

# Einige Gesichtspunkte der Anwendung und Entwicklung von Testverfahren [Schluss]

Autor(en): **Tent, Lothar**

Objektyp: **Article**

Zeitschrift: **Schweizer Erziehungs-Rundschau : Organ für das öffentliche und private Bildungswesen der Schweiz = Revue suisse d'éducation : organe de l'enseignement et de l'éducation publics et privés en Suisse**

Band (Jahr): **35 (1962-1963)**

Heft 11

PDF erstellt am: **12.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-851596>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Einige Gesichtspunkte der Anwendung und Entwicklung von Testverfahren

Dr. Lothar Tent

(Schluß)

## V. Welche Eigenschaften soll ein Test haben?

Unter dem Gesichtspunkt des Verwendungszweckes, also der Aufgaben, die er zu erfüllen hat, ist der Test ein Hilfsmittel der Erkenntnisgewinnung, sei es in der praktischen Diagnostik, in der Forschung oder auch bei der Schulleistungsmessung. Als ein solches Hilfsmittel ist er, wie wir gesehen haben, ein Meßinstrument. Als Meßinstrument muß er ganz bestimmten Anforderungen oder, wie man sagt, *Gütekriterien* genügen, wenn er brauchbar sein soll. Das gilt für jeden Test in ähnlicher Weise wie für alle anderen Meßinstrumente, die in der Psychologie, in anderen Wissenschaften, in der Technik oder im Alltag verwendet werden. Vor der Benutzung eines Tests sollte man sich daher genau über dessen Eigenschaften informieren und seine Wahl und die Interpretation danach einrichten. Ich gehe deshalb auf diese Punkte etwas ausführlicher ein.

1. Teste sollen *objektiv* sein. Das heißt, ihre Auswertung muß eindeutig sein. Sie darf nicht von der subjektiven Meinung des Auswerters beeinflusst werden. Mehrere Auswerter müssen also unabhängig voneinander für jeweils ein und dieselbe Testuntersuchung zu demselben Ergebnis kommen. Ihre Ergebnisse müssen vollständig oder doch nahezu vollständig übereinstimmen. Der Grad der Objektivität eines Tests ist durch das Ausmaß dieser Übereinstimmung bestimmt. Sie ist abhängig von klaren Lösungsmöglichkeiten und genau festgelegten Auswertungsvorschriften. Bei fast allen Leistungstesten und Fragebogenmethoden ist die Voraussetzung genügend großer Objektivität gegeben.

2. Teste sollen *zuverlässig* sein. Mit jedem Test wollen wir — im Sinne der erwähnten Modellvorstellungen — den Ausprägungsgrad eines oder mehrerer Merkmale erfassen. «Merkmal» bedeutet dabei zunächst nichts weiter, als daß etwas da ist, das sich über die Zeit hinweg als konstant erweist. Wir müssen daher von einem Test erwarten, daß er bei wiederholter Anwendung an ein und derselben Personengruppe zu den gleichen Ergebnissen führt. Unabhängig davon, was der Test mißt, muß er als Meßinstrument verläßlich (*reliabel*) sein. Wie von jedem Meßinstrument verlangt wird, daß es genau mißt, muß auch ein Test das, was er mißt, möglichst exakt messen.

Kein Test ist in diesem Sinne hundertprozentig exakt. Jede Messung mit Hilfe eines Tests ist daher,

wie die Ergebnisse anderer Meßinstrumente, mit einem *Meßfehler* behaftet. Das läßt sich nie ganz vermeiden. Wichtig ist allerdings, daß der Meßfehler nicht allzu groß ist und daß wir wissen, *wie* groß er ist.

Es leuchtet ein, daß dieser Meßfehler direkt abhängt von dem Grad der Zuverlässigkeit des Tests in dem oben angedeuteten Sinn. Das heißt, die Ergebnisse eines Tests mit hoher Zuverlässigkeit haben einen vergleichsweise kleinen, die eines Tests mit geringer Zuverlässigkeit einen entsprechend größeren Meßfehler. Dieser Umstand ist — das wird ebenfalls einleuchten, ohne daß wir auf die theoretisch-statistische Herleitung eingehen — bedeutungsvoll für die Aussagen, die wir anhand der Testergebnisse machen. Können wir beispielsweise sagen, daß zwei Jungen, von denen der eine 31, der andere 36 Punkte in einem Gedächtnistest erzielt habe, sich hinsichtlich dieses Merkmals «wirklich» voneinander unterscheiden? Liegt es nicht einfach an der Unvollkommenheit des Meßinstruments, daß sie verschiedene Punktwerte erreicht haben, während ihr Gedächtnis «in Wirklichkeit» gleich gut ist? Mit anderen Worten, ist der Unterschied nicht bloß ein «Zufallsprodukt» der Ungenauigkeit unseres Meßinstruments?

Um das zu entscheiden, müssen wir also, neben anderen Informationen, den Grad der Zuverlässigkeit des Tests genau kennen. Er wird durch das Ausmaß bestimmt, in dem zwei Messungen mit demselben Test an derselben Personengruppe miteinander «korrelieren». Das heißt, es wird der Übereinstimmungsgrad der Ergebnisse beider Messungen in Form eines sogenannten Korrelationskoeffizienten festgestellt, der zwischen  $r = 0$  (keinerlei Zusammenhang) und  $r = \pm 1$  (vollkommener gleichsinniger oder gegenläufiger Zusammenhang) liegen kann.

Manchmal ist es jedoch unzweckmäßig, die Zuverlässigkeit eines Tests durch zweimalige Messung in ein und derselben Form zu ermitteln («Wiederholungszuverlässigkeit»). Bei einem Intelligenztest mit Zeitgrenzen z. B. kann es sein, daß die untersuchten Personen von der ersten Untersuchung unterschiedlich viel behalten haben. Diejenigen, die sich viel gemerkt haben, können beim zweiten Mal allein aufgrund des Gedächtnisses schneller und damit weiter vorankommen als solche, die sich nichts oder nur wenig gemerkt haben. Unsere Ergebnisse würden verfälscht. Beide Messungen bezögen sich nicht mehr in gleicher Weise auf das gleiche Merkmal. Deshalb konstruiert man in vielen Fällen sogenannte Parallelteste, die untereinander gleichwertig sind, und bestimmt die Zuverlässigkeit durch die Untersuchung derselben Gruppe mit beiden Testformen («Paralleltestzuverlässigkeit»). Parallelteste haben außerdem den für Schuluntersuchungen sehr wichtigen praktischen Vorteil, daß man auch auf engem Raum, ohne Gefahr des «Abguckens», gleichzeitig viele Schüler testen kann, indem man dem einen Schüler die Testform A, seinem Nachbarn die Testform B usw. zur Bearbeitung gibt.

Schließlich kann es sein, daß wir Merkmale zu untersuchen haben, für die die Annahme der Kon-

stanz über eine längere Zeit nicht zutrifft, zum Beispiel wenn bestimmte Aspekte der Schulreife oder vorübergehende Stimmungen erfaßt werden sollen. Hier genügt es, wenn der Test nur einmal angewendet und die Übereinstimmung der Hälften des Tests (als zwei Parallelteste) überprüft wurde. Wir beschränken uns damit auf die «*Stabilität*» des Tests, daß heißt auf seine «innere Konsistenz» als Meßinstrument. Die Wiederholungs- oder die Paralleltestzuverlässigkeit kann demnach nie größer sein als die Stabilität. Sie kann ferner, wie leicht einzusehen ist, nicht größer sein als die Objektivität des Verfahrens. Die Objektivität eines Tests ist daher die Grundvoraussetzung für alle seine weiteren Eigenschaften.

Um den Meßfehler in praktisch sinnvollen Grenzen zu halten, soll die Zuverlässigkeit eines für die Individualdiagnostik qualifizierten Tests bei einem  $r$  von etwa 0,90 liegen. Zuverlässigkeitskoeffizienten unter  $r = 0,70$  sind nicht mehr vertretbar. Mangelhafte Zuverlässigkeit kann, den oben gemachten Ausführungen zufolge, einzeln oder kombiniert auf geringer Objektivität und geringer Stabilität beruhen.

3. Teste sollen *gültig* sein. Während wir in der Zuverlässigkeit einen Ausdruck für die Genauigkeit haben, mit der der Test ein (beliebiges) Merkmal mißt, gibt uns die Untersuchung seiner *Gültigkeit* Aufschluß darüber, *was* er mißt. Wir müssen also ein genau definierbares Kriterium haben, das den Untersuchungsgegenstand, der mit dem Test erfaßt werden soll, repräsentiert. Die Gültigkeit eines Tests entspricht dann dem Ausmaß, in dem — bei ein und derselben Personengruppe — die Testergebnisse mit dem Ausprägungsgrad des Kriteriums übereinstimmen.

Diese Beziehung wird ebenfalls in einem Korrelationskoeffizienten ausgedrückt: hohe Gültigkeit liegt vor bei einem Koeffizienten  $r$  von mindestens 0,60, mittlere bei einem  $r$  zwischen 0,40 und 0,60 und geringe bei einem  $r$  unter 0,40. Ein brauchbares Prüfungsverfahren sollte mindestens eine mittlere Gültigkeit haben, wenn diese Eigenschaft in dieser Form für den Test eine Rolle spielt (vgl. LIENERT 1961, S. 15).

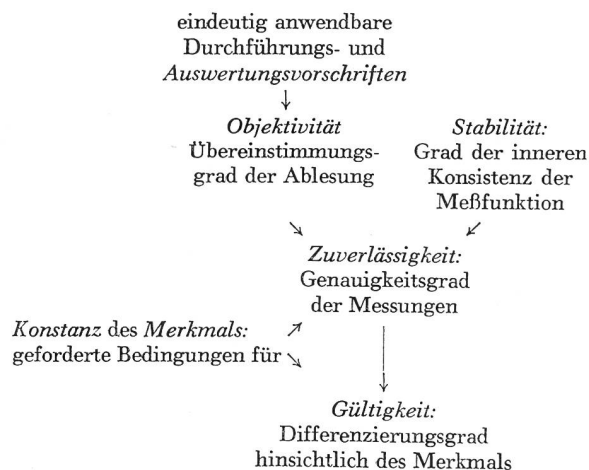
Kommen wir dazu auf das bereits erwähnte Beispiel des Schulreife-tests zurück. Sein Untersuchungsgegenstand ist die «Schulreife», die Eignung der Schulanfänger für die Schule. Werden alle schulpflichtigen Kinder aufgenommen, so stellt sich im Laufe des Schuljahres heraus, welches Kind «mitkommt» und welches nicht. Ein Teil der Kinder bleibt dann erfahrungsgemäß «sitzen» (ungefähr 10—15%). Nachträglich kann man dann sagen, daß die «Sitzenbleiber» aus irgendeinem Grunde nicht schulreif gewesen sein müssen — von Ausnahmen, wie z. B. Krankheitsfällen, abgesehen. — Die «Schulreife» wird durch den Schulerfolg im Anfangsjahr repräsentiert. Der Schulerfolg wird damit zum Kriterium, an dem die Gültigkeit des Tests zu messen ist, wenn wir mit ihm eine Vorhersage treffen und so vielen Kindern das Erlebnis des «Nicht-Mitkommens» mit all seinen Folgen ersparen wollen. Es ist also — vor Veröffentlichung des Tests — sicherzustellen, daß seine Ergebnisse ausreichend hoch mit den Noten korrelieren, die die Kinder am Ende des Schuljahres von ihren

Lehrern erhalten. Analoges gilt für den Übergang von der Grundschule auf weiterführende Schulen oder für die Eignung von Lehrlingen und Stellenbewerbern.

Neben dieser empirisch-statistischen («praktischen») Gültigkeit, die für alle Eignungsteste nachgewiesen sein muß, gibt es weitere Aspekte der Gültigkeit, auf die wir hier nicht näher einzugehen brauchen (vgl. dazu LIENERT 1961, Kap. 11). Für Schulleistungsteste gilt ganz allgemein die Forderung nach «Lehrplan-Gültigkeit» (kurrikulare Validität; vgl. auch F. SÜLLWOLD 1960). Das heißt, die Gültigkeit dieser Teste ist dadurch gewährleistet, daß zur Messung des Schulfortschritts Aufgaben herangezogen werden, die den Lehrplananforderungen für die betreffenden Altersstufen repräsentativ entsprechen. — Grundsätzlich müssen sich für alle Teste plausible Gültigkeitsbeziehungen darstellen lassen.

Objektivität, Zuverlässigkeit, (Verlässlichkeit, Reliabilität) und Gültigkeit (Validität) sind die drei Haupt-Gütekriterien eines Tests. Darüber hinaus sind für jeden guten Test einige weitere Eigenschaften wünschenswert, auf deren Darstellung im einzelnen hier aber ebenfalls verzichtet werden kann.

*Die Beziehung zwischen den wichtigsten Güteeigenschaften eines Tests als Meßinstrument. Zusammenfassendes Schema, modifiziert nach LIENERT 1961, vgl. S. 17—18*



## VI. Wie entsteht ein Test?

Ehe ein Test fertiggestellt ist und zur Benutzung veröffentlicht werden kann, muß eine umfangreiche und vielseitige Arbeit geleistet sein. Diese Aufgabe ist heute praktisch nur durch das Zusammenwirken mehrerer Mitarbeiter zu bewältigen. Zu ihnen gehört, neben dem Psychologen und dem Statistiker, auch der Fachmann des Gebietes, auf dem der Test angewendet werden soll. Für industrielle Eignungsteste z. B. wird man Werkfachleute heranziehen; bei Schulleistungstesten wird man eng mit erfahrenen Pädagogen zusammenarbeiten, die mit den Grundlagen der Testmethodik ein wenig vertraut

sind (vgl. auch SÜLLWOLD 1960). Das Ziel dieser Bemühungen ist es, mit dem neuen Test ein standardisiertes Instrument zu schaffen, das praktische Bedürfnisse befriedigt und dabei den beschriebenen Anforderungen vollauf genügt.

Die ersten Stadien der Testentwicklung sind die *Planung* und der *Testentwurf*. Hier geht es darum, den inhaltlichen Gültigkeitsbereich des Tests abzustrecken und dabei zu überlegen, an welchen Gültigkeitskriterien diese Gültigkeit zu überprüfen wäre. Im Zusammenhang damit muß auch festgelegt werden, auf welchen Personenkreis der Test anwendbar sein soll (Geltungsbereich).

Je nach Zweckmäßigkeit ist außerdem die Art des formalen Testaufbaus und der Aufgaben zu wählen. Sollen es beispielsweise Alternativantwort- (Richtig-Falsch-, Ja-Nein-) Aufgaben sein (etwa: Unterstreich, ob das Ergebnis richtig oder falsch ist: 281 — 117 — 174 R F) oder Mehrfachantwort-Aufgaben (etwa: Kreuze den richtigen Ausdruck an: «rüstig» bedeutet a, alt, b. ausgeruht, c. reisefertig, d. kräftig, e. rostig) oder Umordnungs-Aufgaben (etwa: Gib die richtige Wortfolge an: seinen ein rettete treuer Herrn Hund) oder Ergänzungs-Aufgaben (etwa: Der absolute Nullpunkt liegt bei ... Grad Celsius. — Oder: Setze das angefangene Muster fort!); vielleicht ist auch eine Mischung verschiedener Aufgabentypen erwünscht.

Auf Grund einer vorläufigen psychologischen *Merkmalsanalyse* werden dann Vorschläge für spezielle Testaufgaben zusammengestellt. Für das Merkmal Schulreife beispielsweise hat man gefunden, daß dafür auch die Konzentration eine Rolle spielt, mit der sich die Kinder ungewohnten Arbeiten eine Zeitlang zuwenden können. Man wird also Aufgaben einplanen, die etwa verlangen, Muster steigender Schwierigkeit auf einem Punktegitter nachzuzeichnen. Manchen Kindern fällt es nämlich gar nicht leicht, solch eine Arbeit über eine Reihe von Minuten hin fortzuführen. — Nachdem ein grober Testentwurf, einschließlich der vorläufigen Durchführungs- und Auswertungsvorschriften, zusammengestellt ist, müssen dann die einzelnen Aufgaben genau «konstruiert» werden. Das in Frage kommende Aufgabenmaterial ist auszusuchen, und die gewählten Aufgaben müssen ganz konkret abgefaßt und eingerichtet werden. Von der Erfahrung und dem Geschick des Testverfassers und seiner Mitarbeiter hängt es bereits in diesem Stadium entscheidend ab, ob das Unternehmen erfolgreich sein wird und die Mühe sich lohnt oder nicht. Das Ergebnis der Aufgabenkonstruktion ist die *Testvorform*, die in der Regel noch erheblich mehr Aufgaben enthält als die Testendform.

Die Vorform wird im nächsten Schritt, der *Aufgabenanalyse*, einer Gruppe von Personen vorgelegt, die dem Geltungsbereich des Tests möglichst genau entsprechen soll («repräsentative Stichprobe»).

Ziel der Aufgabenanalyse ist es, durch Ausschaltung und Verbesserung weniger gut geeigneter Aufgaben und durch eine Überprüfung der Testpunkteverteilung die spätere Zuverlässigkeit und Gültigkeit des Tests so weit wie eben möglich zu erhöhen.

Dazu wird die Schwierigkeit, die Trennschärfe und oft auch die Gültigkeit jeder einzelnen Aufgabe untersucht. Trennscharf ist eine Aufgabe dann, wenn diejenigen, die diese Aufgabe lösen, auch im Test insgesamt überdurchschnittlich abschneiden und umgekehrt. Im allgemeinen sind Aufgaben umso wertvoller, je trennschärfer sie sind.

Anhand der Analysedaten wird dann die *Aufgabenselektion* und *-revision* vorgenommen: ungeeignet erscheinende Aufgaben werden ausgeschieden, verbesserungsbedürftige werden geändert. Schließlich wird der Test in seiner endgültigen Form zusammengestellt. Dabei werden das Testmaterial und seine Aufmachung ebenso festgelegt, wie alle Durchführungs- und Auswertungsanweisungen.

Spätestens an der Testendform — und da auf jeden Fall — wird die *Kriterienkontrolle* durchgeführt. Der Test wird an einer wiederum möglichst repräsentativen Stichprobe auf seine Zuverlässigkeit und Gültigkeit überprüft. Nur wenn wir dabei, als Ergebnis aller Mühen, genügend hohe Werte erhalten, ist die Veröffentlichung des Tests gerechtfertigt.

Bevor es so weit ist, muß der Test abschließend jedoch noch «geeicht» werden. Die sogenannten *Testnormen* werden erstellt. Dazu wird der Test an einer relativ großen, für den Geltungsbereich repräsentativen Personengruppe, der Eichstrichprobe, ausprobiert. Die dabei erhaltene Verteilung der Testwerte ist die Grundlage für die Normen, die es nachher, bei der allgemeinen Anwendung des Tests, gestatten, etwas über den Stellungswert eines individuellen Testergebnisses auszusagen. Ein Testpunktwert erhält seine Bedeutung erst dadurch, daß wir wissen, wie sich alle Testwerte verteilen, wo ihr Mittelwert liegt und wie sie «streuen». Testergebnisse werden meist in irgendeine Art von Standardwerten (z. B. IQ-Punkte) transformiert oder in Prozenträngen ausgedrückt, so daß wir immer genau sagen können, wie jemand im Vergleich zu anderen Personen oder im Vergleich zwischen verschiedenen Testverfahren abgeschnitten hat.

Dabei ist ein wichtiger Umstand zu beachten. Während bei den Standardskalen gleiche Punktdifferenzen gleichen Leistungsdifferenzen entsprechen, hängt die Bedeutsamkeit der Unterschiede in der Prozentrangskala von der Verteilung der Ausgangswerte ab. So entspricht z. B. der Unterschied zwischen den IQ-Werten 70 und 80 dem zwischen den Werten 95 und 105. Dagegen sind zwei Leistungen z. B. mit den Prozenträngen 90 und 95 in der Regel sehr viel stärker voneinander verschieden als die mit den Rängen 47 und 52 (vgl. LIENERT, Kap. 12).

## VII. Was können Testverfahren in der Schule leisten?

Abschließend soll an zwei Beispielen aus der Schulpraxis gezeigt werden, was mit Hilfe des systematischen und breit angelegten Einsatzes von geeigneten Testverfahren erreicht werden kann.

1. Zahlreiche Untersuchungen mit Schulreifesten haben in den vergangenen Jahren unsere Erkenntnisse über die Beziehungen zwischen dem Reifestand beim Schuleintritt und dem späteren Schulerfolg wesentlich erweitert und vertieft. Man ist dabei übereinstimmend zu dem Ergebnis gekommen, daß Kinder nicht schematisch nach dem Lebensalter eingeschult werden sollten, sondern daß man in erster Linie den Stand ihrer allgemeinen Entwicklungsreife zugrundelegen müsse (vgl. insbesondere A. KERN 1954, HETZER und TENT 1958, L. SCHENK-DANZINGER in HETZER 1959). Da sich vor allem ein großer Teil der jüngsten Schulanfänger als noch nicht schulreif erwies, wurde in der Folge fast überall das gesetzliche Einschulungsalter um ein Vierteljahr heraufgesetzt, meist verbunden mit der Möglichkeit, jüngere Kinder bei nachgewiesener Reife aufzunehmen. Diese Maßnahmen haben zweifellos entscheidend dazu beigetragen, daß sich in den letzten Jahren die Sitzbleiberzahlen in vielen Gebieten erheblich verringert und die Leistungen erhöht haben. Man vergleiche dazu die folgenden Tabellen.

### Auswirkungen von organisatorischen Maßnahmen auf den Anteil der Sitzbleiber

	Jahrg.	Sitzbleiber im		
		1.	2.	3. Schuljahr
Heidelberg	1949	8,0 %	3,8 %	6,3 %
Saarbrücken	1951	3,8 %	1,8 %	2,4 %
(N ca. je 1000)	1952	3,6 %	2,8 %	—
	1953	2,3 %	—	—

(Nach KERN in M. J. HILLEBRAND 1955)

### Repetentenanteile in den Anfangsklassen in Österreich

Jahrgang	Anteil
1952/1953	9,9 %
1954/1955	8,1 %
in Wien	
1956/1957	3,5 %

(Nach SCHENK-DANZINGER 1959)

### Auswirkungen von Zurückstellungen auf die Schulleistungen

	spätere Schulleistungen	
	mangelhaft oder schlecht	gut oder sehr gut
von 8740 Sitzbleibern	54 %	46 %
von 2259 Zurückgestellten	13 %	87 %

(Nach H. FROMMBERGER 1955, zit. nach SCHENK-DANZINGER 1959)

2. Über den Voraussagewert (die Gültigkeit) der verschiedenen Aufnahmeverfahren für den Übergang auf weiterführende Schulen sind ebenfalls eine Reihe von Untersuchungen angestellt worden. Dabei hat sich ergeben, daß der Wert der von Schulen veranstalteten Aufnahmeprüfungen sehr unterschiedlich ist. Vielfach ist er praktisch gleich null; das heißt, man hätte die Schüler auch nach dem Zufall auswählen können (vgl. U. UNDEUTSCH 1960). Das Entscheidende aber ist, daß man zur Zeit der Aufnahmeprüfungen selbst — wenn also über die Zulassung oder Zurückweisung eines Kindes entschieden wird — gar nichts über den prognostischen Wert des konkreten, jeweils zusammengestellten Prüfverfahrens weiß. Demgegenüber besitzen bereits verfügbare und erprobte Tests einen nachgewiesenermaßen hohen und über Jahre hinaus stabilen Voraussagewert (vgl. J. HITPASS 1961). Bei der Anwendung einer geeigneten Kombination solcher Testverfahren kann der Anteil von Fehlentscheidungen ganz erheblich verringert und damit manches Unrecht vermieden werden:

### Gültigkeitskoeffizienten bei der Aufnahme in weiterführende Schulen

Schulprüfung (Diktat, Aufsatz, Rechnen)	Korrelation mit dem Schulerfolg (biserial)				
	1/2 J.	1 J.	2 J.	3 J.	4 J. 5 J.
Schulprüfung (Diktat, Aufsatz, Rechnen)	0,14	0,08	0,06	-0,05	0,09 0,04
Kombination verschiedener Tests	0,63	0,60	0,54	0,56	0,57 0,52

(N = 293 Schüler, Oberhausen) (Nach HITPASS 1961)

Es ist heute nicht mehr daran zu zweifeln, daß Tests — verständig ausgesucht und durchgeführt — auch für die Schule nützliche Hilfsmittel sind. Jedoch, «wer Tests anwendet, sie lobt und fördert, ohne beurteilen zu können, was sie tatsächlich leisten, macht sich zu einem schlechten Anwalt einer an sich guten Sache» (LIENERT 1961, S. 1).

### Literaturhinweise

1. *Gesamtdarstellung der Testtheorie und Testkonstruktion:* LIENERT, G. A.: Testaufbau und Testanalyse. Verlag J. Beltz, Weinheim 1961. XXXII und 502 Seiten.
2. *Informationen über gebräuchliche Tests und Testliteratur:* DIAGNOSTICA, Zeitschrift für psychologische Diagnostik, Verlag für Psychologie, Göttingen, seit 1955  
HILTMANN, Hildegard: Kompendium der Psychodiagnostischen Tests. Verlag H. Huber, Bern und Stuttgart 1960  
Hochschule für Internationale Pädagogische Forschung: Hilfsmittel zur Schülerbeurteilung (ausführlicher Prospekt). Frankfurt am Main
3. *Einzelfragen:* HETZER, Hildegard (Herausgeberin) Handbuch der Psychologie Band 10: Pädagogische Psychologie. Verlag für Psychologie, Göttingen 1959



HETZER, Hildegard und L. TENT: Der Schulreifetest. Auslesemittel oder Erziehungshilfe? Verlag Dr. Piorkowski, Lindau 1958  
HILLEBRAND, M. J. (Herausg.): Zum Problem der Schulreife. München 1955  
HITPASS, J.: Vergleichende Untersuchungen über den Voraussagewert von Aufnahmeprüfung und Testprüfung zur Erfassung der Eignung für die weiterführenden Schulen. Schule und Psychologie 1961, 8, S. 65—71

KERN, A.: Sitzbleiberelend und Schulreife. Verlag Herder, Freiburg 1954  
SÜLLWOLD, F.: Mitarbeit des Psychologen bei der Konstruktion von Schulleistungstests. Psychologie und Praxis 1960, 4, S. 173—175  
UNDEUTSCH, U.: Auslese für und durch die höhere Schule. Bericht über den XXII. Kongreß der Deutschen Gesellschaft für Psychologie. Verlag für Psychologie, Göttingen 1960

## Pour un rapprochement interculturel

*Quelle éducation convient-il de donner à notre jeunesse?\**

par Louis Meylan

J'ai déjà indiqué comment l'école, et plus généralement l'éducation, peut enraciner en l'enfant un comportement exclusif de cette compréhension intelligente et sympathique d'autrui, qui est la clef de notre problème pratique; et comment la vanité, la sécheresse, la dureté, l'égoïsme de l'adulte peuvent blesser cette confiance, tuer cette curiosité humaines, si vives chez l'enfant.

J'ai rappelé la féroce satire de ce qu'est trop souvent l'éducation domestique, développée par Bernard Shaw dans son *Treatise on education*: L'enfant, un trouble fête et toute l'«éducation», un effort pour réduire au minimum cette «nuisance». D'autre part, on sait que, d'une façon générale, ceux qui devaient, dans la suite, ouvrir à l'humanité des voies nouvelles, dans le domaine de la philosophie, de la poésie, des sciences ou des arts, se sont sentis gênés par l'école et se sont formés en dehors d'elle ou même contre elle. Cette constatation, généralisée, n'est d'ailleurs que plaisante. On se rappelle cette boutade: Comment expliquer que, l'enfant étant généralement si intelligent, l'adulte soit presque toujours si bête? L'école, peut-être?

L'école n'est pas toujours sage. Les éducateurs ne sont pas toujours humains. Le poids de la routine pèse terriblement sur l'institution scolaire. Il n'en reste pas moins que, de même que les Israélites, mordus par les serpents dans le désert, n'ont trouvé la guérison qu'en érigeant au milieu d'eux le serpent d'airain, de même l'éducation pourra, seule, assurer l'ordre et la paix entre les hommes, mais une éducation élevée à la dignité d'œuvre d'art, comme le serpent d'airain.

Michelet, que le problème de l'éducation intéressait au plus haut point, a prononcé un jour ces mots: «Quelle est la première tâche de l'Etat? Et il répond: l'éducation. Quelle est la deuxième tâche de l'Etat? encore l'éducation. Quelle est la troisième tâche de

l'Etat? toujours l'éducation.» L'éducation est le service le plus rentable de l'Etat — à la longue! L'Etat ne saurait y consacrer ni trop de soins ni trop d'argent. Parce que, comme le déclare Kant, dans son *Cours de pédagogie*, si l'homme est le seul être qui puisse être transformé par l'éducation, c'est aussi le seul être qui ait absolument besoin de l'éducation par l'adulte, pour «devenir celui qu'il est.»

Donc le pouvoir de l'éducation est, en droit, sinon absolu, du moins considérable. L'immuable nature humaine, qu'on dit, n'est pas si immuable! L'histoire objective, l'histoire des formes qu'a prises successivement et des contenus qu'ont reçus successivement la vie sociale et la vie individuelle, le prouve.

Dans ces conditions, la collectivité, pratiquement: le ministère de l'éducation ou de l'instruction publique à qui elle délègue ses pouvoirs, a le devoir d'exiger de ceux à qui elle confie cette mission des garanties absolues, du moins en ce qui concerne l'exercice de cette fonction. Et elle a le droit de les prendre.

Actuellement, on s'assure que le futur maître sait ce qu'il devra enseigner. Admettons! Encore que ce que l'on vient d'apprendre ou de redécouvrir, on l'enseigne souvent d'une façon plus vivante et plus efficace que ce que l'on enseigne «depuis trente-six ans». Mais l'éducation n'est pas essentiellement la communication de quelque chose; c'est la rencontre de quelqu'un: un contact intime d'âme à âme, un contact entre l'âme de l'éducateur et l'âme en bouton de l'enfant, qui doit s'épanouir.

Les fonctionnaires de la direction de l'éducation doivent puiser dans l'amour de l'enfant, dont il est normal de penser qu'il les anime, et dans l'amour de l'avenir qui se prépare en lui, le courage de se demander et d'aller voir ce que vaut l'âme de ceux qui prétendent à l'honneur d'être des maîtres.

On prend de grandes précautions avant d'engager les fonctionnaires de certaines administrations. Et, si le médecin décide une intervention et qu'elle tourne

\* Interventions au «Colloque Orient-Occident», Bruxelles, 1958.