

ADÄQUATE LEISTUNGSTESTS BEWERTEN LERNERFOLG NACH WISSENSCHAFTLICHEN KRITERIEN

Helmut Lukesch

Mit jedem Lehrgang, jedem Training und jeder Bildungsmaßnahme ist ein Erfolgsversprechen verbunden. Damit sollen Lernprozesse bei den Teilnehmern angeregt werden, deren intendierte Ziele – allgemein gesagt – im Aufbau von deklarativem oder prozeduralem Wissen liegen. Wissen, Fertigkeiten und Handlungsbereitschaft können sich nahe liegender Weise auf unterschiedlichste inhaltliche Bereiche beziehen, wie auch die entsprechenden Verkündigungen andeuten (z.B. Erwerb fundierten Grundwissens oder berufsfeldspezifischer Kompetenzen, Training von »Soft skills«, Verbesserung des Konfliktlösepotenzials von Führungskräften, Kooperationsförderung in Gruppen, interkulturelle Kompetenz, Kommunikationsfähigkeit). Leistungstests könnten prinzipiell für alle diese Inhalte konzipiert werden.

Jede Aussage über einen Lernerfolg (oder vielmehr »Lehr«erfolg) setzt zuerst die Spezifizierung von Lehrzielen als Erfolgskriterien voraus; und damit dies nicht Verkündigungslyrik bleibt, muss als nächstes eine systematische Evaluation hinsichtlich der Erreichung dieser Ziele mit einer objektiven Methodik vorgenommen werden. Damit aber nicht genug, zur sachgerechten Feststellung des Lehrerfolges müsste im Grunde mittels einer Vorher-Nachher-Messung und unter Einbezug einer Kontrollgruppe nachgewiesen werden, dass gerade diese Maßnahme eine Veränderung in Richtung intendierter Lehrziele bewirkt. Die Aufgabenstellung ist damit prinzipiell die gleiche wie bei der Überprüfung der Wirksamkeit eines neuen Medikaments: Der Produzent muss einen wissenschaftlichen Kriterien genügenden Nachweis führen, dass die Anwendung seines Produkts zu den gewünschten Wirkungen führt und dass unerwünschte Nebenwirkungen minimal ausfallen. Und wohl kaum ein Patient wird sich bewusst einer Be-

handlung unterziehen, für die es keinen Wirknachweis gibt, Ausnahmen seien zugestanden.

Die Realität entspricht bekanntlich keineswegs diesen Ambitionen: Betrachtet man beispielsweise die über das Internet recherchierbaren Angebote bezüglich des multimedialen Lernens in Unternehmen, so findet sich eine Unzahl von Versprechungen, welche inhaltlichen Ziele in kognitiver, motivational-affektiver, psychomotorischer oder sozial-kommunikativer Hinsicht erreicht werden sollen und welche ökonomischen Vorteile mit diesem Lehrmedium verbunden sind; aber es wird keinerlei empirische Evaluation erwähnt, mit der nachgewiesen würde, ob diese Versprechungen auch tatsächlich eingelöst werden (*Nikolaus 2002*). Die angebotenen Experteneinschätzungen sind für einen Wirksamkeitsnachweis nicht hinreichend.

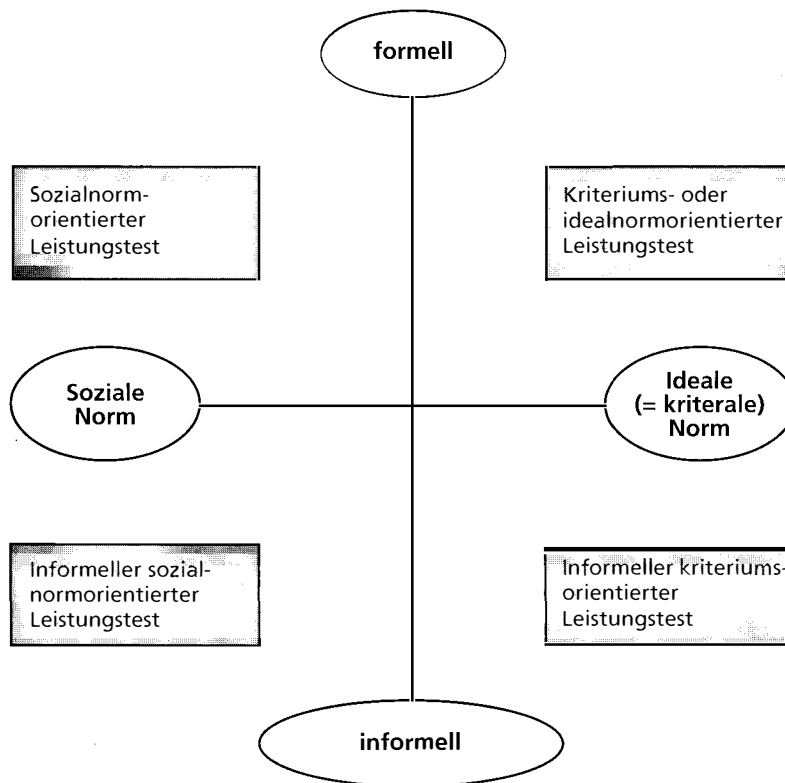
AKZEPTANZ VON LEISTUNGSKONTROLLEN

Auch im Schulsystem wurde die Frage, wie und ob überhaupt die Erreichung von Lehrzielen systematisch überprüft werden könnte, sowohl in der pädagogisch-psychologischen Forschung als auch bei der Ausbildung von Lehrkräften lange Zeit eher vernachlässigt. Aber die unermüdlichen Anstrengungen vor allem von *Karl-Heinz Ingenkamp* (1962) haben hier ein deutliches Umdenken bewirkt, das sicherlich auch durch eine Änderung des gesellschaftlichen Meinungsklimas begünstigt war: Während in Deutschland Ende der 80er Jahre überregionale Leistungskontrollen noch für undurchsetzbar galten (*Ingenkamp/Schreiber 1989*), ist es heute eine Selbstverständlichkeit, dass Deutschland bei internationalen Leistungsvergleichsstudien (z.B. TIMMS, PISA, IGLU) mit im Boot sitzt.

Nahe liegender Weise werden solche Vergleiche aber nur dann akzeptiert, wenn sie auf der Basis einer objektiven, messgenauen und die angezielten Inhaltsbereiche valide abbildenden Methodik durchgeführt werden. Vorgehensweisen, die subjektiven Verfälschungstendenzen unterliegen, deren Reliabilität unbekannt ist und die keine nachgewiesene Validität besitzen, sind hierbei nicht zu gebrauchen.

Damit ist für die summative Evaluation von Lern-

ABB. 1: KLASSIFIZIERUNG VON LEISTUNGSTESTS (NACH LUKESCH 1998, S. 500)



erfolg (z.B. zum Abschluss eines Trainings oder Lehrganges) ein Leistungstest immer das Instrument der Wahl (bei einer formativen und den Lernprozess begleitenden Evaluation mag das hingegen anders sein). Zu beachten ist, dass im Unterschied zum alltäglichen Sprachgebrauch eine beliebige Aufgabensammlung kein Test ist. Vielmehr impliziert die Bezeichnung »Test«, dass das Verfahren daraufhin überprüft wurde, ob es den wesentlichen diagnostischen Gütekriterien (Lukesch 1998) genügt und in objektiver und reliabler Weise tatsächlich das misst, was es zu erfassen gilt.

ARTEN VON LEISTUNGSTESTS

Man kann Leistungstests nach zwei Dimensionen klassifizieren, nämlich (1) nach dem dem Verfahren zugrunde liegenden Normbezug (der Lernerfolg kann durch einen sozialen Vergleich mit den Ergebnissen anderer Lerner oder durch den Vergleich mit kriterial definierten Anforderungen festgestellt werden) und (2) nach dem Grad der vorgenommenen Standardisierung (ein Verfahren kann auf der Basis einer repräsentativen

Stichprobe oder einer nicht-repräsentativen Gelegenheitsstichprobe überprüft werden). Insgesamt ergeben sich somit vier verschiedene Arten von Leistungstests (vgl. Abb. 1).

Diese vier Möglichkeiten zur Erstellung von Leistungstests weisen bestimmte Gemeinsamkeiten, aber natürlich auch Unterschiede auf:

(1) Die Testaufgaben (= Items) müssen immer das (oder die) Lehrziel(e) repräsentieren. Diese so genannte inhaltliche Validität der Items muss auf Grund einer Lehrzielanalyse und entsprechender Expertenratings bestimmt werden. Zudem sollen die inhaltlichen Bereiche adäquat auf die verschiedenen Stufen der Lehrzielhierarchie, also auf die nach Komplexität gestuften Anforderungsebenen, auf denen der Lehrerfolg erfasst werden soll, verteilt sein.

(2) Für jeden Leistungstest muss in einem meist mehrschrittigen Verfahren überprüft werden, ob die bereits genannten Gütekriterien erfüllt sind. Für einen sozialnormorientierten Leistungstest werden die Gütekriterien zumeist vor dem Hintergrund der klassischen

oder seltener der probabilistischen Testtheorie berechnet (immerhin beruhen die bei der PISA-Studie eingesetzten Tests auf dem probabilistischen Testmodell von Rasch); bei einem kriteriumsorientierten Verfahren sind die Routinen der kriteriumsorientierten Testtheorie anzuwenden. Ein Verfahren, das solchen Überprüfungen nicht unterzogen wurde oder diesen nicht standgehalten hat, ist kein Leistungstest, selbst wenn mit seinen Ergebnissen weit reichende Konsequenzen verbunden sind.

(3) Die mit den Items geforderte Leistungsart (z.B. Auswahlentscheidung bei Multiple-Choice-Aufgaben, Zu- und Umordnungsaufgaben, Kurzantwortaufgaben) ist nicht grundlegend, sie kann sich allerdings auf die Einhaltung des Kriteriums der Objektivität in Bezug auf Durchführung, Auswertung und Interpretation und somit in der Folge auch auf Messgenauigkeit und Gültigkeit auswirken. Auch die Frage, ob ein Paper-and-Pencil-Test oder eine Prüfung über ein Rechnerdisplay stattfindet, ob komplexe Lösungen entwickelt werden sollen oder eine Verhaltensprobe erwartet wird, ist für einen Leistungstest nicht entscheidend. Wichtig ist nur, dass der Nachweis geführt wurde, dass die Gütekriterien der Diagnostik bei der Erstellung des Verfahrens gewährleistet wurden.

(4) Lehrzielorientierte Leistungstests könnten bei anderer Standardisierung auch als normorientierte Tests eingesetzt werden. Der Tendenz nach sind bei lehrzielorientierten Verfahren vermehrt Aufgaben enthalten, die nach einem entsprechenden Lehrangebot auch bewältigt werden können sollten, also solche, die eher einem geringeren Schwierigkeitsgrad entsprechen.

(5) Sozialnormorientierte Tests enthalten zumeist Aufgaben auf mittlerem Schwierigkeitsniveau, da mit solchen Aufgaben eine maximale Differenzierung zwischen Probanden erreicht werden kann.

(6) Informelle Tests unterscheiden sich von formellen nur durch den Bezug auf nicht repräsentative Stichproben bei der Überprüfung der Gütekriterien (z.B. von anfallenden Kursteilnehmern), nicht aber dadurch, dass man sich die Überprüfung der Gütekriterien erspart hat. Die bei manchen Trainings, Selbstlernkursen oder Büchern enthaltenen Fragenkataloge mögen für die Steuerung des Lernens von Hilfe sein, sie stellen aber keine Leistungstests im eigentlichen Sinn dar.

(7) Leistungstests können für alle denkbaren Zielgruppen (Erwachsene, Schüler, Führungskräfte, Auszubildende) entwickelt werden. Sie müssen nahe liegender Weise aber immer auf die Ziele einer Weiterbildungsmaßnahme zugeschnitten sein.

VOR- UND NACHTEILE VON LEISTUNGSTESTS

Mit Hilfe adäquat entwickelter Leistungstests kann eine Lernerfolgsdiagnose nach wissenschaftlichen Kriterien erfolgen. Dadurch können eine Reihe weiterführender Fragen auf einer sachlichen Basis entschieden werden, z.B.: Sind die mit dem Lehrangebot verbundenen Versprechungen hinsichtlich eines Lehrerfolgs nachweisbar oder bleibt die Abschätzung des Lehrerfolgs eine Frage des Dafürhaltens (euphemisierend auch als »diskursive Validierung« des Erfolgs eines Trainings bezeichnet)? Eine Erfolgsdiagnose auf der Basis von Leistungstests liefert eine objektive Grundlage sowohl für Platzierungs- als auch für Selektionsentscheidungen. Werden dabei Lücken im Leistungsprofil eines Kursteilnehmers erkannt, können gezielt Empfehlungen für weiterführende Trainings gegeben werden. Steht eine Auswahlentscheidung an, so liefert ein Leistungstest ein objektives Datum für den Vergleich von Bewerbern.

Die Nachteile von Leistungstests liegen vorwiegend auf der Kostenseite, da die Entwicklung entsprechender Verfahren ein kreatives Unterfangen ist, das fachliches Know-how voraussetzt und auch entsprechend zeitaufwendig ist. Durch die Verschriftlichung der Aufgaben besteht allerdings die Gefahr, dass sich die Testautoren auf die Abfrage einfacher Leistungen begrenzen und in einer Lehrzielhierarchie höher stehende Aufgaben nicht einbeziehen.

Leistungstests machen Unterschiede in den Lernergebnissen zwischen Lernenden überdeutlich. Dies kann einer einseitigen Attribuierung der Verantwortlichkeit für den Lernerfolg auf die Lernenden Vorschub leisten und damit können andere, für den Lehrerfolg ebenso wichtige Bedingungen auf der Seite der Anbieter (Lehrkompetenzen, Ausstattungsfaktoren, Aspekte der Kursgestaltung) aus dem Auge verloren werden.

Ein Leistungstest ist für viele Lerner ein Angstauslöser. Mit jedem Test geht eine potenzielle Gefährdung des eigenen Selbstkonzepts einher und es können auch als bedrohlich wahrgenommene soziale Vergleichsprozesse initiiert werden. Deshalb sollten im Kontext jeder Leistungsüberprüfung auch entsprechende kompetenz-erhöhende und emotionsregulierende Bewältigungsmethoden vermittelt werden, um jedem Probanden die Chance zu geben, sein bestes Ergebnis erreichen zu können.

Wegen des Kostenaufwandes könnte es sein, dass sich Trainings- und Lerninhalte ungerechtfertigter Weise verfestigen, da sich Kursteilnehmer an bereits bekannten Aufgaben und ihren Lösungen orientieren.

Für die Qualität des Produktes gibt es zwar gut de-

finierte Kriterien, ob diese erfüllt werden, kann aber zu Beginn einer Verfahrensentwicklung nicht sicher zugesagt werden. Es besteht also ein Entwicklungsrisiko, das aber zu bewältigen ist.

Letztendlich kommt es auf die Auftraggeber von Bildungsmaßnahmen (z. B. Unternehmer, Bundesagentur für Arbeit) an, ob sie bereit sind, ein Training oder eine Weiterbildungsmaßnahme zu akzeptieren, auch wenn für diese kein Erfolgsnachweis vorgelegt wurde. Weiterbildung wird dann zum ebenso blinden wie teuren Aktionismus, da kein messbarer Return zurückkommt. Solange dies den Auftraggebern gleichgültig ist, wird sich an der weiten Verbreitung irrationaler (im Sinne nicht positiv evaluierter) Bildungsangebote nichts ändern!

Literatur

Ingenkamp, K.: Die deutschen Schulleistungstests: Kritische Betrachtungen und Untersuchungen zur pädagogisch-psychologischen Grundlegung. Weinheim 1962 ■ *Ingenkamp, K./Schreiber, W.H.*: Was wissen unsere Schüler? Überregionale Lernerfolgsmessung aus internationaler Sicht. Weinheim 1989 ■ *Nikolaus, U.*: Multimediales Lernen in Unternehmen. Wiesbaden 2002 ■ *Lukesch, H.*: Einführung in die pädagogisch-psychologische Diagnostik. Regensburg 1998².

Prof. Dr. Helmut Lukesch, Institut für Experimentelle Psychologie, Universität Regensburg,
E-Mail: helmut.lukesch@psychologie.uni-regensburg.de