht, fangen Sie bitte an. Sie haben 15

Seite 19
Anhang C
Minuten Zeit. (Nach 15 Minuten sagt der Versuchsleiter: “Dankeschön, die Zeit ist um”.)
Untertest 5: Bitte blättern Sie die erste Übungsaufgabe auf. Bei dieser Aufgabe kommt es im Gesamtfeld auf die gleiche Anzahl von Sternen an. So sind im Gesamtfeld zweimal zwei Sterne, zweimal ein Stern, aber nur einmal drei Sterne. Dies soll zweimal vorkommen, damit das Muster ergänzt wird. Die richtige Antwort ist die Zwei. Wenn man sich in Gedanken anstelle des leeren Ringes Ring-Nr. Zwei vorstellt, so ist das Muster ergänzt. Im Gesamtfeld ist dann alles zweimal vertreten, nämlich zweimal eins, zweimal zwei und zweimal drei. Haben Sie noch Fragen?
Blättern Sie bitte weiter zur zweiten Übungsaufgabe. Bei dieser Aufgabe kommt es ebenfalls auf die gleiche Anzahl der Sterne an. Es soll aber noch die Lage der Halbmonde mit

berücksichtigt werden. Es kommt alles dreimal vor. Die richtige Antwort ist die Eins. Hat jemand eine Frage?
Nun zur dritten Übungsaufgabe. Hat jemand die Aufgabe gelöst? (Der Versuchsleiter bestätigt die Nr. Fünf als richtige Antwort.)
Nun gehen wir zur vierten Übungsaufgabe. Hat jemand die richtige Anwort [sic] gefunden? (Der Versuchsleiter bestätigt: Ja, die Fünf ist wieder richtig.) Hat jemand noch

eine Frage dazu? Wenn nicht, können Sie anfangen. Sie haben 20 Minuten Zeit. (Nach 20 Minuten sagt der Versuchsleiter: “Dankeschön, die Zeit ist um”.)
Untertest 6: Bitte blättern Sie die erste Übungsaufgabe auf. Bei dieser Aufgabe kommt es auf die gleiche Anzahl von Sternen an. Im Gesamtfeld soll alles dreimal vorkommen. Die richtige Antwort ist die Nr. 1. Wenn man sich anstelle des leeren Ringes in der Mitte die Eins vorstellt, dann ergeben sich dreimal ein Stern, dreimal zwei Sterne, dreimal drei Sterne und dreimal fünf Sterne. Bitte, überzeugen Sie sich. Hat jemand dazu noch eine Frage?

Nun gehen wir zur zweiten Übungsaufgabe. Es soll wieder alles dreimal vorkommen. Außerdem soll noch die Lage der Halbmonde mit berücksichtigt werden. Die richtige Antwort ist wieder eins. Hat jemand noch eine Frage?
Jetzt blättern Sie bitte weiter zur dritten Übungsaufgabe. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt die Vier als richtige Antwort.)
Jetzt gehen wir zur vierten Übungsaufgabe. Hat jemand die richtige Antwort gefunden? (Der Versuchsleiter bestätigt: Ja, die Zwei ist richtig.) Hat jemand eine Frage? Wenn nicht, fangen Sie bitte an. Sie haben 15

Minuten Zeit. (Wenn die Zeit um ist, sagt der Versuchsleiter: “Dankeschön”.)
Untertest 7: Bitte blättern Sie die erste Übungsaufgabe auf. Bei dieser Aufgabe handelt es sich um eine Abzieh- oder Subtraktionsaufgabe. Wenn Sie von den Sternen, die innerhalb der Ringe über den Halbmonden stehen, die Sterne unterhalb der Halbmonde abziehen, bleibt das Gleiche übrig.

In diesem Falle eins. Die richtige Antwort ist die Vier. Hat jemand dazu eine Frage?
Blättern Sie bitte zur zweiten Übungsaufgabe weiter. Diesmal ist es etwas anders. Diesmal wird die Lage der Halbmonde mit berücksichtigt. Wenn Sie von den Sternen, die sich vor den Halbmonden befinden, die hinter den Halbmonden stehenden Sterne abziehen, bleibt eins übrig. Die richtige Antwort ist die Fünf. Hat jemand dazu eine Frage?
Blättern Sie jetzt bitte zur dritten Aufgabe weiter. Hat jemand eine richtige Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: Ja, die Drei ist richtig.)
Nun zur vierten Übungsaufgabe. Hat jemand eine richtige Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: Ja, die Drei ist richtig.) Gibt es

Seite 22
Anhang C
noch Fragen zu dieser Übungsaufgabe? Wenn nicht, fangen Sie bitte an. Sie haben 20 Minuten Zeit. (Nach 20 Minuten sagt der Versuchsleiter: “Dankeschön, die Zeit ist um”.)
Untertest 8: Bitte blättern Sie die erste Übungsaufgabe auf. Bei dieser Aufgabe ist es

etwas anders. Wenn Sie in jeder Reihe oder Spalte die Sterne über den Halbmonden und die Sterne darunter zusammenzählen und dann voneinander abziehen, bleiben drei Sterne übrig. Die richtige Antwort ist die zwei. Bitte zählen Sie nach, ob es stimmt. Hat jemand dazu eine Frage?

Blättern Sie bitte weiter zur zweiten Übungsaufgabe. Wenn in jeder Reihe oder Spalte die über den Halbmonden stehenden Sterne zusammengezählt und davon die Sterne unter den Halbmonden abgezogen werden, bleiben diesmal sechs Sterne übrig. Die richtige Antwort ist die Drei. Bitte, zählen Sie nach und überzeugen Sie sich.

Jetzt gehen wir zur dritten Übungsaufgabe weiter. Hat jemand eine richtige Antwort gefunden? (Der Versuchsleiter bestätigt die richtige Antwort: Ja, die Drei ist wieder richtig.) Diesmal bleibt nichts übrig.

Blättern Sie nun bitte weiter zur vierten Übungsaufgabe. Hat jemand eine richtige Antwort gefunden? (Der

Versuchsleiter bestätigt die richtige Antwort: Ja, die Fünf ist richtig.) Haben Sie noch Fragen? Wenn nicht, können Sie anfangen. Sie haben 15 Minuten Zeit. (Nach 15 Minuten sagt der Versuchsleiter: “Dankeschön, die Zeit ist um”.)
Nun kommen wir zum letzten Teil des Tests, zur Übungsaufgabe 6. Bitte prägen Sie sich die Figuren gut ein. (Nach einer halben Minute sagt der Versuchsleiter:) Jetzt blättern Sie um. Suchen Sie die Figuren aus, die Sie eingeprägt haben. Hat jemand sie gefunden? (Der Versuchsleiter bestätigt:) Ja, die richtigen Antworten sind die zwei und die drei. Hat jemand noch eine Frage? Bitte blättern Sie um und prägen Sie sich die Figuren, die Sie sehen, ein. (Nach drei Minuten sagt der Versuchsleiter: “Dankeschön.”) Blättern Sie bitte um und suchen Sie die Figuren, die Sie gesehen haben und streichen Sie die Nummern der Figuren in Ihrem Antwortbogen.

Literaturverzeichnis

Amelang, M. und Langer, I. — Zur Kritik der Divergenzhypothese der Intelligenz – in: Archiv für die gesamte Psychologie Bd 120, 1968, S. 203–217

Amthauer, R. — Intelligenz-Struktur-Test. Handanweisung für die Durchführung und Auswertung. Göttingen: Hogrefe 1955, 2.Aufl.

Andersen, E.B. — A goodness of fit test for the Rasch model. Psychometrika 38, 123, 1973

Birnbaum, A. — Statistical theory of tests of mental ability. Ann. Math. Statist. 29, 1284, 1958

Boesch, E.E. und Eckensberger, L.H. — Methodische Probleme des interkulturellen Vergleichs. In: Handbuch der Psychologie, 7.Band, Sozialpsychologie, 1. Halbband, Göttingen: Hogrefe 1969

Burt, C. — The Structure of the Mind: A Review of the Results of Factor Analysis. Brit.J.Educ.Psychol. 19, 1949, S.100-11, 176-199

Cattel, R.B. — Culture Free Intelligence Test, Scale 3, Champaign, Illinois, USA 1961

Dague, P. — Development, application and interpretation of tests for use in French-speaking black Africa and Madagascar. In: Cronbach and drenth, mental tests and cultural adaption. Hague 1972, S.63-74

Daniel, J.C. — Figure Reasoning Test, Handanweisung. Hrsg. Testzentrale Stuttgart o.J.

Daniel, J.C. — Figure-Reasoning-Test (FRT), London: Lickwood 1949

Drenth, P.J.D. — Der psychologische Test, Barth München 1969, S. 16-233

Drenth, P.J.D. and van der Flier, H. — Cultural differences and comparability of test scores. In: International Review of Applied Psychology, Liverpool, University Press 1976, S.137-144

Fischer, G.H. — Einführung in die Theorie psychologischer Tests. Bern: Huber 1974

Fischer, G.H. — Neue Entwicklungen in der psychologischen Testtheorie. In: Fischer, G.H.

(Hrsg.): Psychologische Testtheorie. Bern: Huber 1968

Fischer, G.H., Scheiblechner, H. — Algorithmen und Programme für das probabilistische Testmodell von Rasch. Psychologische Beiträge 12, 1970, S. 23-51

Fischer, G.H. und Spada, H.F. — Die metrische Analyse eines projektiven Formdeutetests. Psychologische Beiträge, Bd 12, 1970

Groffman, K.J. — Die Entwicklung der Intelligenzmessung. In: Heis (Hrsg.): Psychologische Diagnostik, Hdb. der Psychologie, Bd.6, Göttingen 1971, S. 147-199

Guilford, J.P. — Persönlichkeit. Weinheim: Beltz, 1964, S. 353-397

Guilford, J.P. — The nature of human intelligence Mc Graw-Hill 1967, S. 60-68

Merz, F. — Der Einfluß des Verbalisierens auf die Leistung bei Intelligenzaufgaben. 14. Bericht aus dem Institut für Psychologie, Marburg 1968

Merz, F. und Kalveram, K.Th. — Kritik der Differenzierungshypothese der Intelligenz. In: Archiv für die gesamte Psychologie, Bd. 117, 1965, S.287-295

Merz, F. und Stelzl, I. — Einführung in die Erbpsychologie, Stuttgart, Berlin, Köln, Mainz: Kohlhammer 1977, S. 80-83

Merz, F. und Stelzl, I. — Modellvorstellungen über die Entwicklung der Intelligenz in Kindheit und Jugend, Bericht aus dem Institut für Psychologie der Philipps-Universität Nr. 33, November 1972

Rasch, G. — An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19, 49-57

Rasch, G. — On general laws and the meaning of measurement in psychology. Berkeley symposium on mathematical statistics and probability. University of California Press, 1961

Rasch, G. — Probabilistic models for some intelligence and attainment tests. The Danish Institute for Educational Research, Kopenhagen 1960

Raven, J.C. — Advanced Progressive Matrices, Set 1 und 2, London 1971

Raven, J.C. — Guide to the Standard Progressive matrices, London 1960

Raven, C. — Progressive Matrices. H.K.Lewis & Co., London 1938

Scheiblechner, H. — Das Lernen und Lösen komplexer Denkaufgaben. Z.Exp.angew.Psychol. 1972, 19, 476-506

Schwebcke, A./Lück, H.E./Jandron, E. — Probleme und Erfahrungen der Adaption fremdsprachiger Tests. In: Psychologische Beiträge, Bd.15, 1973, S.434-470

Spada, H.F. — Intelligenztheorie und Intelligenzmessung. Psychologische Beiträge, Bd. 12, 1970, S.83-96

Spearman, C. — Abilities of man. New York: Macmillan 1927

Spearman, C. — General Intelligence, objectively Determined and Measured. Amer.J.Psychol., 15, 1904, 201-293

Stapf, K.H. — Praktische Erfahrungen mit dem Meßmodell von Rasch bei der Konstruktion von erziehungsspezifischen Einstellungsskalen. In: Psychologische Beiträge, Bd. 12, 1970, S.105-114

Stelzl, I. — Ist der Modelltest des Rasch-Modells geeignet, Homogenitätshypothese zu prüfen? ein Bericht über Simulationsstudien mit inhomogenen Daten. Z.exp.angew.Psychol., 1979, Band 26, Heft 4, S.652-672

Stelzl, I. — Was bringt das Rasch-Modell für die Praxis? In: Psychologische Beiträge, Bd.14, 1872 [wohl 1972]

Stene, J. — Einführung in Raschs Theorie psychologischer Messung. In: G.H. Fischer (Hrsg.): Psychologische Test-theorie. Bern: Huber 196

Anhang D

بناغلی محترم د کونړ د پوهنې مدیر صیب !

بناغلی عبدالکبیر (لعل) افغاني محصل په المان کښی
دی، غواړي چې خپل علمي تحقیقات په بنوونځیو کښی
سرته ورسوي او د بناغلي محترم پوهنې وزیر صیب څخه
"اجازه اخستې ده
لطفا پدې برخه کښې لازمه مرسته ورسره وکړئ
په درناوي
دثانوي تدریساتو رئیس
(ډوکتور عبدالواحد ملکیار)

Kultusministerium — 30.6.1976
Abteilung für höhere Ausbildung

Sehr geehrter Herr Kultusdirektor der Provinz Kuner!

Herr Abdul Kabir Lal ist afghanischer Student in Deutschland. Er will seine wissenschaftliche Untersuchung in den Schulen durchführen, und er hat vom Kultusminister die Erlaubnis bekommen. Bitte leisten Sie dabei die notwendige Hilfe.

Chef der Abteilung für höhere Ausbildung
Dr. Abdul Wahid Malakyar

(Schulkollegium Düsseldorf)

SCHULKOLLEGIUM
BEIM REGIERUNGSPRÄSIDENTEN
IN DÜSSELDORF

Az:

4000 DÜSSELDORF 1, den 22. Juni 1979
Bastionstraße 39
Postschließfach 7620
Fernsprech-Sa.-Nr. 8 89 11
Durchwahl 8891……..

Herrn
Diplom-Psychologe A. Kabir Lal
Kaygasse 1

5000 Köln 1

Betr.: Empirische Testuntersuchung an Schulen in Köln und Umgebung
Bezug: 1) Runderlaß des KM vom 17.2.77
2) Ihr Antrag vom 21.5.1979

Sehr geehrter Herr Lal!

Gem. Ziff. 2.2 des o. a. Runderlasses des Kultusministeriums sind wir mit der von Ihnen geplanten Testuntersuchung unter folgenden Bedingungen einverstanden:

1. Vor der endgültigen Genehmigung ist uns ein Verzeichnis der Schulen (Gymnasien) vorzulegen, an denen eine Untersuchung durchgeführt werden soll;
2. ebenfalls vorzulegen sind die Angaben, wann die Untersuchungen stattfinden sollen.

Wir weisen schon jetzt darauf hin, daß die Bestimmungen der Ziff. 2.7 des o. a. Erlasses genau einzuhalten sind. Außerdem müssen wir darauf bestehen, daß das Einverständnis des Schulleiters, der betroffenen Lehrer, der volljährigen Schüler bzw. der Erziehungsberechtigten (bei minderjährigen Schülern) vorliegt.

Die Dauer der Untersuchung, die mit drei Stunden angegeben wurde, darf keinesfalls überschritten werden.

Sofern eine Genehmigung nach Eingang der noch ausstehenden Unterlagen ausgesprochen werden kann, wobei allerdings u. U. der Umfang

Anhang D

– 2 –

der Befragung eingeschränkt werden müßte, müssen wir diese mit der Auflage verbinden, uns die Ergebnisse der Untersuchungen schriftlich mitzuteilen.

Mit freundlichen Grüßen
Im Auftrage:
gez. Feldmann

S C H U L K O L L E G I U M B Ü R O B E I M REGIERUNGSPRÄSIDENTEN IN DÜSSELDORF]

Beglaubigt:
Reg.-Angest.

Eidesstattliche Erklärung

Eidesstattliche Erklärung

Ich versichere, daß ich meine Dissertation

Entwicklung eines kulturfairen Intelligenz-Tests

selbständig ohne unerlaubte Hilfe angefertigt habe und daß ich sie in der jetzigen oder ähnlichen Form noch keiner anderen Fakultät eingereicht habe.

4.2.1985

[Stelzl]

Mit teilweiser finanzieller Unterstützung der Vereinten Nationen (United Nations) gedruckt.

*****

{ www.kabirstori.com} [www.pashtoonkhwa.com]