

Der Mythos „Wirkungskette“ in der Weiterbildung – empirische Prüfung der Wirkungsannahmen im „Four Levels Evaluation Model“ von DONALD KIRKPATRICK

KURZFASSUNG: DONALD KIRKPATRICK publizierte sein „Four Levels Evaluation Model“ erstmals in den Jahren 1959 und 1960. Das Vier-Ebenen-Evaluationsmodell fokussiert Output-Dimensionen von Weiterbildungs- bzw. Trainingsmaßnahmen. Es blieb in seiner Grundstruktur bis heute unverändert und gilt in der betrieblichen Weiterbildung als das einflussreichste Evaluationsmodell. Die vier Evaluationsebenen („Levels“) sind (1) Zufriedenheit (reaction), (2) Lernerfolg (learning), (3) Transfererfolg (behavior) und (4) Geschäftserfolg (results). KIRKPATRICK vermutet, dass diese Ebenen über Wirkungsketten miteinander verbunden seien: Eine höhere Teilnehmerzufriedenheit führe zu einem besseren Lernerfolg, der einen besseren Transfererfolg ermögliche, welcher sodann in einen höheren Geschäftserfolg münde. Diese Annahmen wurden in den vergangenen 50 Jahren vielfach geprüft; eindeutig bestätigt oder widerlegt wurden sie bislang nicht. Nachfolgend werden die Ergebnisse zweier Studien vorgestellt ($N_1=335$, $N_2=187$), welche die Prüfung der Wirkungsannahmen zum Gegenstand haben.

ABSTRACT: DONALD KIRKPATRICK first published his „Four-Level Evaluation Model“ by the years 1959 and 1960. The Four Levels Evaluation Model focuses on the output-dimensions of further education and training. It has remained unchanged in its original structure until the present and has been even considered in the field of industrial training as the most influential model. Its four levels are reaction, learning, behavior and results. Kirkpatrick assumes that these levels are causally linked: The participants' reactions leads to learning success, which leads to behavior that leads to results. These assumptions have been widely examined and criticized in the last 50 years. However, they have been neither explicitly confirmed nor rejected. Below, are the results of two studies that investigate the effect chain.

1 Kontext, Problemstellung und Forschungsfragen

1.1 Kontext

Die Ermittlung der Teilnehmerzufriedenheit (Level 1 nach KIRKPATRICK) hat in der beruflichen Weiterbildung einerseits Tradition (vgl. CARNEVALE/SCHULZ 1990; GÜLPEN 1996; VAN BUREN/ERSKINE 2002) und wird andererseits von den in der Weiterbildung etablierten Qualitätsmanagementsystemen explizit eingefordert. So ist z.B. nach DIN EN ISO 9001 die „Kundenzufriedenheit“ zu erheben, um die Leistung eines Qualitätsmanagementsystems zu überwachen (vgl. DIN EN ISO 9001:2008, S. 40). Diese Forderung findet sich ebenfalls in Qualitätsmanagementsystemen, die für den Weiterbildungsbereich entwickelt wurden, wie z.B. dem „Schweizerischen Qualitätszertifikat für Weiterbildungsinstitutionen“: „Die Zufriedenheit der Kundinnen und Kunden ist eines der wichtigen Masse für die Einschätzung der Qualität erbrachter Dienstleistungen. [...] Bitte zeigen Sie, wie Sie die allgemeine Kundenzufriedenheit auf institutioneller Ebene überprüfen.“ (EDUQUA 2004, S. 33). Das Modell „LQW Lernorientierte Qualitätstestierung“, das insbesondere in der deutschen Weiterbil-

dungslandschaft verbreitet ist, verpflichtet Einrichtungen ebenfalls zur Evaluation und zur Erfassung von Zufriedenheitsaspekten: Eine Bildungseinrichtung muss für eine erfolgreiche Qualitätstestierung die Evaluationsgegenstände und -verfahren sowie den Evaluationsrhythmus und -umfang beschreiben und den Bezug zum Leitbild begründen. Als „maßstabsbildend“ werden die Indikatoren „Zufriedenheit der Teilnehmenden“, ggf. die „Zufriedenheit der Auftraggeber“, aber auch der „Lernerfolg“ benannt (vgl. ZECH 2006, S. 68). In der DIN ISO 29990 „Lerndienstleistungen für die Aus- und Weiterbildung“ findet sich gleichfalls ein entsprechender Absatz: „Der LDL [Lerndienstleister] muss sicherstellen, dass zu den verwendeten Methoden und Ressourcen sowie ihrer Effektivität beim Erreichen der vereinbarten Lernergebnisse Feedback von den Lernenden eingeholt wird.“ (DIN ISO 29990:2010, S. 10). Die Erfassung der Teilnehmerzufriedenheit in der Weiterbildung kann heute als ein etablierter Standard betrachtet werden.

Im LQW-Modell wird über diesen Standard hinaus als „maßstabsbildend“ empfohlen, weitere Werte, wie z.B. den Lernerfolg, zu erheben. Hierzu verpflichtet sind die Anwärter auf eine Qualitätstestierung nicht. Eine Analyse von Selbstbeschreibungen, die im Rahmen der Qualitätstestierung nach dem LQW-Modell angefertigt wurden, ergab sodann, dass die Erfassung der „Zufriedenheit der Teilnehmenden“ den zentralen Evaluationsgegenstand bildet. „Der Lernerfolg etwa wird laut Selbstbeschreibung nur indirekt über die Zufriedenheit erfasst. »Wer zufrieden ist, hat auch erfolgreich gelernt«, lautet die Devise, die zugleich zeigt, dass die Evaluationspraxis in Volkshochschulen auf der ersten Stufe des Modells von Kirkpatrick, also den direkten Reaktionen (reaction), zu verorten ist.“ (RÄDIKER 2010, S. 147).

Hieraus ergeben sich Fragen: Ist die Erhebung der Teilnehmerzufriedenheit, wie sie von normativen Qualitätsmanagementsystemen gefordert und der Evaluationspraxis erbracht wird, als Maß für die Einschätzung der Qualität erbrachter Dienstleistungen ausreichend? Ist die zugrunde liegende Wirkungsannahme, „wer zufrieden ist, hat erfolgreich gelernt“, welche die Fokussierung auf Zufriedenheitswerte legitimieren könnte, zutreffend?

1.2 Problemstellung

Die im Zitat benannte Wirkungsannahme zwischen Output-Dimensionen wurde erstmals von DONALD KIRKPATRICK in seinem Vier-Ebenen-Evaluationsmodell formuliert (vgl. KIRKPATRICK 1960). Das Modell sowie dessen implizite Wirkungsannahmen bilden seit fast 50 Jahren die Grundlage für die Entwicklung von Evaluationsmodellen in der Weiterbildung (vgl. WARR/BIRD/RACKHAM 1970; HAMLIN 1974; BRINKERHOFF 1987; KAUFMAN/KELLER 1994; PHILLIPS 1994; BEWYL/GEITER 1996; PHILLIPS/SCHIRMER 2005; KELLNER 2006).

Das zentrale Verdienst von KIRKPATRICK ist, dass er bereits in den 1960er Jahren vier Ebenen definierte, die Gegenstand einer Output-Evaluation sein sollten: Zu-

friedenheit¹, Lernerfolg, Transfererfolg² und Geschäftserfolg³ (vgl. KIRKPATRICK 1960). Obwohl KIRKPATRICK den konkreten Zusammenhang zwischen den Ebenen nur vage beschreibt, beinhaltet das Modell kausale Wirkungsannahmen, die selbst in der neuesten Fassung von 2006, verfasst zusammen mit seinem Sohn, aufzufinden sind: "If training is going to be effective, it is important that trainees react favorably to it." (KIRKPATRICK/KIRKPATRICK 2006, S. 27) und "Without learning, no change in behavior will occur." (ebd., S. 50). Die expliziten und impliziten Annahmen des Modells fassen ALLIGER & JANAK (1989, S. 332–334) wie folgt zusammen:

- "Assumption 1: Each succeeding level is more informative than the last." (ebd., S. 332)
- "Assumption 2: Each level is caused by the previous level." (ebd., S. 333).
- "Assumption 3: Each succeeding level is correlated with the previous level; or, more generally, there exists a „positive manifold“: all correlations among levels are positive." (ebd., S. 334)

Die Annahme einer Wirkungskette bildete wiederum die Basis für weiterführende Annahmen (vgl. BATES 2004):

- Die Teilnehmerzufriedenheit am Ende eines Trainings/ Lehrgangs/ Kurses gibt prospektiv Aufschluss über den Erfolg einer Weiterbildungsmaßnahme.
- Die Teilnehmerzufriedenheit am Ende eines Trainings/ Lehrgangs/ Kurses gibt retrospektiv Aufschluss über die Qualität einer Weiterbildungsmaßnahme.
- Aufgrund dieser Schlüsselstellung der Teilnehmerzufriedenheit ist es ausreichend und zudem wirtschaftlich, allein die Teilnehmerzufriedenheit zu erheben.

Insofern die Wirkungskette gültig ist und die Teilnehmerzufriedenheit tatsächlich prospektiv und retrospektiv die Evaluation anderer Dimensionen zu ersetzen vermag, wäre die aufgezeigte Praxis der Weiterbildungs-evaluation gerechtfertigt. Die Erhebung der Teilnehmerzufriedenheit wäre damit nicht nur ein ökonomisches, sondern zudem ein hinreichendes Instrument, um die Qualität von Weiterbildungsleistungen zu erfassen.

Die Wirkungsannahmen wurden in den vergangenen 50 Jahren vielfach geprüft (vgl. ALLIGER/JANAK 1989; ALLIGER ET AL. 1997); zweifelsfrei bestätigt oder widerlegt wurden sie bislang nicht. Die Problemstellung unserer Untersuchung lautet nunmehr: Inwiefern handelt es sich beim angenommenen Wirkungszusammenhang um einen empirisch nachweisbaren Zusammenhang oder nur um einen seit 50 Jahren tradierten Mythos? Unsere Untersuchung hat die Überprüfung der Wirkungskette bzw. „Auswirkungskette“ (PHILLIPS/SCHIRMER 2005, S. 29) zum Gegenstand.

- 1 KIRKPATRICK nennt diese Ebene im Original „reaction“. Die Übersetzung „Zufriedenheit“ ist u.E. zutreffender als „Reaktion“: „Evaluating reaction is the same thing as measuring customer satisfaction.“ (KIRKPATRICK/KIRKPATRICK 2006, S. 27).
- 2 KIRKPATRICK nennt diese Ebene im Original „behavior“. Die Übersetzung „Transfererfolg“ ist u.E. zutreffender als „Verhalten“: „How much transfer of knowledge, skills, and attitudes occur? That is what level 3 attempts to evaluate.“ (ebd., S. 52).
- 3 KIRKPATRICK nennt diese Ebene im Original schlicht „results“ bzw. „final results“, da vielfältige Resultate/Erfolge/Arbeitsergebnisse gemeint sind (u.a. Qualitätsverbesserung, Verbesserung der Produktivität). KIRKPATRICK benennt in seinen Beispielen jedoch ausschließlich betriebliche Erfolge, weshalb u.E. der Begriff „Geschäftserfolg“ zutreffend ist.

1.3 Forschungsfragen

Geprüft werden soll insbesondere folgende Annahme: Die Teilnehmerzufriedenheit am Ende eines Trainings/ Lehrgangs/ Kurses gibt prospektiv Aufschluss über den Erfolg einer Weiterbildungsmaßnahme.

In Abbildung 1 sind das CIPP-Modell (Context, Input, Process und Product) von STUFFLEBEAM (2003) zur Verortung der Untersuchung, das Vier-Ebenen-Evaluationsmodell von KIRKPATRICK & KIRKPATRICK (2006) sowie der Fokus der Untersuchung dargestellt.

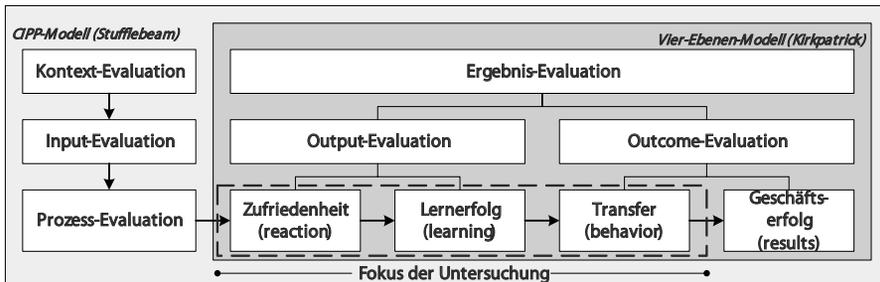


Abbildung 1: Fokus der Untersuchung

Die drei Forschungsfragen unserer Untersuchung lauten:

1. Inwiefern besteht ein Zusammenhang zwischen der Teilnehmerzufriedenheit am Ende einer Weiterbildungsmaßnahme sowie dem Lernerfolg der Teilnehmenden nach Abschluss der Weiterbildungsmaßnahme?
2. Inwiefern besteht ein Zusammenhang zwischen der Teilnehmerzufriedenheit am Ende einer Weiterbildungsmaßnahme sowie dem Transfererfolg der Teilnehmenden nach Abschluss der Weiterbildungsmaßnahme?
3. Inwiefern besteht ein Zusammenhang zwischen dem Lernerfolg der Teilnehmenden nach Abschluss der Weiterbildungsmaßnahme und deren Transferleistung nach Abschluss der Weiterbildungsmaßnahme?

Das erkenntnisleitende Interesse der Untersuchung ist, inwiefern eine Annahme zutreffend ist, die sich in den eingangs vorgestellten Zertifizierungsforderungen der DIN EN ISO 9001 und der EduQua findet: „Die Zufriedenheit der Kundinnen und Kunden ist eines der wichtigen Masse für die Einschätzung der Qualität erbrachter Dienstleistungen.“ (EduQUA, S. 33).

2 Das Vier-Ebenen-Evaluationsmodell von KIRKPATRICK

KIRKPATRICK entwickelte im Rahmen seiner Dissertation bereits 1954 die Grundlagen des Vier-Ebenen-Modells, das er 1959 und 1960 erstmals in vier Artikeln publizierte (vgl. KIRKPATRICK 1960).

Die Weiterbildungsevaluation auf Ebene 1, Zufriedenheit der Teilnehmenden, stellt heute, wie eingangs dargestellt, weitgehend den Normalfall dar. KIRKPATRICK

begründet die Notwendigkeit, die Reaktion der Teilnehmenden bzw. deren Zufriedenheit zu erfassen, wie folgt: "Reactions of participants should be measured on all programs for two reasons: to let the participants know value their reaction, and to measure their reactions and obtain suggestions for improvement." (KIRKPATRICK/KIRKPATRICK 2005, S. 5).

Auf der zweiten Ebene betrachtet KIRKPATRICK den Lernerfolg der Teilnehmer (learning): „Three things can be accomplished in a training program: (1) Understand the concepts, principles, and techniques being taught. (2) Develop and/or improve skills. (3) Change attitudes.“ (ebd. S. 5).

Auf der dritten Ebene wird untersucht, in welcher Weise sich das Verhalten am Arbeitsplatz aufgrund des Trainings verändert (behavior). Falls das Gelernte am Arbeitsplatz nicht angewendet wird, obwohl auf der zweiten Ebene ein guter Lernerfolg nachgewiesen werden konnte, gilt das Trainingsprogramm als gescheitert. „Therefore, measuring behavior change is necessary, not only to see if behavior has changed, but also to determine the reasons why change has not occurred“ (ebd., S. 6).

Auf der vierten Ebene werden die Geschäftsergebnisse erhoben (results). Die Fragestellung lautet: In welcher Weise haben sich die Geschäftsergebnisse aufgrund des Trainings verändert? Kirkpatrick nennt verschiedene Beispiele, die als Parameter den Erfolg des Trainings belegen können: „Results could be determined by many factors including less turn over, improved quantity of work, improved quality, reduction of waste, reduction in wasted time, increased sales, reduction in costs, increase in profits, and return on investment (ROI).“ (ebd., S. 7). Diese Ergebnismessung ist für soziale Themen besonders schwierig: "it is difficult if not impossible to measure final results for programs on such topics as leadership, communication, motivation, time management, empowerment" (KIRKPATRICK 1998, S. 23). Hinzu kommt die Wirkung weiterer (Stör-)Variablen, wie z.B. anderer Programme im Unternehmen, der Performanzentwicklung der Konkurrenz oder konjunktureller Entwicklungen, die das Ergebnis beeinflussen. Um den Effekt des Trainings bestimmen zu können, müssten diese Effekte isoliert werden (vgl. GESSLER 2005).

Das skizzierte Modell verdeutlicht, dass KIRKPATRICK selbst nicht im Sinn hatte, dass sich Weiterbildungs-evaluation allein auf die Ebene der Teilnehmerzufriedenheit begrenzen sollte. Sein Modell ist vielmehr eine Aufforderung, die nächste mögliche Ebene anzustreben. So fordert er beispielsweise für die Ebene „Lernerfolg“: „Strive for a 100% response rate.“ (KIRKPATRICK/KIRKPATRICK 2005, S. 6).⁴

3 Stand der Forschung

ALLIGER & JANAK (1989) identifizieren in einer Meta-Analyse 203 Studien, die über Ergebnisse von Weiterbildungs-evaluationen berichten. Untersucht wurden Weiterbildungs-journals „from 1959 to present, if possible.“ (ebd., S. 336). Von diesen 203 Studien beinhalteten nur 8 Studien Korrelationsanalysen, sodass ausgehend von

4 Phillips und Schirmer (2005), die sich explizit auf das Modell von Kirkpatrick berufen und in Deutschland im Bereich Bildungscontrolling häufig rezipiert werden, empfehlen für die Stufe 2, dass nur 50 % der Weiterbildungsmaßnahmen zu evaluieren seien (ebd., S. 227). Die Empfehlung wird, außer mit dem allgemeinen Hinweis auf eine bestehende und nicht weiter beschriebene Praxis, nicht begründet.

diesen Studien erneut gesucht wurde, wodurch schließlich 16 auswertbare Artikel vorlagen. Vier Studien wurden anschließend wieder ausgeklammert, da sie nicht in der beruflichen Weiterbildung angesiedelt waren. 12 Studien blieben letztendlich übrig.

Problematisch an der Auswertung (ebd., S. 339) ist, dass die uneinheitlichen Effektstärken ohne Signifikanzen sowie ohne Testpower angegeben sind. Aufgeführt wird z.B. eine Studie von NOE & SCHMIDT mit Korrelationen zwischen Zufriedenheit und Lernerfolg, gleichwohl diese nach NOE & SCHMIDT (1986, S. 518) nicht signifikant sei.⁵ ALLIGER & JANAK (1989) konstatieren schließlich: „The data from our literature search certainly do not warrant a definite answer“ (vgl. S. 337).

1997 veröffentlichte ALLIGER zusammen mit TANNENBAUM, BENNETT, TRAVER und SHOTLAND eine weitere Meta-Analyse. Während die ältere Studie der 1980er Jahre nur 12 Untersuchungen identifizierte, konnte die Zahl der verwertbaren Studien fast verdreifacht werden. Der Meta-Analyse lagen diesmal 34 Studien mit insgesamt 115 Korrelationen zugrunde. Die Analyse auf Basis aggregierter Korrelationskoeffizienten ergab einen Zusammenhang zwischen den Ebenen Zufriedenheit und Lernerfolg von nur $r = .07$ sowie zwischen Lernerfolg und Transfererfolg von nur $r = .09$. Die Autoren halten jedoch kritisch fest, dass trotz der diesmal größeren Anzahl an Studien die ausgewiesenen Korrelationen letztendlich nur auf wenigen Studien gründen: „In any case, it needs to be stressed that many of the correlations [...] are based on only a few studies, so that any firm conclusions about relative strength of transfer prediction by certain criteria (e.g., reactions vs. learning) are precluded.“ (ebd., S. 353).

Die Aussagekraft der vorliegenden Studien ist, insgesamt betrachtet, wenig zufriedenstellend.

4 Untersuchungsdesign und -ergebnisse

4.1 Untersuchungsdesign

Untersucht wurden insgesamt 67 ($U_1=43$, $U_2=24$) betriebliche Managementlehrgänge, die auf eine Personenzertifizierung vorbereiten. Alle betrachteten Lehrgänge fanden berufsbegleitend statt und dauerten ca. 4 Monate. Das Curriculum ist aufgrund der Zertifizierung weitgehend standardisiert. Die Trainer ($N=24$) sind rechtlich und wirtschaftlich voneinander unabhängig und verfügen über durchschnittlich 10 Jahre Trainingspraxis sowie eine vergleichbare Kompetenz, da zur Durchführung der Lehrgänge eine Zulassungsprüfung erforderlich ist, in der die Kompetenz der Trainer sowohl fachlich als auch didaktisch überprüft wird.

In der ersten Studie von 2006/07 wurden die Daten von 335 Teilnehmer/-innen ($m=254$, $w=81$) und in der zweiten Studie von 2008/09 die Daten von 187 Teilnehmer/-innen ($m=143$, $w=44$) im Alter von durchschnittlich 41 Jahren ausgewertet. Betrachtet werden jeweils drei Datenstämme:

- Teilnehmerzufriedenheit: Feedbackbogen zum Abschluss eines Lehrgangs;
- Lernerfolg: Ergebnisse der schriftlichen und mündlichen Prüfung;
- Transfererfolg: Ergebnis der Transferprüfung.

5 “Trainee satisfaction with training (reaction criteria) was not related to learning, and learning was not related to behavior change.” (NOE & SCHMIDT 1986, S. 518).

Teilnehmerzufriedenheit: Die Teilnehmerzufriedenheit wurde mithilfe eines Feedbackbogens am Ende der Lehrgänge erhoben (2 Items, Crombachs Alpha: .82; Item 1: „Inwiefern wurden Ihre Erwartungen erfüllt?“, Item 2: „Wie zufrieden sind Sie insgesamt betrachtet mit dem Lehrgang?“).

Lernerfolg: Der Lernerfolg wurde ca. drei Wochen nach Abschluss des letzten Workshops im Rahmen der Personenzertifizierung geprüft. Die Zertifizierung bestand aus einer schriftlichen Klausur im Umfang von zwei Stunden sowie einem mündlichen Interview im Umfang von 30 Minuten.

Transfererfolg: Der Transfererfolg wurde mittels eines Transfernachweises geprüft. Die Teilnehmer erstellten hierfür eine Dokumentation, in der sie die gelernten Methoden auf ein Problem ihrer beruflichen Praxis anwendeten. Der Transfernachweis wurde mittels eines Merkmalkatalogs standardisiert ausgewertet und war Bestandteil der Personenzertifizierung.

4.2 Untersuchungsergebnisse

Analyse der Zusammenhänge: Die Ergebnisse der bivarianten Korrelationsanalyse nach Pearson sind in Tabelle 1 dargestellt:

Tabelle 1: Ergebnisse der Untersuchungen

		Lernerfolg	Transfererfolg
Zufriedenheit	Studie 1: N=331	-.00	-.03
	Studie 2: N=187	.09	.04
Lernerfolg	Studie 1: N=335		.29 *** .99 ¹
	Studie 2: N=187		.22 ** .93 ²

***p<.001 (2-seitig); **p<.01 (2-seitig); ¹ Power 1-β, α<.001; ² Power 1-β, α<.05

Die Korrelationsanalyse ergibt, dass zwischen Zufriedenheit und Lernerfolg in beiden Untersuchungen kein signifikanter Zusammenhang besteht. Zwischen Lernerfolg und Transferleistung besteht hingegen bei der ersten und zweiten Untersuchung ein signifikanter Zusammenhang mit kleiner bis mittlerer Effektstärke. Der Lernerfolg erklärt allerdings in der ersten Untersuchung nur 8,5 % und in der zweiten Untersuchung nur 4,9 % der Varianz der Transferleistung. Als Ergebnis kann festgehalten werden:

1. Es besteht kein signifikanter Zusammenhang zwischen der Teilnehmerzufriedenheit am Ende einer Weiterbildungsmaßnahme sowie dem Lernerfolg der Teilnehmenden nach Abschluss der Weiterbildungsmaßnahme.
2. Es besteht kein signifikanter Zusammenhang zwischen der Teilnehmerzufriedenheit am Ende einer Weiterbildungsmaßnahme sowie dem Transfererfolg der Teilnehmenden nach Abschluss der Weiterbildungsmaßnahme.
3. Es besteht ein kleiner bis mittlerer signifikanter Zusammenhang zwischen dem Lernerfolg der Teilnehmenden nach Abschluss der Weiterbildungsmaßnahme und deren Transferleistung nach Abschluss der Weiterbildungsmaßnahme.

5 Fazit

Die empirische Überprüfung der Wirkungsannahmen im Vier-Ebenen-Evaluationsmodell von KIRKPATRICK ergab, dass die Einschätzung, die Zufriedenheit der Teilnehmenden sei ein hinreichendes (Output-)Maß, um die Qualität erbrachter Dienstleistungen einschätzen zu können, unzutreffend ist, da die Zufriedenheitswerte keinen Zusammenhang aufweisen mit dem Lernerfolg, respektive Transfererfolg, und damit keinen Aufschluss geben über den tatsächlich erreichten Lernerfolg, respektive Transfererfolg.⁶ Die derzeitige Praxis der Weiterbildungs-evaluation sowie die Vorgaben der normativen Qualitätsmanagementsysteme sind zur Erfassung des Outputs von Weiterbildung bislang unzureichend und wären entsprechend weiterzuentwickeln.⁷

Eine einfache und an die bestehende Feedback-Praxis anschlussfähige Möglichkeit der Weiterentwicklung wäre, neben den retrospektiven Zufriedenheitsbewertungen zudem prospektive Nutzwerturteile zu erfassen (z.B. „Von den Inhalten des Lehrgangs kann ich __ % in meiner Arbeit verwenden.“). Laut ALLIGER ET AL. (1997, S. 349) bestehen zwischen prospektiven Nutzwerturteilen sowie dem Lern- bzw. Transfererfolg signifikante Zusammenhänge mit allerdings nur kleinen Effektstärken.

Weitere Anregungen liefert u.a. KIRKPATRICK. Für die Ebene Lernerfolg definiert er beispielsweise folgende „Guidelines“: (1) Messe Fähigkeiten, Fertigkeiten, Kenntnisse und Einstellungen vor und nach dem Training. (2) Verwende einen Papier- und-Bleistift-Test für die Messung von Kenntnissen und Einstellungen. (3) Verwende einen Performanz-Test für die Messung von Fähigkeiten und Fertigkeiten. (4) Strebe eine 100 % Antwortquote an. (5) Wenn möglich, verwende eine Kontrollgruppe (vgl. KIRKPATRICK/KIRKPATRICK 2005, S. 6). Ein Inventar, das die Verbesserung des Transfererfolgs ermöglicht, hat HOLTON, BATES & RUONA (2000) entwickelt; übertragen ins Deutsche wurde dieses von KAUFFELD et al. (2008). Inwiefern schließlich die monetäre Bewertung des trainingsinduzierten Geschäftserfolgs Sinn macht, ist umstritten. Solche Werte können zwar errechnet werden (vgl. PHILLIPS/SCHIRMER 2005), valide sind diese u.E. jedoch in der Regel nicht.⁸

- 6 Die Ergebnisse der zwei durchgeführten Studien bestätigen damit die Annahmen entsprechender Vorgängerstudien (vgl. NOE & SCHMIDT 1986, ALLIGER ET AL. 1997, ARTHUR ET AL. 2003).
- 7 Diese Kritik gilt in gleicher Weise für das neueste Modell, das im Dezember 2010 veröffentlicht wurde: die DIN ISO 29990. Einerseits sieht das Modell vor, dass Verantwortlichkeiten für die Durchführung der Evaluation des Transfererfolgs festgelegt werden müssen. Andererseits soll mittels „Feedback von den Lernenden“ die „Effektivität beim Erreichen der vereinbarten Lernergebnisse“ (DIN ISO 29990:2010, S. 10) ermittelt werden. Diese Formulierung legitimiert die Fortsetzung der bestehenden und unbefriedigenden Praxis „Evaluation mittels Zufriedenheitsdaten“.
- 8 Die Wirtschaftskrise hat beispielsweise branchenübergreifend zu massiven Einnahmeausfällen geführt, weshalb jegliches Training in dieser Zeit einen negativen Return on Investment (ROI) aufweist. Wenn überhaupt, dann kann u.E. der finanzielle Geschäftserfolg immer nur spezifisch und nicht generell ermittelt werden. Es könnte z.B. ermittelt werden, wie sich die Performanz eines Projekts nach einem Training verändert. Instrumente des Projektcontrollings, wie z.B. die „Earned Value Analyse“, ermöglichen Trendanalysen und monetäre Bewertungen mittels der Kombination von Termin-, Kosten- und Leistungskennzahlen auf Basis von Planzahlen, also unabhängig von der konkreten Marktentwicklung. Doch selbst diese Berechnungsform birgt Unwägbarkeiten, da die Verbesserung (oder Verschlechterung) der Performanz auch andere Gründe haben kann (z.B. ein geändertes Informations- und Berichtswesen aufgrund von Vorgaben der Stammorganisation). KIRKPATRICK selbst subsumiert Verschiedenes unter „results“ (siehe Fußnote 3). Zur Frage, inwiefern der Erfolg von Training monetär bewertbar ist, positioniert sich KIRKPATRICK eindeutig: „There is no way it can be done!“ (KIRKPATRICK/KIRKPATRICK 2006, S. 64).

Während der Grad der Kundenzufriedenheit in vielen Branchen einen validen Indikator für die Erfüllung von Anforderungen darstellen kann, ist die Erfassung der Zufriedenheit der Teilnehmenden in prosumentiven Weiterbildungsveranstaltungen nur eine notwendige, jedoch keine hinreichende Bedingung für die Sicherung und Entwicklung der Qualität von Weiterbildung.

Literaturverzeichnis

- ALLIGER, G.M. & JANAK, E.A. (1989): Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology*, 42, 331–342.
- ALLIGER, G.M., TANNENBAUM, S.I., BENNETT, W., TRAVER, H. & SHOTLAND, A. (1997): A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50, 341–358.
- ARTHUR, W., TUBRE, T., PAUL, D. S., & EDENS, P. S. (2003): Teaching effectiveness: The relationship between reactions and learning evaluation criteria. *Educational Psychology*, 23, 275–285.
- BATES, R. (2004): A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27, 341–347.
- BEYWL, W. & GEITER, C. (1996): Evaluation - Controlling - Qualitätsmanagement in der betrieblichen Weiterbildung: kommentierte Auswahlbibliographie. Bielefeld: Bertelsmann.
- BRINKERHOFF, R. (1987): Achieving Results from Training. How to Evaluate Human Resource Development to Strengthen Programs and Increase Impact. San Francisco: Jossey-Bass.
- DIN ISO 29990:2010: Lerndienstleistungen für die Aus- und Weiterbildung – Grundlegende Anforderungen an Dienstleister. Berlin: Beuth Verlag.
- DIN EN ISO 9001:2008: Qualitätsmanagementsysteme – Anforderungen (ISO 9001:2008); Dreisprachige Fassung EN ISO 9001:2008. Berlin: Beuth Verlag.
- CARNEVALE, A.P. & SCHULZ, E.R. (1990): Return on Investment. *Accounting for Training*. *Training and Development Journal*, 7, 1–32.
- EDUQUA (2004): Handbuch. Information über das Verfahren. Anleitung zur Zertifizierung. http://www.eduqua.ch/pdf/eduqua_handbuch.pdf (31.10.10).
- GESSLER, M. (2005): Gestaltungsorientierte Evaluation und der Return on Investment von Weiterbildungsprogrammen. *bwp@*, Nr. 9, http://www.bwpat.de/ausgabe9/gessler_bwpat9.pdf (31.10.10).
- GÜLPEN, B. (1996): Evaluation betrieblicher Verhaltenstrainings unter besonderer Berücksichtigung des Nutzens. Mering: Hampp.
- HAMBLIN, A.C. (1974): *Evaluation and Control of Training*. New York: McGraw-Hill.
- HOLTON, E.F., III (1996): The flawed four level evaluation model. *Human Resource Development Quarterly*, 7 (1), 5–21.
- HOLTON, E.F., III, BATES, R. & RUONA, W. (2000): Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly*, 11, 333–360.
- KAUFFELD, S., BATES, R., HOLTON E.F. III & MÜLLER, A.C. (2008): Das deutsche Lerntransfer-System-Inventar (GLTSI): Psychometrische Überprüfung der deutschsprachigen Version. *Zeitschrift für Personalpsychologie*, 7 (2), 50–69.
- KAUFMAN, R., & KELLER, J. M. (1994): Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly*, 5, 371–380.
- KELLNER, H. (2006): *Value of Investment. Neue Evaluierungsmethoden für Personalentwicklung und Bildungscontrolling*. Offenbach: Gabal.
- KIRKPATRICK, D.L. (1960): Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 4, 13–32.
- KIRKPATRICK, D.L. (1998): *Evaluating Training Programs. The Four Levels*. San Francisco: Berrett-Koehler.
- KIRKPATRICK, J.D. & KIRKPATRICK, D.L. (2005): *Transferring Learning to Behavior. Using the Four Levels to Improve Performance*. San Francisco: Berrett-Koehler.
- KIRKPATRICK, J.D. & KIRKPATRICK, D.L. (2006): *Evaluating Training Programs. The Four Levels*. 3. Auflage. San Francisco: Berrett-Koehler.

- NOE, R.A. & SCHMIDT, N. (1986): The influence of trainee attitudes on training effectiveness: test of a model. *Personnel Psychology*, 39, 497–523.
- PHILLIPS J. (Hrsg.) (1994): *In Action. Measuring Return on Investment. Volume 1*. Alexandria, VA: American Society for Training and Development.
- PHILLIPS, J. & SCHIRMER, F. (2005). *Return on Investment in der Personalentwicklung. Der 5-Stufen-Evaluationsprozess*. Berlin: Springer.
- RÄDIKER, S. (2010): Professionelle Erfolgskontrolle. Wie beschreiben Weiterbildungsorganisationen ihre Evaluation der Bildungsprozesse? In: ZECH, R., DEHN, C., TÖDT, K., RÄDIKER, S., MRUGALLA, M. & SCHUNTER, J. (Hrsg.), *Organisationen in der Weiterbildung: Selbstbeschreibungen und Fremdbeschreibungen*. Wiesbaden: VS Verlag, 132–168.
- STUFFLEBEAM, D.L. (2003): The CIPP Model for Evaluation. In: STUFFLEBEAM, D.L. & KELLAGHAN, T. (Hrsg.), *The International Handbook of Educational Evaluation*. Boston: Kluwer Academic Publishers, 31–62.
- VAN BUREN, M. & ERSKINE, W. (2002): *ASTD State of the Industry Report*. Washington D.C.: ASTD.
- WARR, P.B., BIRD, M. & RACKHAM, N. (1970): *The Evaluation of Management Training*. London: Gower Press.
- ZECH, R. (2006): *Lernerorientierte Qualitätstestierung in der Weiterbildung. Leitfaden für die Praxis*. Hannover: Expressum.

Anschrift der Autoren: Prof. Dr. Michael Gessler, Universität Bremen, Fachbereich Erziehungs- und Bildungswissenschaften (FB 12), ITB Institut Technik und Bildung. Am Fallturm 1, 28359 Bremen. E-Mail: mgessler@uni-bremen.de.

Dipl.-Päd. Andreas Sebe-Opfermann, Universität Bremen, Fachbereich Erziehungs- und Bildungswissenschaften (FB 12), ITB Institut Technik und Bildung. Am Fallturm 1, 28359 Bremen. E-Mail: andreas-so@uni-bremen.de.