

Sonderdruck aus:

Zeitschrift f. Entwicklungspsychologie u. Pädagogische Psychologie

1976, Band VIII, Heft 1, S. 1—12

Erscheint vierteljährlich im Verlag für Psychologie, Dr. C. J. Hogrefe, Göttingen

Untersuchungen zur Fairneß von Intelligenztests

Heribert Simons und Claus Möbus

Universität Heidelberg

Die Annahme, Intelligenztests seien als fair einzustufen, wenn sie keine Leistungsdifferenzen zwischen Angehörigen unterschiedlicher Populationen sichtbar machen, wird als inadäquat zurückgewiesen. Statt dessen werden auf der Basis einer Fairness-Definition von Cleary die Zusammenhänge von Intelligenz und Schulleistung zwischen mehreren Gruppen von Schülern analysiert. Es stellt sich heraus, daß bestimmte Merkmale der Lernumwelt die differentielle Validität von Intelligenztests und damit die Gefahr unfairer Prognosen fördern, wenn diese Merkmale nicht in die Prädiktion einbezogen sind.

Von einem fairen Test wird häufig erwartet, daß er jedem die Chance einräume, sich in einer Art intelligent zu verhalten, die seiner individuellen Lerngeschichte entspreche. Unfair ist nach dieser Auffassung ein Test, wenn er zwischen Gruppen mit unterschiedlichen Lernerfahrungen trennt. Es wird dabei unterstellt, daß der Test nur deshalb zwischen Gruppen diskriminiert, weil seine Items an einem Intelligenzkonzept orientiert sind, das für die Gruppe mit den niedrigen Testwerten inadäquat ist.

Es setzt sich jedoch immer mehr die Auffassung durch (vgl. Guthke, 1972), daß die traditionellen Intelligenztests nicht aufgrund der durch die Testitems definierten Anforderungen an die kognitiven Fähigkeiten gegenüber bestimmten Gruppen unfair sind, sondern daß die aus den Testergebnissen gezogenen Schlüsse und Entscheidungen unfair sein können, wenn Prognose und Klassifikation die Lerngeschichte des Individuums unberücksichtigt lassen. Insofern stellt auch die Konstruktion sprachfreier, sog. culture fair oder status free Intelligenztests keine Lösung dar, da sich das Problem der potentiellen Unfairneß von Intelligenztests von der Betrachtung von Mittelwertsdifferenzen auf den Prädiktoren auf die Analyse der Prädiktionsgleichung verlagert. Die Frage nach der fairen Testanwendung ist identisch mit der Frage, ob für mehrere Untergruppen einer Population eine *gemeinsame* Regression zu akzeptieren ist oder ob mehrere *getrennte* Regressionen angenommen werden müssen. Dabei schließen nicht nur sich

kreuzende, sondern auch parallel versetzte gruppenspezifische Regressionslinien die Annahme einer für alle Gruppen gültigen gemeinsamen Regression aus.

Genau dieses Problem ist in der Definition der Testfairness von Cleary (1969, S. 115) angesprochen, wonach ein Test gegenüber Untergruppen einer Population dann verzerrt ist, „wenn die Kriteriumsschätzung auf der Basis einer gemeinsamen Regression für die Angehörigen einer Gruppe konsistent zu hoch oder zu niedrig ist. Diese Definition eines ‚test-bias‘ kann auch als Definition der Unfairneß eines Tests verstanden werden, vor allem wenn die Kriteriumsschätzungen zu niedrig sind“.

Untersuchungen

1. Arbeiter- und Akademikerkinder mit unterschiedlich hohen Intelligenz- und Schulleistungen

Aus einer Gesamtstichprobe von 310 Schülern des vierten Grundschuljahres wurden 58 Arbeiter- und 63 Akademikerkinder ausgewählt.¹⁾

Zur Intelligenzmessung wurde die PMA-Testbatterie in der deutschen Version von Kemmler (1967) angewandt. Der Untertest ‚Zahlenverständnis‘ wurde jedoch wegen seiner zu großen inhaltlichen Überlappung mit in der Schule gelernten Kenntnissen als ‚achievement test‘ bewertet und hier wie auch in den folgenden Untersuchungen nicht berücksichtigt. Indikator der Schulleistung war der Gesamtfehlerwert im Rechtschreibtest RST 4 sowie die Note im Fach Deutsch. Alle Variablen wurden für die Gesamtgruppe auf dem Mittelwert 50 und die Standardabweichung 10 standardisiert.

Tab. 1. Mittelwerte und Standardabweichungen für 6 Untertests des PMAT, einen Rechtschreibtest und die Deutschnote bei 58 Arbeiter- und 63 Akademikerkindern

Variablen	Arbeiterkinder		Akademikerkinder	
	M	S	M	S
Bilderwortschatz	44.1	8.8	55.5	7.5
Wortschatz	43.5	10.1	56.4	8.2
Raumerfassung	46.9	9.7	54.2	9.0
Figurengruppe	43.7	10.0	53.9	9.5
Wortgruppe	46.8	10.7	52.3	9.2
Wahrnehmung	44.6	11.3	54.9	8.1
RST 4	56.3	11.1	46.2	8.6
Deutschnote	58.2	10.6	45.0	7.6

1) Die Problematik und Begründung dieser „statischen und soziologischen“ gegenüber der „psychologischen“ Schichtzuweisung ist bei Simons (1973 a) diskutiert.

In Tab. 1 sind die Verteilungskennwerte für Arbeiter- und Akademikerkinder aufgeführt. Die Differenzen sind in allen 8 Vergleichen statistisch sehr bedeutsam ($p < 1\%$). Auch in den sog. „milieuresistenten“ Tests schneiden die Arbeiterkinder schlechter ab.

Fragestellung. Kann man zur Schätzung der Rechtschreibleistung und der Deutschnote eine für Arbeiter- und Akademikerkinder gemeinsame Regression benutzen oder würde ein solches Vorgehen zu einer konsistenten Über- bzw. Unterschätzung der Schulleistung in den Gruppen führen?

Ergebnisse. Wie der Test für den Vergleich der Regressionskoeffizienten²⁾ bei Arbeiter- und Akademikerkindern zeigt (vgl. Tab. 2), weichen die Schätzungen durch die gemeinsame und die getrennten Regressionen nicht signifikant voneinander ab. Nach der Definition von Cleary ist der Intelligenztest fair. Allerdings ist ein Hinweis darauf gegeben, daß die Verwendung einer gemeinsamen Regression zur Schätzung der Note problematischer ist als zur Schätzung der im Test erfaßten Rechtschreibleistung.

Tab. 2. Quadrate der multiplen Korrelationskoeffizienten für den Zusammenhang von 6 PMA-Untertests als Prädiktoren und zwei Schulleistungskriterien bei 58 Arbeiter- und 63 Akademikerkindern sowie F-Test für den Vergleich der Regressionen

Kriterium	Arbeiter-	Akademiker-	F (df = 7/107)
	kinder R ²	kinder R ²	
RST 4	.26	.28	0.10
Deutschnote	.40	.26	1.54

2. Arbeiter- und Akademikerkinder mit vergleichbaren Intelligenzleistungen

Aus einer Gesamtstichprobe von 268 Gymnasiasten wurden 74 Arbeiter- und 52 Akademikerkinder ausgewählt.

Zur Messung der Intelligenz wurde gegen Ende der 5. Klasse der PMAT angewandt. Indikatoren der Schulleistung waren die Jahresnoten in den Fächern Deutsch, Englisch und Mathematik in der 5. und 6. Klasse. Alle Variablen wurden für die Gesamtstichprobe auf den Mittelwert 50 und die Standardabweichung 10 standardisiert. Die entsprechenden Verteilungskennwerte sind in Tab. 3 aufgeführt.

2) Die Signifikanzprüfung auf Ungleichheit der Regressionslinien wurde mit einem Test von Chow (1960) durchgeführt. Programm REV, Programmautor Dr. Gebhardt. Gerechnet wurde auf der IBM 360/44 des RZ der Universität Heidelberg.

Interessierte Leser können bei den Autoren eine Auflistung aller Regressionsgleichungen anfordern, die den in dieser Arbeit angegebenen F-Werten zugrunde liegen.

Arbeiterkinder, die ein Gymnasium besuchen, bilden eine hochselegierte Gruppe. So überrascht nicht, daß sie sich in den durch den PMAT definierten intellektuellen Eingangsvoraussetzungen von den Akademikerkindern nicht unterscheiden. Die Differenzen im Fach Mathematik sind in beiden Schuljahren ebenfalls unbedeutend. Dagegen lassen sich die Unterschiede in den sprachlichen Fächern, vor allem in der Fremdsprache, statistisch absichern. Die beobachteten Differenzen vergrößern sich von der 5. zur 6. Klasse.

Tab. 3. Mittelwerte, Standardabweichungen und signifikante t-Werte für den Vergleich von Intelligenz- und Schulleistungen bei 74 Arbeiter- und 52 Akademikerkindern

Variablen	Arbeiter-		Akademiker-		t
	M	S	kinder M	S	
Bilderwortschatz	49.2	10.3	50.2	7.7	—
Wortschatz	48.2	11.6	50.7	9.9	—
Raumerfassung	51.5	10.5	52.2	8.6	—
Figurengruppen	50.6	9.7	49.7	11.6	—
Wortgruppen	49.8	8.8	50.9	11.5	—
Wahrnehmung	51.6	8.1	50.0	10.9	—
Deutsch, 5. Kl.	50.7	9.9	46.7	12.1	2.05*
Englisch	52.1	10.7	46.7	11.8	2.66**
Mathematik	50.6	10.6	48.4	12.3	—
Deutsch, 6. Kl.	51.2	9.1	45.8	11.3	2.95**
Englisch	52.8	9.4	46.4	12.0	3.33**
Mathematik	49.3	9.9	48.0	11.7	—

* $p < .05$; ** $p < .01$.

Fragestellung. Beeinflußt dieser hier sichtbar werdende Schereneffekt die Fairneß des verwendeten Intelligenztests? Sind längerfristige Prognosen des Schulerfolgs bei Arbeiter- und Akademikerkindern trotz vergleichbaren Intelligenzniveaus auf der Basis einer gemeinsamen Regression unfair?

Ergebnisse. Wie Tab. 4 zeigt, hängt eine faire Schätzung der Schulleistungen auf der Basis einer gemeinsamen Regression vom Schulfach und vom Prognosezeitraum ab. Zur Schätzung der Mathematiknote können — zumindest für den hier berücksichtigten Prognosezeitraum — die Intelligenzleistungen von Arbeiter- und Akademikerkindern mit denselben Regressionskoeffizienten gewichtet werden. Ein solches Vorgehen würde bei der Schätzung der Englischnote zu einer unfairen Kriteriumsschätzung im Sinne von Claery führen. Ein ähnlicher Trend ist für die Schätzung der Deutschnote erkennbar. Insgesamt zeigt sich, daß mit zunehmender Dauer des Schulbesuchs die Verwendung einer gemeinsamen Regression problema-

tischer wird. Dieser Befund wird durch die Analyse der Regression der Note in der zweiten Fremdsprache in der 7. Klasse auf die Intelligenzleistungen bei einer Teilgruppe der Arbeiter- und Akademikerkinder erhärtet. Der F-Test zum Vergleich der getrennten und gemeinsamen Regression ergab hier einen Wert von $F = 3.04$ ($df = 7/47$).

Tab. 4. Quadrate der multiplen Korrelationskoeffizienten für den Zusammenhang zwischen 6 PMA-Untertests als Prädiktoren und verschiedenen Jahresnoten in der 5. und 6. Klasse des Gymnasiums bei 74 Arbeiter- und 52 Akademikerkindern sowie F-Werte für den Vergleich der Regressionen

Kriterien	Arbeiter-	Akademiker-	F (df = 7/112)
	R ²	kinder R ²	
Deutsch, 5. Kl.	.24	.39	1.64
Englisch	.17	.31	2.29*
Mathematik	.23	.28	0.19
Deutsch, 6. Kl.	.29	.28	1.94
Englisch	.15	.24	2.85**
Mathematik	.17	.32	0.50

* $p < .05$; ** $p < .01$.

Dikussion. Die Ergebnisse der 1. Untersuchung zeigen, daß das Merkmal ‚Schichtzugehörigkeit‘ für sich allein genommen keine Bedingung für die Unfairneß von Intelligenztests ist. Die Befunde der 2. Untersuchung legen vielmehr die Vermutung nahe, daß bestimmte Randbedingungen des Unterrichtsprozesses sowie Ausmaß und Qualität der außerschulischen Unterstützung die nicht durch die Intelligenz erklärbaren Leistungsdifferenzen von Arbeiter- und Akademikerkindern verursachen (vgl. Simons, 1973 a). Bei Gültigkeit dieser Annahme sind es Instruktionsbedingungen und nicht die soziale Herkunft, die die Gefahr unfairer Verwendungen von Intelligenztests nahelegen.

In einer dritten Stichprobe finden wir eine Bestätigung für diese Vermutung. Simons, Weinert & Ahrens (1975) haben festgestellt, daß die Qualität des Unterrichts darüber entscheidet, welche Rechenfertigkeiten bei gegebener Intelligenz der Schüler erreicht werden. Durch diesen plausiblen Tatbestand wird allerdings die Gültigkeit von Intelligenztests als Prädiktoren der Schulleistung erheblich reduziert. Im Grunde können ohne Kenntnis der schulischen Lernumwelt keine fairen Schlußfolgerungen aus Ergebnissen von Intelligenztests über den zu erwartenden Schulerfolg gezogen werden.

3. Schüler aus unterschiedlich gut geförderten Klassen mit vergleichbaren Intelligenzleistungen

Aus der Überlegung heraus, daß bei vergleichbaren intellektuellen Eingangsvoraussetzungen durchschnittlich bessere Rechenleistungen in einer Schulklasse auf gute Förderung (z. B. durch systematische Konsolidierung relevanter Vorkenntnisse), durchschnittlich schlechtere Schulleistungen dagegen auf einen eher ineffektiven Unterricht schließen lassen, wurden aus insgesamt 12 Schulklassen des 4. Grundschuljahres (297 Schüler) drei Klassen ($N = 74$) mit durchschnittlich hohen und drei Klassen mit durchschnittlich niedrigen Rechenleistungen ($N = 83$) identifiziert. Das Niveau der Allgemeinen Intelligenz aller 6 Klassen war vergleichbar.

Prädiktoren sind auch in dieser Untersuchung die 6 PMA-Untertests. Die Rechenleistungen waren durch einen informellen Test ermittelt worden, dessen Aufgaben auf den Unterrichtsstoff des 4. Grundschuljahres bezogen waren. Außerdem waren noch die Schulangst der Schüler mittels einer deutschen Version der Test Anxiety Scala (Sarason u. a., 1960), das Geschlecht und das Alter der Klassenlehrer sowie deren Einstellung gegenüber Kindern und ihr bevorzugter Erziehungsstil mittels einer Kurzform des MTAI (Minnesota Teacher Attitude Inventory; vgl. Weinert, 1966, S. 104) ermittelt worden. Alle Schülermerkmale waren für die Gesamtstichprobe auf den Mittelwert 50 und die Standardabweichung 10 standardisiert worden. Die entsprechenden Verteilungskennwerte sind in Tab. 5 aufgeführt.

Die Daten in Tab. 5 zeigen erwartungsgemäß keinen Unterschied zwischen beiden Gruppen in den 6 Untertests des PMAT. Andererseits ist die hochsignifikante Differenz in den Rechenleistungen ebenfalls durch die Definition der beiden Gruppen gegeben. Weiterhin zeigte sich jedoch, daß

Tab. 5. Mittelwerte, Standardabweichungen und signifikante t -Werte für den Vergleich von schlecht und gut geführten Klassen ($N = 83$ vs. $N = 74$ Schüler) in Schüler- und Lehrermerkmalen

Variablen	schlecht geförderte Klassen		gut		t
	M	S	M	S	
Bilderwortschatz	52.2	9.9	49.0	11.2	—
Wortschatz	50.5	11.4	51.6	10.1	—
Raumerfassung	52.0	9.1	50.8	9.7	—
Figurengruppen	50.6	9.8	49.7	10.9	—
Wortgruppen	51.8	9.5	48.9	10.1	—
Wahrnehmung	50.8	10.2	50.8	8.7	—
Schulangst	52.0	9.2	47.7	9.4	2.88*
Rechenleistung	45.9	9.3	54.2	9.5	5.52*
Alter der Lehrer	53.1	11.9	39.0	12.2	7.33*
MTAI, Lehrer	1.2	6.4	4.7	3.7	4.15*

* $p < .01$.

(a) die Lehrer der gut geförderten Klassen deutlich jünger sind; (b) die Lehrer in den gut geförderten Klassen tolerantere Einstellungen gegenüber Kindern angeben; (c) möglicherweise als Folge der genannten Charakteristika die Schüler in den gut geförderten Klassen eine geringere Schulangst äußern; (d) alle drei Klassenlehrer der gut geförderten Klassen und nur einer der drei schlecht geförderten Klassen weiblichen Geschlechts sind. Es liegt nahe, diese vier Variablen (a—d) als Indikatoren motivationaler und emotionaler Komponenten des Unterrichtsprozesses bzw. des Lernklimas einer Schulklasse aufzufassen. Da die Variable ‚Geschlecht des Lehrers‘ in der gut geförderten Gruppe nicht variiert, konnte sie bei der Berechnung gruppenspezifischer Regressionen nicht berücksichtigt werden. Ihre Bedeutung kann jedoch durch einen Vergleich der gemeinsamen Regression mit und ohne Ausschluß dieser Variable abgeschätzt werden.

Fragestellung. Kann eine mögliche Unfairneß des verwendeten Intelligenztests durch Berücksichtigung von Indikatoren des ‚Lernklimas‘ so reduziert werden, daß die Schätzung der Rechenleistung auf der Basis einer für beide Schülergruppen gemeinsamen Gewichtung des Prädiktoren vorgenommen werden kann?

Ergebnisse. Zunächst weist der hochsignifikante F-Wert für den Vergleich von gemeinsamer und getrennten Regressionen der Rechenleistung auf die Intelligenzleistungen darauf hin, daß die Schätzungen der Rechenleistung für beide Gruppen getrennt vorgenommen werden müssen (vgl. Tab. 6, 1. Zeile). Die geringe Korrelation zwischen Rechenleistung und Intelligenz in der Gruppe der gut geförderten Schüler läßt es darüber hinaus wenig sinnvoll erscheinen, überhaupt die Intelligenz in dieser Gruppe

Tab. 6. Quadrate des multiplen Koorelationskoeffizienten für den Zusammenhang zwischen verschiedenen Prädiktorkombinationen und der Rechenleistung für Schüler aus gut geförderten und aus schlecht geförderten Klassen und für die Gesamtgruppe sowie F-Werte für den Vergleich von gemeinsamer und getrennten Regressionen

Prädiktorkombination	gut	schlecht	Gesamt	F	df
	geförderte Klassen	R ²			
	R ²	R ²	R ²		
Intelligenz, Schüler	.13	.38	.19	6.44**	7/143
plus Alter, Lehrer	.15	.38	.26	3.84**	8/141
plus MTAI, Lehrer	.16	.38	.26	3.50**	9/139
plus Angst, Schüler	.23	.41	.34	2.41*	10/137
plus Geschl., Lehrer	—	—	.38	—	—

* $p < .05$; ** $p < .01$.

als Prädiktor zu verwenden ($R^2 = .13$). Der Abfall der F-Werte (Tab. 6) demonstriert aber, daß durch die Einbeziehung möglicher Indikatoren des Lernklimas die Unfairneß von Schulleistungsprognosen auf der Basis einer gemeinsamen Regression erheblich reduziert wird. Da durch die Berücksichtigung des Geschlechts der Lehrer sich der Anteil der aufgeklärten Kriteriumsvarianz bei einer für beide Gruppen gemeinsamen Regression von 33.5 % auf 38 % erhöht ($F = 5.23$, $df = 1/72$; $p < 1 \%$), andererseits diese Variable nur zwischen den Gruppen und nicht innerhalb der Gruppen variiert, erweist sich das Geschlecht des Lehrers in unserer Stichprobe als ein relevanter Faktor zur Aufklärung gruppenspezifischer Differenzen in den Rechenleistungen. Es ist anzunehmen, daß mit dem vollständigen Prädiktorsatz die Rechenleistungen von Schülern aus gut und schlecht geförderten Klassen auf der Basis einer gemeinsamen Regression geschätzt werden können.

Durch eine graphische Veranschaulichung der unterschiedlichen Veränderungen der multiplen Korrelationen in den beiden Schülergruppen bei Einbeziehung der Indikatoren des Lernklimas können die wichtigsten Befunde noch einmal verdeutlicht werden (zur Art der graphischen Veranschaulichung vgl. Pawlik, 1968, S. 485). Berücksichtigt man diese Variablen, so finden die interindividuellen Differenzen in den Rechenleistungen eine bessere Erklärung. Die multiplen Korrelationskoeffizienten werden größer, der Winkel zwischen den Vektoren für die tatsächliche und geschätzte

Abb. 1. Darstellung der Zusammenhänge zwischen tatsächlicher Rechenleistung Y und geschätzter Rechenleistung.

Die Schätzungen erfolgen auf der Basis dreier Prädiktormengen:

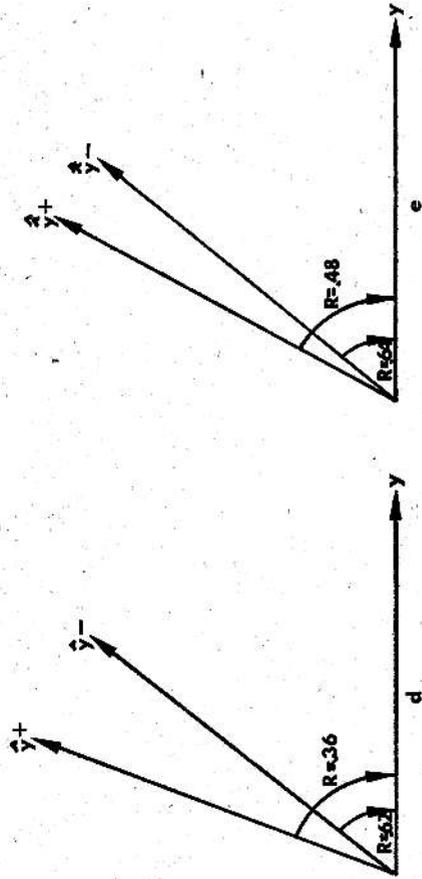
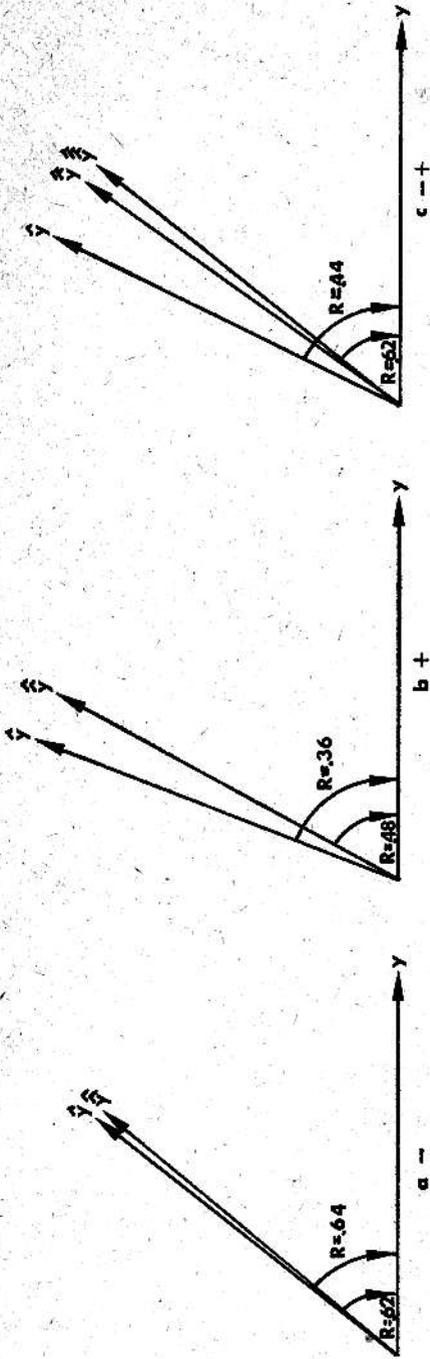
\hat{y} = Schätzung des Kriteriums mit 6 PMA-Tests

$\hat{\hat{y}}$ = Schätzung des Kriteriums mit 6 PMA-Tests, der Schulangst sowie dem Alter und der Erziehungseinstellung der Lehrer

$\hat{\hat{\hat{y}}}$ = Schätzung des Kriteriums mit 6 PMA-Tests, der Schulangst, dem Alter, der Erziehungseinstellung und dem Geschlecht der Lehrer

- a - : Die Schätzung erfolgt für $N = 83$ Schüler aus schlecht (-) geförderten Klassen
- b + : Die Schätzung erfolgt für $N = 74$ Schüler aus gut (+) geförderten Klassen
- c - + : Die Schätzung erfolgt für die Gesamtgruppe ($N = 157$)
- d: Die Schätzung erfolgt auf der Basis der 6 PMA-Tests für gut (+) und schlecht (-) geförderte Klassen
- e: Die Schätzung erfolgt auf der Basis des erweiterten Prädiktorsatzes für gut (+) und schlecht (-) geförderte Klassen

Die Kosinus der Winkel zwischen den Kriterien- und Schätzvektoren entsprechen dabei den multiplen Korrelationen R in Tabelle 6: Je größer R ist, desto kleiner ist der Winkel zwischen den Vektoren Y und \hat{Y} .



Rechenleistung wird kleiner (vgl. Fig. a, b und c). Diese Steigerung der Validität ist nur bei den Schülern aus gut geförderten Klassen bemerkenswert (vgl. Fig. a und b). Aus diesem Grunde wird die Schere zwischen gut und schlecht geförderten Schülern durch Berücksichtigung der Indikatoren des Lernklimas kleiner (vgl. Fig. d und e).

Erörterung

Unsere Ergebnisse bestätigen insgesamt die schon häufig formulierten Forderungen nach Berücksichtigung von Persönlichkeitsmerkmalen und von situativen Merkmalen der Lernumwelt zur Prognose und Kontrolle von Schulleistungen. So schreibt z. B. Brown (1962):

“Description of these environments and the dynamic interaction of personality structure with the culture found at various colleges constitutes an essential sort of data for both the prediction and the control of academic performance” (S. 553).

Kontrolle und Vorhersage setzen nicht nur Kenntnisse über jene Faktoren voraus, die in unausgelesenen Schülerpopulationen mit der Leistung korrelieren; „man muß vielmehr auch die unterschiedlichen Lernumwelten kennen, in denen sich derselbe Faktor unterschiedlich auswirken kann“ (Sanford, 1959, zit. nach Brown, 1962, S. 555).

Auch wenn unsere Befunde diese Notwendigkeit demonstrierten, ist nicht zu übersehen, daß die für eine Fairneß von Tests vorrangige Forderung nach Gleichheit der Validität der Prädiktoren in verschiedenen Gruppen noch nicht voll erfüllt werden konnte. Wesentliche Bedingungen der Variation der Rechenleistungen der gut geförderten Schüler sind unberücksichtigt geblieben. Wir sind auch keineswegs davon überzeugt, daß die einbezogenen Indikatoren des Lernklimas bei einer Kreuzvalidierung zu stabilen Resultaten führen. Die Identifizierung von Merkmalen des schulischen Sozialisationsprozesses, die als Bedingung jenes Anteils der Leistungsvariation in Frage kommen, der nicht durch die Intelligenz der Schüler aufgeklärt wird, ist noch nicht sehr weit fortgeschritten. Die Kenntnis dieser Bedingung ist jedoch nicht nur für eine Optimierung des Unterrichts, sondern auch für eine faire Erfolgsprognose eine entscheidende Voraussetzung.

Man müßte erwarten, daß die Analyse erwartungswidriger Schulleistungen zu hier verwertbaren Resultaten gekommen ist. Gerade die under- und overachievement-Forschung ist auf die Bedingungen der nicht durch die Intelligenz aufklärbaren Leistungsvariation zentriert. Aber die bekannt gewordenen widersprüchlichen und inkonsistenten Ergebnisse demonstrieren u. E., daß das Problem der differentiellen Validität und Unfairneß von Intelligenztests im Sinne der Definition von Cleary in der Regel übersehen wurde. Da die Identifizierung der underachievers und overachievers mittels

regressionsanalytischer Verfahren vorgenommen wird (Thorndike, 1963; Asbury, 1974), richtet sich die Chance eines Schülers, zu einer der beiden Gruppen oder zu den Schülern mit erwartungsgemäßen Leistungen gezählt zu werden, danach, wie seine Intelligenztestleistungen gewichtet werden. Wo statt angezeigter schulklassenspezifischer Regressionen eine für alle Schüler gemeinsame Regression verwendet wird, muß eine unfaire Klassifikation der Schüler resultieren. Die Verwendung einer gemeinsamen Regression hat in solchen Fällen nur dann einen forschungsstrategischen Sinn, wenn man prüfen will, ob die Residuen Muster aufweisen (z. B. Häufung von *underachievers* und *overachievers* in bestimmten Schulklassen), um diese Muster zur Identifizierung von *Schulklassen* zu benutzen, in denen besondere Verhältnisse die Beziehungen zwischen Prädiktor und Kriterium modifizieren.

Ist man jedoch weiterhin daran interessiert, bestimmte *Schüler* als *under-* oder *overachievers* zu identifizieren und bezweifelt man die Gültigkeit von Intelligenztests als Norm für adäquate und abweichende Schulleistungen nicht (Simons, 1973 b; Wahl, 1974), so muß man entweder einen für alle Schulklassen fairen Test verwenden oder aber die Auswahl der *under-* und *overachievers* schulklassenspezifisch vornehmen und das Phänomen der erwartungswidrigen Schulleistungen auch auf schulklassenspezifische Bedingungen relativieren.

Summary

The assumption that intelligence tests can be regarded as fair if they do not show any performance differences between members of different populations is rejected as inadequate. On the basis of the fairness definition of Cleary relationships of intelligence and school achievement between several groups of students were analyzed. It was found that certain characteristics of the learning environment influence the differential validity of intelligence tests and create, thus, the danger of unfair prognoses if these characteristics are disregarded in the prediction.

Literatur

- Asbury, Ch. A., Selected factors influencing over- and underachievement in young school-age children. *Review of Educational Research*, 1974, 44, 409—428.
- Brown, D. R., Personality, college environment, and academic productivity. In: N. Sanford (Hg.), *The American College*. New York: Wiley, 1962.
- Chow, G. C., Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 1960, 28, 591—605.
- Cleary, T. A., Test bias: prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement*, 1968, 5, 115—124.
- Guthke, J., *Zur Diagnostik der individuellen Lernfähigkeit*. Berlin: Verlag Erziehung und Bildung, 1972.

- Kemmler, L., Erfolg und Versagen auf der Grundschule. Göttingen: Hogrefe, 1967.
- Sanford, N., Motivation of high achievers. In: O. David (Hg.), The education of women. Washington: ACE, 1959.
- Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R. & Ruebush, D. L., Anxiety in elementary school children: a report of research. New York: Wiley, 1960.
- Simons, H., Intelligenz- und Schulleistungen von Arbeiter- und Akademikerkindern auf der Unterstufe des Gymnasiums. In: H. Nickel & E. Langhorst (Hg.), Brennpunkte der pädagogischen Psychologie. Bern/Stuttgart: Huber/Klett, 1973 a.
- Simons, H., Fähigkeits- und Kenntnisunterschiede zwischen Schülern. 10. Studienbegleitbrief zum Funkkolleg Pädagogische Psychologie. Weinheim: Beltz, 1973 b.
- Simons, H., Weinert, F. E. & Ahrens, H. J., Untersuchungen zur differentialpsychologischen Analyse von Rechenleistungen. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 1975, 7, 153—169.
- Thorndike, R. L., The concept of over- and underachievement. New York: Teachers College Press, 1963.
- Wahl, D., Erwartungswidrige Schulleistungen. Weinheim: Beltz, 1974.
- Weinert, F. E., Erziehungsstile in ihrer Abhängigkeit von der individuellen Eigenart des Erziehers. In: Th. Herrmann (Hg.), Psychologie der Erziehungsstile. Göttingen: Hogrefe, 1966.

Anschrift der Verfasser:

Prof. Dr. Heribert Simons
Dr. Claus Möbus
Psychologisches Institut
der Universität
Hauptstraße 47—51
6900 Heidelberg