

## TESTVALIDITÄT IM UNTERRICHT: EINE CHECKLISTE FÜR REFLEKTIERTE(RE) BEURTEILUNGEN

Cosa valutare? E come? Per rispondere a queste domande in modo esaustivo l'autrice mette a disposizione una tabella pronta ad esser consultata per compiti in classe, ma anche e soprattutto per test con maggiore visibilità come esami d'ammissione o di maturità. Oltre a ciò, il presente contributo traccia brevemente un quadro dello sviluppo storico della discussione sulla 'validità' di un esame ed espone quale sia lo stato attuale della discussione sulla validità di un test – attenta alla situazione in cui si svolge, alle competenze delle persone testate e a criteri di qualità come la sua usabilità e la sua attendibilità.

Von den vielen Tests in meiner Schulkarriere ist mir eine Geschichtsprüfung in Erinnerung geblieben, die ich in der 6. Klasse abgelegt habe: Völlig unerwartet bestand ich die Prüfung nur knapp und war damit noch unter den besten meiner ansonsten recht erfolgsverwöhnten Klassenkameraden. Die Prüfung war von einer jungen Praktikantin konzipiert worden, die offensichtlich vorausgesetzt hatte, dass wir nicht nur im Unterricht aufgepasst, sondern auch die entsprechenden Seiten im Lehrbuch gelesen hatten. Damals hielten wir diese Erwartung für unfair und protestierten leidenschaftlich. Wir hatten, wie wohl die meisten Schülerinnen und Schüler, erwartet, dass in der Prüfung nur Inhalte beurteilt würden, die wir zuvor im Unterricht von der Lehrperson vermittelt bekommen hatten. Heute weiss ich, dass auch eine andere Lesart möglich ist: Für die junge Praktikantin war das selbstständige Nachlesen der Unterrichtsinhalte wohl schlicht ein integraler Bestandteil der Lernziele des Geschichtsunterrichts und konnte damit geprüft werden. Was auch immer zutrifft – die Anekdote ist vor allem deshalb interessant, weil

sie gut illustriert, warum Testvalidität relevant ist: Beurteilungen sollten gut überlegt, geplant und kommuniziert werden, um ihre Rolle in Unterricht, Schule und Gesellschaft erfüllen zu können. Der vorliegende Artikel soll dazu einen Beitrag leisten: Eine Checkliste soll dazu anregen, die Ziele und Eigenschaften von Beurteilungsinstrumenten und Tests systematisch(er) zu erfassen und deren Schwächen (besser) zu erkennen. Einige kurze Informationen zum Begriff der Testvalidität sollen dann einen Einblick in die Forschung geben und Ausgangspunkt für weitere Lektüren sein. Die Checkliste orientiert sich inhaltlich am *Assessment Use Argument* von Bachman & Palmer (2010). Die darin enthaltenen Kriterien wurden übersetzt und für den schulischen Kontext neu strukturiert. Fünf konkrete Fragen sollen dabei helfen, eine Beurteilung in ihrer Komplexität zu erfassen. So könnte man ausgehend von einem Beurteilungsziel überlegen, wie dieses am besten zu erreichen ist (z.B. eine Prüfung am Ende einer Unterrichtseinheit über Mark Twains *Tom Sawyer* oder eine Übersicht über die

Katharina Karges | ●  
Fribourg



Katharina Karges ist Wissenschaftliche Mitarbeiterin am Kompetenzzentrum für Mehrsprachigkeit in Fribourg/Freiburg. Sie arbeitet an Forschungsprojekten u.a. zu den Themen kompetenzorientierte Beurteilung, computerbasiertes Testen, Entwicklung fremdsprachlicher Kompetenzen.

Die Autorin dankt Malgorzata Barras und Peter Lenz für ihre unendlichen Ideen, Patrick Karges und Katrin Berger für ihren gesunden Menschenverstand und Mathias Picenoni für das Auf-den-Punkt-Bringen.

Ein Sprachtest ist in dem Ausmass valide, wie es ihm gelingt, Sprachkompetenzen im Hinblick auf einen bestimmten Verwendungszweck zuverlässig, fair und inhaltlich aussagekräftig zu messen.

mündlichen Fertigkeiten der Lernenden zu Beginn eines Schuljahres). In diesem Fall dienen die ersten drei Fragen dazu, das Ziel klarer zu formulieren und zu begründen (geht es in der Prüfung über *Tom Sawyer* zum Beispiel darum, etwas über den Autor, das Buch oder die Figur zu wissen, die Erzählung zu interpretieren oder die Geschichte in der Fremdsprache nachzuerzählen). Die Fragen zu Form und Inhalt des Tests können dann dabei helfen, diese Ziele bei der konkreten Umsetzung der Beurteilung nicht aus den Augen zu verlieren. Umgekehrt sollen die Fragen aber auch dazu dienen, bestehende Beurteilungsinstrumente kritisch zu reflektieren, indem z.B. gefragt wird, ob ihr Einsatz in einer konkreten Lernsituation angemessen ist, welche Anpassungen nö-

tig wären oder ob es alternative Formen gibt, um die gewünschten Ziele zu erreichen. Dabei ist es nur schwer möglich, allgemeingültige Regeln aufzustellen, denen man nur noch folgen muss. Deshalb kann es in der Checkliste auch nicht darum gehen, möglichst viele Punkte anzukreuzen. Vielmehr muss der Zusammenhang zwischen der Beurteilung, ihren Zielen und ihren Konsequenzen für jede Anwendungssituation und jede Gruppe von Lernenden neu hergestellt werden, indem man sich über jene Punkte Gedanken macht, die für eine konkrete Beurteilungssituation relevant sind. Nur so kann dem komplexen Gefüge von Lehren, Lernen und Beurteilen Rechnung getragen werden, welches die Grundlage unseres Schulsystems bildet.

### Checkliste zur Überprüfung der Validität eines Beurteilungsinstruments

<p><b>Welche positiven Konsequenzen soll die Beurteilung haben?</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Den Lernprozess der Lernenden fördern.</li> <li><input type="checkbox"/> Die Lernenden in das passende Unterrichtsniveau einordnen.</li> <li><input type="checkbox"/> Die Qualität des Unterrichts verbessern.</li> <li><input type="checkbox"/> Über den Lernstand der Lernenden informieren: <ul style="list-style-type: none"> <li><input type="checkbox"/> die Lernenden selber</li> <li><input type="checkbox"/> die Lehrpersonen</li> <li><input type="checkbox"/> Eltern</li> <li><input type="checkbox"/> Schulbehörden</li> <li><input type="checkbox"/> Arbeitgeber</li> </ul> </li> <li><input type="checkbox"/> ...</li> </ul>		<b>Ziele der Beurteilung</b>
<p><b>Wofür wollen Sie die Beurteilung konkret benutzen?</b> Sie möchten <b>Entscheidungen</b> treffen über...</p>		
<p><b>den Erfolg der Lernenden:</b></p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> zukünftige Unterrichtsinhalte.</li> <li><input type="checkbox"/> die Formulierung von Lernempfehlungen.</li> <li><input type="checkbox"/> die Einteilung der Testteilnehmenden in Gruppen: <ul style="list-style-type: none"> <li><input type="checkbox"/> für die Erteilung von Förderunterricht,</li> <li><input type="checkbox"/> für die Zuteilung zu Unterrichtsniveaus,</li> <li><input type="checkbox"/> für eine Gruppenarbeit oder ein Projekt.</li> </ul> </li> <li><input type="checkbox"/> die Vergabe einer Teilnote.</li> <li><input type="checkbox"/> die Vergabe einer Note oder eines (Abschluss)Zertifikats.</li> <li><input type="checkbox"/> ...</li> </ul>	
<p><b>den Erfolg des Unterrichts:</b></p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> das Beibehalten, Ändern oder Abschaffen einer Unterrichtseinheit.</li> <li><input type="checkbox"/> das Beibehalten, Ändern oder Abschaffen eines Unterrichtskonzepts.</li> </ul>	
<p><b>Welche Informationen benötigen Sie, um die Entscheidungen (s. oben) zu treffen?</b> Die Lernenden müssen zeigen, dass ... Die Lernenden müssen zeigen, wie gut ...</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> sie etwas <i>mitteilen</i> können.</li> <li><input type="checkbox"/> sie etwas <i>verstehen</i>.</li> </ul> <p style="text-align: right;">} <b>savoir-faire</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> sie über etwas <i>reflektieren</i> können.</li> <li><input type="checkbox"/> sie ein Problem <i>lösen</i> können.</li> <li><input type="checkbox"/> sie eine Strategie <i>anwenden</i> können.</li> </ul> <p style="text-align: right;">} <b>savoir apprendre / savoir être</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> sie etwas <i>wissen</i>.</li> <li><input type="checkbox"/> ...</li> </ul> <p style="text-align: right;">} <b>savoir</b></p>		

**Wie können Sie am besten die Informationen erheben, die Sie für die Entscheidungen (s. oben) benötigen?**

- Die Lernenden (regelmässig) im Unterricht beobachten.
- Im Unterricht konkrete Aufgaben stellen
  - und die Lernenden bei der Bearbeitung beobachten.
  - und das Produkt der Arbeit beurteilen.
- Ein Portfolio erstellen lassen und/oder zur Selbstbeurteilung anregen.
- Einen Test oder eine Prüfung durchführen.

**(Wie) Haben Sie Folgendes sichergestellt?**

- Die Schlussfolgerungen über die Fähigkeiten Ihrer Lernenden sind ...
  - relevant und ausreichend** für die zu fällende Entscheidung,
  - gerecht** für alle Testteilnehmenden,
  - evtl. **verallgemeinerbar** auf andere Aufgaben inner- und ausserhalb des Unterrichts.
- Die Beurteilungskriterien werden für die Lernenden und andere interessierte Personen transparent kommuniziert.
- Die Beurteilungskriterien werden bei allen Lernenden gleichermassen und vollständig angewendet.
- Die Aussagen über die Kompetenzen der Lernenden sind (für die Adressaten der Aussagen) verständlich formuliert.
- Niemand wird aufgrund von Faktoren benachteiligt, die nichts mit der Beurteilung zu tun haben (z.B. Lesekompetenzen in einem Test zur mündlichen Produktion).
- Niemandem entstehen wegen der Beurteilung unverhältnismässige Nachteile.

Inhalte und Form  
der Beurteilung

## „Testvalidität“ in der Bildungsforschung

Das Konzept der „Testvalidität“ gewann zu Beginn des 20. Jahrhunderts im Zusammenhang mit der Verbreitung von Intelligenztests an Bedeutung. Damals ging man davon aus, dass man Validität nachweisen kann, indem man die Resultate eines Messinstrumentes (also z.B. eines Intelligenztests) mit den Resultaten anderer Messungen des gleichen Sachverhalts (z.B. eines anderen Intelligenztests) vergleicht. Diese Vorstellung basierte allerdings auf der Hoffnung, man könne klar definierbare Kompetenzen eindeutig messen, was sich bis heute als problematisch erweist: Es besteht Uneinigkeit darüber, welche Fähigkeiten und Kenntnisse zu einer Kompetenz beitragen und wie sie dafür zusammenwirken. Im Laufe der Jahrzehnte wurde der Validitätsbegriff daher mehrmals neu konzeptualisiert (Newton & Shaw, 2014).

Das heute vorherrschende Verständnis von Validität entstand Mitte der 1950er Jahre in der Psychologie (Cronbach & Meehl, 1955) und wurde 1989 von Samuel Messick soweit ausgearbeitet, dass es seitdem in der Testforschung vorherrschend ist: Konstruktvalidität (Messick, 1989). Die darin enthaltene Idee eines zu testenden ‚Konstrukts‘ hat wissenschaftstheoretische Wurzeln und ist nicht immer leicht zu erfassen, sie hat aber zwei ganz konkrete Konsequenzen: a) sie setzt eine bewusste Formulierung

dessen voraus, was getestet werden soll (des Konstrukts); b) der Zusammenhang zwischen den Testinhalten, den Testaufgaben, den Resultaten und den daraus gezogenen Schlussfolgerungen muss hergestellt und nachgewiesen werden (Bachman & Palmer, 2010; Kane, 2006; Mislevy & Yin, 2012).

Moderne Testvalidität verlangt nicht mehr danach, einen Test in all seinen Facetten zu validieren. Stattdessen ist seine Verwendung in einem bestimmten Kontext und mit einer bestimmten Zielgruppe zu begründen. Dabei ist insbesondere zu zeigen, dass die Konsequenzen, die sich aus den Testresultaten ergeben, gerechtfertigt sind.

Die heutige Bildungsforschung verfolgt im Zusammenhang mit Testvalidität vor allem drei Fragestellungen:

- > Welche Funktion erfüllen konkrete Tests in welchen Anwendungssituationen (vgl. z.B. Bachman & Palmer, 2010)?
- > Spiegeln formale und inhaltliche Merkmale von Testaufgaben die Kompetenzen der Teilnehmenden wider (z.B. in diesem Heft: Barras, Lenz & Karges)?
- > Was ist eigentlich ‚Validität‘, wie ist sie nachzuweisen und in welchem Verhältnis steht sie zu anderen Qualitätsmerkmalen von Tests wie Reliabilität, Fairness, Usability u.Ä. (Kane, 2006; Markus & Borsboom, 2013; Newton & Shaw, 2014)?

## Bibliografie

- Bachman, L. & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, New York: Oxford University Press.
- Cronbach, L. J. & Meehl, P. E. (1955). „Construct validity in psychological tests“. *Psychological bulletin*. 52 (4), 281-302.
- Kane, M. (2006). „Validation“. In: R. L. Brennan/ National Council on Measurement in Education/American Council on Education (Hrsg.), *Educational measurement*. 4. Aufl. Phoenix, AZ: Greenwood, pp. 17-64.
- Markus, K. A. & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York: Routledge.
- Messick, S. (1989). „Validity“. In: R. L. Linn (Hrsg.), *Educational measurement*. 3. Aufl. New York: American Council on Education & Macmillan Publishers, pp. 13-103.
- Mislevy, R. J. & Y. Chengbin (2012). „Evidence-centered design in language testing“. In: G. Fulcher & F. Davidson (Hrsg.), *The Routledge handbook of language testing*. London, New York: Routledge, pp. 208-222.
- Newton, P. E. & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. London, Cambridge: SAGE & Cambridge Assessment.