

Welche Gütekriterien sind bei der Bewertung von Lernleistungen wichtig?

Anne Spensberger; Ramona Dutschke; überarbeitet von Susanne Narciss

Eine gerechte Bewertung von Lernleistungen setzt voraus, dass beim Erfassen, Messen und Bewerten der Lernleistungen psychologische Erkenntnisse zu den **testtheoretischen Gütekriterien** berücksichtigt werden.

Nach Lienert und Raatz (1984) unterscheidet man hierbei die folgenden **Haupt- und Nebengütekriterien**:

Hauptgütekriterien	Nebengütekriterien
<ul style="list-style-type: none">• Objektivität• Reliabilität• Validität	<ul style="list-style-type: none">• Ökonomie• Nützlichkeit• Normierung• Vergleichbarkeit

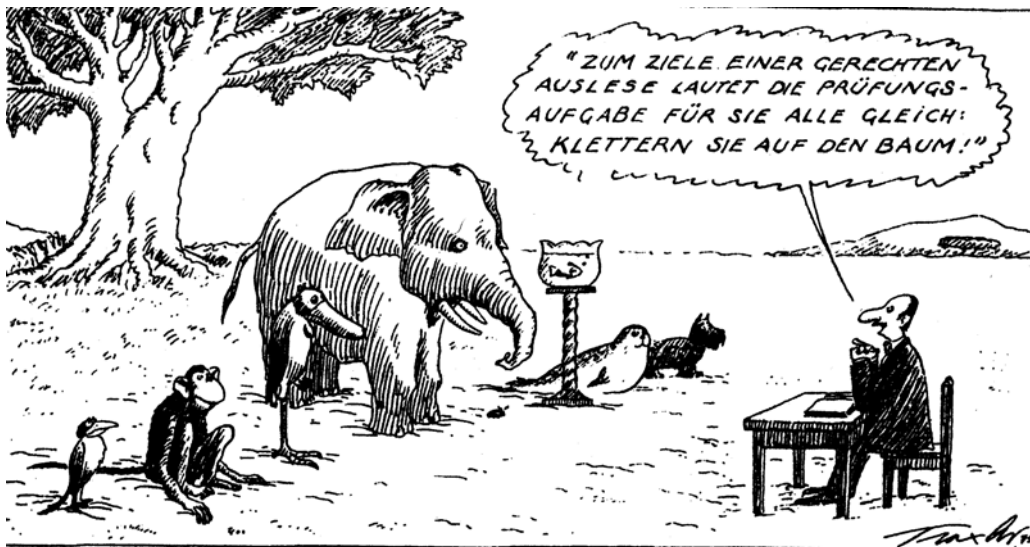
(Vgl. Raatz, L., S.11).

Die drei **Hauptgütekriterien** stehen in einem engen Zusammenhang. Die Objektivität ist eine notwendige Voraussetzung für die Reliabilität diese eine notwendige Voraussetzung für die Validität einer Bewertung. Mit anderen Worten: Ist ein Verfahren nicht objektiv, kann es auch nicht reliabel oder valide sein. Objektivität und Reliabilität sind also notwendige, aber nicht hinreichende Voraussetzungen für die Validität einer Bewertung.



Objektivität

Was ist gerecht?



Rudolf WILKE 1907 im "Simplizissimus"

Die **Objektivität** ist ein Hauptgütekriterium für die Messung der Qualität eines Testes.

Ein Test ist dann objektiv, wenn die Testergebnisse unabhängig sind von der Person oder den Personen, die den Test durchführen, auswerten und interpretieren.

Um **Objektivität** zu erreichen muss man bei der **Durchführung** von (Leistungs-)Tests, der **Auswertung** und der **Interpretation** der Testergebnisse möglichst standardisiert, d.h. nach vorher fest gelegten und begründeten Standards vorgehen. Man kann demnach drei Arbeitsebenen für das Erreichen objektiver Ergebnisse unterscheiden:

Durchführungsobjektivität	Alle Lernenden bearbeiten dieselben Testaufgaben unter gleichen Bedingungen
Auswertungsobjektivität	Die bearbeiteten Aufgaben werden nach denselben, vorher festgelegten und begründeten Kriterien ausgewertet.
Interpretationsobjektivität	Die Zuordnung der bei der Auswertung festgestellten Leistungen (z.B. richtige Lösungen, Anzahl von Fehlern, Punkte) zu Leistungswerten (z.B. Noten) erfolgt nach vorher eindeutig festgelegten Klassifizierungsregeln.

Bei der Bewertung von Lernleistungen im Unterricht wird auf vielfältige Weise gegen diese Objektivitäts-Prinzipien verstoßen. Dies liegt unter anderem daran, dass es weder eindeutige Auswertungsrichtlinien, noch eindeutige Regeln oder Vorschriften für Zuordnungen von Leistungen zu Noten gibt. Ein besonderes Problem ist die Neigung vieler Pädagogen, die Messung der Schülerleistungen nicht konsequent genug von deren Wertung und auch Interpretation zu trennen.

Reliabilität oder Zuverlässigkeit

"Heute hatte ich einen wundervollen Traum: Ich gab Cicero eine Fünf in Latein."



(R. Wilke, 1907)

Rudolf WILKE 1907 im "Simplizissimus"

Ein Test ist dann **reliabel**, wenn er **zuverlässig** und **genau** misst.

Eine reliable Messung von Lernleistungen muss, wenn sie unter denselben Bedingungen, an denselben Aufgaben und mit denselben Personen durchgeführt wird, zu möglichst identischen Ergebnissen kommen.

Beim testtheoretischen Konzept der Reliabilität wird davon ausgegangen, dass bei jeder Messung auch gewisse Messfehler auftreten. Messwerte enthalten daher aus der Sicht der klassischen Testtheorie immer einen wahren Varianz-Anteil und einen Fehlervarianz-Anteil. Der Quotient zwischen der wahren Varianz und der beobachteten Varianz (= wahre Varianz + Messfehler-Varianz) wird als **Reliabilitätskoeffizient** bezeichnet.

Um die Reliabilität eines Tests zu untersuchen werden im Rahmen der klassischen Test-Theorie die folgenden Methoden angewandt (vgl. Rost, 1996):

- **Paralleltest-Methode**
- **Retest-Methode**
- **Halbtest-Methode**
- **Cronbachs Methode der internen Konsistenz-Bestimmung**

Paralleltestmethode Paralleltestreliabilität	Die Testergebnisse eines Tests werden verglichen, d.h. korreliert mit den Ergebnissen eines parallel konstruierten Tests, von dem man annimmt, dass er dieselben Merkmale misst.
Retest-Methode Retest-Reliabilität	Derselbe Test wird nach einem zeitlichen Abstand bei denselben Personen durchgeführt. Man korreliert die Testergebnisse der beiden Messzeitpunkte. Dieses Verfahren ist nur bei der Messung relativ stabiler Merkmale sinnvoll.
Halbtest-Methode Halbierungsreliabilität	Die Aufgaben eines Tests werden in zwei Gruppen aufgeteilt (z.B. alle Aufgaben mit geraden Nummern ergeben Testhälfte A, alle Aufgaben mit ungeraden Nummern Testhälfte B). Man korreliert die Testergebnisse der beiden Testhälften. Dieses Verfahren ist dann sinnvoll, wenn die Aufgaben dasselbe Merkmal messen.
Interne Konsistenz	Jede Aufgabenantwort eines Tests wird als Messwert betrachtet. Man korreliert alle Messwerte miteinander. Dieses Verfahren ist dann sinnvoll, wenn die Aufgaben dasselbe Merkmal messen.

Validität

Das Anlegen von Maßstäben.



Die **Validität** oder **Gültigkeit** ist das dritte Hauptkriterium.

Ein Test ist dann valide, wenn er das misst, was er zu messen vorgibt.

Man unterscheidet die folgenden Arten der Validität:

- **Inhalts- oder Kriteriumsvalidität**
- **Übereinstimmungsvalidität,**
- **Vorhersagevalidität**
- **Konstruktvalidität.**

Inhaltsvalidität Kriteriumsvalidität	Das Testergebnis erlaubt Schlussfolgerungen darüber, wie gut vorher festgelegte, inhaltlich begründete Kriterien erfüllt sind.
Übereinstimmungsvalidität Externe Validität	Das Testergebnis stimmt mit einem Außenkriterium überein, z.B. mit dem Ergebnis eines anderen Tests, von dem man annimmt, dass er dasselbe Merkmal misst.
Vorhersagevalidität	Das Testergebnis lässt Vorhersagen auf künftige Ereignisse zu.
Konstruktvalidität	Das Testergebnis stimmt mit den theoretischen Vorhersagen zu dem zu messenden Konstrukt überein.

Bei der Bewertung von Lernleistungen in der Schule geht es vor allem darum, die kriteriumsorientierte Validität oder Inhaltsvalidität anzustreben (vgl. Klauer, 1987).

Die Nebengütekriterien

Grundsätzlich gibt es für die Nebengütekriterien keinen zahlenmäßigen Kennwert.

Unter der **Normierung** eines Testes versteht man, „dass über einen Test Angaben vorliegen sollen, die als Bezugssystem für die Einordnung des individuellen Testergebnisses dienen können“

(Raatz, L., S. 11).

Ein Test, der zwar die Hauptgütekriterien erfüllt, jedoch nicht normiert ist, besitzt keine oder nur eine sehr geringe Brauchbarkeit.

Vergleichbar ist ein Test, wenn ein oder mehrere Paralleltestformen vorhanden sind und wenn validitätsähnliche Tests verfügbar sind.

Ersteres ermöglicht sozusagen einen Vergleich des Testes mit sich selbst. Zweiteres prüft dasselbe oder ein nahe verwandtes Persönlichkeitsmerkmal.

Die **Ökonomie** eines Testes ergibt sich aus folgenden Punkten: Er ist ökonomisch, wenn er

- eine kurze Durchführungszeit beansprucht
- wenig Material verbraucht
- einfach zu handhaben ist
- als Gruppentest durchführbar ist
- schnell und bequem auszuwerten ist.

Sehr ökonomisch ist ein Test also, wenn er alle oder die wichtigsten dieser Punkte erfüllt.

Nützlich ist ein Test, wenn er eine Verhaltensweise oder ein Persönlichkeitsmerkmal misst oder vorhersagt, für dessen Untersuchung ein praktisches Bedürfnis besteht.

Eine hohe Nützlichkeit entsteht, wenn ein Test in seiner Funktion durch keinen anderen Test vertreten werden kann.

Literatur:

Ingenkamp, K.(1959). Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte. München: PsychologieVerlagsUnion.

Lienert, Gustav A. (1989). Testaufbau und Testanalyse. München: PsychologieVerlagsUnion.

Lienert R. (1994). Testaufbau und Testanalyse. Beltz Psychologie Verlags Union.

Roth, E. (Hrsg)(1993). Sozialwissenschaftliche Methoden. Oldenbourg Verlag München.

Knörzer, W./ Grass, K. (1994): Den Anfang der Schulzeit pädagogisch gestalten. Ein Studien- und Arbeitsbuch für den Anfangsunterricht. Beltz Verlag.

<http://www.stangl-taller.at/TESTEXPERIMENT/testguetekriterien.html>

<http://www.schule.suedtirol.it/blick/angebote/reformpaedagogik/rp55075.htm>

<http://bebis.cidsnet.de/weiterbildung/sps/allgemein/bausteine/beurteilung/objektiv.htm>

http://www.phil.uni-sb.de/~jakobs/paedpsych/noten/guetekriterien_von_noten.html