

DISKUSSION

Wilfried Reisse

Anmerkungen zur Frage der Bewertung von Prüfungsleistungen

(aus der Sicht der pädagogisch-psychologischen Diagnostik)

Ausgehend von der Unterscheidung zwischen der Feststellung und der Bewertung einer Lernleistung werden drei Bewertungsmodelle kurz skizziert, die sich besonders hinsichtlich des verwendeten Vergleichsmaßstabs unterscheiden: die kriteriumsorientierte (ausbildungszielorientierte), die normorientierte und die prozeßorientierte (ausbildungsprozeßorientierte) Leistungsbewertung. Für berufliche Prüfungen wie Zwischen- und Abschlußprüfungen in der beruflichen Erstausbildung sind eher ausbildungszielorientierte Bewertungen angebracht, ein normorientiertes Vorgehen ist etwa bei Eigentumsuntersuchungen bzw. Ausleseprüfungen notwendig, während Lernfortschrittskontrollen mehr eine ausbildungsprozeßorientierte Bewertung erfordern. Diese Modelle führen auch zu einem unterschiedlichen Vorgehen, wenn man Lernleistungen in der beruflichen Bildung bewerten muß.

Bei einer Prüfung werden zunächst Leistungen von Auszubildenden oder anderen Prüfungsteilnehmern ermittelt. Solche Leistungen können zum Beispiel sein: 53 Aufgaben richtig gelöst, alle in der mündlichen Prüfung gestellten Fragen zutreffend beantwortet, eine Arbeitsprobe mit der vorgeschriebenen Genauigkeit hergestellt. Wie ein solches Ergebnis zu bewerten ist, stellt eine andere Frage dar: Ob man es beispielsweise als gut, schlecht, oder mittelmäßig einstuft, ob vorgegebene Anforderungen erreicht werden oder auch nicht. Daher unterscheidet man zweckmäßigerweise zwischen Leistungsfeststellung und Leistungsbewertung; ähnlich wie man in der psychologischen Diagnostik schon immer die Erhebung von Testwerten und deren Interpretation voneinander abgegrenzt hat.

Ebenso wie zur Leistungsfeststellung gibt es auch zur Leistungsbewertung unterschiedliche Verfahrensansätze (d. h. verschiedene Interpretationsarten von Testwerten, vgl. FRICKE 1974, S. 11—15). Beispielsweise könnte man ein Prüfungsergebnis einfach als „bestanden — nicht bestanden“, „mit Auszeichnung bestanden — gut bestanden — befriedigend bestanden — usw.“ bezeichnen. Man kann selbstverständlich auch Noten oder ein irgendwie geartetes Punktbewertungssystem verwenden. Es wäre weiterhin zu unterscheiden zwischen einfachen Bewertungsvorgängen und komplexen Bewertungsverfahren, bei denen etwa Punkte oder Noten als Teilergebnisse in komplexer Weise zu einem Gesamtergebnis zusammengefaßt werden.

Im vorliegenden Beitrag soll aufgezeigt werden, wie sich einige Probleme der Bewertungspraxis auf dem Hintergrund neuerer Entwicklungen in der pädagogisch-psychologischen Diagnostik darstellen; es wird also gewissermaßen versucht, Praxisprobleme zu „übersetzen“. Von daher können sich auch Hinweise auf mögliche Lösungen ergeben.

Ausgangspunkt ist die Überlegung, daß jede Bewertung einer Leistung einen Vergleich erforderlich macht, wobei unterschiedliche Arten von Vergleichsmaßstäben verwendet werden können. Zusammen mit den jeweiligen theoretischen Grundlagen, den daraus abgeleiteten Verfahren der Bewertung (sowie der Leistungsfeststellung) und den mit der Prüfung oder Leistungsbewertung angestrebten Zielen definieren sie „Bewertungsmodelle“. Hier sollen drei Modelle kurz skizziert werden, die von praktischer Bedeutung für berufliche Prüfungen und Lernleistungsmessungen sind: die normorientierte, die kriteriumsorientierte (ausbildungszielorientierte) und die prozeßorientierte (ausbildungsprozeßorientierte) Leistungsbewertung.

Bei der normorientierten Leistungsbewertung dient die durchschnittliche Leistung in einer Bezugsgruppe (etwa alle Auszubildenden eines Berufes) als (relativer) Vergleichsmaßstab, bei der kriteriumsorientierten oder — nach NIBBRIG 1976 — ausbildungszielorientierten Bewertung ein von vornherein feststehender und auch nicht variabler (absoluter) Standard (z. B. präzise Prüfungsanforderungen). Bei der prozeßorientierten Messung ergibt sich schließlich eine Bewertung aus einem Vergleich mit vorhergehenden und späteren Leistungen des gleichen Auszubildenden.

Es leuchtet ein, daß diese unterschiedlichen Modelle auch in der Bewertungspraxis zu Konsequenzen führen: so muß man bei dem normorientierten Vorgehen zuerst die Verteilung der Leistungen in der Bezugsgruppe feststellen, ehe man eine Einzelleistung bewerten kann, bei der ausbildungszielorientierten Bewertung wird u. a. ein Bewertungsschlüssel benötigt, dessen Anwendung genaue Aussagen ermöglicht, ob oder inwieweit das Ziel der Ausbildung (das Kriterium) erreicht wurde. Verfahrensunterschiede ergeben sich auch bei der Konstruktion, Erprobung und Auswahl von Aufgaben. Zur theoretischen Grundlegung und zur praktischen Erprobung dieser drei Ansätze sei nur erwähnt, daß normorientierte Messungen wohl am umfassendsten durch die traditionelle Testtheorie und Praxis der Testkonstruktion (vgl. z. B. LIENERT 1969) fundiert sind, daß für ein kriteriumsorientiertes Vorgehen in den letzten 10 Jahren beachtliche Ansätze entwickelt wurden (vgl. u. a. FRICKE 1974, KLAUER et al. 1971, HERBIG 1976) und die Prozeßdiagnostik als neuestes Konzept in zunehmendem Maße wissenschaftliches Interesse findet (PAWLIK 1976, RÜDIGER 1976).

Welche Bedeutung haben nun diese Modelle für die Bewertungspraxis bei beruflichen Prüfungen und Lernleistungsmessungen? Dazu ist zunächst zu fragen, ob es sich bei beruflichen Prüfungen (wie Zwischen- oder Abschlußprüfungen) eher um kriteriumsbezogene oder mehr um normorientierte Leistungsmessungen handelt. Einiges spricht dafür, daß hier eher eine kriteriumsorientierte (**ausbildungszielbezogene**) **Leistungsmessung** im Vordergrund steht:

- Formulierungen in Prüfungsregelungen stellen häufig einen Vergleich mit einem vorgegebenen Standard in den Mittelpunkt. „Durch die Abschlußprüfung ist festzustellen, ob der Prüfling die erforderlichen Fertigkeiten beherrscht, die notwendigen praktischen und theoretischen Kenntnisse besitzt und mit dem ihm im Berufsschulunterricht vermittelten, für die Berufsausbildung wesentlichen Lehrstoff vertraut ist. Die Ausbildungsordnung ist zugrunde zu legen“ (BBiG, § 35); „Während der Berufsausbildung ist zur Ermittlung des Ausbildungsstandes mindestens eine Zwischenprüfung entsprechend der Ausbildungsordnung durchzuführen“ (BBiG, § 42).
- Bei Abschlußprüfungen mit Hilfe sogenannter „programmierten“ Prüfungsmaterialien erhält man häufig eine Verteilungsform der Ergebnisse, die für kriteriumsbezogene Messungen typisch ist (der überwiegende Teil der Prüfungsteilnehmer löst weitaus mehr als die Hälfte der angebotenen Aufgaben) (vgl. BRAND 1975, LETH 1975).
- Die ausgedehnte Verwendung von Punktsystemen mit vorgegebenen Bewertungsgrenzen (z. B. dem 100-Punkte-System) ist selbst ein Indiz für kriteriumsorientierte Leistungsmessungen.

— Der bisherige Kenntnisstand über die Ursachen für das Versagen bei Abschlußprüfungen erlaubt noch keine konkreten Aussagen darüber, daß die zentrale Funktion beruflicher Abschlußprüfungen in der Steuerung mittels Selektion zu sehen ist (vgl. LENNARTZ 1976). Eine normorientierte Leistungsmessung hingegen dürfte nur dann im Vordergrund stehen, wenn die Bewertungsverfahren für eine Selektionsentscheidung im o. g. Sinne herangezogen werden sollten.

Aus diesen Überlegungen ergeben sich erhebliche Konsequenzen auch für die Leistungsbewertung. So wäre es beispielsweise nicht gerechtfertigt, bei beruflichen Zwischen- und Abschlußprüfungen — als ausbildungszielorientierten Leistungsbewertungen — Aufgabensätze so zu konstruieren oder bereits vorliegende Ergebnisse so zu transformieren, daß eine Normalverteilung (die vielgenannte Gauß'sche Glockenkurve) der Punkte oder Noten entsteht. Dagegen sind Verfahren von Interesse, mit denen aufgrund neuerer teststatistischer Ansätze Bewertungsschlüssel in Form einfacher Tabellen konstruiert werden können, ohne daß es erforderlich ist, zur Gewinnung eines Vergleichsmaßstabs vorher die entsprechenden Leistungen bei einer größeren Gruppe von Prüfungsteilnehmern festzustellen und auszuwerten. Bei einem solchen Bewertungsschlüssel werden die Zahl (oder der Anteil) der gelösten Aufgaben und der Grad der Zielerreichung („Erreichung“ oder „Nichterreichung“ oder Noten als unterschiedliche Grade der Zielerreichung) in Beziehung gesetzt. Beispiele finden sich bei KLAUER 1972 und HERBIG 1975, eine Übersicht über den Forschungsstand hinsichtlich von Bewertungsmodellen für die kriterienorientierte Leistungsmessung enthält das Sammelreferat von MESKAUSKAS 1976.

Während somit eine ausbildungszielorientierte Bewertung auf die Frage abzielt, ob bzw. inwieweit ein Jugendlicher — unabhängig von anderen — das jeweils vorgegebene Ausbildungsziel erreicht hat, geht es bei der **normorientierten Leistungsmessung** darum, Unterschiede zwischen den Auszubildenden möglichst genau zu ermitteln. Eine solche Bewertung kann etwa bei eignungsdiagnostischen Untersuchungen Verwendung finden.

Begriffe wie Lernfortschrittskontrollen, Lernerfolgskontrollen und Lernzuwachsmessung weisen auf eine **prozeßorientierte Leistungsbewertung** hin. Wenn man berufliche Bildung als Veränderungsprozeß ernstnimmt und die Leistungsmessung

in diesem Zusammenhang sieht, müssen solche Bewertungen in den Mittelpunkt des Interesses treten. Um beispielsweise Unterschiede zwischen Eingangleistungen und Ergebnissen von Zwischen- und Abschlußprüfungen bei einem Auszubildenden zu bewerten, waren spezielle Verfahren notwendig, die noch weitgehend fehlen.

Zusammenfassend ist festzuhalten, daß — je nach den Zielen, die mit der Erfassung von Lernleistungen in der beruflichen Bildung verfolgt werden — auch unterschiedliche Bewertungsverfahren erforderlich sind. Es würde den Rahmen dieser Anmerkungen sprengen, die zur Zeit praktizierten Verfahren (100-Punkte-System, Notensystem) auf diesem Hintergrund zu analysieren und — falls es erforderlich ist — Änderungsvorschläge zu entwickeln und auf ihre Brauchbarkeit hin zu prüfen.

Literatur

- Brand, Peter, 1975: Programmierte Prüfungen im Spiegel der Erfahrung und der Bewahrung. In: Programmierte Prüfungen: Problematik und Praxis. Schriften zur Berufsbildungsforschung, Band 35. Schroedel Hannover S. 181—252
- Fricke, Rainer, 1974: Kriteriumsorientierte Leistungsmessung Kohlhammer. Stuttgart
- Herbig, Manfred, 1975: Zensurierung nach dem binomialen Testmodell. In: Lernzielorientierter Unterricht, H 1, S 1—12
- Herbig, Manfred, 1976: Praxis lehrzielorientierter Tests. Schwann: Düsseldorf.
- Klauer, Karl Josef, 1972: Zur Theorie und Praxis des binomialen Modells lehrzielorientierter Tests. In: Klauer et. al.: Lehrzielorientierte Tests Schwann Düsseldorf, S. 161—195.
- Klauer, Karl Josef; Fricke, Rainer; Herbig, Manfred; Ruprecht, Helmut, Schott, Franz, 1972: Lehrzielorientierte Tests Schwann Düsseldorf.
- Lennartz, Dagmar, 1976: Zur Bestimmung von Determinanten des Prüfungsversagens — Überlegungen zu einem Untersuchungs-Ansatz. Manuskriptdruck Bundesinstitut für Berufsbildungsforschung Berlin
- Leth, August, 1975: Programmierte Prüfungen mit dem Normtest-Electronic-System. In: Programmierte Prüfungen: Problematik und Praxis. Schriften zur Berufsbildungsforschung, Bd. 35. Schroedel Hannover S 255—331
- Lienert, Gustav A., 1969: Testaufbau und Testanalyse 3. Aufl. Beltz. Weinheim
- Meskauskas, John A., 1976: Evaluation Models for Criterion-Referenced Testing. Views Regarding Mastery and Standard-Setting. In: Review of Educational Research, Jg 46, H. 1, S. 133—158
- Nibbrig, Bernhard, 1976: Gutachten zum Problem der programmierten schriftlichen Kaufmannsgehilfenprüfung. Köln: Institut für Berufs-, Wirtschafts- und Sozialpädagogik der Universität zu Köln
- Pawlik, Kurt (Hrsg.), 1976: Diagnose der Diagnostik Klett: Stuttgart
- Rüdiger, Dietrich, 1976: Prozeßdiagnose als neues Konzept der Lernfähigkeitsdiagnose. Referat auf dem 30. Kongreß der Deutschen Gesellschaft für Psychologie vom 19.—23. September 1976 in Regensburg.

Ernst Hoffmann

Zur Frage von Punktbewertungsrichtlinien

Seit vielen Jahren sind Versuche großer Industrieunternehmen der Metall- und Elektroindustrie bekannt, die Leistungen ihrer im Bundesgebiet verstreuten Lehrwerkstätten anhand der Prüfungsergebnisse vergleichen. Die Konzernzentrale kann Ergebnisse durchaus konkret erfassen, so daß der Gedanke solcher Vergleiche naheliegt; vor allem, wenn die Zentrale davon überzeugt ist, daß für diese Gruppe von Berufen bundeseinheitliche Aufgaben und bundeseinheitliche Bewertungsrichtlinien verwendet werden. Außer Ärger erbringen aber solche Vergleiche nichts. Sie fußen schlicht auf einer Überschätzung von Prüfungsergebnissen, die anders als in Technik und Sport mit Stoppuhr und Bandmaß weder exakt erfassbar noch exakt vergleichbar sind.

Oder sollte das falsch sein? Sollte es möglich sein, mit genauen und einheitlichen Bewertungsvorschriften Prüfungsergebnisse von Ort zu Ort, von Lehrberuf zu Lehrberuf, von

Lehrbetrieb zu Lehrbetrieb sowie von Termin zu Termin vergleichbar zu machen? Langjährige Erfahrungen lassen da zweifeln.

Bewertungsregelungen sind notwendiger Bestandteil des Prüfungswesens und haben vielerlei Sachverhalte zum Gegenstand. Überlegungen über ihre Bestgestaltung füllen z. B. beim Deutschen Industrie- und Handelstag viele dicke Aktenordner. Bewertungsrichtlinien müssen auch aus naheliegenden Gründen für Metallberufe andere Sachverhalte regeln als für Bau-, Labor- oder reine Büroberufe. Somit sind Vergleiche von einander fremden Berufsgruppen ohnehin nur Zahlenspielerien. Großstadtkammern mit einem sozusagen anonymen Prüfungsklima haben schlechtere Ergebnisse als Kammern mit z. B. 300 km Längenausdehnung, wo sich das Prüfungsgeschehen in vielen kleinen Städten abwickelt und zudem die Auslese durch weiterführende Schulen weniger wirksam ist.