

Was macht eine gute Prüfung aus – Gütekriterien

Prüfen ist eine Art von Messen oder Testen. Um die Qualität einer Prüfung sicherzustellen, ist es hilfreich, verschiedene Gütekriterien, die aus der psychologischen Testtheorie hervorgehen, zu berücksichtigen.

1. OBJEKTIVITÄT

Definition: Eine Prüfung ist objektiv, wenn die Ergebnisse (Punkte und Benotung) unabhängig vom Prüfenden sind. Wir unterscheiden zwischen der Objektivität der Durchführung, der Bewertung/Benotung und der Interpretation der Ergebnisse.

Beispiel: Sogar ein kurzes Lächeln oder die Namen der Studierenden (Kevin wird wahrscheinlich schlechter benotet als Christian) können Einfluss auf das Ergebnis einer Prüfung haben!

2. RELIABILITÄT

Definition: Grad, in dem die Prüfung eine Kompetenz/Leistung genau (ohne Fehler) „misst“. Je kleiner der Messfehler, desto höher die Zuverlässigkeit! Mit anderen Worten: Wenn eine Prüfung reliabel ist, hat sie ein hohes Maß an Messgenauigkeit.

Beispiel: Wenn alle Fragen einer Prüfung entweder einfach oder schwierig sind, zeigen die Ergebnisse der Prüfung nicht den wirklichen Unterschied in den Fähigkeiten der Studierenden. Die Zuverlässigkeit ist dann gering. Wenn die Fragen schlecht formuliert oder sehr kompliziert sind oder ungewöhnliche Begriffe verwenden, provozieren sie Missverständnisse und Fehler, die die Messgenauigkeit verringern.

3. VALIDITÄT

Definition: Eine Bewertung ist valide, wenn sie das bewertet, was sie bewerten soll. In der Regel wird dies durch die angestrebten Lernergebnisse spezifiziert.

Beispiel: Sie haben eine reduzierte oder geringe Gültigkeit/Validität, wenn zum Beispiel ...

- fünf angestrebte Lernergebnisse in einer Vorlesung behandelt, aber nur vier bewertet werden.
- die Studierenden lernen sollen, eine Bilanz zu analysieren, und nur gebeten werden, Definitionen zu nennen.
- Fragen sehr kompliziert formuliert sind. Dann beurteilen Sie wahrscheinlich die Lesekompetenz und das Allgemeinwissen und nicht unbedingt das angestrebte Lernergebnis.

Die ersten drei Kriterien bauen aufeinander auf. Eine Prüfung kann nicht valide sein, wenn sie nicht reliabel ist, und nicht reliabel, wenn sie nicht einigermaßen objektiv ist. Letztlich ist die Validität das wichtigste Kriterium. Wenn eine Prüfung nicht die gewünschten Kompetenzen bewertet, ist sie nicht nur wertlos, sondern kann sogar schädlich sein, wenn Entscheidungen auf ihr basieren.

4. FAIRNESS

Definition: Alle Studierenden haben die gleiche Chance, ihre Kompetenzen zu zeigen und ein entsprechendes Ergebnis zu erzielen. Es gibt keine systematische Diskriminierung bestimmter Personen oder Gruppen. Dieses Kriterium steht in engem Zusammenhang mit dem Kriterium der Objektivität, aber in diesem Fall liegt der Schwerpunkt auf der Seite der Studierenden und auf der Gestaltung der Prüfung.

Beispiel: Wenn Sie Studierende aus dem Ausland haben und Fragen mit einem ihnen unbekannten lokalen Kontext formulieren, ist die Fairness gefährdet. Im Falle einer Behinderung, die dazu führt, dass der/die Studierende nicht in der Lage ist, die Prüfung wie vorgeschrieben abzulegen, kann ein Nachteilsausgleich gewährt werden, z.B. eine Schreibzeitverlängerung, um die Fairness zu gewähren.

5. ÖKONOMIE (AUFWAND VS. NUTZEN)

Definition: Der gesamte Prüfprozess wird effizient mit einem günstigen Verhältnis zwischen Aufwand und Nutzen (Qualität) gestaltet. Die Erstellung und Benotung von Prüfungen fordern in der Regel die meisten Ressourcen in Bezug auf Zeit oder Personen, gelegentlich auch die Durchführung, z.B. bei mündlichen Prüfungen.

Beispiel: Wenn Sie eine große Studierendengruppe zu prüfen haben, wäre es ziemlich zeitaufwändig, offene Fragen zu verwenden. Effizienz könnte bedeuten, mehr Zeit in die Konstruktion guter Fragen und Musterlösungen zu investieren. Es besteht ein Zielkonflikt zwischen Validität (insbesondere bei höheren Stufen der Lerntaxonomie) und Ökonomie. Aber wenn Sie eine Prüfung mit ein paar hundert Studierenden haben, ...

GÜTEKRITERIEN IM BEWERTUNGSPROZESS

Die beschriebenen Gütekriterien beziehen sich auf die Prüfung als Ganzes. Auf den ersten Blick könnte man vermuten, dass bei der Bewertung von Prüfungen nur die Objektivität von besonderer Bedeutung ist. Dem ist aber nicht so.

So kann ein zu grober Notenschlüssel beispielsweise die Messgenauigkeit verringern und damit die Reliabilität.

Die Validität der Prüfung ist gefährdet, wenn während der Bewertung andere Kompetenzen wie z.B. Textverständnis, saubere Schrift oder sprachliche Flüssigkeit von der eigentlich zu prüfenden inhaltlichen Kompetenz ablenken.

Zudem kann es sein, dass ich die zugrundeliegende Kompetenz nicht umfassend und sauber erhebe, da die Bewertung später zu aufwändig wäre. Dann beeinflusst die spätere Bewertung schon während der Prüfungserstellung die Validität der Prüfung. Was hier vorliegt, ist ein Konflikt zwischen den beiden Gütekriterien Validität (die Qualität der aus der Prüfung resultierenden Informationen) und Ökonomie (der zur Durchführung der Prüfung nötige (Zeit-)Aufwand). Leider ist es in der Regel nicht möglich, sowohl die Qualität als auch den Aufwand zu optimieren, sodass Sie als Lehrende ein passendes Gleichgewicht zwischen beiden finden müssen. Je aussagekräftiger die Prüfung wird, desto zeitaufwändiger wird es auch, sie zu entwickeln, durchzuführen und zu bewerten. Dieser Balanceakt ist eher eine Kunst als eine Wissenschaft, aber das Verständnis der Faktoren trägt zu besseren Entscheidungen bei.

Je komplexer oder kreativer die zu erbringende Leistung der Studierenden, desto schwerer wird es, objektiv zu bewerten. Bei sehr standardisierten Prüfungen (z.B. Multiple-Choice Fragen) oder einem eng umrissenen Antwortraum, wird der Einfluss des Prüfenden bei der Bewertung schon durch die Prüfungsart reduziert. Können die Antworten aber nicht mehr nach dem Schema wahr oder falsch bewertet werden, wird der Interpretationsspielraum des Bewertenden größer und Urteilsverzerrungen spielen eine größere Rolle. Dies ist insbesondere bei höheren Taxonomiestufen (Anderson & Krathwohl, 2001) der Fall - insbesondere bei kreativen, komplexen oder praktischen Leistungen. Hier steht der Versuch zur Standardisierung in Durchführung und Auswertung mit dem Interesse, aussagekräftige, holistische Ergebnisse zu erhalten, im Widerspruch.

Quelle:

Anderson, L.W. u. Krathwohl, D.R. (2001). *A Taxonomy of Learning, Teaching and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. Addison Wesley Longman.