BERICHT
aus dem
PSYCHOLOGISCHEN INSTITUT DER UNIVERSITÄT HEIDELBERG

Zur Faierness psychologischer Intelligenztests

gegenüber Ethnischen und sozialen Gruppen:

Kritik klassischer Konzepte

C l a u s   M ö b u s

und

H e r i b e r t   S i m o n s

Diskussipnspapier Nr. 2

Oktober 1975

6900 HEIDELBERG 1, HAUPTSTR. 47-51, TEL. 06221/547346

Info. I
15

Zur Fairness psychologischer Intelligenztests gegenüber ethnischen und sozialen Gruppen: Kritik klassischer Konzepte

Claus Möbus
und
Heribert Simons

Diskussionspapier Nr.2
Oktober 1975

Zusammenfassung:

Ziel der vorliegenden Arbeit ist es, die (in letzter Zeit häufig aufgegriffene) Frage, wann ein Test gegenüber ethnischen und kulturellen Gruppen f a i r ist, einer Beantwortung näher zu bringen. Dazu werden die bekanntesten und miteinander konkurrierenden Testfairnesskonzepte (CLEARY, THORNDIKE, COLE) in einem einheitlichen Rahmen dargestellt und ihre Widersprüche auf bestimmte Annahmen eines allgemeinen "Fehler-in-den-Variablen & Fehler-in-der-Gleichung" - Modells zurückgeführt. Die sich daraus ergebenden Konsequenzen für die Konstruktion eines fairen Tests werden inhaltlich und methodologisch diskutiert. Dabei ergibt sich die Notwendigkeit eines probabilistischen Fairnesskonzeptes. Dieses Konzept integriert in idealer Weise Reliabilität, Kriteriums- und Konstruktvalidität und Populationsunabhängigkeit aller strukturellen Parameter.

Summary:

The often controversially defined topic of test-bias and test - f a i r n e s s is discussed under a unifying point of view. Definitions of CLEARY, THORNDIKE, COLE and DARLINGTON are treated as special cases of a more general linear structural system: the "errors-in-variables & errors-in-equation" - model which describes the relations between criterion and test. Psychological and methodological consequences for test-construction and fair test-administration are discussed in the last parts of the paper, where a probabilistic concept of testfairness is presented. This concept is a direct consequence of the many difficulties with the classical definitions. It combines in an ideal manner reliability, criterion-, construct validity and independence of the structural parameters of subpopulations.

1. Einleitung

"C.Murphy Archibald, 30, ein Vietnam-Heimkehrer aus Alabama ...hatte sich, wie er meinte, durch gute akademische Leistungen die Aufnahme und ein Stipendium an der Villanova-Universität verdient. Doch als er an der Universität auftauchte, fand sich seine Akte unter "Minoritäten-Bewerber" mit einem B wie BLACK darauf. Als die Verwalter erkannten, daß der Mann aus Alabama nicht schwarz, sondern weiß war, wurde ihm das Stipendium entzogen.

Exstipendiat Archibald verdankt sein Scheitern den guten Absichten der Bürgerrechtsgesetzgebung des Präsidenten JOHNSON: Der Civil Rights Act von 1964 verbietet in Absatz 7 jedwede Benachteiligung amerikanischer Bürger auf Grund von Rasse, Geschlecht, Hautfarbe, Religion oder nationaler Herkunft."

"Bei der Equal Employment Opportunity Commission, die in Washington über die Chancengleichheit amerikanischer Bürger wachen soll, haben sich fast tausend Klagen von weißen Männern angesammelt, die sich über die neue, die "umgekehrte Diskriminierung" beschweren" (Zitate aus: Der Spiegel, 1975, Nr.7, S.93).

Diese Meldungen machen deutlich, daß die Fragen nach Chancengleichheit, Diskrimination von Minoritäten, Fairness psychologischer Aussagen und Untersuchungen in Zukunft wohl noch drängender werden als man bisher gemeinhin angenommen hat.

Unter einem fairen Intelligenztest wurde meist ein Test verstanden, der jedem eine Chance einräume, sich seiner individuellen Lerngeschichte entsprechend intelligent zu verhalten. Folglich sei ein unfairer Test so konstruiert, daß er Gruppen mit unterschiedlichen Lernerfahrungen diskriminiere. Der Test bevorzuge oder benachteilige Gruppen nur, weil die Auswahl seiner items nach einem Intelligenzkonzept erfolge, das nur einer Gruppe gerecht werde. Von unerwünschten Unterschieden im Kriterium (z.B.Schulnoten) schloß man auf die Unerwünschtheit von Gruppenunterschieden im Test ("Identitätskonzept" der Testfairness). J e d e Mittelwertsdifferenz im Test zwischen sozialen Schichten und ethnischen Gruppen wurde auf die diskriminierende Konstruktion des Tests zurückgeführt. Konsequent ging man daran die "Unfairness" der traditionellen Intelligenztests durch die Konstruktion von "culture fair" oder "status free" Tests auszuschalten. Die Nivellierung der Mittelwertsunterschiede gelang jedoch auch hier nicht vollständig. Darüber hinaus scheint es aber auch geradezu unsinnig, zu gleichen Populationsmittelwerte, die die intellektuelle Lerngeschichte der Angehörigen unterschiedlicher Populationen widerspiegeln.

sollten durch den Testkonstrukteur ausgeschaltet werden,

"...da a l l e geistigen Fähigkeiten sich erst in der unter bestimmten äußeren Bedingungen verlaufenden Tätigkeit des Individuums herausbilden (vg. LOMPSCHER, 1970; KEISER, 1969). Auch die sogenannten Handlungs- oder performance-Tests...sind natürlich bildungsabhängig.

... Zweitens gehören die verbal-theoretischen Fähigkeiten zu den wesentlichsten und auch für die Lebensbewährung (vgl. u.a. ORLIK, 1967a) sehr entscheidenden Kriterien der Intelligenz eines Menschen, so daß deren Eliminierung aus intelligenzdiagnostischen Verfahren letztlich zu einer einseitigen, verzerrten Widerspiegelung der intellektuellen Potenz führen würde.

Wir fordern also eine möglichst alle wesentlichen Dimensionen des im Alltag zu beobachtenden intellektuellen Leistungsverhaltens berücksichtigende Leistungsdiagnostik..." (GUTHKE, 1972, S.25f.).

Es steht nicht die Fairness des Tests sondern seines G e b r a u c h s zur Diskussion:"Sind die auf seiner Basis angestellten Schlußfolgerungen fair gegenüber Individuen, Gruppen und Institutionen?" Dient der Test der Prognose und Entscheidungsbildung muß die Beziehung zwischen Test und Kriterium auf ihre Fairness überprüft werden. Es verlagert sich also die Betrachtung von Mittelwertsunterschieden im Test auf die Analyse der Regression zwischen Kriterium und Test: Kann für mehrere Populationen eine gemeinsame Regression akzeptiert werden oder müssen mehrere getrennte Regressionen angenommen werden, die populationsspezifische Aussagen treffen? Dieses Konzept ist umfassender als das der "differentiellen Validität" (SAUNDERS, 1956; KIRKPATRICK, EWEN, BARRETT & KATZEL, 1968). Darüberhinaus sind bei der praktischen Testanwendung noch die "differentielle Prädiktibilität" (BANAS, 1964; GHISELLI, 1956, 1960, 1963; HOBERT & DUNNETTE, 1967) und die "differential utility" (CRONBACH & GLESER, 1965) von Bedeutung. Die differentielle Prädiktibilität tritt bei der Prognose von Kriterien- r o h w e r t e n bei unterschiedlichen gruppenspezifischen Standardschätzfehlern auf. Die differential utility ist mehr ins Kalkül des Testanwenders und nicht so sehr in das des Testkonstrukteurs gerückt, wenn er gleichen Kriteriums- oder Testwerten unterschiedlichen Nutzen oder Bonuspunkte zuweisen will (vgl. DARLINGTON, 1971). Die differentielle Validität wird von uns als Störfaktor angesehen, der durch entsprechende Testverlängerung ausgeschaltet werden muß. Dieses klingt vielleicht überraschend,ist aber leicht einzusehen. Für jede Gruppe ist ein gleich enger Zusammenhang zwischen Test und Kriterium zu fordern: Ist ein Test für eine Gruppe weniger valide, sollte durch Hinzunahme weiterer Variabler mittels optimierter

Linearkombination eine neue "Testvariable" geschaffen werden, die innerhalb jeder Gruppe gleich hoch mit dem Kriterium korreliert (vgl. SIMONS & MÖBUS, 1975).

2. Darstellung der Testfairnesskonzepte von CLEARY, THORNDIKE und COLE

In den letzten Jahren haben sich in der psychologischen Forschung neben dem "Identitätskonzept" drei Definitionen der Testfairness herausgeschält (vgl. DARLINGTON, 1971; LINN, 1974; FLAUGHER, 1975), die sich bis auf die unrealistischen Spezialfälle der perfekten Validität bzw. der vollkommenen kulturellen oder ethnischen Unabhängigkeit des Tests zu widersprechen schienen:

a) die Definition von CLEARY:

"A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup. With this definition of bias, there may be a connotation of "unfair", particularly if the use of the test produces a prediction that is too low." (CLEARY, 1968,S.115)

Zwar ist diese Definition "fair" gegenüber Institutionen (nur Personen mit den höchsten vorhergesagten Kriteriumswerten werden akzeptiert) und relativ "fair" gegenüber Individuen (die Vorhersage individueller Leistungen ist innerhalb einer Gruppe nicht konstant zu hoch oder zu niedrig), jedoch kann sie, wie THORNDIKE (1971) gezeigt hat, unfair gegenüber ganzen Gruppen sein, weil unter bestimmten Umständen bei Gruppen mit schlechteren Testwerten keine Person akzeptiert wird, obwohl etliche Mitglieder dieser Gruppe im Kriterium erfolgreich wären. Aus diesem Grunde machte THORNDIKE einen Alternativvorschlag:

b) die Definition von THORNDIKE:

"An alternate definition (of fairness) would specify that the qualifying scores on a test should be set at levels that will qualify applicants in the two groups in proportion to the fraction of the two groups reaching a specified criterion performance" (THORNDIKE, 1971, S.63).

Nach THORNDIKE soll also in jeder Gruppe der Prozentsatz

der Akzeptierten gleich dem Prozentsatz der im Kriterium Erfolgreichen sein. Nicht ganz so bekannt ist die dritte Definition geworden:

c) die Definition von COLE:

"The selection procedure is considered fair according to COLE's definition, if the probability of being selected, given satisfactory criterion performance, is the same for minority and majority groups"(zit.n.LINN, 1974,S.151f.).

Wie wir im folgenden noch zeigen, kann man diese Definition als direktes Gegenstück zur Version von CLEARY auffassen, während THORNDIKE's Position einen Kompromiß darzustellen scheint.

3. Die Beziehung zwischen Fairnesskonzepten und Kriterium-Test-Regressionen

Die Definition von CLEARY ist in jedem Fall erfüllt, wenn die Regressionskoeffizienten der Regression von Y auf X (=Test) für alle Gruppen nicht signifikant voneinander weichen. Darüber hinaus ist sie aber auch erfüllt, wenn sich zwar die gruppenspezifischen Regressionslinien untereinander unterscheiden aber dennoch nicht signifikant von der g e m e i n s a m e n Regression abweichen. Zu den Regressionskoeffizienten rechnen wir auch die Konstante, die als Gewicht einer Dummyvariablen mit dem konstanten Wert 1 angesehen werden kann. Die Definition ist nicht mehr erfüllt, wenn es zu einem "Strukturbruch" (s.a. SCHNEEWEISS, 1971) gekommen ist. Eine graphische Darstellung dieser Definition findet sich in FIGUR 1.

Hier bitte FIGUR 1 einsetzen

Auch THORNDIKE's Definition läßt eine regressionsorientierte Deutung zu, die bisher aber noch nicht entdeckt worden ist. Die Definition ist in jedem Fall erfüllt, wenn die interdependente bzw. orthogonale Regression zwischen Y und X oder X und Y für alle Gruppen bis auf Zufallsfehler gleich ist. Bei der orthogonalen Regression werden die Fehler nicht parallel zur Y-Achse sondern orthogonal zur Regressionslinie oder Regressionshyperfläche gemessen. Weil diese Form der Regression ziemlich unbekannt zu sein scheint, obwohl sie fast jeder in der Forschung tätige Psychologe schon einmal unbemerkt gerechnet hat, sei sie im Anhang A1. kurz skizziert (vgl.a.FIGUR 2 u. 8).

Hier bitte FIGUR 2 einsetzen

Die Definition von COLE ist - invers zu CLEARY - dann erfüllt, wenn die Regressionskoeffizienten der Regression von X auf Y (Y = Kriterium, das jetzt die Rolle des Regressors spielt) nicht signifikant in den Gruppen voneinander abweichen (vgl.a. FIGUR 3).

Hier bitte FIGUR 3 einsetzen

4. Die verschiedenen Fairnesskonzepte und das korrelative Netz zwischen Test, Kriterium und ethnischer bzw. soziologischer Drittvariabler

Unter einigen einschränkenden Annahmen (Gleichheit von Prädiktor- und Kriteriumsstreuungen in allen Gruppen; keine differentielle Validität des Tests; bivariate Normalverteilungen in allen Gruppen; Polung der Variablen, sodaß alle Korrelationen positiv sind) kann man, wie DARLINGTON (1971) gezeigt hat, eine korrelative Darstellung der verschiedenen Definitionen wählen. Beschränkt man sich auf die Betrachtung der drei Variablen X (=Test), Y (=Leistungsvariable) und Z (=kulturelle oder ethnische Drittvariable, die die mögliche Unfairness von X bedingt) können wir die Korrelationen r_{yx}, r_{xz} und r_{yz} beobachten (vgl. a. FIGUR 4).

Hier bitte FIGUR 4 einsetzen

wobei: r_{yx} = Validität des Tests (sollte möglichst groß sein)
r_{xz} = Maß für die kulturelle oder ethnische Abhängigkeit des Tests (k u r z f r i s t i g durch den Testkonstrukteur variierbar)
r_{yz} = Maß für die kulturelle oder ethnische Abhängigkeit des Kriteriums (nur l a n g f r i s t i g veränderbar)

Unter den schon erwähnten Annahmen können wir die Testfairnesskonzepte in ein korrelationsstatistisches Bezugssystem bringen, in dem r_{yz} und r_{yx} als unabhängiger angesehen werden als r_{xz}. Die Forderung nach hoher Validität kann man logischerweise nicht fallen lassen und r_{yz} muß man als Testkonstrukteur im Moment als nicht veränderbar annehmen. Sollte eine Veränderung wünschenswert sein (vgl.a. die Argumentation von GUTHKE), kann dieses nur langfristig erfolgen. Wir haben also:

$$r_{xz} = f(r_{yz}, r_{yx})$$

Je nach Definition der Testfairness muß der Test einen bestimmten Grad an kultureller oder ethnischer Abhängigkeit besitzen, um als fair zu gelten:

CLEARY's Konzept fordert das Verschwinden der Korrelation zwischen Kriterium und kultureller Drittvariabler nach Ausschaltung des Einflusses der Testvariablen:

(4.1) $r_{xz} = (r_{yz}/r_{yx})$ bzw. $r_{yz.x} = r_{y(z.x)} = 0.0$

Z trägt nicht zu einer verbesserten Schätzung von Y bei, wenn man X kennt oder konstant hält. Ist nämlich CLEARY's Definition erfüllt, sind die Regressionen von Y auf X in jeder Gruppe Z_i $(i=1,\ldots,j,\ldots,G)$ gleich. Damit sind auch die bedingten Erwartungswerte $E(Y_{z_i}|X) = E(Y_{z_j}|X)$ einander gleich. Aus diesem Grund muß $r_{yz.x} = 0.0$ sein. Daraus ergibt sich:

$$r_{yz.x} = \frac{r_{yz} - r_{yx}r_{xz}}{\text{Nenner}} \stackrel{!}{=} 0.0 \quad \text{bzw.} \quad r_{xz} = \frac{r_{yz}}{r_{yx}}$$

THORNDIKE's Konzept fordert, daß kulturelle Abhängigkeit von Kriterium und Test gleich sein sollten:

(4.2) $r_{xz} = r_{yz}$

THORNDIKE's Definition ist bei Variablen mit gleichen gruppenspezifischen Varianzen nur dann erfüllt, wenn die Mittelwertsdifferenz auf X und Y für alle Gruppenpaare gleich ist. Damit gilt (4.2).

Nach der Definition von COLE sollte die Korrelation zwischen Test und kultureller Drittvariabler verschwinden, wenn man Y kennt oder konstant hält:

(4.3) $r_{xz} = r_{yz}r_{yx}$ bzw. $r_{xz.y} = r_{x(z.y)} = 0.0$

Ist nämlich die Regression von X auf Y für jede Gruppe Z_i $(i=1,\dots,j,\dots,G)$ gleich, sind damit auch die bedingten Erwartungswerte $E(X_{z_i}|Y) = E(X_{z_j}|Y)$ einander gleich. Für diesen Fall ist dann[1]:

$$r_{xz.y} = \frac{r_{xz} - r_{yx}r_{yz}}{\text{Nenner}} \stackrel{!}{=} 0.0 \quad \text{bzw.} \quad r_{xz} = r_{yx}r_{yz}$$

Die kulturelle Abhängigkeit eines fairen Tests r_{xz} dürfe nur über den Umweg über das Kriterium zustande kommen. Der Test X unterliege keinem d i r e k t e n Einfluß von Z!

Natürlich lassen sich diese Kausalannahmen nicht mit den Mitteln der Korrelationsrechnung überprüfen, jedoch bilden sie die Basis für die zu fordernden Korrelationsmuster.

Auch das altbekannte Identitätskonzept läßt sich korrelativ darstellen:

$$(4.4) \qquad r_{xz} = 0.0 \quad \text{bzw.} \quad r_{y(x.z)} = r_{yx.z} = r_{yx}$$

Die Ausschaltung der Drittvariablen darf keinen Einfluß auf die Validität des Tests haben, bzw. die Kenntnis der Drittvariablen hilft nicht bei der Vorhersage der Testergebnisse X. Vertritt man ernsthaft das Identitätskonzept dürften damit die meisten gesellschaftlich relevanten Kriterien- und Prädiktorbereiche einer testpsychologischen Untersuchung und Prognose entzogen sein.

Die Zusammenhänge zwischen den Definitionen sind in FIGUR 5 dargestellt. Nur in den beiden unrealistischen Fällen perfekter Validität ($r_{yx} = 1$) und der kulturellen und ethnischen Unabhängigkeit des Kriteriums ($r_{yz} = 0.0$) stimmen die Definitionen überein. Für alle anderen Validitätsbereiche $0 < r_{yx} < 1$ mit der kulturellen Abhängigkeit des Kriteriums $r_{yz} > 0$ widersprechen sich die Konzepte. Die geforderte kulturelle Abhängigkeit des Tests ist bei CLEARY am größten und beim Identitätskonzept ($r_{xz} = 0.0$) am niedrigsten (s.a. FIGUR 5).

Hier bitte FIGUR 5 einsetzen

Auch hier zeigt sich die Mittelstellung der THORNDIKEschen Formulierung. Bildet man nämlich das geometrische Mittel $\bar{r}$ aus den Definitionen bzw. Forderungen von CLEARY und COLE ($\bar{r}^2 = r_{xz(CLEARY)} \cdot r_{xz(COLE)}$), erhält man die Konzeption von THORNDIKE in korrelativer Darstellung ($\bar{r} = r_{xz(THORNDIKE)}$). Für DARLINGTON scheint THORNDIKE's Mittelstellung wohl mehr einen f a u l e n Kompromiß darzustellen, wenn er schreibt:

"Definition 2 represents a compromise position midway between these two (CLEARY's u. COLE's); the value of r_{xz} specified is the geometric mean of these values specified by definitions 1 and 3. As in so many cases, however, a compromise may end up satisfying nobody; psychometricians are not in the habit of agreeing on important definitions or theorems by compromise" (DARLINGTON, 1971, S.76f.).

Wir werden weiter unten zeigen, daß das THORNDIKEsche Konzept doch fundierter ist als es bisher von DARLINGTON, LINN und FLAUGHER gesehen wurde.

5. Die verschiedenen Fairnesskonzepte und Akzeptanz- bzw. Erfolgsraten

Betrachten wir die Verhältnisse im kombinierten Kriteriums-Prädiktorraum, können wir durch entsprechende cut-offs im Prädiktor und Kriterium zwischen Akzeptierten und Nichtakzeptierten bzw. zwischen Erfolgreichen und Nichterfolgreichen trennen. Wir können also den Variablenraum in Quadranten zerlegen. Sind die regressionsorientierten Fairnessdefinitionen erfüllt, decken sich die gruppenspezifischen cut-offs. Damit sind die Testwerte zwischen den Gruppen vergleichbar (vgl.a. FIGUR 6).

Hier bitte FIGUR 6 einsetzen

Oberstes Ziel sollte es sein, die Validität zu erhöhen. Das geschieht durch die Maximierung der Zahl der richtigen Prognosen und Minimierung der Zahl der Fehlprognosen:

$$(5.1) \qquad \begin{aligned} \Sigma\, II_i + \Sigma\, IV_i &= \max! \\ \Sigma\, I_i + \Sigma\, III_i &= \min! \end{aligned}$$

Summe aller sich in Quadranten II und IV befindlichen Personen
Summe aller sich in I und III befindlichen Personen
i = Gruppenindex

Nach CLEARY sollten die cut-offs so gewählt werden, daß der

Prozentsatz der Erfolgreichen unter den Akzeptierten in allen Gruppen gleich ist:

(5.2) $II_i/(II_i + III_i) \stackrel{!}{=} II_j/(II_j + III_j) \quad (i,j = 1,\ldots,G)$

Dahingegen bedeutet Fairness nach THORNDIKE, daß innerhalb jeder Gruppe der Prozentsatz der Erfolgreichen gleich dem Prozentsatz der Akzeptierten ist:

(5.3) $(I_i + II_i) \stackrel{!}{=} (II_i + III_i)$ bzw. $I_i \stackrel{!}{=} III_i$

Nach Maximierung der Trefferquote soll das Verhältnis von Leistungsüber- und unterschätzungen in jeder Gruppe gleich sein.

Nach COLE können wir wieder CLEARY's Definition invertieren: ein Test ist nach COLE fair, wenn der Prozentsatz der Akzeptierten unter den Erfolgreichen in allen Gruppen gleich ist:

(5.4) $II_i/(I_i + II_i) \stackrel{!}{=} II_j/(I_j + II_j) \quad (i,j = 1,\ldots\ldots,G)$

6. Bewertung der verschiedenen Regressionskonzepte

Die faire Anwendung eines Tests setzt also die Gleichheit einer von der jeweiligen Definition abhängigen Regression in allen relevanten Populationen voraus. Die Regressionsparameter müssen populationsunabhängig sein.

Zur Bewertung und Auswahl einer bestimmten Regression und damit eines Fairnesskonzepts müssen wir die jeweiligen Modellannahmen und Fehlertheorien untersuchen. Sind im Regressionsmodell Spezifikations- ("errors in equations"; "shock model") und Meßfehler ("errors in variables") berücksichtigt (vgl.a.GRILICHES, 1974)? Zur Untersuchung dieser beiden Fragen wollen wir uns die Annahmen des klassischen Regressionsmodells, auf der CLEARY's Definition basiert, in Erinnerung rufen (vgl.a. SCHÖNFELD, 1969,53ff):

Für das klassische lineare Regressionsmodell der Matrixgleichung

(6.1) $$y = X\beta + \varepsilon$$

mit y,ε als Zufallsvektoren, X als Beobachtungswertmatrix und β als Regressionsparametervektor sind folgende Annahmen zu treffen:

A1: Die Wahrscheinlichkeitsverteilung des Zufallsvektors ε hat für die gegebene Beobachtungswertmatrix X und alle a priori zulässigen Parameterwerte β folgende Eigenschaften:
A1.1 $\quad E(\varepsilon) = 0$
A1.2 $\quad E(\varepsilon\varepsilon') = \sigma^2 I$

Die Varianz/Kovarianzmatrix der Fehlervariablen ist eine Skalarmatrix.

A2: Die Beobachtungswertmatrix ist exogen determiniert und fest vorgegeben. Sie hat einen Rang, der ihrer Spaltenzahl entspricht (Zeilenzahl = Personenzahl)

A3: Über die festliegenden Werte der Regressionsparameter ist keine andere als die in der Beobachtungswertmatrix (yX) enthaltene Information verfügbar.

A4: Die Größe X ist e x a k t beobachtet und enthält daher k e i n e Meß- oder Beobachtungsfehler.

Für die Praxis ist es in den meisten Fällen unerheblich, ob man y als meßfehlerfrei und daher e als Gleichungs- oder Spezifikationsfehler ansieht ("shock model"), oder ob man y als meßfehlerbelastet (incl. oder excl. Gleichungsfehler) bezeichnet. In j e d e m Fall aber wird X als meßfehlerfrei gefordert. Am meisten wird den Psychologen A4 stören müssen, weil die mangelnde Reliabilität der psychologischen Tests in direktem Widerspruch zu dieser Forderung steht. Ignoriert man die Meßfehlerbelastetheit der Prädiktoren, können die Parameter im klassischen Regressionsmodell nicht mehr konsistent geschätzt werden (JOHNSTON, 1963, 148ff.).

Im folgenden Abschnitt soll der Effekt der Meßfehlerbelastung im bivariaten Fall kurz demonstriert werden. Der deskriptiven Regression

(6.2) $$Y = b_o + b_1 X + e$$

steht die Linearbeziehung zwischen den Konstrukten oder latenten Dimensionen gegenüber:

(6.3) $$\tau_y = \beta_o + \beta_1 \tau_x + \varepsilon_\tau \quad \text{(bei } \varepsilon_\tau = 0\text{: funktional)}$$

Dabei denken wir uns die Meßwerte zusammengesetzt aus latenten "wahren" Anteilen und Meßfehlern, die z.T. auf nicht gelungene Operationalisierungen zurückzuführen sind:

(6.4) $$Y = \tau_y + \varepsilon_y \qquad X = \tau_x + \varepsilon_x$$

Die Hoffnung, die Meßfehlervarianz des Prädiktors ungestraft ignorieren zu können, wird schon im bivariaten Fall zerstört:

(6.5) $$\text{plim } b_1 = \beta_1/(1 + \sigma^2_{\varepsilon_x}/\sigma^2_{\tau_x}) = \beta_1/(1 + (1-\rho_{xx'})/\rho_{xx'})$$
$$= \beta_1 \rho_{xx'}$$ wobei: plim = Wahrscheinlichkeitslimes

Der Wahrscheinlichkeitslimes des Parameters b_1 der deskriptiven Regression (6.2) ist gleich dem Strukturparameter β_1 multipliziert mit der Reliabilität des Prädiktors. Ignoriert man die Meßfehler des Prädiktors, unterschätzt man den Zusammenhang der latenten Dimension, benachteiligt Personen mit besseren Prädiktorwerten und bevorzugt Personen mit schlechteren Prädiktorwerten in jeder Population. Man hat aber vor allem u.a. mit ε_x prognostiziert. Diese Werte aber sind durch schlechte Operationalisierung wesentlich mitbedingt.

Wir wollen jetzt zeigen, daß die drei Fairnesskonzepte sich nur bezüglich ihrer Fehlerannahmen unterscheiden und ineinander überführbar sind. Geht man von einer "wahren" linearen Beziehung zwischen den latenten Variablen τ_x und τ_y aus, kann man je nach Fehlerannahmen verschiedene bivariate Datensätze erzeugen (FIGUR 7). Umgekehrt kann man von e i n e m Datensatz je nach Fehlerannahmen auf verschiedene "wahre" (hier: lineare) Zusammenhänge zwischen τ_x und τ_y schließen.

Hier bitte FIGUR 7 einfügen

Nehmen wir jetzt den stochastischen Zusammenhang (6.3) mit meßfehlerbehafteten manifesten Variablen (6.4) an. Die Störgrößen $\varepsilon_x, \varepsilon_y, \varepsilon_\tau$ seien voneinander unabhängig verteilt mit Varianzen $\sigma^2_{\varepsilon_x}, \sigma^2_{\varepsilon_y}, \sigma^2_{\varepsilon_\tau}$. Der Störterm $(\varepsilon_y + \varepsilon_\tau)$ hat ebenfalls Erwartungswert Null und Varianz $\sigma^2_{\varepsilon_{y\tau}} = (\sigma^2_{\varepsilon_y} + \sigma^2_{\varepsilon_\tau})$. Ferner sei uns noch das Verhältnis λ_1:

(6.6) $\lambda_1 = \sigma^2_{\varepsilon_x} / (\sigma^2_{\varepsilon_y} + \sigma^2_{\varepsilon_\tau}) = \sigma^2_{\varepsilon_x} / \sigma^2_{\varepsilon_{\tau y}}$

bekannt. Setzen wir den wahren Wert für Y aus (6.4) in (6.3) ein, haben wir den Kriteriumswert Y in latente Komponenten zerlegt:

(6.7) $Y = \beta_o + \beta_1 \tau_x + \varepsilon_\tau + \varepsilon_y$

Setzen wir zusätzlich den wahren Wert für X in (6.7) ein:

(6.8) $Y = \beta_o + \beta_1 X + (\varepsilon_\tau + \varepsilon_y - \beta_1 \varepsilon_x)$

Danach bilden wir cov(X,Y) und cov(Y,Y) = var(Y):

(6.9a) $\mathrm{cov}(X,Y) = \beta_1 m_{xx} + \mathrm{cov}(\tau_x + \varepsilon_x, \varepsilon_\tau + \varepsilon_y - \beta_1 \varepsilon_x)$

(6.9b) $\mathrm{cov}(Y,Y) = \beta_1 m_{xy} + \mathrm{cov}(\beta_1 \tau_x + \varepsilon_\tau + \varepsilon_y, \varepsilon_\tau + \varepsilon_y - \beta_1 \varepsilon_x)$

und erhalten zwei Schätzgleichungen mit drei Unbekannten, von denen wir eine wegen (6.6) eliminieren können:

(6.10a) $m_{xy} \hat{=} \beta_1 m_{xx} - \beta_1 \sigma^2_{\varepsilon_x} = \beta_1 m_{xx} - \beta_1 \sigma^2_{\varepsilon_{\tau y}} \lambda_1$

(6.10b) $m_{yy} \hat{=} \beta_1 m_{xy} + \sigma^2_{\varepsilon_{\tau y}} = \beta_1 m_{xy} - 1 \sigma^2_{\varepsilon_{\tau y}} \lambda_2$

mit $\lambda_2 = 1 = (\sigma^2_{\varepsilon_{\tau y}} / \sigma^2_{\varepsilon_{\tau y}})$. Die Gleichungen (6.10) lassen sich kurz fassen:

(6.11) $\left[M - \sigma^2_{\varepsilon_{\tau y}} \cdot \mathbf{\Lambda}_1\right] \left[\hat{\beta}\right] = \left[0\right]$ bzw $\left[M - \mathbf{\Lambda}_2\right] \left[\hat{\beta}\right] = \left[0\right]$

mit: $\mathbf{M}$ = Matrix der zentralen Produkt-Momente (=Varianz-Kovarianzmatrix)

$\hat{\mathbf{\beta}}' = (\hat{\beta}_1 \quad -1)$ = Regressionsparametervektor mit $\hat{\beta}_2 = -1$

$\hat{\beta}_o$ wird geschätzt: $\hat{\beta}_o = \bar{Y} - \hat{\beta}_1 \bar{X}$

$\mathbf{\Lambda}_1 = \mathrm{diag}(\lambda_1, \lambda_2)$
$\mathbf{\Lambda}_2 = \mathrm{diag}(\sigma^2_{\varepsilon_x}, \sigma^2_{\varepsilon_{\tau y}})$
Diagonalmatrizen

$\mathbf{0}$ = Nullvektor

An (6.10) und (6.11) lassen sich jetzt die den drei Fairness-definitionen zugrunde liegenden Regressionskonzepte verdeutlichen:

a) bei $\sigma^2_{\varepsilon_x} = 0$ (Meßfehlerfreiheit des Prädiktors) genügt (6.10a) zur Schätzung von β_1, denn $\lambda_1 = 0$.

(6.12) $\hat{\beta}_1 = \dfrac{m_{xy}}{m_{xx}}$ Der Parameter wird durch die klassische Regression von Y auf X geschätzt

CLEARY's Regressionskonzept beruht auf der Vorstellung eines meßfehlerfreien Prediktors ($\rho_{xx} \hat{=} 1$: perfekte Reliabilität!). Ist dagegen die Meßfehlervarianz von X bekannt und größer als Null, kann man aus $(M - \mathbf{\Lambda}_2) \cdot \beta = 0$ in (6.11) die Schätzung

(6.13) $\hat{\beta}_1 = \dfrac{m_{xy}}{m_{xx} - \sigma^2_{\varepsilon_x}}$ herleiten.

b) bei $\sigma^2_{\varepsilon_{\tau y}} = 0$ (stochastischer Zusammenhang der wahren Werte) oder bei $\sigma^2_{\varepsilon_y} = 0$ (beim funktionalen Zusammenhang) genügt (6.10b) zur Schätzung von β_1 ($\lambda_1 = \infty$).

(6.14) $\hat{\beta}_1 = \dfrac{m_{yy}}{m_{xy}}$ Der Parameter wird durch die klassische Regression von X auf Y geschätzt

COLE's Regressionskonzept beruht auf der Vorstellung eines störtermfreien Kriteriums ($\rho_{yy} \hat{=} 1$: perfekte Reliabilität!)

c) bei $\sigma^2_{\varepsilon_x} = \sigma^2_{\varepsilon_{\tau y}}$ ist $\lambda_1 = 1$ und damit $\mathbf{\Lambda}_1 = I$ (I=Einheitsmatrix) bzw. $\mathbf{\Lambda}_2$ eine Skalarmatrix, deren unbekannter Skalarfaktor sich aus der Matrix, die dann gleich I wird, herausziehen läßt. Normieren wir den Parametervektor $\mathbf{\beta}' = (\beta_1, -1)'$ in (6.11) auf $\mathbf{\beta}'\mathbf{\beta} = 1$ und setzen $\sigma^2_{\varepsilon_{\tau y}} = \mu$, erhalten wir das gleiche Eigenwertproblem wie bei der orthogonalen Regression oder der Hauptkomponentenanalyse (s.a. Anhang A1):

(6.15) $\left[M - \mu \cdot I\right] \left[\hat{\beta}\right] = \left[0\right]$ mit $\mu = \mu_{min}$

Wir können jetzt THORNDIKE's Konzept erklären: Ist der Zusammenhang zwischen τ_y und τ_x funktional und sind die Meßfehlervarianzen $\sigma^2_{\varepsilon_x} = \sigma^2_{\varepsilon_y}$ oder ist der Zusammenhang zwischen τ_y und τ_x stochastisch und sind die Störvarianzen $\sigma^2_{\varepsilon_x} = \sigma^2_{\varepsilon_{\tau y}}$, entspricht THORNDIKE's Regressionskonzept der orthogonalen Regression. Wir haben diese Regression auch interdependente Regression genannt, weil sie Prediktor und Kriterium gleichwertig behandelt: Die Regressionslinie der orthogonalen Regression bleibt von der Wahl des Kriteriums in ihrer Lage im Variablenraum unberührt. Die Regressionslinie ist dabei identisch mit der ersten Hauptkomponente, während die Normale der Regressionslinie der zweiten Hauptkomponente entspricht, Soll wie im klassischen Fall auch hier das Kriterium den Regressionskoeffizienten 1 (vgl. 6.7) bekommen, sind alle Komponenten des zweiten Eigenvektors der Varianz/Kovarianzmatrix durch die Komponentedes Kriteriums zu dividieren. Für den Fall zweier standardisierter Variabler beträgt die so normierte Steigung $\beta_1 = 1$ (also die 45° Gerade). Der Regressionsvergleich beschränkt sich dann - unter der Voraussetzung gleicher populationsspezifischer Validitäten - auf den Vergleich der Konstanten: Sind die populationsspezifischen Regressionen parallel versetzt?

Eine gruppenspezifische Standardisierung der Prädiktor- und Kriteriumsvariablen (z-Werte) würde in jedem Falle zu einem fairen Test im Sinne von THORNDIKE führen, wenn die Störvarianzen in Prädiktor und Kriterium gleich sind(gleiche Validitäten in den Gruppen wäre zusätzlich erstrebenswert). So bräuchte man bei Verwendung von IQ-Werten nur noch die Kriterienwerte für jede Altersgruppe auf z-Werte zu transformieren, um einen nach THORNDIKE fairen Test zu erhalten. Die nachträgliche statistische Ausschaltung gruppenspezifischer Differenzen in den Varianzen und Mittelwerten der Prädiktor- und/oder Kriteriumsvariablen ist unseres Erachtens jedoch unzulässig, wenn gerade die Berücksichtigung der faktischen Ungleichheit der Gruppen entscheidend für eine faire Prognose ist. So führt z.B. die Verwendung des IQ-Maßes, was ja einer Standardisierung des Prädiktors pro Altersgruppe gleichkommt, bei verschiedenen Kriteriumsmittelwerten für die Altersgruppen in j e d e m Fall zu einer u n f a i r e n Testanwendung: Die altersspezifischen Regressionslinien können nicht zur Deckung gebracht werden! Somit führt die Einführung des IQ-Maßes zu einem Effekt, den man durch den IQ gerade abwenden wollte.

Es lassen sich jetzt die Annahmen für die verschiedenen Fairnesskonzepte zusammenfassen (TABELLE 1).

TABELLE 1: Voraussetzungen der drei Fairnesskonzepte

CLEARY	THORNDIKE	COLE
1. g l e i c h e Validität r_{yx} in allen Populationen, d.h.: A u s s c h a l t u n g jeglicher differentieller Validität		
2. $_i\sigma^2_{\varepsilon_x} = {}_j\sigma^2_{\varepsilon_x} = 0$	$\sigma^2_{\varepsilon_x} / \sigma^2_{\varepsilon_{\tau y}} = 1$	$_i\sigma^2_{\varepsilon_{\tau y}} = {}_j\sigma^2_{\varepsilon_{\tau y}} = 0$
3. g l e i c h e Regressionsparameter in allen Populationen (incl.Konstante) ermöglichen gleiche cut-offs auf den Tests		
bei der Regression von Y auf X	bei der orthogonalen bzw. interdependenten Regression	bei der Regression von X auf Y
zusätzlich ist es noch sinnvoll, gleiche Prädiktibilität zu fordern mit: $_i\sigma^2_{\varepsilon_{\tau y}} = {}_j\sigma^2_{\varepsilon_{\tau y}} = \text{const.}$		
Populationen: $i, j = 1, 2, \ldots, G$ mit $i \neq j$		

Wir sind der Ansicht, daß mit Meßfehlern behafteten Variablen nur dann prognostiziert werden darf, wenn die Prognoseverfahren eine entsprechende Fehlertheorie besitzen. CLEARY's und COLE's Fairnessdefinitionen gehen aber von unrealistischen Regressionskonzepten aus. Zumindest sind diese Konzepte in nichtexperimentellen Fragestellungen zweifelhaft. Dahingegen scheint THORNDIKE's Konzept einige interessante Denkanstöße zu geben:

1. Prädiktor und Kriterium sollen gleichberechtigte Indikatoren einer latenten Dimension sein (faktorielle Validität).
2. Die Trennung zwischen endogenen und exogenen Variablen ist weitgehend aufgehoben, zumal Prädiktor und Kriterium auch von der kulturellen Drittvariablen abhängig sein können. Die manifesten Indikatoren der latenten Dimension unterscheiden sich jetzt hauptsächlich nach dem Grad ihrer kurzfristigen Manipulierbarkeit bei der Testkonstruktion. Während die Kriteriumsitems fest ausgewählt sind (zB. "die Anforderungen des Berufs oder des Studiums"), können oder müssen die Testitems solange ausgetauscht werden bis sie die gleiche latente Dimension messen.
3. In diesem Sinne konvergieren Reliabilität und Validität des Tests: Alle items (Kriteriums- wie auch Prädiktor-items) messen die gleiche latente Dimension. Ein Test hätte demnach nicht mehr viele Validitäten und eine Reliabilität sondern nur e i n e Validität und e i n e Reliabilität.

gebauten statistischen Theorie zu fußen, multiple Kriterien zu erlauben, Reliabilität und Validität zu integrieren und darüber hinaus Kriteriums- und Konstruktvalidität nicht mehr zu trennen. Eine ausführlichere Darstellung findet sich bei MÖBUS & SIMONS (1976).

~~Auf der Basis unseres und THORNDIKE's Fairnesskonzept wollen wir zunächst wünschenswerte Eigenschaften eines fairen Tests bzw. Forderungen an einen fairen Test zusammenstellen:~~
~~gebauten statistischen Theorie zu fußen, multiple Kriterien zu erlauben, Reliabilität und Validität zu integrieren und darüber hinaus Kriteriums- und Konstruktvalidität nicht mehr zu trennen. Eine ausführlichere Darstellung findet sich bei MÖBUS & SIMONS (1976).~~

Auf der Basis von THORNDIKE's Fairnesskonzept wollen wir zunächst wünschenswerte Eigenschaften eines fairen Tests bzw. Forderungen an einen fairen Test zusammenstellen:

F1. Zu einem Kriterium wird theoriegeleitet ein Test entwickelt (nicht in umgekehrter Richtung!).

F2. Das Kriterium wird durch mehrere Kriteriumsindikatoren (KI) beschrieben.

F3. KI und Testitems (TI) sind homogen;d.h.messen die gleiche latente Dimension. Sie sind kongenerische Tests(JÖRESKOG, 1974).

F4. Gleiche Testwerte führen in allen relevanten Populationen zu gleichen Vorhersagen; d.h. alle Parameter (Strukturparameter im Sinne von NEYMAN & SCOTT) auf denen Vorhersagen beruhen, müssen in allen Populationen gleich sein: gleiche Testwerte führen zu gleichen latenten Werten (=incidentelle Parameter) und diese dann zu gleichen Vorhersagen.

F5. KI und TI werden im Prädiktionsmodell als meßfehlerbehaftet angesehen.

F6. Es gibt keine unabhängigen Variablen im experimentalpsychologischen Sinne: Prognosen sind richtungsunabhängig.

F7. Die Vorgabe der TI erlaubt die Messung der den KI als auch den TI unterliegenden latenten Dimension und damit die Prognose des Verhaltens in den KI ("Konzept der multiplen Kriterien").

F8. Das gesamte Meßmodell ist empirisch prüfbar.

Diese Forderungen werden von den spezifisch objektiven Meßprozeduren (RASCH, 1960; FISCHER, 1974) erfüllt. Dabei muß man nicht nur die Populations-(bzw. Stichproben)unabhängigkeit der Itemparameterschätzungen sondern auch die der Personenparameterschätzungen ausnutzen. Während die Populationsunabhängigkeit der Itemparameter häufig für die Modelltests des RASCH-Modells herangezogen wird, hat die Stichprobenunabhängigkeit der Personenparameter gegenüber Itemstichproben bisher nur theoretischen Wert gehabt. Zwar wäre die Zulassung multipler Kriterien auch im Rahmen der klassischen Testtheorie möglich (HORST, 1960), jedoch ohne die spezifische Objektivität kontrollieren und gewährleisten zu können. Nur bei einem Test, der mit den KI eine spezifisch objektive Meßprozedur bildet, können die interindividuellen Differenzen so gemessen werden, wie sie bei den KI meßbar wären. Zu berücksichtigen wäre dabei allerdings die unterschiedliche Feinheit der Personenparameterschätzung, die abhängig von der Zahl der items ist. Nur eine große Anzahl von TI und/oder KI läßt eine große Zahl verschiedener Personenparameter zu. Je geringer die Zahl der möglichen verschiedenen Rohwertsummen, desto geringer die Zahl möglicher verschiedener Personenparameter.

Die Forderung nach der Integration von Validität und Reliabilität, wie sie in F1 - F7 zum Ausdruck kommt, ist wegen der starken Reliabilitätsorientierung der Testtheorie nicht in ausreichendem Maße beachtet worden. Weit verbreitet ist die Ansicht, ein Test habe nur e i n e Reliabilität aber m e h r e r e Validitäten. Demgegenüber billigt unser Konzept dem Test nur e i n e Validität zu, weil der Test, mit dem wir prognostizieren wollen z u m Kriterium und nicht unabhängig davon konstruiert ist. Validität und Reliabilität sind i d e n t i s c h e Begriffe geworden. Darüber hinaus lassen sich bei uns auch die Kriteriums- und die Konstruktvalidität nicht mehr voneinander trennen, da a l l e Indikatoren (TI + KI) e i n Konstrukt messen.

So stehen wir hier im Gegensatz zu der landläufigen Ansicht, daß sich die Test- oder Meßtheorie nur mit dem Studium der Reliabilität und Konstruktvalidität im Sinne des Zusammenhangs zwischen latenten und manifesten Dimensionen zu befassen habe, die testtheoretische Behandlung der Kriteriumsvalidität aber erfolglos bleiben müsse bzw. nur vom rein pragmatischen Standpunkt legitim sei. Nach unserer Auffassung ist auch weiter einsichtig, daß die Kriteriumsvalidität eine Konstruktvalidität nicht ausschließt (vice versa), wenn man bedenkt, daß man die alten KI und TI gegen neue substituieren kann, sofern sie die gleiche latente Dimension messen.

Die praktische Konstruktion eines fairen Tests mit multiplen Kriterien hätte dann im Rahmen des probabilistischen Validitäts- und Fairnesskonzepts folgendermaßen abzulaufen (s.a. MÖBUS & SIMONS, 1976):

A. Validierungs- und Konstruktionsphase

A1. Bildung eines Pools von KI und TI. Die TI sollen die gleiche latente Dimension wie die KI messen, jedoch sollten sie leichter applizierbar sein, um die Testkonstruktion zu rechtfertigen.

A2. Bildung homogener Subpools. Subpools, die nur TI enthalten sind unbrauchbar. Subpools, die nur KI enthalten, fordern die weitere Suche nach TI. Für die Testkonstruktion sind nur die aus KI und TI vermischten Pools brauchbar.

A3. Schätzung aller Itemparameter der KI und TI an einer Population, die alle relevanten ethnischen und sozialen Gruppen einschließt. Schätzalgorithmen hierzu gibt FISCHER (1974) an.

A4. Modelltest der spezifisch objektiven Meßprozedur (s.a. FISCHER, 1974, Kap.15). Läßt sich die Modellgültigkeit nicht zurückweisen, bedeutet die Gleichheit der Itemparameter (strukturelle Parameter) in allen relevanten ethnischen und sozialen Gruppen T e s t f a i r n e s s gegenüber diesen Gruppen. Die Strukturparameter sollen bei den spezifisch objektiven Verfahren aber auch noch in allen anderen Personengruppierungen gleich sein, sodaß der Test auch gegenüber diesen Gruppen fair ist, wenn er den Modelltest passiert hat.

A5. Ausschluß aller modellunverträglicher TI und KI. Bleibt kein KI in der Meßprozedur enthalten, hat der Test seine Validität eingebüßt.

B. Prognosephase

B1. Vorgabe der TI an eine zu untersuchende Gruppe von Personen

B2. Schätzung der Personenparameter (vgl.FISCHER,S.239) bei festgehaltenen TI-Itemparametern (gewonnen in A3.). Da der Modelltest die Annahme der Populationsunabhängigkeit (bzw. Stichprobenunabhängigkeit) von Item- und Personenparametern nicht zurückgewiesen hat, müssen die Personenparameterschätzungen auf der Basis der TI annähernd mit denen auf der Basis KI + TI übereinstimmen. Die Übereinstimmung wird dabei durch Stichprobenfehler und die verringerte Anzahl von Rohwertsummen eingeschränkt. Daher sollte die Anzahl der TI gegenüber der Anzahl der KI groß sein.

B3. Die Personenparameter aus B2. werden in die Modellgleichungen bei festgehaltenen KI-Itemparametern (aus A3.) eingesetzt und die theoretisch zu erwartenden Lösungs- und Reaktionswahrscheinlichkeiten für die KI berechnet.

B4. Evaluation der Prognosen an Hand der Lösungs- oder Reaktionswahrscheinlichkeiten, die prognostiziert werden.

Diese Prozedur könnte dazu führen, daß man für jede Gruppe andere TI einbezieht, die trotz ihrer gruppenspezifischen Relevanz im Sinne der ökologischen Validität (PAWLIK 1975) gemeinsam mit den KI dieselbe latente Dimension messen. Das Resultat wäre dann nicht nur die f a i r e Messung einer spezifischen Kompetenz, sondern diese Kompetenz würde auch in einer für die jeweilige Gruppe angemessenen, maßgeschneiderten Weise erfaßt. Diese Vorgehensweise des Testens ist als "tailored testing" (s.a. LORD, F.M.,1974) bekannt geworden. Dabei tritt allerdings das Problem auf, in jedem Einzelfall die Gruppenzugehörigkeit des Individuums und damit die für sie inhaltlich angemessenen und relevanten TI festzulegen.

Die Diskussion der klassischen Fairnesskonzepte und unsere Vorschläge zielen darauf ab, das traditionelle Verständnis intellektueller Prädiktoren aufzugeben, wonach 'intellectual aptitude' (bzw. intellectual abilities) Tests eine zwar nur mittlere, dafür aber generelle Validität besitzen, während 'achievement' Tests eine hohe, aber spezifische Validität aufweisen (LEVINE,1958). Wir sind dagegen mit McCLELLAND der Auffassung, es gäbe keine gesicherten Hinweise dafür, daß die bisher verwendeten Intelligenztests K o m p e t e n z vorhersagten. "Man war bisher so davon überzeugt, mit Intelligenztests die echte Kompetenz zu messen, daß man nicht auf die Idee kam, einmal zu prüfen, ob dem auch wirklich so ist" (McCLELLAND,1973, S. 6).

Die Realisierung der im Verlauf dieser Arbeit begründeten Vorschläge wird unserer Meinung nach zu einer gültigeren und faireren Messung und Vorhersage spezifischer Kompetenzen von Individuen und Gruppen führen, als es durch die bisherige Konstruktion und Interpretation von Intelligenztests erreicht wurde.

Anhang A: Die orthogonale bzw. interdependente Regression

Die in der klassischen Regression verwendete Geradengleichung $x_2 = x_1 \tan\alpha + c$ kann nach einigen Umformungen in die Richtungskosinusgleichung (A.1), die der HESSEschen Normalform entspricht, überführt werden:

(A.1) $x_2 \cos\alpha + x_1 \cos\beta - p = 0$ (vgl.a. FIGUR 8)

wobei: $-p = -c\cos\alpha = b_o$
$\cos\alpha$ = Richtungskosinus (Normale, x_2) = b_1
$\cos\beta$ = Richtungskosinus (Normale, x_1) = b_2

Hier bitte FIGUR 8 einfügen

(A.1) kann verkürzt werden zu:

(A.2) $b_o + x_1 b_1 + x_2 b_2 = 0$ unter der Bedingung $\sum_{j=1} b_j^2 = 1$.

Zur Vorhersage von x_2 oder x_1 dividieren wir (A.2) einfach durch b_2 oder b_1, sodaß der Regressionskoeffizient des betreffenden Kriteriums in (A.2) gleich eins ist:

(A.3) $x_2 = x_1(-b_1/b_2) - b_o/b_2$ bzw. $x_1 = x_2(-b_2/b_1) - b_o/b_1$

Nun werden empirische Meßwerte nicht exakt auf der Regressionsgeraden (A.3) liegen. Bei der orthogonalen Regression wird der senkrechte Abstand u_i des Punktes $P(x_{1i}, x_{2i})$ als Residuum angesehen:

(A.4) $b_o + x_{1i} b_1 + x_{2i} b_2 = u_i$ bzw.: $\mathbf{X}_e \mathbf{b} = \mathbf{u}$

wobei: $\mathbf{X}_e$ eine erweiterte Rohdatenmatrix $\mathbf{X}$ von der Ordnung (Personen, Variable+1) ist: die erste Spalte von $\mathbf{X}_e$ besteht aus Komponenten mit dem Wert 1
$\mathbf{b}' = (b_o, b_1, b_2)$ vollständiger Parametervektor
$\mathbf{u}'$ = Zeilenvektor mit den Residuen

Das Regressionsproblem besteht jetzt in der Bestimmung des Koeffizientenvektors $\mathbf{b}$, sodaß die Fehlerquadratsumme $Q_u = \mathbf{u}'\mathbf{u} = \mathbf{b}'\mathbf{X}_e'\mathbf{X}_e\mathbf{b}$ unter der Nebenbedingung $b_1^2 + b_2^2 = 1$ ein Minimum erreicht:

(A.5) $Q = \mathbf{b}'\mathbf{X}_e'\mathbf{X}_e\mathbf{b} - \lambda(\mathbf{b}_1'\mathbf{b}_1 - 1)$

$= Nb_o^2 + 2b_o\mathbf{e}'\mathbf{X}\mathbf{b}_1 + \mathbf{b}_1'\mathbf{X}'\mathbf{X}\mathbf{b}_1 - \lambda(\mathbf{b}_1'\mathbf{b}_1 - 1) \overset{!}{=} \min$

wobei: N = Zahl der Personen
λ = ein Lagrangescher Multiplikator
$\mathbf{b}_1' = (b_1, b_2)$ = der um b_o reduzierte Parametervektor
$\mathbf{e}' = (1, 1, 1, ..)$ ein Vektor mit Komponenten 1
$\mathbf{X}$ = Rohdatenmatrix

Die partiellen Ableitungen von Q nach $\mathbf{b}$ gleich Null setzend erhält man das Gleichungssystem:

(A.6a) $\frac{\partial Q}{\partial b_o} = 2\tilde{b}_o N + 2\mathbf{e}'\mathbf{X}\tilde{\mathbf{b}}_1 = \tilde{b}_o + \bar{\mathbf{x}}'\tilde{\mathbf{b}}_1 = 0$

(A.6b) $\frac{\partial Q}{\partial \mathbf{b}_1'} = 2\tilde{b}_o\mathbf{e}'\mathbf{X} + 2\tilde{\mathbf{b}}_1'\mathbf{X}'\mathbf{X} - 2\lambda\tilde{\mathbf{b}}_1' = \mathbf{0}'$

wobei: $\bar{\mathbf{x}}' = (\bar{x}_1, \bar{x}_2)$ Mittelwertsvektor

(A.6) läßt sich noch verkürzen zu:

(A.7a) $\tilde{b}_o = -\bar{\mathbf{x}}'\tilde{\mathbf{b}}_1$ (die Regression läuft durchs Centroid; (A.7a) dient der Bestimmung von $\tilde{b}_o$)

(A.7b) $\tilde{b}_o\bar{\mathbf{x}}' + (1/N)\tilde{\mathbf{b}}_1'\mathbf{X}'\mathbf{X} - (\lambda/N)\tilde{\mathbf{b}}_1' = \mathbf{0}'$

und beim Übergang zu Abweichungswerten:

(A.8a) $\tilde{b}_o = 0$ (die Konstante ist Null)

(A.8b) $\left[\mathbf{M} - \frac{\lambda}{N}\cdot\mathbf{I}\right]\left[\tilde{\mathbf{b}}_1\right] = \left[\mathbf{0}\right]$ (einfaches Eigenwertproblem)

(A.8b) hat nur dann eine nichttriviale Lösung, wenn die Koeffizientendeterminante gleich Null ist. Von den verschiedenen Eigenwerten wählen wir den kleinsten, da $\lambda = Q_u$ ist (was wir bei (A.8b) bezogen auf Abweichungswerte zeigen wollen):

$$(\mathbf{M} - \lambda/N\cdot\mathbf{I})\tilde{\mathbf{b}}_1 = ((1/N)\cdot\mathbf{X}'\mathbf{X} - (\lambda/N)\cdot\mathbf{I})\tilde{\mathbf{b}}_1 = \mathbf{0}$$

$$\mathbf{X}'\mathbf{X}\tilde{\mathbf{b}}_1 - \lambda\mathbf{I}\tilde{\mathbf{b}}_1 = \mathbf{0}$$

$$Q_u = \tilde{\mathbf{b}}_1'\mathbf{X}'\mathbf{X}\tilde{\mathbf{b}}_1 = \lambda\tilde{\mathbf{b}}_1'\tilde{\mathbf{b}}_1 = \lambda$$ (wegen der Normierung von $\tilde{\mathbf{b}}_1$)

Setzen wir die Varianz der Residuen $Q_u/N = \lambda/N = \mu$, stimmt das Eigenwertproblem (A.8b) mit (6.15) überein. Die Schätzung der Parameter in der orthogonalen Regression ist identisch mit der Parameterschätzung im "Fehler-in-den-Variablen"-Modell, wenn gleiche Störvarianzen vorliegen.

Die praktische Berechnung der orthogonalen Regression bei zwei auf Mittelwert 0 und Streuung 1 standardisierten Variablen erübrigt sich, weil die Regressionsgerade bei positiver Korrelation der 45° Geraden durch den Ursprung entspricht. Bei Abweichungs- oder Rohwerten müssen wir die Eigenwerte und Eigenvektoren der Varianz-Kovarianzmatrix bestimmen (bei Rohwerten noch $\tilde{b}_o$ nach (A.7a)). Dieses geschieht am einfachsten mit einem Programm zur Hauptkomponentenanalyse. Ein etwas komplizierter Weg wurde von SPÄTH (1974) vorgeschlagen, der auch ein FORTRAN-Programm angibt. Lösen wir (A.8b), erhalten wir die Eigenvektormatrix der Varianz-Kovarianzmatrix **M** (bei Roh- und Abweichungswerten) oder der Korrelationsmatrix **R** (bei standardisierten Werten). Die Spalten sind auf 1 normiert.

	Eigenvektor I des "Faktors" I mit λ_{max}	Eigenvektor II des "Faktors" II mit λ_{min}
Var. x_1	$f_1 = \cos\alpha$	$b_1 = \cos\beta = -f_2$
Var. x_2	$f_2 = \cos(90^\circ - \alpha) = -\cos\beta$	$b_2 = \cos\alpha = f_1$
	Richtungskosinus der Regressionslinie	Richtungskosinus der Normalen (=Koeffizienten der HESSEschen Normalform (A.2))

Im Gegensatz zur Hauptkomponentenanalyse sind wir am kleinsten Eigenwert und dem dazugehörigen Eigenvektor interessiert. Das ist dennoch nicht erstaunlich, denn die orthogonale Regressionslinie minimiert die Störvariation Q_u (korrespondiert zur Varianzaufklärung der 2.Hauptkomponente) und maximiert die Varianzaufklärung der 1.Hauptkomponente (korrespondiert zur Störvarianz der Normalen oder 2.Hauptkomponenten).

Faßt man die beiden Variablen als Teile eines Kompositums auf, daß den latenten trait messen soll, der sowohl Kriterium wie auch Prediktor unterliegt, kann man ein rein intuitiv formuliertes deskriptives Maß für die Güte der Anpassung der Regression überführen in ein Maß für die Reliabilität (interne Konsistenz) dieses Kompositums. Als Anpassungsmaß bietet sich, da es normiert und skaleninvariant ist, L^2 an:

(A.10) $$0 \leq L^2 = 1 - 2\,(\lambda_{min}/(\lambda_{min} + \lambda_{max})) \leq 1$$

was sich bei standardisierten Variablen vereinfacht zu

(A.11) $$0 \leq L^2 = 1 - \lambda_{min} \leq 1$$

Multiplizieren wir aber (A.10) mit $(\lambda_{max} + \lambda_{min})/\lambda_{max}$ und damit (A.11) mit $2/\lambda_{max}$, erhalten wir Maße, die mit der oberen Grenze von CRONBACH's α übereinstimmen:

(A.12) $$0 \leq \alpha \leq \theta^2 = \left[1 - 2\,\frac{\lambda_{min}}{\lambda_{min} + \lambda_{max}}\right] \cdot \left[\frac{\lambda_{max} + \lambda_{min}}{\lambda_{max}}\right] \leq 1$$

und bei standardisierten Daten:

(A.13) $$0 \leq \alpha \leq \theta^2 = \left[1 - \lambda_{min}\right]\left[\frac{2}{\lambda_{max}}\right] \leq 1$$

was der Vergleich mit der von LORD (1958) und BUTLER (1968) angegebenen Originalformel für θ bestätigt (vgl.a.ARMOR, 1974):

(A.14) $$\theta^2 = (p/(p-1))(1 - (1/\lambda_{max})) \qquad p = \text{Anzahl der Var.}$$

Diese obere Grenze wird dann erreicht, wenn alle items des Kompositums optimal gewichtet werden, wie es ja hier geschehen ist. Wir sehen also, daß die Güte der Anpassung und die interne Konsistenz bzw. Reliabilität des Kompositums in monotoner Beziehung zueinander stehen.

Erweitern wir unsere Betrachtung auf p Variable. Das Kompositum besteht aus 1 Kriterium und (p-1) Prediktoren. Dann können wir mit MALINVAUD (1970,S.38) definieren:

> "Die orthogonale Regression der Ordnung m ist die (p-m)-dimensionale Hyperfläche, die durchs Centroid $\bar{x}$ verläuft und orthogonal zu den m letzten Hauptkomponenten ist"

Es wird danach klar, daß wir die orthogonale Regression 1.Ordnung betrachtet haben, die bei zwei Variablen eine 1-dimensionale Fläche (=Linie) bildet, die auf der Normalen senkrecht steht.

Anhang B: Das "Fehler-in-den-Variablen" Modell

Zur Vereinfachung der Betrachtungen seien die Variablen standardisiert (z-Werte). Die Rohwerte lassen sich auch hier entsprechend (6.4) und (C.1) zerlegen:

(B.1) $$\mathbf{Z} = \mathbf{T}_z + \mathbf{E}_z$$

$\mathbf{T}_z$ = Matrix der wahren Werte
$\mathbf{E}_z$ = Matrix der Meßfehler (Ordnung= Personen, Variablen)

Für jede Person seien die Fehler unabhängig vom Personenindex und von den wahren Werten mit dem Nullvektor als Erwartungswertvektor und regulärer Varianz/Kovarianzmatrix $\boldsymbol{\Omega}^2$ verteilt:

$$\boldsymbol{\Omega}^2 = \text{p lim}\,\frac{1}{N}\,\mathbf{E}_z'\mathbf{E}_z \qquad \text{p lim} = \text{Wahrscheinlichkeitslimes}$$

Ferner sei die Beziehung zwischen den wahren Werten fehlerfrei (SCHÖNFELD, 1971, II, S.115ff):

(B.2) $$\mathbf{T}_z\boldsymbol{\beta}_z = \mathbf{0}$$

(keine Fehler in der Gleichung)
$\boldsymbol{\beta}_z$ = (Regressions)Parametervektor

Die Schätzung der strukturellen Parameter $\boldsymbol{\beta}_z$ erfolgt nach der Methode der verallgemeinerten Quadrate, von der die orthogonale Regression ein Spezialfall ist.

Wir minimieren jetzt die gewogene Meßfehlerquadratsumme Q unter der Nebenbedingung (B.2) und unter $\boldsymbol{\beta}'\boldsymbol{\Omega}^2\boldsymbol{\beta} = 1$.

(B.3) $$Q = \text{Spur}((\mathbf{Z}-\mathbf{T}_z)\boldsymbol{\Omega}_z^{-2}(\mathbf{Z}-\mathbf{T}_z)')/n + 2\boldsymbol{\lambda}'\mathbf{T}_z\boldsymbol{\beta}_z/n - \mu(\boldsymbol{\beta}_z'\boldsymbol{\Omega}_z^2\boldsymbol{\beta}_z - 1) = \text{min!}$$

Im ersten Schritt werden wir bei festgehaltener $\boldsymbol{\beta}$ die incidentellen Parameter $\mathbf{T}_z$ eliminieren, um im zweiten Schritt die Regressionsparameter zu schätzen.

(B.4) $$\frac{\partial Q}{\partial \mathbf{T}_z} = -2(\mathbf{Z}-\tilde{\mathbf{T}}_z)\boldsymbol{\Omega}^{-2} + 2\boldsymbol{\lambda}\boldsymbol{\beta}_z' = \mathbf{0} \quad \text{mit } \boldsymbol{\lambda}' = (\lambda_1, \ldots, \lambda_n)'$$

Durch Umstellung von (B.4) und Einsetzung von (B.1) ist:

(B.5) $$\mathbf{E}_z = (\mathbf{Z}-\tilde{\mathbf{T}}_z) = \boldsymbol{\lambda}\boldsymbol{\beta}_z'\boldsymbol{\Omega}^2$$

und mit (B.3):

(B.6) $$(\mathbf{Z}-\tilde{\mathbf{T}}_z)\boldsymbol{\beta}_z = \boldsymbol{\lambda}\boldsymbol{\beta}_z'\boldsymbol{\Omega}^2\boldsymbol{\beta}_z = \boldsymbol{\lambda} \quad \text{bzw.:} \quad \mathbf{Z}\boldsymbol{\beta}_z = \boldsymbol{\lambda}$$

An (B.6) sieht man, daß alle Störungen des Systems auf Meßfehler zurückgeführt wird. Q kann jetzt unabhängig von den $\mathbf{T}_z$ ausgedrückt werden:

(B.7) $$Q^0 = \text{Spur}(\boldsymbol{\lambda}\boldsymbol{\beta}_z'\boldsymbol{\Omega}^2\boldsymbol{\Omega}^{-2}\boldsymbol{\Omega}^2\boldsymbol{\beta}_z\boldsymbol{\lambda}')/n - \mu(\boldsymbol{\beta}_z'\boldsymbol{\Omega}^2\boldsymbol{\beta}_z - 1)$$
$$= \text{Spur}(\boldsymbol{\lambda}\boldsymbol{\lambda}') - \mu(\boldsymbol{\beta}_z'\boldsymbol{\Omega}^2\boldsymbol{\beta}_z - 1) = \boldsymbol{\beta}_z'\,\mathbf{Z}'\mathbf{Z}\boldsymbol{\beta}_z/n - \mu(\ldots)$$
$$= \boldsymbol{\beta}_z'\,\mathbf{R}\,\boldsymbol{\beta}_z - \mu(\boldsymbol{\beta}_z'\boldsymbol{\Omega}_z^2\,\boldsymbol{\beta}_z - 1)$$

(B.8) $$\frac{\partial Q^0}{\partial \boldsymbol{\beta}'} = 2\,\mathbf{R}\,\tilde{\boldsymbol{\beta}}_z - 2\,\mu\,\boldsymbol{\Omega}^2\tilde{\boldsymbol{\beta}}_z = \mathbf{0}$$

(B.9) $$(\mathbf{R} - \mu\boldsymbol{\Omega}^2)\tilde{\boldsymbol{\beta}}_z = \mathbf{0} \qquad \text{mit } \mu = \tilde{\boldsymbol{\beta}}_z'\mathbf{R}\tilde{\boldsymbol{\beta}}_z$$

Das allgemeine Eigenwertproblem (B.9) läßt sich, wie es auch bei C.6 geschieht, in ein einfaches transformieren. Sind alle Meßfehlervarianzen gleich, geht das "Fehler-in-den-Variablen"-Modell in die orthogonale Regression über.

Anhang C: Das kombinierte "Fehler-in-den-Variablen" & "Fehler-in-der-Gleichung" Modell

Wir wollen die Parameterschätzung zuerst für den einfachsten Fall der standardisierten Werte angeben. Die z-Werte lassen sich analog zu (6.4) zerlegen:

(C.1) $$\mathbf{Z} = \mathbf{T}_z + \mathbf{E}_z$$
$\mathbf{Z}$ = Z-Wertmatrix (N x p)
$\mathbf{T}_z$ = Matrix der latenten Werte
$\mathbf{E}_z$ = Matrix der Meßfehler

Die Fehler sollen wieder unabhängig verteilt sein mit Erwartungswert Null und Varianz/Kovarianzmatrix $\boldsymbol{\Omega}^2 = \text{diag}(\sigma^2_{\varepsilon_x})$. Auf Grund dieser Annahmen läßt sich die Korrelationsmatrix zerlegen bzw. reduzieren:

(C.2) $$\mathbf{R} = \mathbf{Z}'\mathbf{Z}/n = (\mathbf{T}_z'\mathbf{T}_z + \mathbf{E}_z'\mathbf{E}_z) = \mathbf{R}^0 + \boldsymbol{\Omega}^2$$

Unsere Regressionsgleichung zwischen den latenten Variablen hat den Charakter einer sogenannten Strukturgleichung (Van de GEER, 1971):

(C.3) $$\mathbf{T}_z \cdot \boldsymbol{\beta}_z = \mathbf{v}_z$$
$\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$
$\mathbf{v}_z$ = Vektor mit den stochastischen Abweichungen

Wir minimieren $Q_v = \mathbf{v}'\mathbf{v}/n$ unter einer Nebenbedingung, die in ähnlicher Weise im Diskrimininanzanalysenkriterium auftaucht. wenn man gewillt ist, $\boldsymbol{\Omega}^2$ mit der gepoolten Binnenvarianzmatrix $\mathbf{W}$ zu vergleichen. Man müßte $\mathbf{W}$ (im Gegensatz zur Zwischenvarianzmatrix) als Meßfehlervarianzmatrix ansehen, die die Gruppenzugehörigkeiten der Personen verschleiert (vgl. Van de GEER, S. 272, 234).

(C.4) $$Q_v = \mathbf{v}_z'\mathbf{v}_z/n = \boldsymbol{\beta}_z'\,\mathbf{T}_z'\,\mathbf{T}_z\,\boldsymbol{\beta}_z - \lambda\,\boldsymbol{\beta}'\,\boldsymbol{\Omega}_z^2\,\boldsymbol{\beta}_z \stackrel{!}{=} \min$$

Partiell ableiten und die Ableitungen gleich Null setzen; so kommen wir zum allgemeinen Eigenwertproblem:

(C.5) $$\frac{\partial Q_v}{\partial \boldsymbol{\beta}_z'} = (\mathbf{R}^0 - \lambda\,\boldsymbol{\Omega}_z^2)\,\tilde{\boldsymbol{\beta}}_z = \mathbf{0}$$

Das allgemeine Eigenwertproblem kann mit der Hilfsvariablen $\tilde{\mathbf{q}}_z = \boldsymbol{\Omega}_z\tilde{\boldsymbol{\beta}}_z$ in ein symmetrisches umgewandelt werden:

(C.6) $$(\boldsymbol{\Omega}_z^{-1}\mathbf{R}^0\,\boldsymbol{\Omega}_z^{-1} - \lambda \cdot \mathbf{I})\,\tilde{\mathbf{q}}_z = \mathbf{0} \quad \text{mit } \boldsymbol{\Omega}^{-1} = \text{diag}(1/\sigma^2_{\varepsilon_{x_j}})$$

Bei g l e i c h e n Meßfehlervarianzen $\boldsymbol{\Omega}^2 = \sigma^2\mathbf{I}$ vereinfacht sich (C.5) zu (C.7):

(C.7) $$(\mathbf{R}^0 - \mu\,\mathbf{I})\,\tilde{\boldsymbol{\beta}}_z = \mathbf{0} \quad \text{mit } \mu = \lambda\,\sigma^2$$

und bei Abweichungswerten ersetzen wir sinngemäß $\mathbf{R}$ durch die reduzierte Varianz/Kovarinazmatrix $\mathbf{M}^0$. Bei gleichen Meßfehlervarianzen vereinfacht sich das "Fehler-in-den-Variablen-und-in-der-Gleichung" Modell auf die Berechnung des kleinsten Eigenvektors der reduzierten Korrelations- oder Varianz/Kovarianzmatrix.

(C.6) und (C.7) laufen jeweils auf die Bestimmung des zum kleinsten Eigenwert gehörenden Eigenvektors im Rahmen der der kanonischen Faktorenanalyse (C.6) und der normalen Faktorenanalyse (C.7) hinaus.

L i t e r a t u r

ARMOR, D.J., Theta reliability and factor scaling. In: H. L. COSTNER (Hg.), Sociological methodology. San Francisco: Jossey-Bass, 1974, 17 - 5o.

BANAS, P.M., An investigation of transsituational moderators. Univ. of Minnesota. Unveröff. Diss. 1964.

BUTLER, P.M., Alpha-maximized factor analysis and its relation to alpha and canonical factor analysis. Psychometrika, 1968, 33, 335 - 346.

CLEARY, T.A., Test bias: Prediction of grades of negro and white students in integrated colleges. J. Educ. Measmt., 1968, 5, 114 - 124

COHEN, N.S., Bias in selection. ACT Research Report No. 51. Iowa: American College Testing Program, 1972.

CRONBACH, L.J. u. GLESER, G.C., Psychological tests and personnel decisions. Urbana: Univ. of Illinois Press, 1965².

DARLINGTON, R.B., Another look at 'cultural fairness'. J. Educ. Measmt., 1971, 8, 71 - 82.

FISCHER, G., Einführung in die Theorie psychologischer Tests. Bern: Huber, 1974.

FLAUGHER, R.L., The new definitions of test fairness in selection: Developments and implications. Educ. Researcher, 1934, 3, 13 - 16.

GNANADESIKAN, R., Development lectures on multivariate Analysis. NUFFIC Int. Summer Session, 197o.

GHISELLI, E.E., Moderating effects and differential reliability and validity. J. Appl. Psychol., 1963, XLVII, 81 - 86.

GUTHKE, J., Zur Diagnostik der intellektuellen Lernfähigkeit. Berlin: VEB Dt. Verlag d. Wissensch., 1972.

HOBERT, u. DUNNETTE, , Development of moderator variables to entrance the prediction of managerial effectiveness. J. Appl. Psychol., 1967, 51, 5o - 64.

HORST, P., Optimal estimates of multiple criteria with restrictions on the covariance matrix of estimated criteria, Psychological Reports, Monograph Supplement, 6-V6, 427-444, 1960

JÖRESKOG, K. G., Analyzing psychological data by structural analysis of covariance matrices. In: KRANTZ, ATKINSON, LUCE & SUPPES (Hg.), Contemporary developments in Mathematical Psychology. San Francisco, 1974, Vol. II, 1 - 56

JOHNSTON, J., Econometric methods. New York: Wiley, 1963.

KEISER, W., Psychologische Grundprobleme der Entwicklung geistiger Fähigkeiten. Wiss. Z. Karl-Marx-Univ. Leipzig, Ges. u. Sprachwiss. Reihe 1, 197o.

KIRKPATRICK, J.J., EWEN, R.B. BARRETT, R.S. u. KATZEL, R.A., Testing and fair employment. New York: New York Univ. Press, 1968.

LEVINE, S., Aptitude versus achievement tests as predictors of achievement. Educ. Psychol. Measmt., 1958, 18, 517 - 525

LINN, R.L., Fair test use in selection. Rev. Educ. Res., 1973, 43, 139 - 161

LORD, F.M., Some relations between GUTTMAN's principal components of scale analysis and other psychometric theory. Psychometrika, 1958, 23, 291 - 296.

LORD, F.M., Individualized testing and item characteristic curve theory, in: KRANTZ, D.H., ATKINSON, R.C., LUCE, R.D. & SUPPES, P. (eds), Contemporary developments in mathematical psychology, Vol.II, 1974, lo6 - 12o

MALINVAUD, E., Statistical methods of econometrics. Amsterdam: North-Holland, 197o.

Mc CLELLAND, D.C., Testing for competence rather than for 'intelligence'. Amer. Psychologist, 1973, 28, 1 - 14.

MÖBUS, C. u. SIMONS, H., Untersuchungen zu einem probabilistischen Fairnessmodell. (in Vorbereitung)

ORLIK, P., Kritische Untersuchungen zur Begabtenförderung. Meisenheim/Glan: Hain, 1967.

RASCH, G., Probabilistic models for some intelligence and attainment tests. Kopenhagen: The Danish Institute for Educational Research, 196o.

SAUNDERS, D.R., Moderator variables in prediction. Educ, Psychol. Measmt., 1956, 16, 2o9 - 222.

SCHNEEWEISS, H., Ökonometrie. Würzburg: PhysienVVerlag 1971.

SCHÖNFELD, P., Methoden der Ökonometrie. Bd. I/II. Berlin: Franz Vahlen, 1969.

SIMONS, H. u. MÖBUS, C., Untersuchungen zur Fairness von Intelligenztests. Z. Entwicklungspsychol. Päd. Psychol. (im Druck).

SPÄTH, H., Algorithmen für multivariable Ausgleichsmodelle. München: Oldenbourg, 1974.

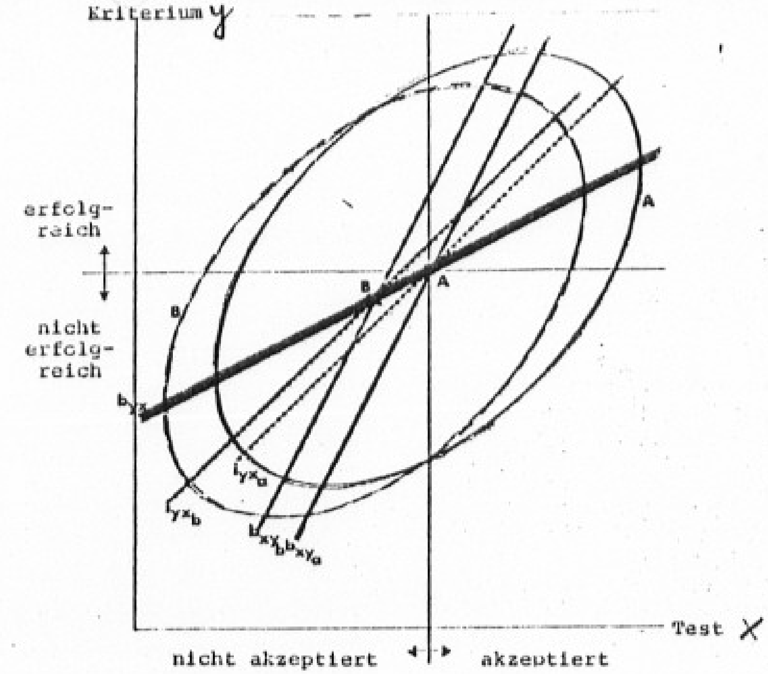
THORNDIKE, R.L., Concepts of culture-fairness. J. Educ. Measmt., 1971, 8, 63 - 7o

VAN DE GEER, J.P., Introduction to multivariate Analysis for the social sciences. San Francisco: Freeman, 1971.

WONNACOTT, R.J. u. WONNACOTT, T.H. , Econometrics. New York: Wiley, 197o.

Nachtrag:

PAWLIK, K., Ist Umwelt Persönlichkeit? Zum Verhältnis von Ökologie, Differentieller Psychologie und Persönlichkeitsforschung. Vortrag, gehalten auf dem Symposion über die ökologische Fragestellung. Schloß Heinsheim/Neckar, Oktober 1975.

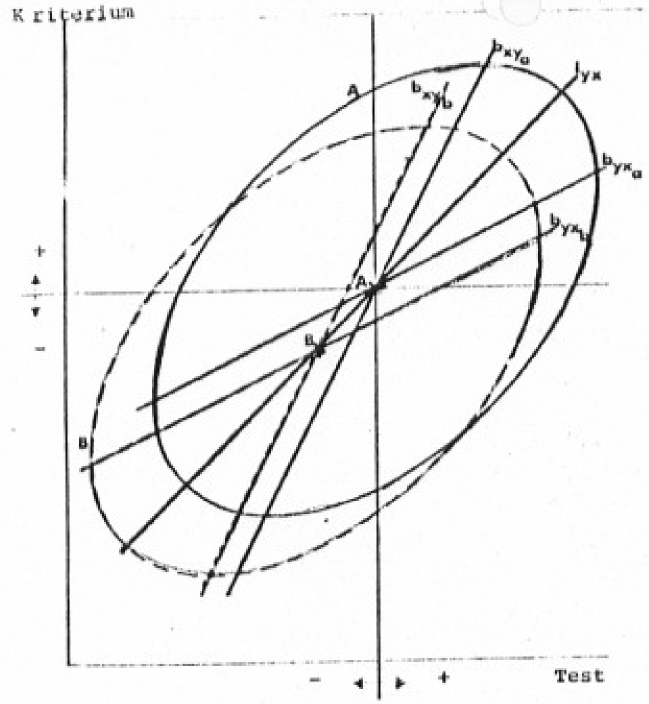


FIGUR 1: Situation in der nach CLEARY derselbe cut-off im Test für beide Gruppen A und B fair ist: Die Regressionen von Y auf X sind für beide Gruppen gleich.

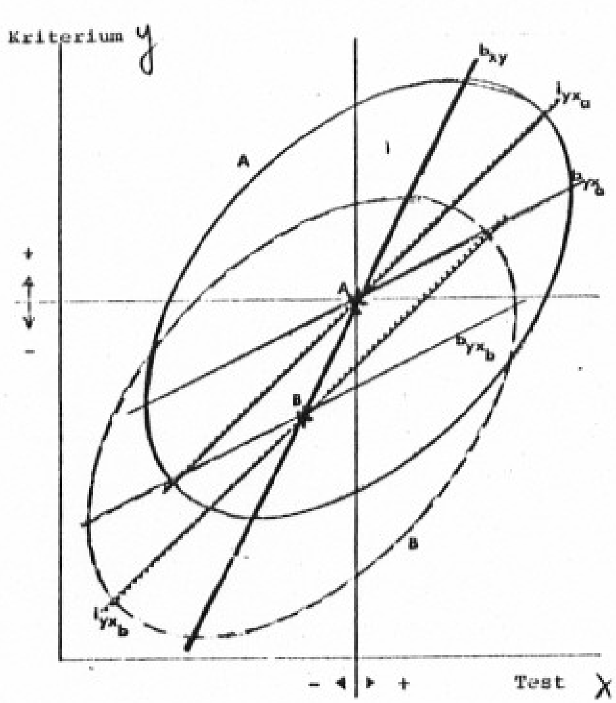
A, B = Centroide der Gruppen A und B im Kriteriums-Prädiktorraum

b_{yx_a} = Regression der Gruppe A mit Y als Kriterium und X als Prädiktor

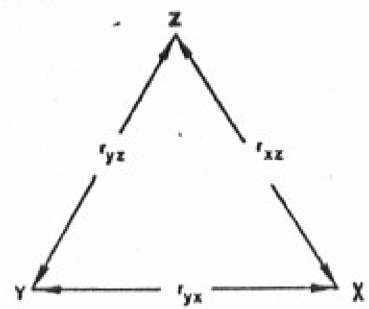
i_{yx_a} = interdependente oder orthogonale Regression in Gruppe A



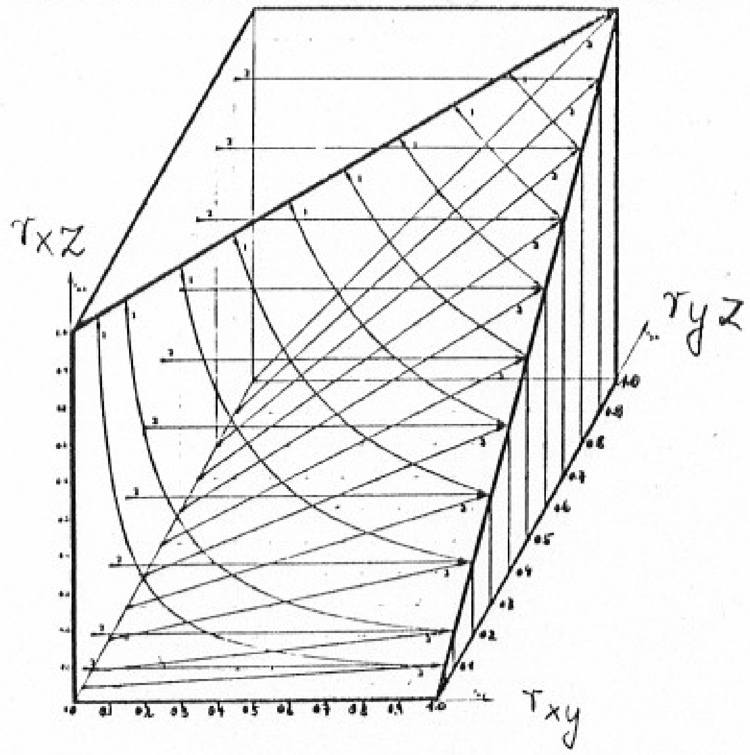
FIGUR 2: Situation in der nach THORNDIKE derselbe cut-off im Test für beide Gruppen A und B fair ist: Die interdependenten oder orthogonalen Regressionen sind für beide Gruppen A und B gleich.



FIGUR 3: Situation in der nach COLE derselbe cut-off im Test für beide Gruppen A und B fair ist: Die Regressionen von X auf Y sind für beide Gruppen gleich.



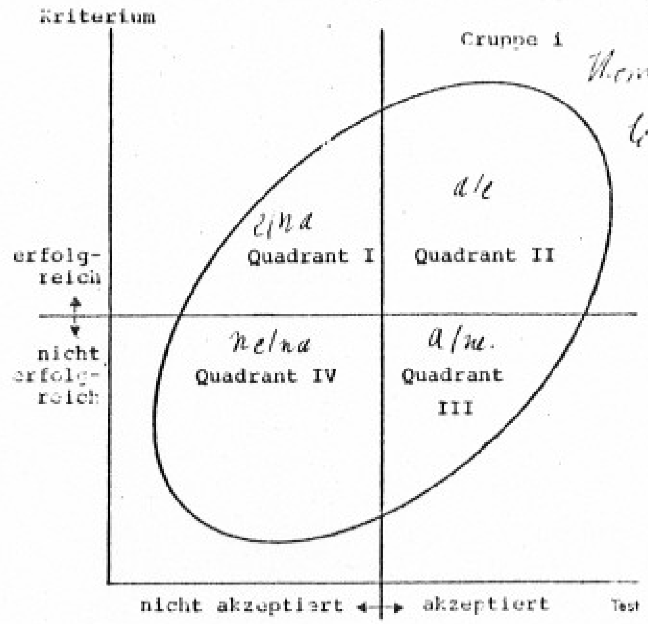
FIGUR 4: Korrelative Beziehungen zwischen Test X, Kriterium Y und Drittvariabler Z



FIGUR 5: Definitionsspezifische Zusammenhänge zwischen der kulturellen Abhängigkeit des Kriteriums r_{yz}, der Validität des Tests r_{yx} und der kulturellen Abhängigkeit des Tests.

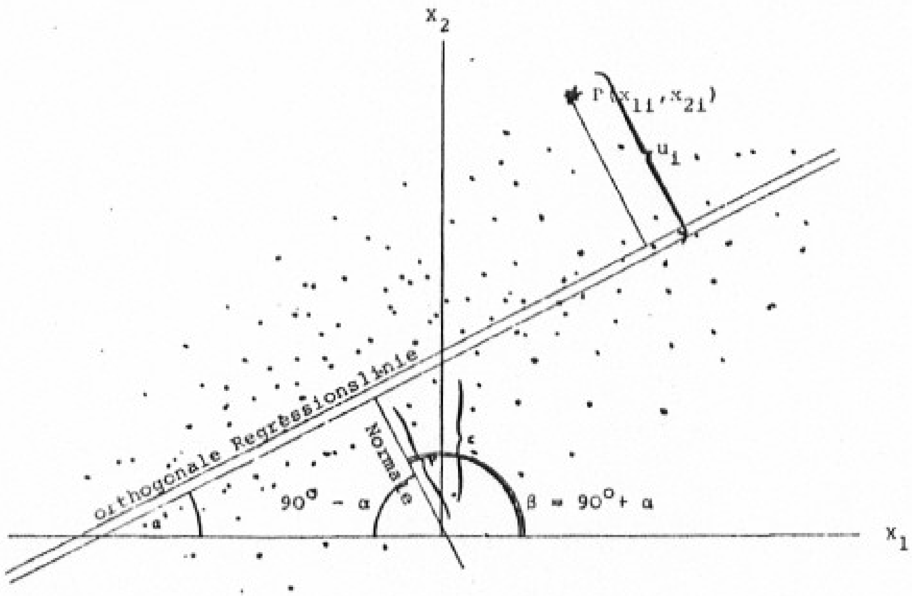
Die vom Testkonstrukteur nicht kurzfristig veränderbare Korrelation r_{yz} und die zu maximierende Korrelation r_{yx} spannen die Grundfläche der FIGUR 5 auf. Je nach Fairnessdefinition muß der Test - unter den Rahmenbedingungen, die ihm durch r_{yz} und r_{yx} auferlegt sind - eine bestimmte kulturelle Abhängigkeit r_{xz} aufweisen, um als fair angesehen zu werden. Diesen bestimmten Wert r_{xz} suchen wir, indem wir vom Punkt (r_{xy}, r_{yz}) in der Ebene senkrecht und parallel zur Achse r_{xz} emporwandern bis wir auf eine Linie 3 (COLE's Definition), eine Linie 2 (THORNDIKE's Definition) oder eine Linie 1 (CLEARY's Definition) stoßen. Die Höhen der Linien 1, 2, 3 entsprechen den Funktionswerten der Fairnessdefinitionen (4.1), (4.2) und (4.3). Das Identitätskonzept fordert $r_{xz} = 0.0$ für alle r_{yx} und r_{yz}, sodaß wir die Fußebene von FIGUR 5 nicht verlassen müssen. Die Definitionen stimmen bei perfekter Validität (r_{xy}=1.0) oder kultureller Unabhängigkeit des Kriteriums überein.

Cleary: $\frac{II}{II + III}$
Thorndike: (I + II) : (II + III)
Cole = $\frac{II}{I + II}$



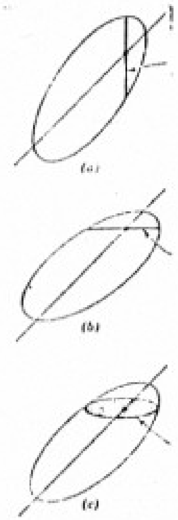
FIGUR 6: durch cut-offs aufgeteilter Test-Kriteriumsraum

Der Anteil von Personen in den Quadranten I und III sollte möglichst niedrig sein, da er die prädiktive Validität des Tests stört.



FIGUR 8: Die orthogonale Regression im Zwei-Variablen-Fall

Die Residuen werden nicht wie bei der klassischen Regression parallel zur Y-Achse sondern orthogonal zur Regressionslinie gemessen.



FIGUR 7: Eine einzige "wahre" Beziehung $\tau_y = \alpha + \beta\tau_x$ kann je nach Fehlerannahme unterschiedliche Rohwertverteilungen erzeugen:

a: Y ist fehlerhaft :
Alle Personen auf der senkrechten Linie besitzen dieselben wahren Werte τ_y und τ_x.

b: X ist fehlerhaft ;
Alle Personen auf der waagerechten Linie besitzen dieselben wahren Werte τ_y und τ_x.

c: Y und X sind fehlerhaft :
Alle Personen innerhalb der Ellipse können dieselben wahren Werte τ_y und τ_x besitzen.