

## DISKUSSIONSBEITRÄGE

JOST REISCHMANN

### Der Einsatz von Tests in der Schule

Stellungnahme zu: W. ZEUCH: Was spricht gegen die Anwendung von Testverfahren? In: Die Deutsche Schule, 65. Jg. (1973), 5, S. 340—348.

*Tests gehören nicht in die Hände von Dilettanten. Tests haben ihre Grenzen; ihre Herstellung und Anwendung unterliegt Gefahren, die der kennen muß, der sie sinnvoll als Hilfsmittel einsetzen will. Wenn Lehrer mit Tests als pädagogisch-diagnostischen Verfahren umgehen, dann müssen sie in ihrer Ausbildung mit den Problemen der Testherstellung und -verwendung vertraut gemacht werden; weder ein „naiver Optimismus“, vor dem Zeuch mit Recht warnt, noch ein naiver Pessimismus tragen zu einer realistischen Stellung gegenüber dem Hilfsmittel Test bei.*

Daß Tests Hilfsmittel mit Grenzen, Gefahren und Schwächen sind, braucht weder eingestanden zu werden, noch dient diese Feststellung der Vorbeugung (S. 340)<sup>1</sup>. Vielmehr ist dies eine Einsicht, ohne die eine sachgerechte Arbeit mit Tests nicht möglich ist. Von daher ist es begrüßenswert, wenn ein Autor auf die begrenzte Aussagekraft von Tests hinweist. Dennoch scheinen uns die Ausführungen ZEUCHS zu einseitig. Er beschränkt sich auf Kritik, die sich mit THORNDIKE / HAGEN in zwei Gruppen zusammenfassen läßt: „They are criticisms either of (a) the incompleteness and imperfection of the information yielded by our measurement procedures or (b) the unwise things that we do with that information“<sup>2</sup>, wobei u. E. das Schwergewicht seiner Beispiele entgegen seiner Intention auf (b) liegt. Zudem leitet er aus seiner Kritik keine positiven Forderungen ab und zeigt auch keine Alternativen auf. Dies aber ist nach unserem Erachten als Beitrag für eine realistische Einschätzung der Möglichkeiten des Tests nicht hinreichend. Wir wollen daher einerseits Information über Tests vermitteln, andererseits zu den Vorwürfen ZEUCHS Stellung nehmen.

Dabei beschränken wir unsere Überlegungen auf den Bereich der *Schultests*. Dies sind Prüfverfahren, durch die bestimmte Effekte von Unterricht und Erziehung festgestellt werden sollen; in einem weiteren Sinne werden hierzu auch Verfahren zur Feststellung der Voraussetzungen von Lehr- bzw. Erziehungszielen gerechnet<sup>3</sup>. Die Einschränkung auf Schultests scheint uns gerechtfertigt, da sich die Ausführungen in diesem Rahmen vor allem an Lehrer wenden und sich die Kritik ZEUCHS auch an diesem Gebiet behandeln läßt.

ZEUCH will seine Kritik „nicht vordergründig gegen die Anwendung noch unvollkommener Techniken richten, sondern in erster Linie gegen die Gefahren, die auch

<sup>1</sup> Die Seitenangaben in Klammern beziehen sich auf Zeuch, a. a. O.

<sup>2</sup> Thorndike, R. / Hagen, E.: *Measurement and Evaluation in Psychology and Education*, 2nd Edition — John Wiley and Sons Inc. New York 41964, S. 13.

<sup>3</sup> Süllwold, F.: *Schultests*. In: Heiss, R. (Hrsg.): *Handbuch der Psychologie*, Band 6. Psychologische Diagnostik. Göttingen 1964, S. 352.

mit einer Perfektionierung der Testtechniken nicht aus dem Weg geräumt werden“ (S. 340). Um diesen Anspruch erfüllen zu können, erscheint es uns notwendig, mit den Überlegungen einen Schritt vor ZEUCH einzusetzen.

### Die Aufgabe von Tests

Ein Test ist – so definiert LIENERT – „ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“<sup>4</sup>. Das bedeutet: Inhaltlich orientiert sich ein solches Routineverfahren an einem oder mehreren Merkmalsgebieten und den dort anzutreffenden Verhaltensvariationen; zur Feststellung der individuellen Ausprägung wird ein bestimmter Merkmalsraster an die zu untersuchenden Individuen angelegt, wodurch sie hinsichtlich dieses Merkmals vergleichbar werden. Das Ziel einer solchen Messung ist die möglichst exakte Feststellung des Ist-Zustandes auf einer Skala.

Daraus ergibt sich: Die Aufgabe von Tests besteht darin, zu messen. Tests sind ein Hilfsmittel, um Information zu gewinnen. Das Ergebnis eines Tests ist ein Verhältnismaß, zumeist in Form eines oder mehrerer Testwerte („score“).

Messungen durch Tests liefern also zunächst nur Information, nicht jedoch ein Urteil oder eine Bewertung. Auf dieses oft anzutreffende Mißverständnis weisen THORNDIKE/HAGEN eindringlich hin: „A test will yield only a score, not the conclusion to be drawn from that score. The information provided in a test score is not a substitute for insight.“<sup>5</sup>

Dieser „test score“ (nicht das, was mit dem Ergebnis gemacht wird), so wird in der Testtheorie gefordert, muß bestimmten Forderungen genügen: Hauptgütekriterien sind Objektivität, Reliabilität und Validität, Nebengütekriterien (von ZEUCH nicht genannt) sind Normierung, Vergleichbarkeit, Ökonomie und Nützlichkeit<sup>6</sup>. Normierung bedeutet dabei zumeist, daß ein Vergleich mit der entsprechenden Population möglich ist (Realnorm), nicht die Angabe eines zu erreichenden Zieles (Idealnorm<sup>7</sup>).

Bei der Messung mit Tests wird demnach die *Leistungsfeststellung* von der *Leistungsbeurteilung* oder *-bewertung* getrennt. Der erste Schritt, die Leistungsfeststellung, erfolgt so schematisch und objektiv (d. h. unabhängig von subjektiven Einflüssen verschiedener Benutzer) wie möglich und dient der Feststellung dessen, was ist. In einem zweiten Schritt muß der Lehrer dann aus seiner pädagogischen Erfahrung und Verantwortung heraus, also subjektiv entscheiden, was werden soll. Vor einer Vermischung dieser beiden Schritte warnt INGENKAMP: „Würden die Arbeitsschritte der Datensammlung und Interpretation nicht getrennt werden, so wären der subjektiven Willkür Tür und Tor geöffnet. Man wüßte nicht, was Fakten und Wertungen sind, und könnte nie rational begründete Stellungnahmen abgeben.“<sup>8</sup>

Objektive Leistungsmessung durch Schulleistungstests stellt also nur die *Voraussetzung für Schülerbeurteilung* dar. Die Beurteilung selbst wird sich an dieser Voraussetzung orientieren, aber zusätzlich Kriterien heranziehen. Ein Test kann dem Lehrer eine Entscheidung wohl erleichtern, nicht jedoch abnehmen.

<sup>4</sup> Lienert, G. A.: *Testaufbau und Testanalyse*. Weinheim 1967, S. 7.

<sup>5</sup> a. a. O., S. 13.

<sup>6</sup> Vgl. a. a. O., S. 12 ff.

<sup>7</sup> Vgl. Klauer, K. J., u. a.: *Lernzielorientierte Tests*. Düsseldorf 1972, S. 23.

<sup>8</sup> Ingenkamp, K. (Hrsg.): *Tests in der Schulpraxis*. Weinheim 1972, S. 29.

Wenn wir die Einschränkung von Tests auf den Gewinn von Information betonen, mag dies den Eindruck erwecken, wir würden eine solche Messung als Selbstzweck betrachten und auf diese Weise den problematischen Verwertungsbereich ausklammern. Dies ist nicht der Fall. Wir halten es für ausgesprochen wichtig, auf die Gefahren bei der Weiterbenutzung von Testergebnissen hinzuweisen und werden darauf noch eingehen. Jedoch ist es für die Verbesserung der Situation zunächst unabdingbare Voraussetzung, zu unterscheiden, ob beobachtete Fehler (a) auf Mängeln des Tests oder (b) auf unsinniger Weiterverwendung der Ergebnisse beruhen.

### Grenzen, Mängel, Gefahren von Tests

Aufgrund seiner Konzeption, die wir kurz zu charakterisieren versucht haben, ist der Test ein Hilfsmittel mit Grenzen, die auch bei einer Perfektionierung der Testtechniken nicht überwindbar sind. Hierzu gehört:

- Tests orientieren sich an definierten Merkmalsbereichen
- Diese Merkmalsbereiche werden isoliert betrachtet
- Gemessen wird ein (erlerntes) Verhalten in konkreter Situation
- Testaufgaben werden so zusammengestellt, daß sie einen Vergleich ermöglichen (differenzieren)
- Es handelt sich um eine zeitlich punktuelle Messung
- Es wird lediglich eine Aufgabenstichprobe gegeben
- Testnormen sind Erfahrungswerte

Zieht man solche Grenzen in Betracht, so relativiert dies die Aussagemöglichkeit von Tests. Da ZEUCH unmittelbar nach dem Verwertungszusammenhang von Tests fragt, entgehen ihm diese Voraussetzungen, die den Einsatzbereich von Tests viel grundsätzlicher einschränken als seine Beispiele. Das Wissen um solche Grenzen schützt vor einer ungerechtfertigten Überschätzung der Verfahren und unrealistischen Erwartungen, denen ZEUCH gerade in seiner Kritik selbst verschiedentlich verfällt.

Unrealistische Erwartungen bestehen oft hinsichtlich der Meßgenauigkeit von Tests. So weist SIMONS unmißverständlich darauf hin, daß trotz der relativen Überlegenheit von Tests "das zu einem bestimmten Zeitpunkt erfassbare Fähigkeits- und Kenntnisniveau eines Kindes keine langfristige Prognose seines Schulerfolges erlaubt, die so sicher ist, daß man eine Schullaufbahnlenkung innerhalb des starren und weitgehend undurchlässigen dreigliedrigen Schulsystems darauf gründen könnte"<sup>9</sup>.

Zu den unausweichlichen Problemen jeder Testkonstruktion gehört auch die *Bestimmung der Validität*<sup>10</sup>. Um sich zu versichern, daß ein Test tatsächlich das mißt, was er messen soll, gibt es zwei Prüfansätze: einen empirischen und einen inhaltlich-logischen. Bei der *empirischen* Validierung wird das Testergebnis mit dem Kriteriumsverhalten, das es vorhersagen soll, verglichen (also z. B. das Ergebnis eines Schulleistetestes oder die Ei – Eierbecheraufgabe mit dem tatsächlichen Schulerfolg nach vier Jahren). Bei der *inhaltlich-logischen* Validierung werden die Aufgaben des Tests daraufhin be-

<sup>9</sup> Simons, H.: In: Studienbegleitbrief 10 zum Funkkolleg Pädagogische Psychologie, Weinheim 1973, S. 50. Vgl. auch Zielinski, W.: In: Studienbegleitbrief 12.

<sup>10</sup> Die Validitätsdefinition bei Zeuch (S. 343 / 4. Abschnitt) ist zum größten Teil („wenn seine Ergebnisse . . . lokalisiert“) ein nicht belegtes Zitat von Lienert, a. a. O., S. 16.

urteilt, ob sie repräsentativ sind für den geforderten Inhalt (so kritisiert z. B. NIEMZ<sup>11</sup> am Erdkundetest Deutschland ETD 5–7 ein von der Geographie her nicht begründendes Übergewicht des norddeutschen Raumes).

Beide Verfahren haben ihre Vor- wie Nachteile: Die empirische Validierung versichert einen tatsächlichen Zusammenhang zwischen Test und späterem Verhalten, ist aber langwierig, daher auch rigide, und überhaupt nur dann durchführbar, wenn ein operational definierbares und quantifizierbares Kriterium vorliegt. Die inhaltlich-logische Validierung dagegen läßt sich relativ kurzfristig und auch bei Inhalten durchführen, für die ein operationales Kriterium nicht unmittelbar ausgemacht werden kann, hat aber den Nachteil, daß die Schätzung der Beurteiler möglicherweise nicht mit den tatsächlichen Gegebenheiten des entsprechenden Gebietes übereinstimmen.

Um die Brauchbarkeit der auf die eine oder andere Weise erstellten Validitätsangaben abschätzen zu können, enthalten die Handanweisungen für Tests Angaben über die in (hoffentlich sorgfältig beschriebenen) Stichproben enthaltenen Werte. Dies sind, wie bereits gesagt, *Erfahrungswerte*<sup>12</sup> und nicht etwa „Hochrechnungen“ auf zukünftige Durchführungen. Hier muß der Testbenutzer selbst abschätzen, inwieweit der Test für seine Situation überhaupt aussagekräftig ist; dies erfordert, daß er entsprechend vorgebildet ist.

Wenn man Zuverlässigkeits- und Gültigkeitskoeffizienten als Erfahrungswerte betrachtet, kann es nicht verwundern, daß in unterschiedlichen Untersuchungen verschiedene Werte ermittelt werden. Hierzu braucht man nicht auf einen über 10 Jahre alten amerikanischen Beleg zu rekurrieren (S. 345, Gross); es gibt genügend Erfahrungsberichte über Testeinsatz in Deutschland<sup>13</sup>. Dabei bestätigt sich jedoch nicht, daß die festgestellten Werte „in der Regel weit niedriger (sind) als die von Testautoren selbst angegebenen“ (S. 345).

Ein perfekter Test würde also zumindest den obigen Beschränkungen unterliegen. Da Tests aber von Menschen ausgedacht werden, können bei der Herstellung auch Fehler unterlaufen, wie ZEUCH mit Beispielen illustriert. Allerdings wird man Tests mit solchen Fehlern im Gegensatz zu ZEUCH wohl kaum als „perfekt“ oder „technisch gut“ bezeichnen können. Konstruktionsfehler ergeben schlechte Tests; ZEUCHS Argumentation ist irreführend, wenn er „Mängel und Gefahren, die auch technisch gute Tests aufweisen“ (S. 341), aufzeigen will, dann aber behebbare Konstruktionsmängel aufzählt und damit den Eindruck erweckt, trotz solcher Mängel handele es sich um gute Tests.

Zu diesen behebbaren Konstruktionsfehlern gehört die *stereotype Aufgaben- und Auswertungsform*. Eine einseitige Festlegung auf einen bestimmten Aufgabentyp übersieht den Zusammenhang von Inhalt und Prüfungsform und verkürzt in unzulässiger Weise den möglichen Prüfungsbereich. Die Aufgabenformen für Testaufgaben<sup>14</sup> sind zahlreich und sollten beweglich gehandhabt werden. Mit entsprechender Auswertungsanweisung lassen sich auch Aufgaben mit ungebundener Antwortform

<sup>11</sup> Niemz, G.: Objektivierte Leistungsmessung im Erdkundeunterricht. In: Geographische Rundschau, 24. Jg. (1972), S. 102–107.

<sup>12</sup> Vgl. Ingenkamp, a. a. O., S. 69.

<sup>13</sup> Marsolek, T. / Ingenkamp, K.: Literatur über Tests im Bereich der Schule. Weinheim u. a. 1968.

Krapp, A. / Mandl, H.: Schulleistetest und Schulerfolg. München 1971.

<sup>14</sup> Rütter, Th.: Aufgabenformen für den lernzielorientierten Test. In: Dohmen, G. (Hrsg.): Forschungstechniken für die Hochschuldidaktik. München 1971, S. 93–109.

hinreichend objektiv auswerten (vgl. z. B. die Nachzeichenaufgaben im Weilburger Schulreifetest, die Oberbegriffsbildung im I-S-T von AMTAEUER). Der Hinweis auf eine mögliche Verschiebung des Inhalts durch eine unbewegliche Aufgabenformulierung ist also richtig, jedoch unter „unvollkommene Konstruktionstechnik“ einzuordnen. Unter die gleiche Kategorie läßt sich die bemängelte Sinnverschiebung bei der notwendigen Operationalisierung des Zieles (S. 344 f.) subsumieren.

Auch Kreativität läßt sich grundsätzlich mit Tests erfassen. Da sich Tests jedoch, wie wir oben ausführten, auf ein einzelnes oder auf einige wenige Merkmalsgebiete beziehen, kann man nicht erwarten, daß ein Test alles mißt. Andererseits ist kaum anzunehmen, daß Kreativität in Leistungstests „häufig negativ bewertet“ (S. 342) wird, da deren Aufgabenformulierung – wie ZEUCH selbst ausführt – zumeist eine kreative Lösung nicht zuläßt; sie wird vielmehr nicht erfaßt. Wer meint, daß die herkömmlichen Test nicht alle wichtigen Aspekte erfassen, muß die Entwicklung neuer, besserer Tests fordern. ZEUCH begnügt sich auch hier mit der Kritik.

Das sich hieran anschließende Beispiel (S. 342) schildert ein *Validitätsproblem*: Ist die Zuordnung von Ei zu Eierbecher eine Aufgabe, die schulreife von nicht-schulreifen Kindern trennt? Und was ist mit den Schülern, die mit ZEUCH die Zuordnung „Ei zu Hut“ als eine sich geradezu anbietende (die Größenordnung sei vernachlässigt) unkonventionelle Lösung verstehen? Bei dieser Aufgabe, wie sie in den Weilburger Testaufgaben und dem Reutlinger Test für Schulanfänger enthalten ist, wäre in der Tat die Aufgabeninstruktion dahingehend abzuändern, daß man fragt: Was paßt am besten zueinander? Es muß aber gesehen werden, daß solche Aufgabenbeispiele – und zwar gerade einfachster Art – zur Vorübung von Tests gehören und damit die Furcht vor „direkten“ Lösungen abgebaut wird. Außerdem besteht ein Test immer aus einem ganzen Satz von Aufgaben, so daß einzelne solche problematische Lösungen sich kaum auswirken (hier könnten sich Überlegungen zur Reliabilität anschließen). Vor allem aber werden die Einzelaufgaben und der Gesamttest vor dem endgültigen Einsatz empirisch überprüft; bei einer Testanalyse werden Aufgaben, bei denen durch Fehler in der Aufgabenformulierung, irreführende Hinweise oder Lösungsvorschläge die Leistungsbesten auf eine falsche Antwort gelenkt werden, jedoch bald entdeckt und überarbeitet oder entfernt. Gerade in dieser Erprobung ist ein großer Vorteil des Tests gegenüber der herkömmlichen Leistungsfeststellung zu sehen.

Kurz wollen wir noch auf einige andere *Gefahren bei der Konstruktion von Schulleistungstests* hinweisen:

- Überbetonung des leicht Meßbaren
- Atomisierung der pädagogischen Aufgabe in relativ sinnarme Fragmente

und *bei normorientierten Tests*:

- Überbetonung des Leistungsprinzips: „oben“ und „unten“ wird den Schülern als „natürliche“ Ordnung geboten.
- Maximierung tatsächlicher Unterschiede
- Keine Aussage über absolutes Leistungsniveau

Über die Brauchbarkeit eines Tests entscheidet also nicht allein das Ausmaß der Fehldiagnosen, sondern zusätzlich die unterrichtliche Brauchbarkeit der inhaltlichen Aussage. Hierzu empfiehlt INGENKAMP: „Ein Test sollte . . . nur entwickelt werden,

wenn er ein Schülerverhalten angemessener erfassen kann als die bisherigen Untersuchungsmethoden oder vorliegenden Tests, wenn die Überprüfung des angestrebten Lernerfolgs bedeutsam genug ist und nicht durch die Existenz eines Tests eine sonst nicht gerechtfertigte Akzentuierung erfährt und wenn die Testkonstruktion in Übereinstimmung mit didaktischen Intentionen gebracht werden kann.“<sup>15</sup>

### Die Weiterverwendung von Testergebnissen

Nicht nur bei der Testkonstruktion, sondern auch bei der Weiterverwendung der Testergebnisse sind Fehler möglich. Und hier liegt ein Problem der ZEUCHSchen Argumentation: Offenbar nimmt er an, daß die Lehrer weder bereit noch in der Lage sind, ihr subjektives Urteil relativieren zu lassen und deshalb gezielt Meßergebnisse verfälschen: „Der Einsatz von Tests verleiht der subjektiven Beurteilung des Lehrers nur den Schein der Objektivität. Er kann somit dem Lehrer zur Verschleierung oder Rechtfertigung seines subjektiven Urteils vor Eltern, Schülern und vor sich selbst dienen“ (S. 348, ähnlich S. 343). Mit einem solchen Lehrerbild wäre es logischer, die Abschaffung der gesamten Schule zu fordern, als nur so ein unbedeutendes Hilfsmittel wie den Test zu geißeln, vor allem, wenn man zusätzlich davon ausgeht, daß „die Schule meist gar keine materiellen Möglichkeiten besitzt, wirksame therapeutische Maßnahmen gegen erkannte Schwächen einzusetzen“ (S. 341). Zudem geht der erhobene Vorwurf am Test vorbei: Die Kritik trifft nicht den Test, sondern die mangelnde Objektivität des Lehrerurteils.

Das Bemühen, subjektive Einflüsse auf die Durchführung und Auswertung von Tests auszuschließen, wird keineswegs dadurch fragwürdig – wie ZEUCH meint (S. 343) –, daß ein festgestellter Ist-Zustand anschließend auf seine Aussagekraft hinterfragt wird („Würden Besonderheiten in der Testsituation beobachtet? Wie groß ist der Meßfehler?“) und dann – möglicherweise nach Heranziehen weiterer Überlegungen – zu einer Entscheidung führt, in die auch subjektive Erfahrungen und Wertungen eingehen<sup>16</sup>. Denn es ist einleuchtend, daß die Zuverlässigkeit objektiver Daten auch bei subjektiver Verwertung noch höher sein dürfte, als wenn von vorne herein nur subjektive Daten herangezogen werden. Auf jeden Fall irrt ZEUCH, wenn er meint, daß nicht kontrollierbar wäre, in welchem Maß ein (objektives) Testergebnis für die (subjektive) Gesamtzensur herangezogen wird (S. 348): Eine einfache Korrelationsberechnung zeigt diesen Zusammenhang.

Um den Unsicherheitsfaktor in der Weiterverwendung von Testergebnissen so gering wie möglich zu halten, kann man die Kriterien der Objektivität, Reliabilität und Validität auch für diesen Schritt fordern. Diese Forderung kann nicht durch Manipulationen am Test erfüllt werden; hierzu muß der Lehrer einerseits das Meßergebnis hinterfragen können, andererseits müssen zusätzliche Beobachtungs- und Schätzverfahren entwickelt und mit den Lehrern trainiert werden. Die Qualität der Verwendung von Testergebnissen ist also von der Kompetenz der Lehrer abhängig, die eine solche Aufgabe zu erfüllen haben.

Daß ZEUCH die Inkompetenz des Testanwenders als den Normalfall annimmt, geht auch aus der Ausgangsüberlegung hervor, es sei für die Beurteilung eines Tests nicht nebensächlich, „in welchem Maße er eine mißbräuchliche Anwendung ermöglicht

<sup>15</sup> a. a. O., S. 44.

<sup>16</sup> Auf das Problem des Zusammenhangs dieser beiden Schritte weist H. Hiltmann (Kompendium des psychodiagnostischen Tests, Stuttgart 1966, S. 23) hin; das Zitat bei Zeuch (S. 343 – ohne Auflage und Seitenzahl) wirkt deshalb sinnentstellend, weil er den im Original enthaltenen Kursivdruck der Begriffe „Diagnostik“ und „Messung“ nicht übernommen hat.

oder sogar provoziert“ (S. 340). Ein Test „provoziert“ dann den Mißbrauch, wenn die Angaben über sein Zustandekommen und seine Erprobung fehlen (nach unseren obigen Überlegungen wäre ein solcher Test mangelhaft<sup>17</sup>), aber auch, wenn ein unkritischer Benutzer trotz dieser Angaben nicht in der Lage ist, die *Grenzen der Testaussage* zu realisieren. Im ersten Fall wäre das Testbeihft neu zu fassen oder auf den Test zu verzichten, im zweiten die Qualifikation des Benutzers zu heben oder ihm von jeder Testbenutzung abzuraten. ZEUCH empfiehlt jeweils das Zweite, unschiene das Erste sinnvoller. Wir würden also meinen, daß Tests nicht per se eine mißbräuchliche Anwendung provozieren, sondern daß dies an der Qualität des Tests und der Qualifikation des Benutzers liegt. Und beides ist einer Verbesserung zugänglich.

Für unsere Überlegungen, wie der Testeinsatz in der Schule verbessert werden kann, bewährt sich also die bei ZEUCH vernachlässigte Trennung von Messung und Bewertung. Damit verstehen wir auch eine Reihe von Vorwürfen bei ZEUCH als Fragen an die Kompetenz des Anwenders, nicht an den Test: Wenn Tests zur Einstufung, Selektion oder Disziplinierung eingesetzt werden (S. 341), wenn Berufseignungstests der Ausbeutung dienen (S. 346), wenn ZEUCH beklagt, daß Schulleistung unabhängig von der Lerngeschichte beurteilt werde (S. 347). Das Vorhandensein eines Tests genügt noch nicht zur Rechtfertigung seines Einsatzes. Nur wenn sich der Anwender über die pädagogische Bedeutung, die didaktische Angemessenheit, die diagnostische Funktion und die möglichen Konsequenzen im klaren ist, wird der Testeinsatz ein sinnvolles Element im Unterricht sein.

### Der Vorwurf der Systemstabilisierung

Ob eine Testanwendung eher der Institution oder eher dem Individuum nützt, wird nach diesen Überlegungen also nur zum Teil vom Test bestimmt; die endgültige Determination liegt beim Verwender. Wenn Tests zum Zweck der Einstufung angewendet werden (z. B. Zuordnung zu einem Stützkurs, Leistungsgruppe, Schularart), dann orientieren sie sich an den bestehenden Institutionen – auch an ihren Mängeln – und den dort vorgefundenen Verhaltensvariationen. ZEUCH verweist in diesem Zusammenhang darauf, daß die sozialen und kulturellen Bedingungen der Gesellschaft wesentlicher Bestandteil der Leistungstests seien<sup>18</sup>. Tests tragen demnach in vielen Fällen zur Systemstabilisierung bei. Dies ist jedoch keine grundsätzliche Aussage: Die Überlegungen zur Validität haben gezeigt, daß Testinhalte nicht nur empirisch, sondern auch durch inhaltlichen Konsens als verbindlich anerkannt werden können. Damit ist es grundsätzlich möglich, Tests für einen gerade erst erfundenen Schultyp zu konstruieren.

Kaum eine Schule kommt ohne *Zuordnungs- und Selektionsentscheidungen* aus. Wie aber werden diese Entscheidungen in der Praxis gefällt? Unabhängig von der Existenz von Tests orientieren sich Lehrer wohl in den meisten Fällen an den vorhandenen Einrichtungen. Systemstabilisierung durch Selektion gibt es also schon vor den Tests. Kritik an unerwünschten Eigenschaften dieser Einrichtungen und Innovationsvorschläge müssen dann an der Einrichtung selbst ansetzen; wenn man statt dessen die Tests kritisiert, schlägt man den Sack, statt den Esel! Es erscheint uns daher zumin-

<sup>17</sup> Vgl. Belsler, H.: Wie beurteilt man Schultests? In: Ingenkamp, a. a. O., S. 135–145.

<sup>18</sup> S. 345. Dabei (vorletzter Absatz „empirische Normen . . . Leistungstest ist“) handelt es sich mit Ausnahme der Hervorhebungen um ein nicht belegtes Zitat von Hiltmann, a. a. O., S. 21.

dest einseitig gesehen, in diesem Zusammenhang Tests als Loyalitätsuntersuchungen zu qualifizieren.

Durch eine Innovation werden Tests möglicherweise überholt – vielleicht nur für einzelne Klassen, Schularten oder aber im ganzen. Schultests müssen deshalb immer wieder auf ihre Übereinstimmung mit Schule und Curriculum geprüft werden. In dieser Überprüfbarkeit und Kontrollierbarkeit von Tests sehen wir einen weiteren bedeutsamen Vorteil dieses Hilfsmittels. Vermutlich ist der Test durch diese Eigenschaften eher einer Systemüberwindung zugänglich als das Lehrerurteil.

### Was ist die Alternative?

Unser Hauptvorwurf an ZEUCH besteht darin, daß er sich damit begnügt, den pädagogischen Zeigefinger gegen den Test zu erheben. Alternativen zeigt er leider nicht auf.

Als Alternative zum Test ist entweder *der Verzicht auf Messungen oder die Lehrerschätzung* denkbar: Im Leistungsbereich kann die Leistungsfeststellung dann ersatzlos gestrichen werden, wenn die Überlegung, wozu ein Leistungsvergleich gebraucht wird und welcher Funktion er dienen soll, zu einer negativen Antwort führt<sup>19</sup>. Jedoch dürften für die tägliche Schularbeit (von der Unterrichtsforschung ganz zu schweigen) genügend Bereiche bleiben, wo auf eine Leistungsfeststellung nicht verzichtet werden kann. Die Alternative wäre hier die herkömmliche Zensierung mit Hilfe von Klassenarbeiten und mündlichen Prüfungen. Diese Verfahren weisen jedoch – wie vielfach nachgewiesen<sup>20</sup> – ganz erhebliche Mängel auf; gerade bei der Zensur kommen die Befürchtungen, die ZEUCH gegen den Test hegt, voll zur Geltung.

Meßverfahren für nicht-kognitive Bereiche kommen derzeit in der Schule kaum zum Einsatz. Motivation, Interesse, Angstlichkeit werden, wenn sie überhaupt Beachtung finden, vom Lehrer subjektiv eingeschätzt. Auch hier läßt sich vermuten, daß diese Schätzverfahren wesentlich unzuverlässiger sind als Tests, da Vorurteile, selektive Beobachtungen und Stereotypen das Lehrerurteil verfälschen<sup>21</sup>.

Bei aller Problematik stellt demnach der Test dort, wo es auf eine möglichst genaue und zuverlässige Aussage ankommt, das derzeit beste Hilfsmittel zum Gewinn von Information dar. Dennoch kann das nicht einen Verzicht auf das Lehrerurteil bedeuten. In der Entwicklung besserer Schätz- und Beobachtungstechniken für den Lehrer liegt ein noch kaum angegangenes Problem der pädagogischen Diagnostik. Diese Verfahren sollen sich jedoch ergänzen, nicht ausschließen.

### Aktuelles Meßgebiet: Unterrichtsevaluation

Wir haben uns bisher der Beschränkung ZEUCHS (S. 341) angeschlossen und uns mit der Einstufungsfunktion von Tests befaßt. Es muß sicher zugegeben werden, daß die Aufgabe der objektiven Leistungsmessung noch vor zehn Jahren fast ausschließlich in der Differenzierung von unterschiedlich leistungsfähigen Schülern gesehen wurde; inzwischen hat sich die Diskussion ganz entschieden verlagert. Mit der Forderung nach kriterien- und lernzielbezogenen Tests, der Operationalisierung von Lehrzielen

<sup>19</sup> Vgl. z. B. die Überlegungen zum Untersuchungsziel bei Ingenkamp, a. a. O., S. 44 f.

<sup>20</sup> Ingenkamp, K. (Hrsg.): Die Fragwürdigkeit der Zensurengabung. Weinheim 1971 (Bibl.).

<sup>21</sup> Vgl. Höhn, E.: Der schlechte Schüler. München 1967.

und ihrer taxonomischen Strukturierung sowie Anweisungen zu informellen Verfahren ist die *Funktion von Tests zur Curriculum- und Unterrichtsevaluation* zunehmend in den Mittelpunkt der pädagogisch-psychologischen Testdiskussion getreten.

In diesem Zusammenhang dient der Test nicht zur Feststellung der „Qualität“ von Schülern, sondern zur Kontrolle von Unterricht. Eine solche Kontrolle wird nötig, wenn zur Bewertung unterrichtlicher Arrangements nicht allein die proklamierten, sondern ihre realisierten Ziele herangezogen werden sollen. Die Kontrolle, ob ein gewünschtes Unterrichtsziel auch tatsächlich erreicht wurde, unter welcher Bedingung und wie dauerhaft dies geschah, wurde in den Unterrichtstheorien in Deutschland lange Zeit zugunsten der Planung und Vorbereitung von Unterricht vernachlässigt. Erst im Didaktikmodell der Kybernetik und der Berliner Schule sowie mit der Curriculumsdiskussion wird die Überprüfung des Unterrichtserfolges zur Steuerung weiteren Unterrichts als integrierter Bestandteil des Unterrichts verstanden. Unter diesem Aspekt haben Prüfungen keinen Selbstzweck, sondern dienen dazu, den Unterrichtsprozeß erfolgreicher zu gestalten.

Wir müssen uns in diesem Rahmen mit einem kurzen Verweis auf dieses aktuelle und auch von ZEUCH am Schluß erwähnte Anwendungsgebiet von Schultests begnügen. Die Auseinandersetzung mit diesem Gebiet würde für die Praxis sicherlich erheblich positivere Anregungen bringen als die Kritik der Auslesefunktion. Voraussetzung auch für den Einsatz von Tests zu Evaluationszwecken ist jedoch, daß die Lehrer die Möglichkeiten, aber auch die Grenzen von Tests kennen.

### Schluß

Wir akzeptieren durchaus die Beispiele, mit denen ZEUCH die Mängel<sup>22</sup> und Gefahren bei der Anwendung von Tests illustriert, sehen allerdings die Ursachenzusammenhänge anders: Die meisten Vorwürfe beziehen sich auf Gefahren, die durch Fehler bei der Konstruktion und Weiterverwendung von Tests den Zweck des Testeinsatzes verfehlen lassen. Im Gegensatz zu ZEUCH halten wir es für grundsätzlich möglich, daß ein Test vom Benutzer analysiert, hinterfragt und auf seine Brauchbarkeit geprüft werden kann. Dies führt zu der unterschiedlichen Bewertung von Testverfahren bei ZEUCH und bei uns: Wir sehen für Diagnose, Prognose und Selektion, für Evaluations- und Forschungszwecke derzeit trotz aller Mängel, Gefahren und Einschränkungen keine realisierbare Alternative, die den Test überflüssig machen könnte.

Für einen sachgerechten und grenzbewußten Einsatz ist es jedoch notwendig, daß die Lehrer lernen, mit diesen Verfahren umzugehen. Wir glauben jedoch, daß unsere Schulen und Lehrer nicht so „systemstabil“ sind, sich die für einen besseren Unterricht unersetzliche Information auf die Dauer entgehen zu lassen. Wir schließen uns daher THORNDIKE/HAGEN an: „We must grant that our measurement procedures are not complete and our actions based on them are not always wise. But the remedy lies in developing better measurement procedures that will give us more complete and more accurate information about the individual. It lies in gaining better understanding of our measures – their strengths and their weakness – so that we may use them with more wisdom. It does *not* lie in getting less information.“<sup>23</sup>

<sup>22</sup> Ausgenommen die Ausführungen zu projektiven Verfahren, die uns mehr Vorurteil statt Urteil zu sein scheinen!

<sup>23</sup> a. a. O., S. 13.