

FACHHOCHSCHULE NORDOSTNIEDERSACHSEN
FACHBEREICH WIRTSCHAFTSPSYCHOLOGIE

Diplomarbeit zur Erlangung des Grades
einer Diplom-Wirtschaftspsychologin (FH),
eines Diplom-Wirtschaftspsychologen (FH)

Thema

**Evaluation einer betrieblichen Personalentwicklungs-
maßnahme: Theoretische Fundierung, praktische
Anwendung und Möglichkeiten zur Implementierung**

Erstgutachter: Prof. Dr. Jürgen Deller

Zweitgutachterin: Dipl.-Ing. Christine Voigtländer

Eingereicht von:

Katharina Niehoff & Marius Werth

Marcus-Heinemann-Straße 7

21337 Lüneburg

E-Mail:

katharina_niehoff@yahoo.de

mar.werth@gmx.de

Lüneburg, den 14. Mai 2004

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis.....	V
Abkürzungsverzeichnis	VI
Zusammenfassung.....	VII
Abstract	VIII
1 Einleitung	1
2 Begriffsdefinitionen.....	8
2.1 Training	8
2.2 Evaluation.....	8
3 Relevante theoretische Konzepte und zugrunde liegende Modelle.....	12
3.1 Das Modell von Kirkpatrick.....	13
3.1.1 Beschreibung der 4 Evaluationsebenen.....	14
3.1.2 Erfassung der unterschiedlichen Evaluationsebenen	19
3.1.3 Beziehungen innerhalb des Modells	20
3.1.3.1 Erweiterung des Modells nach Alliger et al.	23
3.1.3.2 Korrelationen zwischen den Ebenen	25
3.1.4 Kritik am Kirkpatrick-Modell	28
3.1.5 Alternative Evaluationsmodelle	30
3.2 Die Analyse des Return on Investment nach Phillips	32
3.2.1 Möglichkeiten zur Erhebung von Nutzen und Kosten eines Trainings	33
3.2.2 Die Berechnung des ROI.....	35
3.2.3 Vor- und Nachteile des ROI.....	36
3.3 Utility Analysis in der Personalentwicklung.....	38
3.3.1 Formel zur Nutzenerfassung nach Schmidt, Hunter und Pearlman	39
3.3.2 Methoden zur Bestimmung der Parameter.....	40
3.3.2.1 Schätzung der Effektgröße	40
3.3.2.2 Methoden zur Schätzung der SD_y	41
3.3.2.2.1 Das globale Schätzverfahren.....	42
3.3.2.2.2 CREPID-Methode	43
3.3.2.2.3 40-Prozent-Regel.....	44
3.3.2.3 Schätzung der Dauer von Trainingseffekten.....	45
3.3.3 Weiterentwicklungen der Formel zur Nutzenerfassung nach Schmidt, Hunter und Pearlman	46
3.3.3.1 Erweiterung um drei ökonomische Konzepte nach Boudreau	46
3.3.3.2 Berücksichtigung des betroffenen Tätigkeitsbereiches nach Gülpen	47
3.3.4 Kritik an der Utility Analysis	47
4 Auswirkungen verschiedener Transferbedingungen auf den Trainingserfolg	52
4.1 Trainingsdesign und Bedarfsanalyse.....	53
4.2 Berücksichtigung von Traineeigenschaften	53
4.3 Transferkritische Umgebungsfaktoren.....	55
4.4 Konsequenzen für den Evaluationsprozess	56
5 Evaluationsstudie.....	58
5.1 Trainingsbeschreibung	58
5.1.1 Software integriertes Förderprogramm für Teilprojektleiter (SWIFT).....	59

5.1.2	Auswahl der Trainingsteilnehmer	59
5.1.3	Kosten des Moduls	59
5.1.4	Zielsetzung des Moduls.....	60
5.1.5	Derzeitige Evaluation.....	60
5.1.6	Hypothesen.....	60
5.2	Methode.....	61
5.2.1	Stichprobenbeschreibung	61
5.2.2	Zufriedenheitsmessung.....	62
5.2.3	Lernerfolgsmessung	63
5.2.4	Transfererfolgsmessung	63
5.2.5	Auswertungsmethoden	64
5.3	Ergebnisse	65
5.3.1	Zusammenfassende Darstellung.....	65
5.3.2	Überprüfung der Hypothesen.....	67
5.4	Diskussion der Studie.....	69
6	Ausblick.....	75
6.1	Anwendbarkeit und Akzeptanz verschiedener Evaluationsansätze	75
6.2	Trends bei der Gestaltung zukünftiger PE-Programme	77
6.2.1	Implementierung des Trainings in die Organisation	78
6.2.2	Zunehmender Einsatz von Technologien.....	78
6.3	Berücksichtigung der aufgezeigten Trends durch die Evaluation.....	79
6.3.1	Integratives Modell zur formativen Evaluation	79
6.3.2	Folgewirkungen der Evaluation	81
6.3.3	Evaluation von Softskill-Trainings	81
6.3.4	Trainingsstandards	82
6.3.5	Wo sollte Evaluation ansetzen?.....	83
	Literaturverzeichnis.....	IX
	Verzeichnis der Anhänge	XX

Abbildungsverzeichnis

Abb. 1	Angebot und systematische Erfolgskontrolle von Weiterbildungskursen im europäischen Vergleich	5
Abb. 2	Ergebnisse einer Unternehmensbefragung	19
Abb. 3	Prozentuale Verteilung der in der Literatur gefundenen Datensätze für die Ebenen 1-4. Ergebnisse einer Metaanalyse über die veröffentlichte Literatur zur Trainingsevaluation der Jahre 1960-2000	19
Abb. 4	Differenzierte Betrachtung der Zufriedenheit der Trainingsteilnehmer über die sechs Zufriedenheitsfaktoren.....	66
Abb. 5	Erreichte Punktzahl in den Fragebogen zur Lernebene. Vergleich von Pre- und Post-Lernfragebogen.....	68

Tabellenverzeichnis

Tab. 1	Das erweiterte Modell von Alliger et al. (1997)	24
Tab. 2	Evaluationsmodelle im Vergleich	30
Tab. 3	Kosten-Klassifikationsmatrix.....	35

Abkürzungsverzeichnis

Abb.	Abbildung	N	Anzahl der Personen
ASTD	American Society for Training and Development	o. Ä.	oder Ähnliche(s)
Aufl.	Auflage	p.	page
bzw.	beziehungsweise	p. a.	pro anno
β	standardisierter Regressionskoeffizient	pp.	pages
C	Kosten des Trainings pro Teilnehmer	PE	Personalentwicklung
ca.	circa	r	Korrelationskoeffizient
d. h.	das heißt	ROI	Return on Investment
EADS	European Aeronautic Defence and Space Company	S.	Seite(n)
Ed.	Editor	SDy	Standardabweichung der Leistung
ed.	edition	sog.	so genannt
Eds.	Editors	Suppl.	Supplement
erw.	erweiterte	SW	Software
et al.	et alii	SWIFT	Software integriertes Förderprogramm für Teilprojektleiter
etc.	et cetera	Tab.	Tabelle
ggf.	gegebenenfalls	TN	Teilnehmer
HR	Human Resource	u. a.	unter anderem, unter anderen
HRD	Human Resource Development	überarb.	überarbeitete
HRO	Human Resources Operations	ΔU	Nutzen des Trainingsprogramms ausgedrückt in Geldeinheiten
HRP	Human Resources Policy	Verf.	Verfasser
Hrsg.	Herausgeber	vgl.	vergleiche
i. d. R.	in der Regel	Vol.	Volume
i. e.	id est	vollst.	vollständig
IT	Informationstechnologie	z. B.	zum Beispiel
Kap.	Kapitel		

Zusammenfassung

Vor dem Hintergrund des wachsenden Kostendrucks für Unternehmen und der zunehmenden Forderung nach einem Nachweis der Wirtschaftlichkeit von Weiterbildungsmaßnahmen wird in der vorliegenden Arbeit die Notwendigkeit und die Bedeutung der theoriegeleiteten Trainingsevaluation für die Praxis herausgestellt. Es werden relevante theoretische Konzepte und zugrunde liegende Evaluationsmodelle diskutiert und Faktoren genannt, die den Transfer von Trainingsinhalten beeinflussen können.

Evaluationsinstrumente zur Erfolgsmessung wurden entwickelt und beispielhaft in einer Pilotstudie mit 11 Teilnehmern eines Software-Trainings angewendet, um die Praktikierbarkeit einer theoriegeleiteten Trainingsevaluation aufzuzeigen. Neben der Messung von Zufriedenheit und Lernleistung der Teilnehmer wird die Erfassung der Transferleistung diskutiert und die Berechnung des Return on Investment der Trainingsmaßnahme sowie eine Utility Analysis zur Bestimmung des Trainingsnutzens durchgeführt. Die Instrumente zur Datenerhebung sind eigens entwickelte Fragebogen, zur anschließenden Auswertung der Daten werden verschiedene Kenngrößen der deskriptiven Statistik und Methoden der statistischen Testtheorie herangezogen. Bisherige Forschungsergebnisse, die aufzeigen konnten, dass kein Zusammenhang zwischen der Zufriedenheit der Teilnehmer mit dem Training und ihrer Lernleistung besteht, können bestätigt werden. Die durchgeführten Berechnungen können zeigen, dass der Nutzen des Trainings dessen Kosten bei weitem überwiegt.

Erst durch die Offenlegung des monetären Nutzens von Weiterbildungsmaßnahmen können diese als zukunftssträchtige Investitionen kommuniziert werden. Verantwortlichen der Personalentwicklung gelingt es somit, Investitionen in Training und Entwicklung auch aus ökonomischer Sicht zu rechtfertigen. Diskutiert wird die Anwendbarkeit und Akzeptanz der verschiedenen Evaluationsansätze und die bestehende Notwendigkeit eines frühzeitigen Beginns des Evaluationsvorhabens im Personalentwicklungsprozess.

Abstract

Today's business demands of companies to calculate tight and to act cost-effective in every area. Thus also the Human Resource Department has to furnish proof of its training and development programs. Against this background the paper on hand demonstrates the necessity and relevance of theoretically-based training evaluation for practical settings. Relevant concepts and models in the context of evaluation are discussed. Moreover factors influencing the transfer of training are presented.

Evaluation tools to measure the success of training programs are developed. In order to show how these theoretically-based tools can be applied they are used on a study with 11 participants who took part in a software-training. The participants' satisfaction and learning is measured, possibilities to register transfer are discussed and the return on investment as well as a utility analysis are calculated to assess the program's profit. Questionnaires to raise data are developed especially for this purpose; methods out of the field of descriptive statistics and statistical testing theory are used to analyse the data. The study confirms previous research results which found that there is no correlation between the trainee's satisfaction and learning. The calculations demonstrate that the training's profit surpasses its costs by far.

In order to justify current and future personnel programs the training's benefit has to be expressed in monetary values. On this basis it can be proved that investments in training and development are profitable. Furthermore the usability and acceptance of different evaluation models are discussed and the necessity to start with the evaluation in an early stage of the personnel development process is explained.

1 Einleitung

Die Ökonomie zu Beginn des 21. Jahrhunderts zeichnet sich durch hohe Arbeitslosenzahlen, zunehmende Finanznot der öffentlichen Haushalte, reformbedürftige Sozialsysteme und weitreichende demographische Veränderungen aus. Gleichzeitig verkürzen sich Innovationszyklen, der internationale Wettbewerb nimmt zu und neue Märkte – besonders im asiatischen Raum – werden erschlossen. Individuelle und kollektive Wissensbestände verändern sich kontinuierlich und Führungskräfte sowie Mitarbeiter stehen somit vor der Aufgabe, Anpassungsleistungen in immer kürzeren Zyklen erbringen zu müssen (Sonntag, 2002). Aktuelle wirtschaftliche Entwicklungen erfordern kreative, flexible, kompetente und eigenverantwortlich handelnde Mitarbeiter, um immer wieder neuen Anforderungen gerecht werden und das eigene Unternehmen erfolgreich am Markt behaupten zu können. In dieser schnelllebigen Zeit stellt „Brainpower“ (Bassi & Van Buren, 1998, ¶ 7) oftmals die einzige Möglichkeit dar, Wettbewerbsvorteile zu erzielen und Marktanteile auszubauen. Führende Wirtschaftsvertreter bezeichnen aus diesen Gründen ihre Mitarbeiter zunehmend als „the most important asset“ (Ruona, Leimbach, Holton & Bates, 2002, p. 218) des Unternehmens. Zur Sicherstellung organisationaler Effektivität ist ein Ressourcenmanagement gefordert, das auf kontinuierliche Weiterentwicklung der beruflichen Handlungskompetenz ausgerichtet ist (Ford, 1997; Sonntag, 2002). Stehen keine erstklassigen Mitarbeiter mit den künftig erforderlichen Kompetenzen sowie entsprechender Loyalität und Motivation zur Verfügung, so sind die Wachstumsaussichten eines Unternehmens limitiert, womit zeitgleich auch das Bestehen im internationalen Wettbewerb gefährdet ist (Claßen, 2003). Die Personalentwicklungsabteilungen (PE-Abteilungen) der Unternehmen stehen vor der Aufgabe, das Wissen der Mitarbeiter fortlaufend zu aktualisieren und diese somit bestmöglich auf derzeitige und zukünftige Aufgaben vorzubereiten. Thierau (1991) betont in diesem Zusammenhang, dass Investitionen, die dazu dienen, qualifizierte Mitarbeiter heraus- oder weiterzubilden, im Vergleich zu Investitionen in das Sachvermögen nicht als minder wichtig angesehen werden dürfen. Vielmehr sind es laut Mandl (2003) „entscheidende Zukunftsinvestitionen“ (S. 296), die sowohl der Weiterentwicklung des Einzelnen als auch dem Unternehmenserfolg dienen. Unternehmen können es sich daher auch in Zeiten schwächerer Konjunktur nicht leisten, aus Budgetgründen auf Weiterbildung zu verzichten.

Die Rezession der vergangenen Jahre und der verstärkte internationale Wettbewerb sorgten u. a. dafür, dass Konzepte wie Downsizing, Restrukturierung sowie Mergers & Acquisitions

einen festen Platz in der Sprache und im Agieren von Organisationen eingenommen haben, da sie das Potenzial aufweisen, Kostensenkungspotenziale aufzuspüren und somit Einsparungen zu ermöglichen. Ruona et al. (2002) beobachteten im Jahre 2002: „With the economy experiencing its current downturn, it is already clear that organisations are shaving every expense that does not promise a return“ (p. 218). Zwei Jahre später erwarten die großen deutschen Forschungsinstitute zwar wieder ein Wachstum des realen Bruttoinlandsproduktes von 1,5 bis 1,8% (Informationsdienst des Instituts der Deutschen Wirtschaft Köln, 2004), es wird aber dennoch Zeit verstreichen, bis die Stagnation der letzten drei Jahre überwunden ist. Kostenaspekte werden auch in den kommenden Jahren noch im Vordergrund organisationaler Entscheidungen stehen. Als ein Überbleibsel der Konjunkturschwäche weist der für Investitionen zur Verfügung stehende Etat heute einen weitaus geringeren Umfang als noch in Vorjahren auf und es gilt, bei der Verteilung der knappen Ressource Geld, den Rückfluss einer Investition (Return on Investment) genau zu kalkulieren.

Im innerbetrieblichen Verteilungskampf um die begrenzten monetären Mittel müssen alle Abteilungen der Unternehmen den Aufwand, den sie betreiben, legitimieren können – ansonsten droht eine Kürzung der Etats. In einer Zeit, in der Gelder längst nicht mehr nach dem Gießkannenprinzip verteilt werden, wird auch der Wertbeitrag von unterstützenden Organisationseinheiten wie z. B. den PE-Abteilungen, die bisher selten den Erfolg ihrer Handlungen anhand von Daten nachweisen mussten, einer verstärkten Überprüfung durch die Controlling-Verantwortlichen unterzogen (Cascio, 2000), da sie häufig relativ hohe Summen zur Unterstützung ihrer Aktivitäten beanspruchen. Unternehmen in den USA investieren jährlich mehr als \$50 Milliarden in Weiterbildungsmaßnahmen, um das Potenzial ihrer Mitarbeiter optimal zu nutzen (Galvin, 2003). Die jährlichen Aufwendungen deutscher Unternehmen der gewerblichen Wirtschaft für betriebliche Weiterbildung liegen bei knapp 17 Milliarden Euro; dieser Wert blieb seit 1998 fast konstant (Weiß, 2003). Dieser Aufwand mag notwendig sein, um das Ziel zu erreichen, das Wissen und die Fähigkeiten der eigenen Mitarbeiter kontinuierlich auf dem neuesten Stand zu halten, er muss jedoch im Wettstreit um die begrenzten Ressourcen vor der Geschäftsleitung gerechtfertigt werden (Gloger, 2004). „If human resource professionals are to compete with sales, marketing, production and other functional areas for scarce resources, then the expected dollar returns (utility) of personnel programs must be evaluated against the utility of alternative investments“ (Cascio, 1989, p. 67). Abteilungen wie z. B. das Marketing, die Produktion oder der Vertrieb weisen jedoch im Vergleich zur Personalentwicklung häufig effizientere Kontrollsysteme zur Darlegung der Wirksamkeit

ihrer Aktivitäten auf und arbeiten mit Fakten und Kennzahlen, welche den Konzernleitungen vertraut sind. Qualitative Argumente aus dem Personalressort finden jedoch nicht immer eine angemessene Berücksichtigung, da Entscheidungen um knappe Ressourcen zunehmend auf Kosten/Nutzen-Betrachtungen basieren. (Claßen, 2003). Die Geschäftsführung möchte wissen, welchen Return on Investment ein in die Weiterbildung investierter Euro hat (Gloger, 2004; Shapiro, 1995) und trifft auf dieser Basis Verteilungsentscheidungen. Der Stellenwert von Weiterbildung wird von der Unternehmensleitung meist erst dann erkannt und geachtet, wenn auch hier mit glaubwürdigen Zahlen belegt wird, dass PE-Aktivitäten tatsächlich einen Mehrwert kreieren (Aldisert, 2002).

Laut der Studie des Bundesinstituts für Berufsbildung gehen im Jahre 2001 von 474 befragten deutschen Unternehmen 94% von einer zunehmenden Kosten- und Effizienzkontrolle der betrieblichen Weiterbildung in den kommenden Jahren aus (Grünewald, Moraal & Schönfeld, 2003) – das Bewusstsein über die Wichtigkeit der Nutzenerfassung von PE scheint also vorhanden zu sein. Tatsächlich gelingt es dem Human Resource Development (HRD) bisher jedoch nur selten, elaborierte Evaluationstechniken und deren Anwendung aufzuweisen. Über Evaluation wird im Bereich Personal nach wie vor mehr geredet, als dass sie praktiziert wird (Gloger, 2004). Die Verantwortlichen des HRD werden es folglich schwer haben, sich im Wettstreit um die knappen Ressourcen zu behaupten, solange sie nicht auch einen glaubwürdigen Nachweis über den Nutzen von Trainings für das Unternehmensergebnis erbringen und damit den Wertbeitrag ihrer Entwicklungsprogramme aufzeigen. Allein die Offenlegung des Nutzens von PE-Programmen wird zukünftige Investitionen in diese rechtfertigen (Cascio, 2000; Delahaye, 2000; Mattson, 2003; Pfau & Cohen, 2003; Phillips & Phillips, 2001; Salas & Kosarzycki, 2003). „The investment made in training must be justified in terms of improved organizational performance – increased productivity, profit, or salary; reduced error, enhanced market share“ (Salas & Cannon-Bowers, 2001, p. 472). Personalentwicklung muss zukünftig als rentable Investition gesehen werden, wenn sie als potenzieller Anwarter auf Teile des begrenzten Budgets verstanden werden möchte.

Obwohl dem Thema Trainingsevaluation in letzter Zeit starke Beachtung zuteil wird, findet diese in der Praxis bisher kaum Verwendung (Salas & Kosarzycki, 2003). Thierau-Brunner, Stangel-Meseke und Wottawa (1999) sowie Salas und Kosarzycki (2003) führen zahlreiche Gründe für die mangelnde Evaluationsaktivität der Unternehmen an. Diese reichen von mangelndem Interesse und fehlendem Evaluationsbewusstsein über Unkenntnis der Methoden und

deren Möglichkeiten bis hin zu überhöhten Kostenerwartungen sowie Bewertungsangst der Trainer und Teilnehmer. Cascio (2000) fasst das Problem der oftmals unzureichenden Evaluation von Weiterbildungsmaßnahmen wie folgt zusammen:

Unfortunately, billions will be spent providing training and development programs, but little will be spent assessing their social and financial outcomes. . . . Decision makers thus remain unguided by systematic evaluations of past experiments and uninformed about the costs and benefits of alternative HRD programs when considering training efforts in their own organizations. „Billions for training, but not one cent for evaluation” is an exaggerated but not entirely false characterization of present training practice in many organizations. (p. 286)

Wolff (2003) verwendet in diesem Kontext den Begriff der Personalökonomik und fordert, den Nutzen einiger PE-Maßnahmen kritisch zu hinterfragen, um nicht länger fälschlicherweise in ineffiziente Programme zu investieren.

Im Jahre 2000 wurde im Rahmen der zweiten europäischen Weiterbildungserhebung eine Studie mit dem Titel „Assessment of the Second Continuing Vocational Training Survey (CVTS 2)“ in 25 europäischen Staaten durchgeführt (Grünewald, Moraal & Schönfeld, 2003). Die Studie, in der weit über 76.000 Unternehmen (davon 3.184 aus Deutschland) befragt wurden, hatte u. a. zum Ziel, die systematische Erfolgskontrolle von Weiterbildungsprogrammen im Ländervergleich zu erheben. Es wurde zum einen gemessen, welcher Anteil der Unternehmen eines Landes Weiterbildungsmaßnahmen anbietet, zum anderen wurde der prozentuale Anteil der weiterbildenden Unternehmen angegeben, die ihre Maßnahmen evaluieren. Ein Vergleich von acht ausgewählten europäischen Ländern wird in Abbildung 1 aufgezeigt. Um einen Ländervergleich vornehmen zu können, wird in unserer Darstellung die Evaluationstätigkeit der Unternehmen mit deren Angebot an Weiterbildungskursen in Beziehung gesetzt.

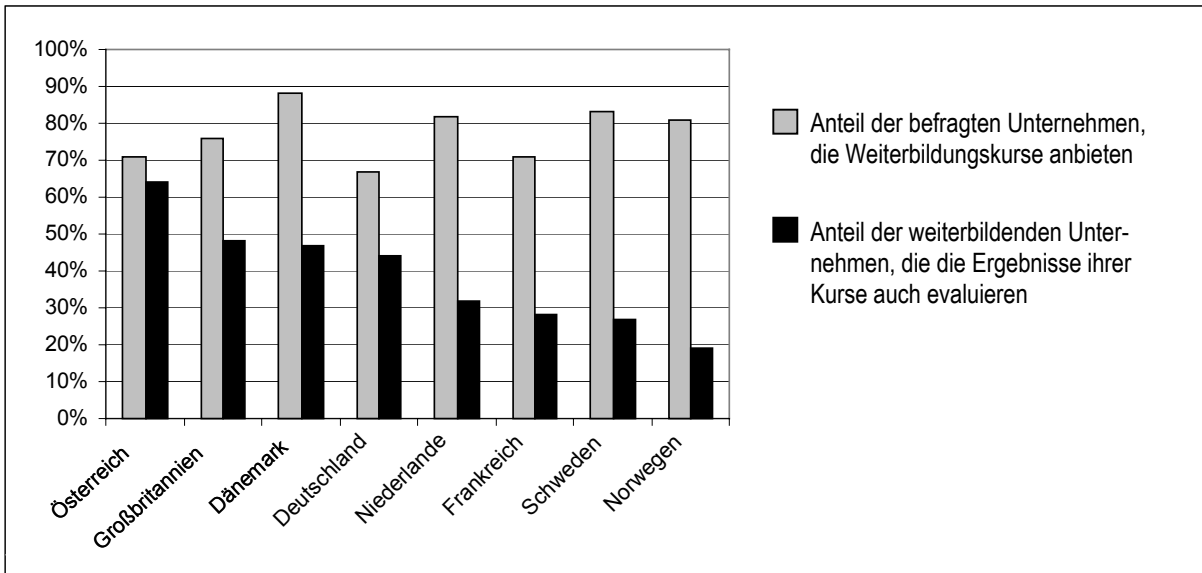


Abbildung 1: Angebot und systematische Erfolgskontrolle von Weiterbildungskursen im europäischen Vergleich.

Quelle: Eigene Darstellung nach: *Betriebliche Weiterbildung in Deutschland und Europa* (S. 20, 66), von U. Grünewald, D. Moraal, G. Schönfeld, 2003, Bielefeld: Bertelsmann.

Anhand der Abbildung 1 wird z. B. deutlich, dass 71% der österreichischen Unternehmen Weiterbildungsprogramme anbieten und 64% dieser Unternehmen ihre Programme auch evaluieren. In Deutschland bieten laut der aufgezeigten Studie 67% der Unternehmen Weiterbildungsprogramme an, allerdings kontrollieren nur 44% dieser Unternehmen laut eigenen Angaben auch den Erfolg ihrer Programme. Die geringsten Evaluationsaktivitäten weist in dieser Auswahl europäischer Länder Norwegen mit 19% auf; ein sehr geringer Anteil, wenn man beachtet, dass 81% der befragten norwegischen Unternehmen Weiterbildungskurse anbieten.

Die mangelnde Erfolgsüberprüfung im HRD-Bereich kann neben der schwachen Konjunktur der vergangenen Jahre als einer der Gründe dafür angeführt werden, warum das Trainingsbudget amerikanischer Unternehmen mit einer Mitarbeiterzahl ≥ 100 im Jahre 2003 zum zweiten Mal in Folge im Vergleich zum Vorjahr um etwa 5% zurückgegangen ist. Betrug im Jahre 2001 laut „2003 Industry Report“ (Galvin, 2003) das jährliche Trainingsbudget der Unternehmen in den USA noch \$56,8 Milliarden, sank es in den Jahren 2002 und 2003 auf \$54,2 bzw. \$51,3 Milliarden. In Deutschland führt das Institut der deutschen Wirtschaft alle drei Jahre eine Erhebung der Weiterbildungstätigkeit der gewerblichen Wirtschaft durch. Die letzte Erhebung fand 2001 statt und konnte bereits zu dem Zeitpunkt einen erheblichen Rückgang der Weiterbildungsaktivitäten deutscher Unternehmen feststellen (Weiß, 2002, 2003). Laut der Studie des Instituts der deutschen Wirtschaft entfielen im Jahre 2001 im Durchschnitt 13,6 Stunden off-the-job-Weiterbildung auf jeden Mitarbeiter, was im Vergleich zu der vorangehenden Erhebung aus dem Jahre 1998 einen Rückgang von 6,4 Stunden pro Mitarbeiter darstellt. Die an der Erhebung beteiligten Betriebe wandten 2001 im Durchschnitt einen Betrag

von 869 Euro je Mitarbeiter für die Weiterbildung auf; dieser Wert liegt um 23% unter dem Wert des Jahres 1998 (1128 Euro). Zur Zeit sind keine Studien verfügbar, die aktuellere Zahlen zu den Weiterbildungsaktivitäten der deutschen Wirtschaft liefern. Diese Entwicklungen verdeutlichen, dass Personalentwickler nach Wegen suchen müssen, den Geldwert von Weiterbildung nachzuweisen, um Etatverantwortliche davon zu überzeugen, auch in wirtschaftlich angespannten Zeiten in Personalentwicklung zu investieren. „The increased importance of training coupled with decreased fixed funding has accentuated the need for effective training evaluation” (Kraiger & Jung, 1997, p. 151). Diese Aussage von Kraiger und Jung (1997) hat unter den heute herrschenden und oben aufgezeigten ökonomischen Bedingungen noch an Bedeutung gewonnen, darüber hinaus bezeichnen Phillips und Phillips (2001) Trainingsevaluation als „truly a global issue“ (p. 241).

Dieser Diplomarbeit liegt die Problemstellung zugrunde, dass trotz vorhandener Modelle und praktischer Richtlinien in der Unternehmenspraxis nach wie vor Weiterbildungsmaßnahmen nur selten umfassend evaluiert werden und Personalverantwortliche somit selten in der Lage sind, den Nutzen sowie den Beitrag der Programme zur Verbesserung der Gewinnsituation des Unternehmens aufzuzeigen. Bereits vor 15 Jahren bezeichnete Fitz-enz (1988) die Annahme, der Nutzwert von Entwicklungsprogrammen könne weder identifiziert noch kalkuliert werden, als einen Mythos. Dennoch ist die Vorstellung, dass der Nutzen einer Trainingsmaßnahme – wenn überhaupt – nur schwer nachgewiesen werden kann, in der Praxis auch heute noch weit verbreitet und führt dazu, dass häufig nur kurzfristige Zufriedenheitsmessungen zur Erfassung des Erfolgs einer Maßnahme durchgeführt werden, wohingegen Transferleistungen und organisationale Auswirkungen in Folge eines Trainings unberücksichtigt bleiben (vgl. Kap. 3.1.2). Eine wachsende Anzahl an Publikationen zu diesem Thema sowie vermehrte veröffentlichte Praxisanwendungen lassen Entscheidungsträger verstärkt erkennen, dass Möglichkeiten zur Trainingsevaluation bestehen. Deren Anwendung wird daher verstärkt von Unternehmensleitungen gefordert (Phillips & Phillips, 2001). Laut Remdisch und Utsch (2004) wird die Nachfrage nach Evaluationen im Bereich der Personal- und Organisationsentwicklung in Zukunft zunehmen. Oft scheitert ein Evaluationsvorhaben jedoch, da es an geeigneten und auf die jeweiligen Organisationsbedürfnisse zugeschnittenen Konzepten mangelt. Ein weiterer Grund für das Scheitern liegt darin, dass sich besonders neuere Methoden zur Evaluation durch eine hohe Komplexität und die Verwendung statistischer Kenngrößen auszeichnen, was nicht selten zu Missinterpretationen und Unverständnis der Ergebnisse führen kann. Die Evaluation von PE-Maßnahmen geht daher oftmals mit hoher Unsicherheit seitens der

Betroffenen einher – ein weiterer Aspekt, der die Durchführung dieser im Unternehmen hemmt.

Aus den oben genannten Gründen wird in der vorliegenden Arbeit die Notwendigkeit der Evaluation von Trainingsmaßnahmen und die Bedeutung der theoriegeleiteten Trainingsevaluation für die Praxis herausgestellt. Unser Ziel ist es, die Anwendbarkeit einer umfassenden Trainingsevaluation am Beispiel eines Trainings der international agierenden European Aeronautic Defence and Space Company (EADS) zu veranschaulichen. Für dieses Vorhaben werden unterschiedliche Konzepte und Modelle herangezogen, welche sich bezüglich ihres Informationsgehaltes unterscheiden und somit ergänzende Aspekte für die Messung der Trainingseffektivität liefern. Die im Unternehmen durchgeführte Evaluationsstudie wird als Handlungsschema verstanden, welches der EADS eine Generalisierbarkeit auf weitere PE-Programme ermöglichen soll, um so in Zukunft Trainingsmaßnahmen anhand der hier vorgestellten Pilotstudie in Eigenregie zu evaluieren.

Zu Beginn der vorliegenden Arbeit werden in Kapitel 2 die Begriffe Training und Transfer sowie Evaluation definiert, da sie zentraler Bestandteil sämtlicher Kapitel sind. Anschließend werden relevante theoretische Konzepte und zugrunde liegende Modelle vorgestellt und diskutiert (Kap. 3). Dazu gehören u. a. das 4-Ebenen-Modell von Kirkpatrick (1959a, 1959b, 1960a, 1960b), die Berechnung des Return on Investment im Trainingskontext von Phillips (1996) sowie die Utility Analysis zur Bestimmung des monetären Trainingsnutzens nach Schmidt, Hunter und Pearlman (1982). Faktoren, die den Transfer von Training beeinflussen werden erörtert und daraus abgeleitete Konsequenzen für den Evaluationsprozess aufgezeigt (Kap. 4). In dem darauf folgenden Kapitel 5 wird eine konkrete Trainingsmaßnahme zum Thema „Systematischer Software-Test“ der Firma EADS beschrieben und die Durchführung einer umfassenden Evaluation auf Basis der aufgezeigten Modelle vorgestellt, um daran anknüpfend die Ergebnisse zu diskutieren. Abschließend wird in Kapitel 6 ein Ausblick auf zukünftige Forschungen auf dem Gebiet der Trainingsevaluation gegeben.

Aus Gründen der Verständlichkeit und Lesbarkeit wird in der vorliegenden Diplomarbeit auf die zusätzliche Nennung der weiblichen Sprachform verzichtet. Wann immer von Teilnehmern, Mitarbeitern, Autoren o. Ä. die Rede ist, sind weibliche Personen mit eingeschlossen.

2 Begriffsdefinitionen

Grundlegende Inhalte der vorliegenden Arbeit sind die Beschreibung einer Trainingsmaßnahme und die Methoden zu deren Evaluation. Zum einheitlichen Verständnis ist es notwendig, wiederkehrende Begriffe zu definieren und wichtige Aspekte angewandter Konzepte zu erklären. Relevante Termini werden nachfolgend erläutert.

2.1 Training

Die Begriffe Training, Personalentwicklungs- und Interventionsmaßnahme sowie Trainings- und Weiterbildungsprogramm werden im Folgenden synonym verwendet. Sie alle umschreiben geplante Programme, in deren Fokus die Optimierung tätigkeitsrelevanter Fähigkeiten steht (Phillips, 1996). Das Ziel derartiger Prozesse besteht darin, Wissen, Fähigkeiten oder Einstellungen teilnehmender Mitarbeiter (sog. Trainees) systematisch und dauerhaft in eine gewünschte Richtung zu lenken (Campbell, Dunnette, Lawler & Weick, 1970). „Training is a set of activities planned to bring about learning in order to achieve goals“ (Hesketh & Frese, 2002, p. 14097). Darüber hinaus beschreiben Arthur, Bennet, Edens und Bell (2003) Training als „one of the most pervasive methods for enhancing the productivity of individuals and communicating organizational goals to new personnel“ (p. 234). Bramley (1996) definiert Training als eine Methode, Mitarbeiter dabei zu unterstützen, effektiver und effizienter in ihrem jeweiligen Tätigkeitsfeld agieren zu können. Hager und Hasselhorn (2000) verstehen unter dem Begriff Training den Aufbau längerfristig verfügbarer Fertigkeiten, Fähigkeiten und/oder Einstellungen, deren Anwendung nicht nur auf Aufgaben- und Problemstellungen innerhalb der Maßnahme beschränkt bleibt, sondern darüber hinaus auf den Arbeitskontext transferiert wird. Die Definition von Noe, Hollenbeck, Gerhart und Wright (2004) fasst die verschiedenen Charakteristika eines Trainings zusammen: „Training consists of an organization’s planned efforts to help employees acquire job-related knowledge, skills, abilities, and behaviors, with the goal of applying these on the job“ (p. 200). Eine umfassende Begriffsbestimmung von Training ist unserer Meinung nach nur durch die additive Zusammenschau verschiedener Definitionsansätze möglich.

2.2 Evaluation

„Definitions of evaluation abound“ (Glass & Ellet, 1980, p. 212). Auch Wottawa und Thierau (1998) beschreiben die Evaluation als einen außerordentlich vielfältigen Begriff, der eine Menge möglicher Verhaltensweisen umfasst und sich somit prinzipiell einer abstrakten, die Wirklichkeit gleichzeitig voll umfassenden Definition entzieht. Laut Schmidt (2004) weist das Verständnis von Evaluation ein Differenzierungsniveau auf, das kaum von Evaluation in

einem allgemeinen Sinne sprechen lässt. Zum Wortfeld der (Trainings-)Evaluation gehören auch Begriffe wie Erfolgskontrolle, Lernfortschrittskontrolle, Bildungscontrolling, Effizienzforschung, Bewertungsforschung und Wirkungs- oder Qualitätskontrolle, die oftmals synonym verwendet werden. Ein allgemeiner Konsens liegt darin, dass all diese Tätigkeiten etwas mit „Bewerten“ zu tun haben. Dabei ist nach Nork (1989) die Evaluation einer Weiterbildungsmaßnahme als ein Prozess zu sehen, der sämtliche Aktivitäten von dem Zeitpunkt an umfasst, zu dem der Weiterbildungsbedarf erkannt wird, bis hin zur Ermittlung des Lern- und Transfererfolges und der Analyse, inwieweit die ursprünglichen Weiterbildungsziele erreicht werden. Für Höft (2001) umfasst die Evaluation im Trainingskontext die Bewertung des eingesetzten Konzeptes, des Untersuchungsplans, der Implementierung und der Wirksamkeit des untersuchten Trainingsprogramms. Rae (2002) bezieht in ihre Definition des Evaluationsbegriffes „the assessment of the total value of a training programme, training system or training course in both value- and *cost-effective* [Hervorhebung durch Verf.] terms (i. e. an effective ROI)“ (p. 2) ein. Damit wird betont, dass der Erfolg einer Trainingsmaßnahme nicht allein durch das Erreichen von Lernzielen oder Verhaltensmodifikationen, sondern auch durch das Verhältnis von aufgewendeten Mitteln zum Nutzen der Maßnahme bestimmt wird. Goldstein und Ford (2002) betonen ebenfalls, dass die Evaluation den Aspekt der „accountability“ (p. 20) eines Trainings herausstellen sollte. Zusammenfassend ist die Trainingsevaluation demnach als planmäßige Datensammlung zu verstehen, die das Ziel verfolgt, die Wirksamkeit einer Trainingsmaßnahme – auch aus ökonomischer Sicht – zu bestimmen, diese Maßnahme zu hinterfragen und ggf. zu modifizieren (Gibb, 2002; Goldstein & Ford, 2002; Noe, 2002).

Dipboye (1997) stellt folgende Anforderungen an eine Trainingsevaluation: Zum einen sollte sie aufzeigen, ob Veränderungen feststellbar sind und ob diese Veränderungen auf das Training zurückgeführt werden können. Des Weiteren sollte erhoben werden, ob die gleichen Veränderungen bei Trainingswiederholung in der gleichen Organisation mit neuen Teilnehmern auftreten. Zudem sollte sie zeigen, ob die gleichen Veränderungen bei Durchführung desselben Trainings in einer anderen Organisation mit neuen Teilnehmern auftreten und damit die Generalisierbarkeit des Trainings gewährleistet ist. Goldstein (1991) und Goldstein und Ford (2002) nehmen eine ähnliche Unterteilung vor und sprechen in diesem Zusammenhang von vier Validitätsebenen, die aufeinander aufbauen: Trainingsvalidität, Transfervalidität, intraorganisationaler Validität und interorganisationaler Validität.

Eine laut Mittag und Hager (2000) weit verbreitete Differenzierung in verschiedene Evaluationsarten stammt von Scriven (1969, 1991). Durch Scriven wurden laut Russ-Eft und Preskill (2001) erstmals die verschiedenen Arten einer Evaluation mit den Begriffen *summativ* und *formativ* gekennzeichnet. Forschungen im Bereich der Trainingsevaluation haben ihren Fokus bisher meist auf summative Evaluation gelegt (Brown & Gerhardt, 2002). „Summative evaluation refers to evaluation conducted to determine the extent which trainees have changed as a result of participating in the training program“ (Noe, 2002, p. 180). Es geht hierbei also um eine Bewertung des Ausmaßes, inwieweit sich die Teilnehmer Wissen, Fähigkeiten, Einstellungs- und Verhaltensveränderungen aneignen oder anderweitige Ergebnisse wie z. B. monetäre Nutzenzuwächse identifiziert werden konnten (Noe, 2002). Diese Art von Evaluation wird daher hauptsächlich nach der Fertigstellung von Programmen geplant und durchgeführt, d. h. wenn die Konzeption des Programms abgeschlossen ist, es eingesetzt werden soll oder bereits implementiert ist (Mittag & Hager, 2000; Scriven, 1991). Es stehen eine zusammenfassende Beurteilung der Wirksamkeit (Bortz & Döring, 2002) und das Bewerten von „outcome measures“ (Goldstein & Ford, 2002, p. 166) der Interventionsmaßnahme im Vordergrund. Zur summativen Trainingsevaluation gehören u. a. die in dieser Arbeit vorgestellten Modelle von Kirkpatrick (1996, 1998a), Alliger, Tannenbaum, Bennet, Traver & Shotland (1997), Kraiger, Ford & Salas (1993) und Phillips (1996) sowie die Utility Analysis (Schmidt, Hunter & Pearlman, 1982). Formative Evaluation „refers to evaluation conducted to improve the training process“ (Noe, 2002, p. 179) und liefert Informationen, wie die Maßnahme optimiert werden kann, damit die im Vorfeld der Maßnahme aufgestellten Trainingsziele erreicht werden können (Goldstein & Ford, 2002; Carnevale & Schulz, 1998). Sie wird meist während der Entwicklung, Ausgestaltung und Erprobung eines Programms durchgeführt (Mittag & Hager, 2000), oftmals sogar mehr als einmal (Scriven, 1991). Erst durch eine formative Evaluation können Gründe für ein bestimmtes Trainingsergebnis aufgezeigt werden (Goldstein & Ford, 2002). Um bestmögliche Evaluationsergebnisse zu erzielen und die Vorteile der jeweiligen Evaluationsart zu nutzen, wird eine Kombination von summativer und formativer Trainingsevaluation empfohlen (Brown & Gerhardt, 2002).

Da unsere Evaluationsstudie nicht dahingehend ausgelegt ist, die Trainingsmaßnahme der Firma EADS zu optimieren, steht dieser formative Aspekt der Evaluation nicht im Mittelpunkt dieser Diplomarbeit. Das primäre Ziel der in Kapitel 5 vorgestellten Studie ist die Überprüfung der Trainingsmaßnahme im Hinblick auf deren Ziel- und Zweckorientierung anhand im Vorfeld festgelegter Kriterien.

Ausgehend von den unterschiedlichen Evaluationsarten hat die Evaluation verschiedene Funktionen, die als Begründungen für deren Einsatz herangezogen werden können. Eine wichtige Rolle spielt nach Nork (1989) ihre *Legitimationsfunktion*, die den Entscheidungsträgern darlegt, dass „die Weiterbildungsmaßnahmen ihre Ziele erreicht und die intendierten Verhaltensmodifikationen herbeigeführt haben, welche die organisatorische Zielerreichung über entsprechende Transferleistungen auch tatsächlich fördern“ (S. 39). Die Evaluation ermöglicht, sowohl die durch die Trainingsprogramme verursachten Kosten als auch Investitionen in zukünftige Weiterbildungsmaßnahmen zu rechtfertigen (Gülpen, 1996), wenn als Ergebnis einer Maßnahme Lernerfolge oder sogar eine Leistungssteigerung im Funktionsfeld (z. B. Steigerungen im Umsatz oder der Kundenzufriedenheit, geringere Krankheitsrate, etc.) festgestellt werden kann. Darüber hinaus besitzt die Evaluation eine *Kommunikationsfunktion*, da sie die Möglichkeit bietet, den Nutzen von Weiterbildung in monetären Größen darzustellen (Gülpen, 1996): „Es kann also in einer Form kommuniziert werden, die allen Entscheidungsträgern – auch Nicht-Psychologen – geläufig ist“ (S. 40). Eine weitere Begründung für die Erfolgskontrolle im Bereich der Personalentwicklung stellt die damit einhergehende Verbesserung gegenwärtiger sowie zukünftiger Maßnahmen dar (*Verbesserungsfunktion*). Die Evaluation liefert eine fundierte Datenbasis, die Informationen über die Effizienz und Effektivität untersuchter Programme bereitstellt und somit eine Basis bietet, Optimierungen mit dem Ziel vorzunehmen, zukünftig einen höheren Nutzen mit den Weiterbildungsinvestitionen zu erzielen (Nork, 1989). Aus den Evaluationsergebnissen können Verbesserungsvorschläge für die Maßnahme abgeleitet werden und ggf. kann auf dieser Basis über die zukünftige Form sowie das Bestehen des Trainings entschieden werden (Wottawa & Thierau, 1998). Dieser letztgenannte Aspekt betrifft die *Entscheidungsfunktion* der Evaluation. Die bereits im Rahmen der Verbesserungsfunktion angesprochene fundierte Datenbasis liefert Fakten, die Unsicherheit reduzieren und Entscheidungsträgern dazu verhelfen, Entschlüsse auf einer rationaleren und objektiveren Grundlage zu treffen. Die Evaluation dient somit dazu, Entscheidungen über die Fortsetzung, die Modifikation oder den Abbruch einer Maßnahme und über den weiteren Trainingsbedarf der Teilnehmer zu unterstützen (Sackett & Mullen, 1993). Demzufolge sei die Evaluation nur dann sinnvoll, wenn sie in die Entscheidungsprozesse einbezogen wird und diese beeinflussen kann (Nork, 1989). In diesem Sinne bezeichnen auch Goldstein und Ford (2002) die Trainingsevaluation als eine „systematic collection of descriptive and judgmental information necessary to make effective training *decisions* [Hervorhebung durch Verf.] related to the selection, adoption, value, and modification of various training activities“ (p. 177). Die Evaluation dient somit als Planungs- und Entscheidungshilfe (Wottawa & Thie-

rau, 1998). Die vier ausgewählten Funktionen (Legitimation, Kommunikation, Verbesserung und Entscheidung) wollen Akzeptanz für das Optimierungsinstrument schaffen und dessen zu erwartenden Nutzen aufzeigen, um Managern die Notwendigkeit und die Vorteile einer Evaluation ihrer betrieblichen Trainingsmaßnahmen zu verdeutlichen (Wottawa & Thierau, 1998).

3 Relevante theoretische Konzepte und zugrunde liegende Modelle

Vor dem oben aufgezeigten Hintergrund eines fehlenden einheitlichen Verständnisses des Evaluationsbegriffes gingen Glass und Ellett (1980) schon vor über 20 Jahren davon aus, dass mindestens 20 unterschiedliche Evaluationsmodelle bestünden. Wottawa und Thierau (1998) gehen von einer noch weitaus größeren Anzahl aus. Inzwischen dürfte sich die Anzahl unterschiedlicher Konzepte zur Evaluation abermals vergrößert haben. Im Folgenden werden die unserer Meinung nach wichtigsten theoretischen Konzepte und Modelle der Trainingsevaluation vorgestellt, welche zur Diskussion der Problemstellung hilfreich sind und erste Lösungsansätze bieten können. Dazu gehört u. a. das 4-Ebenen-Modell von Kirkpatrick (1959a, 1959b, 1960a, 1960b) mit seinen Neuentwicklungen (1994, 1996, 1998a, 1998b). Obwohl es zahlreiche aktuellere Evaluationsmodelle gibt (z. B. Alliger et al., 1997; Day, Arthur & Gettman, 2001; Holton, 1996; Kraiger et al., 1993), wird das Kirkpatrick-Modell im Bereich der Trainingsevaluation nach wie vor als das Modell angesehen, welches am häufigsten angewandt wird und den höchsten Bekanntheitsgrad aufweist (Alliger et al. 1997; Goldstein & Ford, 2002; Kraiger & Jung, 1997; Salas & Cannon-Bowers, 2001). Des Weiteren werden die mit dem Modell von Kirkpatrick vergleichbaren Stufenmodelle von Hamblin (1974) sowie von Warr, Bird und Rackham (1978) vorgestellt. Außerdem wird ausführlich auf die Analyse des Return on Investment (ROI) von Phillips (1996, 1997, 1998a, 1998b, 1998c) eingegangen, da hier erstmals der monetäre Wertbeitrag einer Weiterbildungsmaßnahme direkt mit den entstandenen Trainingskosten verglichen wird (Phillips & Phillips, 2001). Die ROI-Analyse ist somit eine Methode, die es ermöglicht, Investitionen in Trainingsmaßnahmen anderen im Unternehmen getätigten Investitionen gegenüberzustellen. Eine alternative Herangehensweise, den Nutzen von PE-Programmen aufzuzeigen, bietet die Utility Analysis. Schmidt et al. (1982) modifizierten die ursprünglich zur Personalauswahl aufgestellte Nutzengleichung von Brodgen (1949), Cronbach und Gleser (1965) in der Form, dass diese im Bereich der Personalentwicklung Anwendung fand. Im Unterschied zu weiteren Nutzengleichungen anderer Autoren (z. B. Boudreau, 1983; Raju, Burke & Norman, 1990) berücksichtigt die von

Schmidt et al. (1982) empfohlene Gleichung neben Standardfaktoren auch den zeitlichen Aspekt, welcher den Nutzwert eines Trainings beeinflussen kann.

3.1 Das Modell von Kirkpatrick

Gülpen (1996) sieht den Grund für die oben angesprochene anhaltende Popularität des Kirkpatrick-Modells in seiner Plausibilität, der Einfachheit und in der Anwendungsfreundlichkeit. Ursprünglich war es kein Modell im eigentlichen Sinn, sondern lediglich von Kirkpatrick erstellte Richtlinien zur praktischen Durchführung von Trainingsevaluationen, die auf seinen persönlichen Erfahrungen als Trainer beruhten (Kraiger, 2003). Kirkpatrick (1959a) schlug vor, in der Evaluation eines Trainings die folgenden vier Ebenen zu berücksichtigen: Reaction, Learning, Behavior und Results. Grund für diesen Vorschlag zur Einteilung war für Kirkpatrick die bestehende Unklarheit über den Begriff der Evaluation. Er stellte fest, dass Personal- und Trainingsverantwortliche unterschiedliche Auffassungen von diesem Begriff hatten. „By breaking it down in reaction, learning, behavior, and results, the training professional . . . can gradually progress from a simple subjective reaction sheet to a research design that measures tangible results“ (Kirkpatrick, 1996, p. 311). Sein Ziel war es, Trainingsverantwortliche mit einem praktischen Leitfaden zur Trainingsevaluation auszustatten (Kirkpatrick, 1998b). Laut Kirkpatrick (1996) konnten die Richtlinien folgendermaßen zur Klärung des Begriffes beitragen: „Training directors can now begin to pinpoint their efforts at evaluation. They better realize what they are doing, and they recognize the limited interpretations and conclusions that can be drawn from their findings“ (p. 295). Die Unterteilung in Reaction, Learning, Behavior und Results stellte für Tannenbaum und Yukl (1992) „the prevalent framework for categorizing training criteria“ (p. 425) dar. In neueren Veröffentlichungen merkt Kirkpatrick (1998a) jedoch an, dass die Behauptung, seine vier Ebenen seien die alleinigen vier Ebenen der Trainingsevaluation, unzulässig, wenn nicht sogar absolut anmaßend sei. Grund für diese Einschätzung ist, dass andere Autoren neue Modelle zur Evaluation beschrieben haben, die sich von seiner eigenen Herangehensweise unterscheiden.

Im Folgenden werden die vier Ebenen des Kirkpatrick-Modells vorgestellt und die Häufigkeit ihrer Anwendung in Theorie und Praxis wird aufgezeigt. Außerdem werden mögliche Beziehungen zwischen den Ebenen und alternative Evaluationsmodelle beschrieben. Abschließend wird die Kritik verschiedener Autoren am Modell aufgeführt.

3.1.1 Beschreibung der 4 Evaluationsebenen

1. Ebene: Reaction

Kirkpatrick (1959a) erhebt auf dieser Ebene, ob und wie sehr dem Trainee ein bestimmtes Trainingsprogramm gefallen hat. „Evaluating in terms of reaction is the same as measuring the feeling of the conferees. In fact, it is measuring ‘customer satisfaction’” (Kirkpatrick, 1996, p. 295). „They indicate trainees’ perceptions, opinions, feelings, and attitudes about a training event“ (Pershing & Pershing, 2001, p. 77). Hierbei geht es um Effekte, die direkt im Anschluss an das Seminar gemessen werden. Kirkpatrick (1996) betont, dass diese Messung der Zufriedenheit keine Aussage darüber machen kann, ob ein Lernprozess beim Teilnehmer stattgefunden hat: „It is important to recognize that a favorable reaction to a program does not assure learning“ (p. 301). In einer Studie von Morgan und Casper (2000) konnte anhand einer Faktorenanalyse die Reaction-Ebene durch sechs Faktoren abgebildet werden:

1. Zufriedenheit mit dem Trainer
2. Zufriedenheit mit dem administrativen Trainingsmanagement
3. Zufriedenheit mit dem Prozess der Leistungsüberprüfung
4. Nützlichkeit des Trainings
5. Zufriedenheit mit dem Kursmaterial
6. Zufriedenheit mit dem Aufbau des Kurses.

Die Daten, die auf dieser Ebene meist mit Hilfe von Fragebogen, sog. „smile sheets“ (Kirkpatrick, 1998a, p. 67; Parry, 1997, p. 3), kurz nach Ende des Trainings erhoben werden, bezeichnet man oftmals als „happiness index“ (Bell & Kerr, 1987, p. 70). Kommentare und Wertungen können bei der Konzeption zukünftiger und bei der Verbesserung gegenwärtiger Trainings hilfreich sein, es ist jedoch zu berücksichtigen, dass diese Form der Evaluation auf rein subjektiven Daten basiert (Nork, 1989). Laut Nork (1989) sind möglichen Umgestaltungen der Programme jedoch Grenzen gesetzt: Wird davon ausgegangen, dass das Training auf einer Bedarfsanalyse (vgl. Kap. 4.1) basiert und die Lehrziele und -inhalte aus dieser abgeleitet wurden, dann können sich die Veränderungen folglich nur auf den Trainer, die Unterbringung o. Ä. beziehen. Veränderungen des Inhalts oder Ablaufs würden der aus den Trainingszielen und der Bedarfsanalyse abgeleiteten Trainingsmaßnahme die Grundlage entziehen und diese somit überflüssig machen.

2. Ebene: Learning

In seiner ersten Veröffentlichung zur zweiten Ebene seines Modells versteht Kirkpatrick (1959b) unter Lernen „what principles, facts, and techniques were understood and absorbed by trainees“ (p. 22). In neueren Publikationen zum gleichen Thema scheint es, als würde Kirkpatrick die Ausführungen von Kraiger et al. (1993) berücksichtigen, denen zufolge Lernen als multidimensionales Konstrukt (1. Verbesserung der kognitiven Leistung, z. B. bessere Wissensorganisation, 2. verbesserte Fertigkeiten, z. B. Änderung der Arbeitsabläufe, 3. affektive Änderungen, z. B. Einstellungsänderungen) verstanden wird: Kirkpatrick (1998a) bezeichnet Lernen nun „as the extent to which participants change attitudes, improve knowledge, and/ or increase skill as a result of attending the program“ (p. 20).

Kirkpatrick (1996) empfiehlt, bei der Datenerhebung auf Ebene 2 folgende Richtlinien zu berücksichtigen:

1. Quantitative Messung der Lerneffekte bei jedem Trainee.
2. Verwendung von Pre- und Posttest, um erhobene Lerneffekte auf das Training zurückführen zu können.
3. Vergleich der Daten der Experimentalgruppe mit denen einer Kontrollgruppe, um Ergebnisse eindeutig interpretieren zu können (Erhöhung interner Validität).
4. Statistische Analyse der Evaluationsergebnisse, so dass Nachweise über Erfolge in Form von Korrelationen oder Vertrauensintervallen erfolgen können.

Für die Erfassung der Lerndimensionen bieten sich nach Kirkpatrick (1996) zwei Methoden an: „Classroom Performance“ (p. 303) und „Paper-and-Pencil Tests“ (p. 303). „Classroom activities such as demonstration, individual performance of the skill being taught, and discussions following a role-playing situation can be used as evaluation techniques“ (p. 303). Diese Methode der „Classroom Performance“ kann hilfreich sein bei Maßnahmen zur Verbesserung der Präsentationstechniken, bei Interviewertrainings, bei Trainings zum Führen von Mitarbeitergesprächen o. Ä.. Hier findet die Vermittlung von Trainingsinhalten und deren Evaluation zeitgleich statt. Durch eine Messung zu Anfang und am Ende des Trainings kann überprüft werden, ob die Teilnehmer die gelehrteten Techniken und Prinzipien anwenden können (Kirkpatrick, 1996). Die Datenerhebung kann z. B. durch Verhaltensbeobachtungen seitens des Trainers anhand von Rating-Skalen erfolgen. Kirkpatrick (1996) schlägt noch eine weitere Methode vor: „Where principles and facts are taught rather than skills, paper-and-pencil tests can be used“ (p. 303); diese Testverfahren sollten im besten Fall standardisiert sein. Papier-

und-Bleistift-Tests sowie Rollenspiele bzw. Simulationen sind nach Goldsmith und Kraiger (1997) die am häufigsten eingesetzten Methoden zur Überprüfung des Lernerfolgs. Nach Tannenbaum und Yukl (1992) stellt der Lernerfolg eines Teilnehmers eine notwendige, jedoch nicht hinreichende Bedingung für eine Verhaltensänderung am Arbeitsplatz (3. Ebene) dar. Auch Arthur, Bennett et al. (2003) betonen: „They [d. h. learning criteria] are not measures of job performance” (p. 235).

3. Ebene: Behavior

Erst über dieses Kriterium wird die Umsetzung des Gelernten am Arbeitsplatz überprüft (Höft, 2001). Rae (2002) stellt in diesem Zusammenhang die Frage, inwieweit Trainees Trainingsinhalte im Arbeitsalltag umsetzen können: „What did they [die Trainees] do with the training?“ (p. 4). Für Kirkpatrick (1996) gibt es fünf notwendige Bedingungen für das Auftreten einer Veränderung auf der Verhaltensebene: Der Wunsch nach Veränderung (1), das Wissen darüber, was zu tun ist und wie es getan werden kann (2), das richtige Betriebsklima (3), Unterstützung bei der Anwendung des Gelernten (4), Anreize für verändertes Verhalten (5). Bei dieser Evaluationsebene geht es auch um Aspekte, die den „transfer of training“ (Wexley & Latham, 2002, p. 113) betreffen. Diese beziehen sich auf die letzten vier der fünf oben genannten Bedingungen. Laut Baldwin und Ford (1988) wird der Transfer eines Trainings wie folgt definiert: „Transfer of Training is defined as the degree to which trainees effectively apply the knowledge, skills, and attitudes gained in a training context to the job” (p. 63). Für Hesketh und Frese (2002) ist die Kontrolle des Transfer einer Trainingsleistung wichtig, um sicherstellen zu können, dass die gelernten Fähigkeiten „in a range of different contexts and tasks“ (p. 14099) angewendet werden können, denn auch wenn Trainees im Training hervorragende Leistungen erbringen, bedeutet dies noch nicht, dass ihnen nun auch die Bewältigung ähnlicher Aufgaben im Arbeitskontext einfacher fallen wird (Hesketh, 1997). Der Transfer eines Trainings ist laut Huint und Saks (2003) „one of the most pressing concerns facing trainers today” (p. 181). Auf die Transferproblematik wird in Kapitel 4 näher eingegangen.

Nach Nork (1989) bereitet der Nachweis kausaler Zusammenhänge zwischen den Weiterbildungsmaßnahmen und dem Verhalten der Trainees im Vergleich zu den Ebenen 1 bzw. 2 sehr viel größere Schwierigkeiten. Um auch hier fundierte Ergebnisse generieren zu können, empfiehlt Kirkpatrick (1996) folgende Richtlinien:

1. Eine systematische Leistungsbeurteilung im Arbeitskontext sollte vor und nach dem Training stattfinden.

2. Die Leistungsbeurteilung sollte im besten Fall ein Abgleich der Selbsteinschätzung des Trainees mit verschiedenen Fremdeinschätzungen durch Vorgesetzte, Mitarbeiter und Kollegen sein.
3. Eine statistische Berechnung der Leistungsdifferenz (vorher/ nachher) sollte erfolgen, um Trainingseffekte isolieren zu können.
4. Die Post-Training-Performance sollte frühestens drei Monate nach dem Training gemessen werden, um Trainees die Möglichkeit zur Anwendung des Gelernten zu geben. Weitere Follow-up-Messungen zur Validitätserhöhung können durchgeführt werden.
5. Wenn möglich, sollte eine Kontrollgruppe eingesetzt werden.

4. Ebene: Results

Auf dieser Ebene wird versucht, die Ergebnisse des Trainings mit den Zielen der Organisation in Einklang zu bringen (Goldstein & Ford, 2002). Organisationsziele, die ein Training zu erreichen sucht, können nach Kirkpatrick (1996, 1998a, 1998b) z. B. Verringerung der Fluktuation, Reduzierung der Kosten, Steigerung der Effizienz, Rückgang der Beschwerden oder der Unfallhäufigkeit, verbesserte Beziehungen zwischen Führungskräften und deren Mitarbeitern sowie Qualitätszuwachs und Produktionswachstum sein. „Results can be defined as the final results that occurred because the participants attended the program“ (Kirkpatrick, 1998a, p. 23). Laut Kirkpatrick (1998a) ist es bei vielen Trainings nur sehr schwer möglich, auf dieser Ebene zu evaluieren. Daher rät er, jeweils zuerst die Kriterien der Ebenen 1 bis 3 zu messen und dann zu überlegen, ob Ebene 4 machbar und sinnvoll ist. Der Autor unterscheidet zwischen zwei verschiedenen Gruppen von Trainings, die unserer Meinung nach in „hard-skill“- und „soft-skill“-Trainings unterteilt werden können. Die Trainings der ersten Gruppe lassen sich auf allen Ebenen (1-4) evaluieren, die der zweiten Gruppe hingegen nur auf den Ebenen 1-3.

„Hard-skill“-Trainings können relativ einfach auf Ebene 4 evaluiert werden: „Programs that aim at increasing sales, reducing accidents, reducing turnover, and reducing scrap rates can often be evaluated in terms of results“ (Kirkpatrick, 1998a, p. 65). Kirkpatrick (1998a) betont, dass genau derartige „tangible results“ (p. 23), also messbare Ergebnisse, für Unternehmen die Gründe darstellen, warum einige Trainings durchgeführt werden: „The final objectives of the training program need to be stated in these terms“ (p. 23). Einige Autoren aktueller Publikationen (z. B. Höft, 2001; Noe, 2002) gehen in ihrer Beschreibung der 4. Ebene des Kirkpatrick-Modells ausschließlich auf diese „tangible results“ ein. Noe (2002) geht davon aus,

dass die Ergebnisse dieser Ebene dazu verwendet werden, „to determine the training program’s payoff for the company” (p. 186). Für Höft (2001) werden Ergebnisse auf dieser Ebene durch globale organisationale Leistungskriterien wie betriebliche Produktivitätskennzahlen oder Bilanzzahlen beschrieben.

Kirkpatrick (1998a) äußert sich jedoch kritisch darüber, alle Trainingsprogramme auf Ebene 4 evaluieren zu wollen. Eine Schwierigkeit bei der Evaluation von „soft-skill“-Trainings auf Ebene 4 liegt darin, dass die „final results“ (p. 23) dieser Trainings häufig „not tangible“ (p. 23) sind und somit nicht messbar gemacht bzw. erhoben werden können. Daraus ergibt sich Unklarheit darüber, welche Anteile organisationaler Verbesserungen, wie z. B. verringerte Fluktuation oder verbesserte Kommunikationsprozesse, auf das Training zurückgeführt werden können bzw. welchen dagegen andere Faktoren zugrunde liegen. Veränderungen auf der Organisationsebene können nur sehr schwer direkt dem spezifischen Programm zugeordnet werden. „Soft-skill“-Trainings, zu denen Kirkpatrick die Mehrheit aller Programme zählt, sind zwar letztendlich auch auf greifbare Organisationsziele ausgerichtet, weisen allerdings im Vergleich zu „hard-skill“-Trainings einen entscheidenden Unterschied auf: „[They] have these [d. h. tangible results] in mind on a long-term basis“ (Kirkpatrick, 1998a, p. 23). Ein Change-Management-Training beispielsweise zielt darauf ab, dass notwendige Veränderungsprozesse von den teilnehmenden Managern akzeptiert und somit besser implementiert werden. Diese Akzeptanz und die daraus hervorgehende verbesserte Implementierung der Veränderungsprozesse sind für Kirkpatrick keine messbaren Ergebnisse und daher kann deren Auftreten nicht bewiesen werden. „These are not tangible results that can be measured in terms of dollars and cents. But it is *hoped* [Hervorhebung durch Verf.] that tangible results will follow“ (Kirkpatrick, 1998a, p. 23). Messbare Ergebnisse liefert dieses „soft-skill“-Training, nachdem die Manager die im Training erworbenen Fähigkeiten im Arbeitskontext eingesetzt und somit die Veränderungsprozesse unterstützt haben. „Tangible results“ wären dann im Sinne von Kirkpatrick erhöhte Arbeitsqualität, höhere Arbeitszufriedenheit und eine niedrigere Fehlerrate als Folge von erhöhter Akzeptanz und verbesserter Implementierung des Prozesses. Hier ergibt sich jedoch das bereits oben angesprochene Problem: Können diese „tangible results“ dem Training zugeschrieben werden? Kann bewiesen werden, dass diese Ergebnisse tatsächlich eintreten? Laut Kirkpatrick (1998a) ist die Antwort auf diese Fragen ein „resounding no“ (p. 65).

3.1.2 Erfassung der unterschiedlichen Evaluationsebenen

Bei der Erfassung der vier verschiedenen Ebenen scheint es einen Bruch zwischen Praxis und Wissenschaft zu geben. Während laut ASTD 2002 State of the Industry Report (Van Buren &

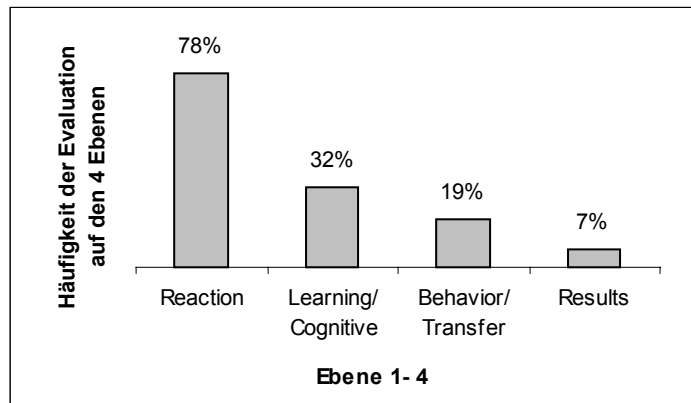


Abbildung 2: Ergebnisse einer Unternehmensbefragung. Mehrfachnennungen waren möglich.

Quelle: Eigene Darstellung nach: *State of the industry: ASTD's annual review of trends in employer-provided training in the United States* (p. 23), von M. E. Van Buren, W. Erskine, 2002, Alexandria, VA: American Society of Training and Development.

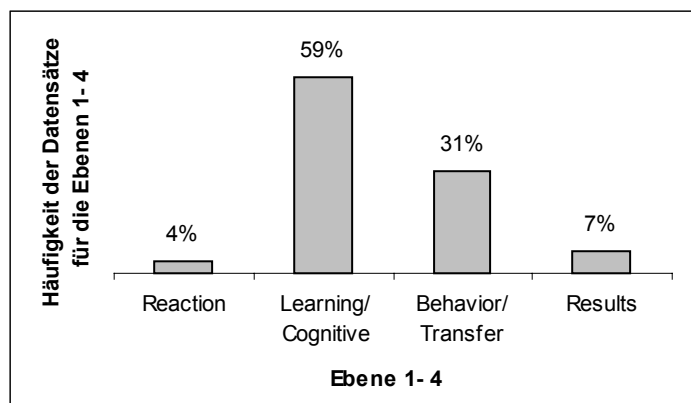


Abbildung 3: Prozentuale Verteilung der in der Literatur gefundenen Datensätze für die Ebenen 1-4. Ergebnisse einer Metaanalyse über die veröffentlichte Literatur zur Trainingsevaluation der Jahre 1960-2000.

Quelle: Eigene Darstellung nach: „Effectiveness of Training in Organizations: A Meta-Analysis of Design and Evaluation Features,” von W. Jr. Arthur, W. Jr. Bennett, P. S. Edens, 2003, *Journal of Applied Psychology*, 88(2), S. 239.

Erskine, 2002) in der amerikanischen Unternehmenspraxis hauptsächlich Zufriedenheitsmessungen (Ebene 1) zur Erfolgsüberprüfung eines Trainings vorgenommen werden (78% der befragten Unternehmen bei möglicher Mehrfachnennung, siehe Abb. 2), und ähnliche Ergebnisse auch die Befragung europäischer Unternehmen (Grünwald et al., 2003) sowie kuwaitischer Organisationen (Al-Athari & Zairi, 2002) ergab, finden sich in den Studien einer metaanalytischen Literaturrecherche zur Trainingseffektivität in Organisationen von Arthur, Bennett et al. (2003) häufiger Evaluationsstudien zu den Ebenen 2-4 (siehe Abb. 3). Lediglich 4% der Angaben behandeln Evaluationen auf Ebene 1.

Eine Metaanalyse wissenschaftlicher Publikationen von Alliger et al. (1997) stützt die Ergebnisse von Arthur, Bennett et al. (2003): In ca. 15% der Studien wurde auf der Ebene

Reaction, in ca. 50% Studien auf der Ebene Learning und in ca. 35% auf der Ebene Behavior evaluiert (keine Angaben über Ebene 4). Arthur, Bennett et al. (2003) kommen zu dem Schluss, dass die überwiegende Evaluation auf Ebene 1 in der Praxis durch die relativ einfache Datengewinnung auf dieser Ebene zustande kommt. Da es sehr unwahrscheinlich sei, dass Trainingsevaluationsstudien, die hauptsächlich Ebene 1 als Maß der Effektivität ansehen, in wissenschaftlichen und empirisch ausgerichteten Zeitschriften publiziert werden, ist der Un-

terschied zwischen Praxis und Wissenschaft nicht überraschend. Auch Höft (2001) sieht in dem vorgeschalteten Reviewverfahren wissenschaftlicher Publikationen einen Grund für diese Diskrepanz.

Schon 1986 wiesen Ralphs & Stephan (1986) darauf hin, dass die meisten Organisationen lediglich auf der Reaction-Ebene evaluieren. Alliger und Janak (1989) konnten diesen Widerspruch zwischen Praxis und Wissenschaft ebenfalls aufzeigen: „. . . in practise, most training is evaluated on the reaction level only, while editors for the reviewed journal may look for evaluation results measuring learning, behavior, or results“ (p. 336). Für Ruona et al. (2002) scheint es, dass viele Organisationen nach wie vor glauben, dass die Zufriedenheitseinschätzungen der Teilnehmer valide und reliable Indikatoren seien, um die Effektivität von Trainingsprogrammen zu überprüfen, obwohl durch zahlreiche Forschungsergebnisse aufgezeigt werden konnte, dass nur geringe Beziehungen zwischen Reaction und Learning und gar keine zwischen Reaction und der Ebene Behavior bzw. Results bestehen. „Even the most recent literature continues to expose the insufficiency of reaction measures as they are currently designed“ (Ruona et al., 2002, pp. 219-220). Die Forschungsergebnisse von Studien zu Korrelationsangaben zwischen den Ebenen werden in Kapitel 3.1.3.2 besprochen.

3.1.3 Beziehungen innerhalb des Modells

Die eingangs beschriebene Einfachheit des Kirkpatrick-Modells kann laut Alliger und Janak (1989) jedoch auch zu Missverständnissen und unzulässigen Verallgemeinerungen führen. „The model’s simplicity is appealing but, as revealed in more recent work, this simplicity is also liability“ (Alliger et al., 1997, p. 342). Eine sog. „implizite Prämisse“ (Gülpen, 1996, S. 32) des Modells, die Kirkpatrick in seinem Modell nicht intendierte, sondern laut Alliger und Janak (1989) z. B. von Newstrom (1978) und von Hamblin (1974) fälschlicherweise dem Modell zugeschrieben worden ist, heißt nach Alliger & Janak (1989) wie folgt: „Each succeeding level is correlated with the previous level“ (p. 334). Dabei erwähnen die Autoren jedoch nicht, dass Hamblin (1974) sich bei der Beschreibung der Kausalkette nicht auf das Kirkpatrick-Modell, sondern auf sein eigenes Modell bezieht (vgl. Kap. 3.1.5). Ebenso postuliert Newstrom (1978) keine Interkorrelation zwischen den Ebenen, sondern weist auf die Gefahren hin, die bei einer solchen Annahme auftreten können. Kirkpatrick (1959b) stellte jedoch selbst klar, dass die Reaction-Ebene noch nichts mit Lernen zu tun hat: „Favorable reactions don’t assure learning“ (p. 22). Ebenso stellte er fest, dass das alleinige Wissen von Techniken nicht schon ihre Anwendung im Arbeitsalltag bedeutet (Kirkpatrick, 1960a).

Laut Kraiger (2003) wird das 4-Ebenen-Modell oftmals – so z. B. auch von Noe und Schmitt (1986) – als ein hierarchisches Modell charakterisiert, womit zwei Aspekte impliziert werden: Höhere Ebenen sollten nicht evaluiert werden, solange nicht zufriedenstellende Ergebnisse auf den unteren Ebenen erzielt wurden, und Veränderungen auf höheren Ebenen seien nützlicher für die Organisation als Veränderungen auf niedrigeren Ebenen. Eine hierarchische Ordnung gilt also demnach für die Trainingserfolgsebenen: Veränderungen auf der Organisationsebene seien z. B. nützlicher als auf der Transfer- oder Verhaltensebene. Alliger und Janak (1989) sowie Russ-Eft und Preskill (2001) halten die Annahme, es handele sich um ein hierarchisches Modell, für ein Missverständnis. „It is not clear that all training in organizations is meant to effect change at all four levels“ (Alliger & Janak, 1989, p. 332). Da den Trainingsprogrammen unterschiedliche Prämissen zugrunde liegen, lassen sich die unterschiedlichen Trainingsziele dementsprechend nicht hierarchisch ordnen.

In diesem Zusammenhang stellt sich die Frage, inwiefern der Unternehmenserfolg direkt oder indirekt bzw. kurz-, mittel- oder langfristig ausnahmslos bei allen Trainingsprogrammen im Vordergrund steht und damit eine Verbesserung auf der Organisationsebene (d. h. Ebene 4) den eigentlichen Grund für sämtliche PE-Maßnahmen darstellt. Dieser Ansicht widersprechen wie oben aufgeführt Alliger und Janak (1989), für die nicht jedes Training zwangsweise auf eine Veränderung der Results-Ebene abzielt. Die verschiedenen Positionen können unserer Meinung nach dadurch erklärt werden, dass unterschiedliche Ansprüche an die Zielvorgabe eines Trainings gelegt werden: Das Verständnis eines hierarchischen Modells ist dann gerechtfertigt, wenn mit jedem Training eine Veränderung auf einer möglichst hohen Ebene erreicht werden soll; eine nicht-hierarchische Betrachtungsweise erscheint dann sinnvoll, wenn das Training auf die Veränderung einer ausgewählten Ebene ausgerichtet ist, welche im Vorfeld definiert worden ist.

Kirkpatrick (1998a) vergleicht nicht die Trainings-, sondern die Evaluationsebenen miteinander: „As you move from one level to the next, the process becomes more difficult and time-consuming, but it also provides more valuable information“ (p. 19), er spricht jedoch nicht explizit von einem hierarchischen Modell. Außerdem rät Kirkpatrick (1998), mit der Evaluation auf Ebene 1 zu beginnen und anschließend nach Möglichkeit mit den anderen Ebenen der Reihe nach fortzufahren. Zur Veranschaulichung dieser Notwendigkeit nennt der Autor ein Beispiel: Sollte sich bei einer Evaluation auf der dritten Ebene herausgestellt haben, dass keine Verhaltensveränderung eingetreten ist, lässt sich ohne vorherige Evaluation auf

Ebene 2 der Grund hierfür nicht aufzeigen. „Therefore, it is important to evaluate at level 2 so you can determine whether the reason for no change in behavior was lack of learning or [z. B.] negative job climate“ (p. 68). Für Gülpen (1996) ist es nicht richtig, Informationen auf einer höheren Stufe als wertvoller anzusehen. Das Evaluationsziel hängt laut der Autorin von der jeweils spezifischen Forschungsfrage ab und müsse bei jeder neuen Trainingsevaluation neu definiert werden. So kann im Interesse einer Trainingsevaluationsstudie lediglich die Erhebung von Lernleistung stehen, ohne dabei auf mögliche weitere Auswirkung auf der Transfer- und/oder Organisationsebene eingehen zu wollen. Somit sind Informationen auf den Ebenen Transfer und/oder Organisation in diesem Fall nicht wertvoller, da sie nicht der Klärung der Forschungsfrage dienen. Daher kann nicht von einer Hierarchie der Evaluationsebenen gesprochen werden, wenn ein Evaluationsprozess nicht darauf abzielt, ein Training auf allen Ebenen evaluieren zu wollen, sondern vielmehr die Evaluation einer bestimmten Ebene als Untersuchungsziel anstrebt.

Unserer Meinung nach können auch hier die unterschiedlichen Positionen mit den verschiedenen Ansprüchen begründet werden – diesmal mit denen, die an das Evaluationsziel gelegt werden. Da einerseits die Annahme besteht, dass die Evaluationsebenen „schrittweise“ qualitativ hochwertigere Informationen liefern können, bilden die vier Ebenen eine Art Evaluationspyramide, wobei die vierte Ebene immer die höchste und wertvollste Stufe darstellt. Ziel des Evaluationsprozesses ist es demnach, ein Training so „hoch“ wie möglich, d. h. auf immer höheren Ebenen zu evaluieren. Wird andererseits der Anspruch gestellt, lediglich auf der Ebene zu evaluieren, auf der forschungsrelevante Informationen gewonnen werden können, kann von gleichrangigen Evaluationsebenen gesprochen werden. Also entscheiden auch in diesem Fall die unterschiedlichen Perspektiven der Herangehensweise darüber, ob die Evaluationsebenen des Kirkpatrick-Modells als hierarchisch oder eben nicht-hierarchisch angesehen werden können.

Wir sind der Ansicht, dass auch der hierarchischen Betrachtungsweise der Evaluation Grenzen gesetzt sind und möchten daher näher auf den Begriff der hierarchisch höchsten Evaluationsebene eingehen. Wir sprechen also nicht von einer Hierarchie der Trainingserfolgsebenen im Sinne von Kraiger (2003), wonach Veränderungen auf höheren Ebenen nützlicher seien als auf niedrigeren Ebenen, sondern von einer Hierarchie der Evaluationsebenen, wonach eine Evaluation auf höheren Ebenen wertvoller und informativer sei als auf niedrigeren Ebenen. Es ist jedoch nicht immer eine Evaluation auf allen Ebenen möglich. Als hierarchisch höchste

Evaluationsebene sollte diejenige angesehen werden, welche das spezifische Trainingsziel erfassen kann. Dabei beziehen wir uns auf das formulierte und in der Bedarfsanalyse definierte Ziel eines Trainings. Verfolgt eine Weiterbildungsmaßnahme beispielsweise das Ziel, Mitarbeitern lediglich die Geschichte des Unternehmens näher zu bringen, so wird die Evaluation auf Ebene 2 angemessen sein, um das Trainingsziel (Wissenserweiterung) abbilden zu können, und ist somit die hierarchisch höchste Evaluationsebene: Die Mitarbeiter sollen ihr Wissen über die Unternehmenshistorie erweitern, es wird jedoch keine explizite Verhaltensveränderung seitens der Mitarbeiter erwartet bzw. angestrebt und auch nicht ausdrücklich auf eine Verbesserung auf Organisationsebene abgezielt. Die hierarchisch höhere Evaluationsebene wäre also nur dann wertvoller, wenn das Training tatsächlich auf dieser höheren Ebene gesetzte Ziele (z. B. Verhaltensveränderung der Mitarbeiter) zu erreichen sucht, was in diesem Training jedoch nicht der Fall ist.

Des Weiteren sollte stets eine Evaluation auf derjenigen Ebene erfolgen, welche der Zielebene hierarchisch untergeordnet ist, um zumindest einige Gründe für ein eventuelles Nichtauftreten von Veränderung auf der Zielebene ausschließen zu können. Gründe für nicht erfolgtes Lernen könnten beispielsweise in einer der Kategorien der Ebene Reaction (vgl. Alliger et al. 1997; Morgan & Casper, 2000; Warr & Bunce, 1995) zu finden sein. So könnten die Teilnehmer das Training als nicht nützlich empfinden oder als zu schwierig wahrnehmen, was dazu führen kann, dass ein Lernprozess nicht stattfindet.

3.1.3.1 Erweiterung des Modells nach Alliger et al.

Alliger et al. (1997) schlugen eine Erweiterung des Kirkpatrick-Modells vor (siehe Tab. 1), „to facilitate a generalized understanding of training evaluation results“ (pp. 342). Das Modell in Tabelle 1 stellt nach Meinung der Autoren keinen Ersatz für Kirkpatrick's Originalmodell dar, sondern beschreibt einen ersten Versuch, die einzelnen Ebenen deutlicher als bisher voneinander abgrenzen zu können.

Alliger et al. (1997) unterteilen die Ebene Reaction in folgende Kategorien: Zum einen in die der affektiven Reaction (1a), welche die Zufriedenheit des Teilnehmers mit dem Training misst. Ein Beispielitem lautet nach Alliger et al. (1997): „I found this training program to be enjoyable“ (p. 344). Eine weitere Kategorie bezeichnet „Reactions as utility judgements“ (p. 344) (2b). Hier wird der vom Trainingsteilnehmer wahrgenommene Nutzenbeitrag des Trainings für eine Verbesserung seiner Arbeitsleistung gemessen. Diese Unterteilung der ersten Ebene nahmen die Autoren unabhängig von einer ähnlichen Einteilung von Warr und Bunce

(1995) vor, welche die Reaction-Ebene in drei Kategorien abbildeten: „Three kinds of reactions may be identified as conceptually separate: reported enjoyment, usefulness, and difficulty“ (Warr & Bunce, 1995, p. 353). Die ersten beiden Kategorien entsprechen inhaltlich denen von Alliger et al. (1997), das Konzept der „difficulty“, d. h. die subjektiv empfundene Schwierigkeit eines Trainings, wurde jedoch in dem erweiterten Modell nicht berücksichtigt, da die Trainingsschwierigkeit laut Alliger et al. (1997) nur selten abgefragt wird und das Datenvolumen dementsprechend unzureichend ist.

Tabelle 1: Das erweiterte Modell von Alliger et al. (1997)

Das Modell nach Kirkpatrick	Erweitertes Modell (Alliger et al., 1997)
Reactions	Reactions (1) Affective reactions (1a) Utility judgements (1b)
Learning	Learning (2) Immediate knowledge (2a) Knowledge retention (2b) Behavior/ skill demonstration (2c)
Behavior	Transfer (3)
Results	Results (4)

Quelle: Eigene Darstellung nach: „A Meta-Analysis of the Relations among Training Criteria,“ von G. M. Alliger, S. I. Tannenbaum, W. Bennett, H. Traver & A. Shotland, 1997, *Personnel Psychology*, 50, p. 343.

Die zweite Ebene Learning unterteilt Alliger et al. (1997) in drei Kategorien: Mit der ersten Kategorie, dem „Immediate post-training knowledge“ (p. 345) (2a) wird gemessen, wie viel der Trainee über die dargestellten Trainingsinhalte weiß. Meist wird diese Kategorie über traditionelle Testverfahren (z. B. Multiple-Choice-Verfahren) erhoben. Die Kategorie „Knowledge retention“ (2b) entspricht inhaltlich der ersten, wird jedoch nach einem längeren Zeitraum gemessen. Für die Autoren differenzierte Kirkpatrick mit dem Begriff „Behavior“ (Ebene 3 in seinem Modell) nicht deutlich genug zwischen dem Verhalten, welches im Training selbst (z. B. Rollenspiele, Simulationen) gezeigt wird und dem späteren Verhalten am Arbeitsplatz. „That is, his level 3 may include both results of behavioral skill tests administered at the conclusion of training (i. e., indications of „can do“) as well as on-the-job performance (i. e., indications of „does do“)“ (p. 345). Daher nehmen Alliger et al. sämtliche Verhaltensindikatoren mit in die dritte Kategorie der Ebene Learning auf, solange sie innerhalb des Trainings gemessen werden (Behavior/ Skill demonstration, 2c).

Auf der dritten Ebene nehmen Alliger et al. (1997) lediglich eine Begriffsänderung vor und bezeichnen Behavior als „Transfer“, um die „on-the-job nature“ (p. 346) des Kriteriums zu betonen. Als weiteren Kritikpunkt nennen die Autoren die vage Beschreibung von Ebene 4.

Einige Indikatoren wie z. B. „employee attendance or scrappage rates“ (p. 347) könnten sowohl Ebene 3 als auch Ebene 4 beschreiben.

Die vierte Ebene bleibt unberücksichtigt und wurde von den Autoren unverändert beibehalten, da in einer Metaanalyse, welche nachfolgend vorgestellt wird, lediglich drei Studien Korrelationen angaben, die womöglich Ebene 4 hätten zugeordnet werden können.

3.1.3.2 Korrelationen zwischen den Ebenen

Wie bereits in Kapitel 3.1.2 aufgezeigt wurde, werden Trainingsprogramme nach wie vor (in der Praxis) hauptsächlich auf der ersten Ebene evaluiert. Daher ist es wichtig, „to understand the extent to which positive reactions are indicative of the other three types of criteria“ (Tannenbaum & Yukl, 1992, p. 425). U. a. werden die Ergebnisse einer umfassenden Literaturrecherche (12 Studien mit 26 Korrelationsangaben) von Alliger und Janak (1989) vorgestellt. Goldstein (1991) ebenso wie Tannenbaum und Yukl (1992) merken an, dass diese Ergebnisse mit Vorsicht betrachtet werden müssen, da sie sich nur auf wenige Studien stützen. Basierend auf dem erweiterten Modell führten Alliger et al. (1997) eine Metaanalyse der einschlägigen Trainingsliteratur durch und fanden insgesamt 34 Studien mit 115 Angaben zu Korrelationen zwischen den 4 Evaluationsebenen. Goldstein und Ford (2002) bemerken dazu, dass die ebenfalls geringe Anzahl der von Alliger et al. (1997) gefundenen Studien weiterführende Analysen notwendig mache. Aus diesem Grund werden im Folgenden die Resultate weiterer Studien mit aufgeführt, bei denen es sich bis auf die Untersuchungen von Campion und Campion (1987), Noe und Schmitt (1986) sowie Ruona et al. (2002) um Metaanalysen handelt.

1. Korrelationen der Ebene 1 mit den Ebenen 2 und 3

Frühere Studien von Alliger und Janak (1989) zeigten nur sehr geringe Korrelationen zwischen Ebene 1 und den übrigen 3 Ebenen. Arthur, Tubre, Paul & Edens (2003) konnten diese Ergebnisse zwischen den Ebenen 1 und 2 bestätigen: „Consistent with the extant organisational training evaluation literature, a small to zero relationship was found between these evaluation criteria“ (p. 282). Ähnliche Ergebnisse fanden auch Colquitt, LePine und Noe (2000). Keine Beziehung zwischen Ebene 1 und 2 ergaben ebenfalls die Messungen von Alliger et al. (1997) zwischen Affective Reaction (1a) und den drei Kategorien der Learning-Ebene ihres erweiterten Modells, wohingegen Utility Judgements (1b) zumindest mit Immediate knowledge (2a) positiv korrelieren ($r=.26$). Goldstein und Ford (2002) bemerken dazu Folgendes: „One could certainly make the argument that positive perceptions of the future utility of a training program should logically lead to stronger learning performance than perceptions of

whether the experience was enjoyable” (p. 163). Überraschenderweise fanden Alliger et al. (1997) in ihrer Studie heraus, dass die Kategorie Utility Judgements (2b) die Ebene Transfer (3) eher vorhersagen kann ($r=.18$) als dies die Ebene Learning (2a; 2b; 2c) zu tun vermag ($r=.11$; $r=.08$; $r=.18$). In Anbetracht der in Kapitel 3.1.2 angesprochenen anhaltenden Popularität von Zufriedenheitsmessungen merken Ruona et al. (2002) aufgrund ihrer Forschungsergebnisse an: „Despite their widespread use, participant reaction do not seem to contribute greatly to predicting transfer of learning nor do they seem to predict actual performance improvement” (pp. 226-227). Außerdem konnten die Autoren die Ergebnisse von Alliger et al. (1997) auch in folgender Hinsicht belegen: „If reaction measures are to be used at all, utility reactions may be of greater value in evaluation outcomes [of training] than traditional affective reactions” (p. 227).

Tan, Hall und Boyce (2003) konnten einen positiven Zusammenhang zwischen „pretraining knowledge“ (p. 406) und Unzufriedenheit der Teilnehmer mit dem Training ermitteln, den die Autoren dadurch erklärten, dass Teilnehmer, die bereits sehr erfahren sind, das Training eher kritisch bewerten, da es ihren hohen Erwartungen häufig nicht entspreche. Bereits hier zeigt sich unserer Meinung nach die Notwendigkeit einer Bedarfsanalyse; Weiterbildungsmaßnahmen sollten individuell auf das Wissen und die Fähigkeiten der Mitarbeiter abgestimmt werden. Außerdem erscheint es sinnvoll, Moderatorvariablen wie z. B. die Erwartungen an das Training oder das Vorwissen der Teilnehmer mit in die Trainingsevaluation und die Interpretation der Ergebnisse aufzunehmen (Tan et al., 2003). Ausführlich wird auf diese Aspekte in den Kapiteln 4.2 und 4.4 eingegangen. Außerdem fanden die Autoren entgegen ihren Erwartungen und früheren Studien anhand von Regressionsanalysen heraus, dass Unzufriedenheit mit dem Training der beste Prädiktor zur Bestimmung der Höhe der Lernleistung zu sein scheint. Trainees, die mit dem Training unzufrieden waren, wiesen in dieser Studie höhere Lernleistungen auf als Trainees, die angaben, mit dem Training zufrieden gewesen zu sein. Für diesen Zusammenhang bieten die Autoren jedoch keine Erklärung an. Dies zeige aber: „This relation between negative evaluation and learning questions the prevailing notion that positive reactions to a training program are evidence for its success” (Tan et al., 2003, p. 406).

2. Korrelationen innerhalb der Ebene 2

Die Ergebnisse von Alliger et al. (1997) zeigen positive Korrelationen zwischen Immediate Knowledge (2a) und Knowledge Retention (2b) ($r=.35$), was sich durch die oben erwähnten

inhaltsidentischen Messungen der beiden Kriterien erklären lässt. Nur schwach positive Zusammenhänge ließen sich zwischen Immediate Knowledge (2a) und Behavior (2c) ($r=.18$) bzw. Knowledge Retention (2b) und Behavior (2c) ($r=.14$) finden. Colquitt et al. (2000) konnten stärkere positive Zusammenhänge innerhalb der Ebene 2 finden. Allerdings lag ihren Untersuchungen die Definition der Learning-Ebene zugrunde, wie sie von Kraiger et al. (1993) vorgeschlagen wurde und welche nicht identisch mit der Definition nach Alliger et al. (1997) ist (vgl. Kap. 3.1.3.1). Für den Vergleich zwischen den Ergebnissen von Alliger et al. (1997) und Colquitt et al. (2000) sind hier jedoch nur zwei der drei von Kraiger et al. aufgestellten Kategorien relevant, nämlich „cognitive outcomes“ und „skill-based outcomes“ (Kraiger et al., 1993, p. 312). Zwischen der kognitiven Komponente und der anwendungsorientierten Komponente der multidimensionalen Ebene konnten Colquitt et al. (2000) einen mittleren Zusammenhang aufzeigen ($r=.44$).

3. Korrelationen der Ebene 3 mit der Ebene 2

Nur äußerst schwach positive Ergebnisse fanden sich nach Alliger et al. (1997) zwischen jeweils allen drei Kategorien der Ebene 2 mit der Ebene Transfer (3) ($r=.11$; $r=.08$; $r=.18$). „This indicates that the current learning measures obtained in these studies are not very predictive of how well a person actually transfers the learning to the job“ (Goldstein & Ford, 2002, p. 164). Damit konnten Ergebnisse von Campion und Campion (1987) bestätigt werden, welche aufzeigen konnten, dass eine positive Verbesserung auf der Ebene Learning keine Verbesserung des Verhaltens am Arbeitsplatz garantiert. Ähnliche Ergebnisse finden sich auch bei Noe und Schmitt (1986). Für Tannenbaum und Yukl (1992) scheint Learning nach einer Sichtung der für sie relevanten Trainingsliteratur von 1987 bis 1991 eine notwendige, aber nicht hinreichende Bedingung für Verhaltensänderungen zu sein. Arthur, Bennet et al. (2003) merken dazu an: „This is because behavioral criteria are susceptible to environmental variables that can influence the transfer or use of trained skills or capabilities on the job“ (p. 235). Auf transferkritische Faktoren wird in Kapitel 4 eingegangen.

Abweichend zu diesen Ergebnissen fanden Colquitt et al. (2000) jedoch folgende Resultate: „The corrected learning-transfer correlations were moderate to large in our analysis . . . but were small in Alliger et al. (1997)“ (p. 695). Die korrigierte Korrelation zwischen der kognitiven Lernkomponente und dem Transfer betrug $r_c=.38$, die korrigierte Korrelation zwischen der anwendungsorientierten Lernkomponente und dem Transfer betrug $r_c=.69$. Diese stark abweichenden Ergebnisse könnten erneut aufgrund der von Colquitt et al. (2000) verwendeten

Definition der Learning-Ebene nach Kraiger et al. (1993) zustande gekommen sein, welche, wie bereits erwähnt, nicht mit der von Alliger et al. (1997) übereinstimmt. Auch muss darauf hingewiesen werden, dass Colquitt et al. (2000) ihre Korrelationen im Gegensatz zu Alliger et al. (1997) um Stichprobenfehler und Unreliabilität korrigiert haben. Allerdings zeigen sich selbst dann, wenn die unkorrigierten Korrelationen der Studie nach Colquitt et al. (2000) herangezogen werden, wesentlich stärkere positive Korrelationen zwischen den Ebenen 2 und 3, als dies von anderen genannten Studien aufgezeigt werden konnte, nämlich $r=.28$ für die kognitive Lernkomponente und Transfer sowie $r=.50$ für die anwendungsorientierte Lernkomponente und Transfer. Des Weiteren konnten Colquitt et al. (2000) anhand einer multiplen metaanalytischen Regressionsanalyse feststellen, dass die anwendungsorientierte Komponente den besten Prädiktor darstellt, wenn es darum geht, von Lernleistung auf Verhalten am Arbeitsplatz zu schließen ($\beta=.59$); dieses Ergebnis der Metaanalyse ist auf einem Niveau von $p<.05$ signifikant. Der kognitiven Komponente konnte hingegen im Zuge der multiplen Regressionsanalyse nur eine wesentlich geringere Bedeutung attestiert werden ($\beta=-.03$).

3.1.4 Kritik am Kirkpatrick-Modell

Das vorgestellte Modell der Trainingsevaluation von Kirkpatrick mit seinen vier Ebenen erlangte schnell einen hohen Bekanntheitsgrad im Personalentwicklungsbereich (Cascio, 1987). In den letzten 15 Jahren gab es etliche Autoren, die das Kirkpatrick-Modell kritisierten (z. B. Alliger & Janak, 1989; Alliger et al., 1997; Holton, 1996; Kraiger et al., 1993; Kraiger & Jung, 1997). Salas und Cannon-Bowers (2001) bemerken dazu:

It has been used, criticized, misused, expanded, refined, adapted, and extended. It has served the training research community well – but a richer, more sophisticated typology is needed. Research needs to continue finding better, more diagnostic and rigorous assessments of learning outcomes. (p. 487)

Kraiger (2003) fasst die kritischen Aspekte zusammen und nennt u. a. die folgenden Punkte: „The approach [of Kirkpatrick] is largely atheoretical, and to whatever extent that it may be theory based, it is based on a 1950s behavioral perspective that ignores modern, cognitively based theories of learning“ (p. 185). Außerdem sei das Modell allzu stark vereinfachend, indem es Reaction und Learning als eindimensionale Konstrukte betrachtet. Mehrere Autoren konnten in ihren Studien die Mehrdimensionalität der Reaction-Ebene (Alliger et al., 1997; Morgan & Casper, 2000; Tan et al., 2003; Warr, Allan & Birdi, 1999; Warr & Bunce, 1995) bzw. der Learning-Ebene (Alliger et al., 1997; Kraiger et al., 1993) aufzeigen. Des Weiteren biete das Kirkpatrick-Modell laut Kraiger und Jung (1997) zwar Unterstützung dabei, *wie*

evaluiert werden könne, aber keine Entscheidungshilfe dabei, „*how to link evaluation results to strategic decision making about training*“ (p. 153). Auf die Kritik von Holton (1996), der dem Modell von Kirkpatrick abspricht, ein Modell und statt dessen eher ein Klassifikationschema zu sein, erwidert Kirkpatrick (1998b): „I don't care whether it's a model or taxonomy as long as training professionals find it useful in evaluating training programs“ (p. 4).

Als einen weiteren Kritikpunkt nennt Delahaye (2000), dass das Modell von Kirkpatrick „output oriented“ (p. 339) sei und wichtige Faktoren, sog. „presage factors“ (p. 354), nicht berücksichtige. „Presage factors are those factors or events which can be examined to predict a particular outcome . . . [and] can keep the program moving in the intended direction“ (p. 354). Auch Rae (2002) äußert sich ähnlich und fordert ein sog. „Level 0“ (p. 4), welches mit dem Evaluationsprozess beginnt, bevor die Teilnehmer am Training teilnehmen. Ein Modell, welches den von Delahaye (2000) geforderten „improvements-oriented focus“ (p. 354) integriert, ist das 6-Stufen-Modell von Brinkerhoff (1987, 1998). Zusätzlich zu den Ebenen 2-4 von Kirkpatrick nimmt Brinkerhoff (1998, p. 79) die Bedarfsanalyse („What's the need?“), das Trainingsdesign („What will work?“) und die Trainingsimplementation („Is it working?“) in den Evaluationsprozess mit auf. Dieses Modell kann den formativen Evaluationsmodellen zugeschrieben werden (Russ-Eft & Preskill, 2001).

Zusätzlich zu den vier Evaluationsebenen bietet Kirkpatrick (1996) zu jeder einzelnen beispielhaft verschiedene Verfahren und Techniken aus der Praxis an, mit Hilfe derer die Ebenen evaluiert werden können, er erwähnt allerdings keine Ergebnisse. „It is important to stress that the described *procedures* and *techniques* can be used in almost any organization. It is also important to stress that the *results* from one organization cannot be used in another organization“ (Kirkpatrick, 1996, p. 295). Außerdem rät Kirkpatrick (1998b): „We *can* borrow evaluation forms, procedures, designs, approaches, techniques, and methods from other people“ (p. X). Nach der Einschätzung von Nork (1991) sind jedoch die Verfahren und Techniken – selbst wenn sie in anderen Untersuchungen zu erfolgreichen Ergebnissen geführt haben – mit Vorsicht zu betrachten und nicht einfach auf die eigene betriebliche Situation übertragbar. Das Einsatzrisiko bei der Übernahme von Instrumenten aus anderen Organisationen, selbst wenn sie mit Erfolg angewendet wurden, ist entsprechend hoch, da deren Unzulänglichkeiten und notwendige Voraussetzungen, die die Anwendbarkeit und Grenzen erst verdeutlichen, nicht bekannt sind. „Die individuelle Situation einer jeden Unternehmung (personelle Ressourcen, Konkurrenzsituation, Marktgegebenheiten etc.) erfordert stets eine maßgeschneiderte

Auswahl und Anpassung von Methoden, Vorgehensweisen und Techniken, die den Zielvorstellungen und Ergebnissen der Bedarfsanalyse jeweils entsprechen“ (Nork, 1991, S. 98). Desgleichen merken Wang, Dou und Li (2002) an: „Besides some general evaluation guidelines . . . the [Kirkpatrick-]model does not offer specific techniques for quantitative evaluation of training programs” (p. 208). Für Wang et al. (2002) liegt der größte Beitrag des Modells in der Tatsache, „that it created a common vocabulary for HRD practitioners to communicate their evaluation efforts” (p. 208).

3.1.5 Alternative Evaluationsmodelle

Die Evaluationsebenen von Kirkpatrick stellen nicht die einzige Untergliederung dar, die in der Literatur und Praxis Verwendung findet. Vergleichbar mit dem Modell Kirkpatrick sind die Stufenmodelle von Hamblin (1974) und von Warr et al. (1978) (siehe Tab. 2).

Tabelle 2: Evaluationsmodelle im Vergleich

Stufen	Kirkpatrick	Hamblin	Warr et al.
1	Reactions	Reactions	Reactions
2	Learning	Learning	Immediate
3	Behavior	Job Behavior	Intermediate
4	Results	Organization effects	Ultimate
5		Ultimate value	

Quelle: Eigene Darstellung nach: *Evaluation and Control of Training* (p. 14), von A. C. Hamblin, 1974, London: Mc Graw-Hill.

Die Einteilung der Evaluationsebenen von Hamblin (1974) entspricht auf den Stufen 1-3 der Untergliederung Kirkpatrick. Die Organisationsebene unterteilt er jedoch in zwei Stufen: Zum einen in „Organization effects“ (p. 14), mit denen erhoben wird, welche Auswirkungen das veränderte Verhalten der Trainingsteilnehmer auf die Organisation bzw. die einzelnen Abteilungen (z. B. Verbesserung der Fehlzeiten) hat und zum anderen in den „Ultimate value“ (p. 14), welcher den Grad, mit dem die Organisation ihre höchsten Ziele (z. B. finanzieller Art: Return on Investment) erreicht, abbildet. Hamblin (1974) betont: „If we are to evaluate training . . . we must collect some information about the changes that have been caused by the training“ (p. 14). Er geht dabei explizit – im Gegensatz zu Kirkpatrick – von einer möglichen „cause-and-effect chain“ (p. 15) zwischen den Ebenen aus: „Training leads to reactions which leads to learning which leads to changes in job behaviour which leads to changes in the organization which leads to changes in the achievement of ultimate goals” (p. 15). Diese Annahme der Kausalkette ist nach Stewart (1999) „the most significant and important” (p. 186) Aspekt dieses Modells.

Das Modell von Warr et al. (1978) wird in Tabelle 2 aus Gründen der Vergleichbarkeit mit den anderen beiden Modellen nur verkürzt wiedergegeben. Vollständig lautet das Modell „CIRO“ (p. 20), wobei dies für die Anfangsbuchstaben der unterschiedlichen Evaluationstypen steht und sich wie folgt darstellt:

- » Context Evaluation: Beinhaltet die Sammlung von Informationen über die gegenwärtige Arbeitssituation, um daraufhin die Bedarfsanalyse vollziehen und Trainingsziele bestimmen zu können.
- » Input Evaluation: Beinhaltet die Sammlung von Informationen über mögliche „training resources“ (p. 20), um zwischen verschiedenen Input-Faktoren (z. B. verfügbares Budget, Hinzuziehen externer Trainingsanbieter, unterschiedliche Trainingsmethoden) für die Durchführung der Maßnahme wählen zu können.
- » Reaction Evaluation: Beinhaltet die Sammlung von Informationen über gegenwärtige und zukünftige Reaktionen der Teilnehmer auf die Trainingsmaßnahme, um die Maßnahme später verbessern zu können.
- » Outcome Evaluation: Beinhaltet die Sammlung von Informationen über die Ergebnisse einer Trainingsmaßnahme, um zukünftige Maßnahmen verbessern zu können. Die Ergebnisse eines Trainings lassen sich in drei Gruppen unterteilen:
 - Immediate Outcomes: Damit bezeichnen Warr et al. (1978) „changes in knowledge, skills or attitude“ (p. 18). Diese können direkt im Anschluss an ein Training gemessen werden, bevor der Teilnehmer diese im Arbeitskontext anwendet.
 - Intermediate Outcomes: Hiermit ist eine Veränderung der Arbeitsleistung „on-the-job“ (p. 18) gemeint. Veränderungen der „Immediate Outcomes“ sind dafür laut den Autoren eine notwendige, jedoch nicht hinreichende Bedingung.
 - Ultimate Outcomes: Damit sind Veränderungen des Arbeitsverhaltens ganzer Abteilungen oder der gesamten Organisation gemeint, was z. B. zu einer Verminderung der Kosten, der Unfallhäufigkeit und der Fehlzeiten führen kann. Warr et al. (1978) sehen hier die größten Schwierigkeiten bei der Evaluation, da die tatsächlichen Auswirkungen von Trainingsmaßnahmen nur schwer zu identifizieren seien.

Damit unterscheiden sowohl Warr et al. (1978) als auch Kirkpatrick zwischen Reaction und Results bzw. Output, wobei Warr et al. (1978) die Context und die Input Evaluation zu dem Evaluationsprozess hinzufügen. Laut Nork (1989) ermöglicht das Einbeziehen der Input Evaluation eine Ermittlung des Return on Investment einer Trainingsmaßnahme, da die Höhe der Investitionen (Input) dem Ergebnis (Output) gegenübergestellt werden kann. Dieser Aspekt

wird auch von Phillips (1996, 1997, 1998a, 1998b, 1998c) in der Berechnung des ROI eines Trainings berücksichtigt, welche nachstehend beschrieben wird.

3.2 Die Analyse des Return on Investment nach Phillips

Laut Phillips (1998b) und Phillips und Phillips (2001) wird mit der Messung des Return on Investment (ROI) einer Trainingsmaßnahme das 4-Ebenen-Modell von Kirkpatrick um eine fünfte Ebene erweitert. Für Phillips (1996) ist eine Evaluation auf vier Ebenen in wirtschaftlichen Zeiten, in denen vermehrt die Messbarkeit von HR Interventionen gefordert wird, nicht mehr ausreichend: „A fifth level is needed . . . requiring business results to be converted to a monetary unit“ (p. 316). Dennoch wird die Ermittlung des ROI einer Trainingsmaßnahme in der vorliegenden Arbeit als ein eigenständiger Unterpunkt der relevanten theoretischen Modelle und Konzepte und nicht als Erweiterung des Kirkpatrick-Modells behandelt, da in diesem Ansatz zur Trainingsevaluation eine explizite Gegenüberstellung von erzieltm Gewinn und getätigten Investitionen in Geldeinheiten stattfindet. Damit wird – im Gegensatz zu der Beschreibung von Ebene 4 in dem Kirkpatrick-Modell – die Anwendung einer bestimmten Technik aufgezeigt. In diesem Zusammenhang stellen Wang et al. (2002), welche die von Kirkpatrick vorgestellten vier Ebenen ebenfalls wie Holton (1996) nicht als klassisches Modell, sondern eher als Klassifikationsschema und „communication tool“ (p. 209) betrachten, fest: „It is inappropriate to parallel a specific approach or application of a formula [d. h. ROI] with a classification scheme [d. h. die 4 Ebenen nach Kirkpatrick]“ (p. 209). Des Weiteren unterscheidet sich die Berechnung des ROI dahingehend vom Kirkpatrick-Modell, dass auch – ähnlich wie von Warr et al. (1978, vgl. Kap. 3.1.5) gefordert – der „Input“ eines Trainings berücksichtigt wird und als Trainingskosten in die Berechnung eingeht.

Studien zur monetären Bewertung eines Trainingsprogramms können bis in die 50er Jahre des 20. Jahrhunderts zurückverfolgt werden (Wang et al., 2002). Mincer (1962) stellte folgende Überlegungen an: „(1) How large is the allocation of resources to the training process? (2) What is the rate of return on this form of investment?“ (p. 50). Damit spricht er wesentliche Aspekte an, die noch heute in die Berechnung des ROI eingehen: Der erste angesprochene Aspekt betrifft die Kosten eines Trainings, der zweite entspricht der eigentlichen Berechnung des ROI. Laut Phillips und Phillips (2001) wird das Konzept des ROI bereits seit Jahrzehnten genutzt, um die Auswirkungen von organisationalen Prozessen und Programmen zu messen. Ende der 90er Jahre war die Berechnung des ROI einer Trainingsmaßnahme einer der herausforderndsten und interessantesten Bereiche der HRD-Abteilungen in Organisationen (Phillips, 1997).

Phillips (1996, 1997, 1998a, 1998b, 1998c) entwickelte ein step-by-step Verfahren, wie der ROI eines Trainings gemessen werden sollte und fasst die einzelnen Schritte in dem „ROI process model“ (1997, p. 35) zusammen, welches seiner Meinung nach eine systematische Vorgehensweise bei der Berechnung des ROI gewährleistet. Er geht dabei auch auf Methoden der Datengewinnung ein und auf Techniken, die es ermöglichen, die Effekte des Trainings von Störfaktoren zu isolieren. Außerdem nennt der Autor Strategien, wie die gewonnenen Daten am besten in „monetary values“ (1996, p. 336) umgewandelt werden können. Zu Beginn werden zwei Schritte des ROI-Modells näher beschrieben: Zum einen wird auf eine Möglichkeit der Datengewinnung eingegangen und zum anderen wird der letzte Schritt in dem Modell, die eigentliche Berechnung des ROI, vorgestellt. Daran anknüpfend werden die mit einer ROI-Analyse verbundenen Vor- und Nachteile aufgezeigt.

3.2.1 Möglichkeiten zur Erhebung von Nutzen und Kosten eines Trainings

Laut Parry (1997) lassen sich Auswirkungen eines Trainings in vier Kategorien unterteilen:

- » Zeit (z. B. Reduktion des Zeitaufwands, um Tätigkeiten auszuführen; schnellere Projektabwicklung)
- » Quantität/ Produktivität (z. B. Verbesserung der Absatzzahlen; Erhöhung der Produktivitätsrate)
- » Qualität (z. B. Verringerung des Ausschusses)
- » Mitarbeiterdaten (z. B. geringere Fehlzeiten; weniger Beschwerden)

Diese Auswirkungen, welche bereits im Vorfeld des Trainings als messbare Ziele kommuniziert werden sollen, können unter Zuhilfenahme von Mitarbeitergehältern oder Kosten pro produzierter Einheit in monetäre Geldeinheiten umgewandelt werden. Eine häufig angewandte Methode, um die Auswirkungen eines Trainings zu bemessen, ist die Follow-Up-Befragung der Trainees, welche ca. 2-3 Monate nach der Trainingsteilnahme stattfinden sollte: „Questionnaires can be used to obtain subjective information about participants, as well as to objectively document measurable business results for an [sic!] ROI analysis“ (Phillips, 1997, p. 45). In dem Fragebogen werden die Teilnehmer gebeten, den erzielten Trainingseffekt aufzuzeigen. Die Teilnehmer schätzen also den Effekt ihrer Verhaltensänderung bezüglich einer der vier oben aufgeführten Kategorien. So kann dieser Effekt z. B. sein, dass Teambesprechungen statt durchschnittlich 75 Minuten nach dem Training noch 55 Minuten dauerten, also um jeweils 20 Minuten verkürzt werden konnten, außerdem konnte die Anzahl der Meetings von

durchschnittlich acht auf fünf pro Woche verringert werden¹. An dieser Stelle bieten sich verschiedene Verfahren an, den monetären Wert dieses Effekts (in diesem Fall die 20 Minuten Zeitersparnis) zu bestimmen. Für das oben genannte Beispiel erscheint am sinnvollsten, diesen Wert durch den Personalverantwortlichen bestimmen zu lassen (Parry, 1997). Ihm stehen die notwendigen Angaben, in diesem Fall die Mitarbeitergehälter, für die Berechnung zur Verfügung. Eine weitere Möglichkeit besteht darin, die Trainingsteilnehmer zu fragen, mit welcher jährlichen Verbesserung in Geldeinheiten dieser Trainingseffekt verbunden ist – laut Phillips (1997) eine der am schwierigsten zu beantwortenden Fragen. Um die geschätzte verbesserte Arbeitsleistung in Geldeinheiten zu korrigieren und somit eine konservative Schätzung zu gewährleisten, wird der Teilnehmer anschließend gefragt, wie sicher er sich dieser monetären Schätzung ist und um eine Angabe zwischen 0 und 100% gebeten. Mit diesem Prozentwert wird die geschätzte jährliche monetäre Verbesserung multipliziert, um so einen korrigierten Wert für diese zu erhalten. In beiden Rechenarten sollte abschließend die ermittelte jährliche Leistungsverbesserung mit einem sog. Isolations-Faktor multipliziert werden, um aufzuzeigen, welcher Anteil der Verbesserung ausschließlich auf das Training zurückzuführen ist. Phillips (1997) schlägt Techniken vor, mit Hilfe derer der Faktor erhoben werden kann, diese sollen hier jedoch nicht näher besprochen werden. Das Produkt aus Isolations-Faktor und ermittelter Leistungsverbesserung in Geldeinheiten stellt den Gesamtgewinn eines Trainings dar.

Die Kosten eines Trainings können laut Phillips (1997) weit einfacher als die „Benefits“ erhoben werden und lassen sich in folgende Kategorien unterteilen:

- » Analysis Costs: Oft übersehene Kosten einer Bedarfsanalyse.
- » Development bzw. Acquisition Costs: Kosten für die Kursentwicklung bzw. Preise für den Erwerb externer Trainingsmaßnahmen und deren Lizenzgebühren.
- » Delivery Costs: Machen den größten Teil der Gesamtkosten aus und umfassen Gehälter, Material und Kosten für Reise, Verpflegung, Unterkunft.
- » Evaluation Costs: Kosten für Entwicklung von Evaluationsinstrumenten, Datengenerierung, Analyse und Präsentation der Ergebnisse.

Mit Hilfe der Kategorien können die Kosten in einer Kosten-Klassifikationsmatrix aufgeschlüsselt werden (siehe Tab. 3).

¹ Eine ausführliche Darstellung des Beispiels findet sich bei Parry, 1997, S. 174ff.

Tabelle 3: Kosten-Klassifikationsmatrix

Expense Account Classification		Process/ Functional Categories			
		Analysis	Development	Delivery	Evaluation
00	Salaries and Benefits – HRD Personnel	X	X	X	X
01	Salaries and Benefits – Other Company Personnel		X	X	
02	Salaries and Benefits – Participants			X	X
03	Meals, Travel, and Incidental Expenses – HRD Personnel	X	X	X	X
04	Meals, Travel, and Accommodation – Participants			X	
05	Office Supplies and Expenses	X	X	X	X
06	Program Materials and Supplies		X	X	
07	Printing and Reproduction	X	X	X	X
08	Outside Services	X	X	X	X
09	Equipment Expense Allocation	X	X	X	X
10	Equipment – Rental		X	X	
11	Equipment – Maintenance			X	
12	Registration Fees	X			
13	Facilities Expense Allocation			X	
14	Facilities Rental			X	
15	General Overhead Allocation	X	X	X	X
16	Other Miscellaneous Expense	X	X	X	X

Quelle: Return on Investment in training and performance improvement programs (p. 144), von J. J. Phillips, 1997, Houston, TX: Gulf.

Die Summe aller Kosten geht als „Program Costs“ (Phillips, 1997, p. 153) in die unten stehende Berechnung mit ein. Zu unterscheiden ist laut Noe (2002) zwischen „one-time costs“ (p. 202), die einmalig in der Analyse und der Programmentwicklung auftreten und „costs per offering“ (p. 202), die bei jeder Trainingsdurchführung erneut anfallen.

3.2.2 Die Berechnung des ROI

Nach Phillips (1997) wird der ROI eines Trainings meist wie folgt beschrieben: Von dem Gesamtgewinn des Trainings werden die Trainingskosten subtrahiert, um so den Nettogewinn zu erhalten, welcher anschließend durch die Trainingskosten dividiert wird. Der ROI stellt somit das Verhältnis zwischen Netto-Gewinn (Net Program Benefits) und Investition (Program Costs) dar und wird durch folgende mathematische Formel ausgedrückt:

Gleichung 1

$$\text{ROI}(\%) = \frac{\text{Net Program Benefits}}{\text{Program Costs}} \times 100$$

„This is the same basic formula used in evaluating other investments where the ROI is traditionally reported as earning divided by investments“ (Phillips, 1997, p. 33). Beträgt der ROI

eines Trainings 50%, so bedeutet dies, dass die Kosten gedeckt und zusätzlich 50% der Kosten als Gewinn erzielt wurden. Dies soll anhand eines Beispiels veranschaulicht werden (Phillips, 1998b): Ein Training zur Lese- und Schreibfertigkeit produzierte einen Gesamtgewinn in Höhe von \$321.600 und Kosten in Höhe von \$38.233. Der Netto-Gewinn beträgt also $\$321.600 - \$38.233 = \$283.367$. Eingesetzt in Gleichung 1 ergibt sich folgender ROI:

$$\text{ROI}(\%) = \frac{\$283.367}{\$38.233} \times 100 = 741\%$$

D. h., dass es für jeden in das Trainingsprogramm investierten Dollar einen Ertrag von \$7,4 Netto-Gewinn nach Kostendeckung gibt. Die Gewinne werden meist als jährliche Gewinne ausgedrückt, „representing the amount saved or gained for a complete year after program completion“ (Phillips, 1997, p. 12). Bereits im Vorfeld sollte von HR- und Trainingsverantwortlichen vereinbart werden, „what level of ROI is acceptable“ (Noe, 2002, p. 204), wofür auch die ROI-Ergebnisse ähnlicher Trainings genutzt werden können.

3.2.3 Vor- und Nachteile des ROI

Als einen Vorteil der Messung des ROI nennt Cascio (2000): „The advantage of ROI is that it blends in one number all the major ingredients of profitability and the ROI statistic can be compared with other opportunities elsewhere inside or outside the company“ (p. 127). Besonders der zweitgenannte Aspekt ist laut Parry (1997) entscheidend: Trainingsetats der Personalentwicklung können nun – basierend auf finanziell aussagekräftigen und vergleichbaren Daten – gerechtfertigt werden. Sie werden nicht länger als „act of faith or a cost of doing business“ (Parry, 1997, p. 156) oder „overhead expense“ (Davidove & Schroeder, 1998, p. 206) gesehen, sondern lassen sich wie andere Investitionsoptionen im Unternehmen legitimieren. Für Shelton und Alliger (1998) ist der ROI einer Trainingsmaßnahme außerdem „easy to compute and almost universally understood in the business arena“ (p. 173). Einen weiteren Vorteil sieht Parry (1997) darin, dass nun ein besseres Commitment der Trainees zur Trainingsmaßnahme geschaffen werden kann. Die Trainees sind auf einmal mitverantwortlich für den Erfolg einer Maßnahme, da Erwartungen an die Maßnahme geknüpft sind, welche im Nachgang überprüft werden. Dieses Bewusstsein kann zur Folge haben, dass Trainees zukünftig verstärkt Verantwortung für die Anwendung des Gelernten im Arbeitskontext übernehmen.

Nach Wang et al. (2002) lässt sich bei der Schätzung der Leistungsverbesserung in Geldeinheiten durch Trainingsteilnehmer bzw. deren Vorgesetzte kritisch anmerken, dass eine gesicherte Grundlage für diese Schätzung – wenn überhaupt – nur sehr vage existiert, da sich vie-

le Indikatoren eines HR-Programmes nicht in monetäre Werte konvertieren lassen. „The ROI results obtained in this fashion can be not only subjective but also misleading“ (Wang et al. 2002, p. 211). Und obwohl Phillips (1997) einige allgemeine Richtlinien nennt, Trainingseffekte zu isolieren, bleibt das Fehlen einer „rigorous methodology for isolating HRD impact and separating it from other variables“ (p. 210) ein Hindernis für die Anwendung des ROI und seiner Berechnung.

Wie zu Beginn dieses Kapitels erwähnt, verfolgt Kirkpatrick (1998a) weder das Ziel, konkrete Techniken für die Evaluation der Ergebnisse eines Trainingsprogramms aufzuzeigen, noch stellt er den Anspruch, die Benefits sowie die Kosten eines Trainings in monetären Einheiten auszudrücken und miteinander zu vergleichen. Zu den Schwierigkeiten einer Berechnung des ROI merkt er an:

I almost laugh when I hear people say that training professionals should be able to show benefits in terms of return on investment (ROI). The same thought occurs to me when they expect trainers to relate training programs directly to profits. Just think of all the factors that affect profits. And you can add to the list when you consider all the things that affect ROI. (p. 63)

Hunter, Schmidt und Coggin (1988) üben Kritik an der Annahme, alle PE-Programme mit Hilfe von betriebswirtschaftlichen Methoden der Finanz- und Investitionsrechnung, zu denen auch die ROI-Analyse gehört, angemessen evaluieren zu können. Nach Meinung der Autoren führt die Ermittlung des ROI im HRD nicht zu aussagekräftigen Ergebnissen, da diese Methode einen im Weiterbildungsbereich entscheidenden Faktor außer Acht lässt: Weiterbildung sei nicht auf einen „one-periode time horizon“ (p. 522) ausgerichtet, sondern müsse meist als „multiperiode investment“ (p. 522) betrachtet werden. Die Daten, die in die Berechnung des ROI einfließen, scheinen den Autoren nicht geeignet, die längerfristigen Benefits zu erfassen, die ein Training einer Organisation liefern kann. Um Investitionen in derart komplexe Unternehmensbereiche wie die Personalentwicklung aufzeigen zu können, bedarf es Methoden, welche die Dauer der Trainingseffekte, d. h. die Zeit, über die hinweg trainingsbedingte Verbesserungen im Arbeitsverhalten der Trainees auftreten, berücksichtigen. Die Nutzwertanalyse stellt eine Methode dar, die diesen Anforderungen gerecht wird und wird daher im nachfolgenden Kapitel vorgestellt. Hunter et al. (1988) sind der Meinung, „that financial accounting methods are frequently inapplicable to human resource programs“ (p. 522). Die Autoren hegen sogar die Befürchtung, dass das Management den im Vergleich zu anderen Investiti-

onsoptionen häufig extrem hoch ausfallenden Ergebnissen der ROI-Analyse oftmals wenig Glauben schenkt – somit hätte diese Evaluationsmethode ihr Ziel, den Wertbeitrag von PE glaubwürdig darzulegen, verfehlt. Dieser Vermutung widersprechen jedoch Cascio und Morris (1990), die darauf verweisen, dass die Glaubwürdigkeit von Evaluationsmethoden erst recht gefährdet sei, wenn anerkannte Methoden wie der ROI verändert würden, um vermeintlich glaubhaftere Ergebnisse zu erzielen.

Abschließend lässt sich sagen, dass der ROI eine Möglichkeit darstellt, mit relativ wenig zeitlichem und finanziellem Aufwand die Kosten und Erträge einer Trainingsmaßnahme gegenüberzustellen und somit PE-Investitionen vergleichbar mit anderen im Unternehmen getätigten Investitionen macht. Aus den oben genannten Gründen sollte er aber nicht als alleiniges Evaluationsinstrument, sondern zusammen mit anderen Methoden wie z. B. der folgenden Utility Analysis Anwendung finden.

3.3 Utility Analysis in der Personalentwicklung

Eine Methode, die sich als komplex und vielversprechend herausgestellt hat, um den monetären Wertbeitrag von PE-Programmen zu erheben, stellt die Nutzwertanalyse (Utility Analysis) dar (Boudreau, 1983; Boudreau & Ramstad, 2003; Cascio 1989, 2000; Schmidt et al., 1982; Whyte & Latham, 1997). Das große Interesse an dieser Methode liegt darin begründet, „that this technique ties human resource interventions to the measuring unit of the business world - dollar value“ (Law, 1995, p. 198). Auch Cascio (2000) betont, dass „the language of business is dollars“ (p. VII) und leitet daraus die Notwendigkeit ab, Weiterbildungsmaßnahmen in Geldeinheiten auszudrücken, damit zukünftige Investitionen in die Personalentwicklung der Unternehmensführung verständlich und gerechtfertigt erscheinen. Die Utility Analysis kann Etatverantwortliche unterstützen, wenn diese entscheiden müssen, welche Weiterbildungsprogramme zukünftig finanzielle Unterstützung erfahren sollen (Latham & Whyte, 1994). „Utility Analysis has been proposed as a way to help managers better understand the substantial value organizations can gain by using information from industrial and organizational psychology, economics, and information theory to make better choices regarding human resource management activities“ (Boudreau, 1991, p. 621).

Der erreichbare finanzielle Nutzen einer Weiterbildungsmaßnahme wird hierbei über verschiedene Kenngrößen geschätzt. Bewertet wird u. a. die Differenz der Arbeitsleistung trainierter gegenüber der untrainierter Mitarbeiter, die Anzahl der Personen, die ein Training durchlaufen sowie die erwartete Zeitspanne, über die hinweg ein Training die Arbeitsleistung

der trainierten Mitarbeiter positiv beeinflusst (Noe, 2002). Diese Parameter werden mit monetären Werten in Beziehung gesetzt und es gelingt somit, Kosten und Nutzen einer PE-Maßnahme zu erheben, diese einander gegenüberzustellen und abzuwägen, ob die Durchführung der Maßnahme aus ökonomischer Sicht gerechtfertigt erscheint.

3.3.1 Formel zur Nutzenerfassung nach Schmidt, Hunter und Pearlman

Eine grundlegende Formel zur quantitativen Erfassung von Weiterbildungsmaßnahmen stammt von Schmidt et al. (1982). Zur Evaluation von Trainingsmaßnahmen wird eine lineare Regressionsgleichung herangezogen, die es erlaubt, den Nutzen der Maßnahme in Geldeinheiten auszudrücken. Die Formel basiert auf dem von Brodgen (1949) sowie Cronbach und Gleser (1965), ursprünglich zur Analyse des Nutzwertes einer Personalauswahlentscheidung entwickeltem Modell, dem sog. B-C-G-Modell. Schmidt et al. (1982) setzten sich zum Ziel, „to illustrate how the linear-regression-based decision theoretic equation previously used to evaluate selection methods in dollars can be adapted to the evaluation of intervention programs in dollars“ (p. 334). Hunter et al. (1988) betonen, dass bekannte Techniken der Finanzplanung, der Investitionsrechnung und des Rechnungswesens oft „conceptually and logically inappropriate“ (p. 522) sind und nicht dazu dienen können, den Nutzen von PE-Maßnahmen zu bestimmen. Die Formel von Schmidt et al. (1982) hingegen ermöglicht eine angemessene Erfassung des Nutzens von Weiterbildungsprogrammen. Wird daher im Folgenden von der Nutzengleichung gesprochen, so nehmen wir jeweils auf die nachfolgend aufgeführte Formel nach Schmidt et al. (1982) Bezug. Der durchschnittliche trainingsbedingte Leistungszuwachs, ausgedrückt in Standardabweichungen der Leistung (d_t), wird erhoben und in jährlichen monetären Wertzuwachs konvertiert (SD_y). Dieser Wert wird mit der Anzahl (N) der Kursteilnehmer multipliziert, außerdem muss die Dauer der trainingsbedingten Leistungssteigerung (T) geschätzt werden. Von dem sich somit ergebenden Produkt werden die Kosten (C) der Weiterbildungsmaßnahme pro Person, welche mit der Anzahl der Teilnehmer (N) multipliziert werden, subtrahiert. Somit ergibt sich die folgende Formel (Darstellung in Anlehnung an Cascio, 2000):

Gleichung 2

$$\Delta U = N \cdot T \cdot d_t \cdot SD_y - N \cdot C$$

wobei

ΔU = Nutzen des Trainingsprogramms ausgedrückt als Leistungszuwachs in Geldeinheiten

N = Anzahl der Teilnehmer

T = erwartete Dauer, über die hinweg die Trainierten verbesserte Leistungen zeigen, die auf das Training zurückgeführt werden können

d_t = wahre Leistungsdifferenz zwischen Trainierten und Untrainierten, ausgedrückt in Standardabweichungen der Leistung

SD_y = Standardabweichung der Leistung der untrainierten Gruppe in Geldeinheiten

C = Kosten des Trainings pro Teilnehmer

3.3.2 Methoden zur Bestimmung der Parameter

Die einzelnen Parameter der Nutzgleichung stellen unterschiedlichste Anforderungen an ihre Anwender. Einige lassen sich recht objektiv bestimmen, andere hingegen weisen einen stark subjektiven Charakter auf. Zur Bestimmung der Parameter d_t , SD_y und T muss oftmals auf Schätzmethoden zurückgegriffen werden um das Gewicht der Faktoren zu bestimmen. Ausgewählte Methoden zur Erhebung dieser nicht direkt bestimmbareren Größen werden nachfolgend beschrieben.

3.3.2.1 Schätzung der Effektgröße

Der Parameter d_t ersetzt in der Formel von Schmidt et al. (1982) das Produkt aus dem Validitätskoeffizienten des Auswahlinstruments (r_{xy}) und dem durchschnittlichen standardisierten Prädiktorwert der ausgewählten Bewerber (\bar{Z}_x), welches im ursprünglichen B-C-G-Modell auftritt. Die Effektgröße d_t beschreibt den Abstand der Mittelwerte zwischen den trainierten und den untrainierten Mitarbeitern in Einheiten der Standardabweichung und somit auch die Abweichung der Traineedaten von der Nullhypothese, die aussagt, das Training habe keinerlei Effekt, bzw. die Job Performance der trainierten Gruppe unterscheidet sich nicht von der Leistung der untrainierten Gruppe (Cascio, 2000).

Gleichung 3

$$d = \frac{\bar{X}_t - \bar{X}_u}{SD_y}$$

wobei

d = Effektgröße

\bar{X}_t = mittlere Arbeitsleistung der trainierten Gruppe (Post-Training-Messung)

\bar{X}_u = mittlere Arbeitsleistung der untrainierten Gruppe (Pre-Training-Messung)

SD_y = Standardabweichung der Arbeitsleistung der beiden Gruppen (sind die SDs der beiden Gruppen nicht identisch, so sollte die SD der untrainierten Gruppe verwendet werden, da dieser Wert repräsentativer für einen Großteil der Arbeitnehmer-Population ist)

Die obige Formel legt zur Schätzung der Effektgröße eine perfekt reliable und objektive Messung der Arbeitsleistung zugrunde. Wird die Arbeitsleistung jedoch über Expertenurteile erhoben, so ist eine Interraterreliabilität von 1.00 bei der Bewertung der Arbeitsleistung unwahrscheinlich und eine Korrektur dieser Reliabilität wird somit erforderlich (Cascio, 2000; Gülpen, 1996; Schmidt et al. 1982). Oftmals wird die von King, Hunter und Schmidt (1980) über quantitative Literaturstudien ermittelte mittlere Korrelation von $r=.60$ zwischen zwei Ratern verwendet. Diese Korrekturen der Effektgröße ergeben die tatsächliche Differenz zwischen der Arbeitsleistung eines durchschnittlichen trainierten und der eines durchschnittlichen untrainierten Mitarbeiters (d_t).

Gleichung 4

$$d_t = d / \sqrt{r_{yy}}$$

3.3.2.2 Methoden zur Schätzung der SD_y

Vielfach diskutiert wird der Parameter SD_y , der die Standardabweichung der Leistung in Geldeinheiten ausdrückt. Diese Konstante wird häufig als das am schwersten zu bestimmende Element der Nutzengleichung und sogar als Achillesferse der Nutzwertanalyse (Cronbach & Gleser, 1965) bezeichnet, da die Gewinnung zuverlässiger Daten zu ihrer Bestimmung sich als außergewöhnlich schwierig herausstellt (Law, 1995; Raju et al., 1990; Siemers, 1995). Cabrera und Raju (2001) beschreiben die SD_y als „crucial variable in the model“ (p. 92); Boudreau und Ramstad (2003) bezeichnen sie als „most esoteric of utility parameters“ (p. 216) und treffen nach Sichtung vieler Studien folgende Aussage: „There is agreement that the performance transformation [into dollar values] requires expert judgement, and the choice of slope-scaling factors (e.g. compensation, sales, full employment costs) remains a challenge for which we have little research and few tools“ (p. 214). Auch Höft (2001) hält fest, dass die SD_y -Schätzung nach wie vor eine Schwierigkeit der Nutzenanalyse darstellt, da sie abhängig von den eingesetzten Schätzmethode relativ großen Schwankungen unterliegt. Cascio (2000) beschreibt die Kalkulation der Standardabweichung der Leistung zwar als schwierig, weist jedoch darauf hin, dass diese mit Hilfe relativ objektiver Daten aus dem Bereich der Humanvermögensrechnung mit hinreichender Genauigkeit geschätzt werden könne.

Bei der Berechnung der SD_y im PE-Bereich ist zu berücksichtigen, dass die Varianz der Arbeitsleistung hier immer geringer ausfallen wird als im Bereich der Personalauswahl, da im Vorfeld bereits eine Selektion aus den potentiellen Mitarbeitern stattgefunden hat (Gülpen, 1996). Außerdem muss beachtet werden, dass Tätigkeiten in Arbeitsfeldern, die eine geringe-

re Komplexität aufweisen, sich gleichzeitig durch eine geringere Variabilität der Arbeitsleistung auszeichnen und die SD_y daher stets geringer ausfällt (Carbrera & Raju, 2001). Ein Training für Inhaber weniger komplexer Berufe muss daher eine sehr hohe Qualität und damit einhergehend eine große Leistungsverbesserung unter den Teilnehmern aufweisen, bis trainingsbedingte Optimierungen (ersichtlich an Leistungsdifferenzen d_t zwischen trainierten und untrainierten Mitarbeitern) ersichtlich und der Nutzen der Maßnahme deutlich werden können (Schmidt et al., 1982). Somit wird erkennbar: „The standard deviation of job performance in dollars (SD_y) must be compatible with the effect size measure [d_t] in order to produce meaningful final utility estimates“ (Law, 1995, p. 197).

Mehrere Autoren haben wiederholt versucht, Methoden aufzuzeigen, mit denen die Schätzung der SD_y optimiert, objektiviert und zudem vereinfacht werden kann (Boudreau & Ramstad, 2003). Die in der Literatur am häufigsten diskutierten Methoden zur Schätzung der SD_y werden im Folgenden vorgestellt.

3.3.2.2.1 Das globale Schätzverfahren

Nachdem sich die Messung der Standardabweichung der Leistung mit Hilfe von Daten aus der betrieblichen Erfolgsrechnung in der Praxis als sehr zeitaufwendig, kostspielig und nahezu unmöglich erwies (Cascio, 2000), lieferten Schmidt, Hunter, McKenzie und Muldrow (1979) mit der Entwicklung des globalen Schätzverfahrens der SD_y laut Siemers (1995) sowie Cabrera und Raju (2001) einen ersten praktikablen Ansatz zur Messung des monetären Wertes von Arbeitsleistung. Schmidt et al. (1979) schlagen vor, mit Hilfe eines standardisierten Fragebogens, Vorgesetzte die Leistungen sowohl durchschnittlicher als auch über- sowie unterdurchschnittlicher Mitarbeiter in Geldeinheiten schätzen zu lassen. Als Hilfestellung empfehlen Schmidt et al. (1979) die Überlegung, welche Kosten ein externer Anbieter für die jeweiligen Leistungen in Rechnung stellen würde: „The mental standard to be used by the supervisor-judges is the estimated cost to the organization of having an outside consulting firm provide the same products and/or services“ (p. 619). Unter Annahme einer Normalverteilung der Arbeitsleistung entspricht die absolute Differenz zwischen dem mittleren Leistungswert eines durchschnittlichen Mitarbeiters (Prozentrang 50) verglichen sowohl mit dem mittleren Wert eines Spitzenmitarbeiters (Prozentrang 85) als auch mit dem durchschnittlichen Leistungswert eines problematischen Mitarbeiters (Prozentrang 15) der SD_y (Cascio, 2000; Schmidt et al., 1979; Schmidt et al., 1982; Siemers, 1995). Die zweifach geschätzte SD_y (Differenz der Prozentränge 15 und 85 zu Prozentrang 50) wird gemittelt und es ergibt sich somit ein Wert, der begründbar und nachvollziehbar ist (Cabrera & Raju, 2001). Voraussetzung für

die Annahme einer Normalverteilung ist nach Schmidt et al. (1979), dass die Abstände der Schätzungen der Prozentränge 15 sowie 85 zum Prozentrang 50 gleich sind: „If the distribution is approximately normal, these two estimates will not differ substantially in value“ (p. 621). Verschiedene Studien (Bobko, Karren & Parkington, 1983; Greer & Cascio, 1987; Schmidt et al., 1979) konnten die Annahme einer normalverteilten Arbeitsleistung nachweisen und damit zeigen, dass die Schätzungen mit der Methode von Schmidt et al. (1979) zu zuverlässigen Ergebnissen führen. Kritik erfährt diese Methode häufig aufgrund der Abstraktheit der Prozentrang-Bezugsbasis und den damit einhergehenden subjektiven Urteilen (Siemers, 1995). Darüber hinaus ist das globale Schätzverfahren bei der Bewertung von Leistungen, für die keine Marktpreise zur Orientierung herangezogen werden können, nur schwer anwendbar. Studien, die das globale Schätzverfahren mit zwei weiteren Schätzmethoden vergleichen, ergaben, dass dieses Verfahren durchgängig höhere Schätzungen liefert, als die beiden im Anschluss vorgestellten Methoden (CREPID-Methode und 40-Prozent-Regel) dies tun (Bobko et al., 1983).

3.3.2.2 CREPID-Methode

Cascio und Ramos (1986) schlagen als Alternative zum Ansatz von Schmidt et al. (1979) vor, die Leistungsbeurteilung eines jeden Mitarbeiters zur Schätzung der SD_y heranzuziehen und somit den „Cascio-Ramos Estimate of Performance in Dollars“ (CREPID) zu berechnen. Zu Beginn des Verfahrens, welches aus acht Schritten besteht, werden Einzeltätigkeiten individueller Mitarbeiter identifiziert, auf siebenstufigen Likert-Skalen bezüglich der beiden Arbeitsanalyse-Skalen Zeitaufwand/Häufigkeit und Wichtigkeit² beurteilt und anschließend anhand dieser Attribute gewichtet (Cascio & Ramos, 1986; vgl. Cascio, 2000). Entsprechend der jeweiligen Gewichtung wird das Bruttojahresgehalt eines Mitarbeiters auf die Einzeltätigkeiten verteilt. Somit wird der individuelle Wert jeder Tätigkeit in Geldeinheiten ausgedrückt. In einem nächsten Schritt wird die Qualität der Mitarbeiterleistung für jede Einzeltätigkeit auf einer Skala von 0-200 durch den jeweiligen Vorgesetzten eingeschätzt, wobei davon ausgegangen wird, dass 100 Punkte einer durchschnittlichen Leistung entsprechen. Cascio und Ramos (1986) definieren durchschnittliche Leistung dabei folgendermaßen: „This employee is better than 50% of those I’ve seen do this activity“ (p. 22). Folglich wird einem überdurchschnittlichen Mitarbeiter mit einem Prozentrang von 75 häufig der Skalenwert 150 zugewiesen. Bei der Konstruktion der Skala wurde von der Annahme ausgegangen, dass selbst die

² Ursprünglich arbeiteten die Autoren mit vier Arbeitsanalyse-Skalen: „time/frequency, level of difficulty, importance, and consequence of error“ (Cascio & Ramos, 1986, p. 21). Die Skalen „consequence of error“ und „level of difficulty“ konnten jedoch aufgrund von Interkorrelationen vernachlässigt werden. „Only the time/frequency and importance of each principal activity need to be rated in order to derive and assign an overall weight to each principal activity“ (Cascio & Ramos, 1986, p. 27).

besten Mitarbeiter nicht mehr als die doppelte Leistung der durchschnittlichen Mitarbeiter erbringen: „Second-level managers indicated that the very best first-level manager was typically not more than twice as productive as the average first-level manager. Thus, the 0- to 200-point scale was used” (Cascio & Ramos, 1986, p. 22). Weiterführend erfolgt eine Division der Leistungsbeurteilung durch 100 und eine Multiplikation dieses Wertes mit den gewichteten Einzeltätigkeitswerten in Geldeinheiten. Die somit erhaltenen Netto-Wertbeiträge der Einzeltätigkeiten werden summiert und ergeben so den ökonomischen Wert der Arbeitsleistung eines jeden Mitarbeiters, welcher i. d. R. über dessen Gehalt liegt. Cascio und Ramos (1986) bezeichnen das Ergebnis dieses vorletzten Schrittes der CREPID-Methode als die Bestimmung des „overall net worth of each individual to the organization“ (p. 22). Abschließend werden die umfassenden Werte der Arbeitsleistung aller untersuchten Mitarbeiter zusammengefasst, um den Mittelwert sowie die Standardabweichung der in Geldeinheiten ausgedrückten Arbeitsleistung zu berechnen.

Die CREPID-Methode wird in der Praxis oft dem globalen Schätzverfahren vorgezogen, da sie in ihrem Aufbau praxisüblichen Verfahren der Leistungsbeurteilung ähnelt (Siemers, 1995) und daher eine höhere Inhaltsvalidität aufweist (Greer & Cascio, 1987). Kritiker dieser Methode merken an, dass das Verfahren nicht anwendbar sei, wenn es für eine Position die SD_y zu berechnen gilt, die nur einmal im Unternehmen existiert, da erst die Berechnung des Mittelwerts und der Standardabweichung der monetär bewerteten Arbeitsleistung über eine genügend große Stichprobe zur SD_y führe (Gerlich, 1999).

3.3.2.2.3 40-Prozent-Regel

Der Aufwand der CREPID-Methode und auch der des globalen Schätzverfahrens nach Schmidt et al. (1979) kann für Unternehmen in der Praxis zu hoch sein. Für den Fall, dass Zeit und Ressourcen eine sehr detaillierte Schätzung von SD_y nicht erlauben, empfehlen Schmidt und Hunter (1983) eine Schätzmethode, die von der Annahme eines proportionalen Zusammenhangs zwischen Arbeitsleistung und Gehalt ausgeht. Demnach liegt die Standardabweichung der Leistung in Geldeinheiten bei 40-70% des durchschnittlichen Bruttojahresgehaltes einer Zielposition (Schmidt et al., 1982). Von Schmidt und Hunter (1983) durchgeführte Untersuchungen stützen die These, dass richtige bzw. sogar konservative Schätzungen erzielt werden, wenn die Varianz der Leistung mit 40% des Jahresgehaltes in die Berechnungen eingeht. Eine konservative Schätzung der SD_y im PE-Bereich sollte für Unternehmen Bedeutung haben, deren Ziel es ist, den Wertbeitrag bedeutender Trainings eher zu niedrig als zu hoch zu erfassen, um sich somit erhobener Effekte sicher sein zu können. Fehlerhafte Entscheidungen,

die hohe Kosten zur Folge haben, können somit minimiert werden. Gleichzeitig sollte die Varianz der Leistung laut Boudreau (1991) aber auch nicht unterschätzt werden, da „severely understated utility estimates“ (p. 657) dazu führen können, dass sinnvolle PE-Maßnahmen fälschlicherweise abgelehnt werden. Der von Schmidt und Hunter (1983) empfohlene Wert von 40% wird in der aktuellen Literatur zum Thema Training und Evaluation allgemein anerkannt.

Ein großer Vorteil der 40-Prozent-Regel im Unterschied zu den anderen beiden vorgestellten Schätzmethode ist, dass diese nicht von Expertenurteilen abhängig ist und daher sowohl mit den geringsten Kosten als auch dem geringsten Aufwand einhergeht (Gülpen, 1996). Hazer und Highhouse (1997) stellen fest, dass in einem Vergleich der bekanntesten drei Methoden zur Schätzung der SD_y die 40-Prozent-Regel die höchste Akzeptanz unter Praktikern erfährt. Wurde diese Methode zur Berechnung des SD_y -Parameters der Nutzengleichung herangezogen, so erschien die Utility Analysis den Entscheidungsträgern glaubwürdiger und fand vermehrt Anwendung.

3.3.2.3 Schätzung der Dauer von Trainingseffekten

Einen weiteren kritischen Parameter in der Nutzengleichung stellt die erwartete Dauer (T) der Trainingseffekte auf die Leistung der Trainierten in Jahren dar. Parry (1998) bezeichnet diesen Faktor als „payback period“ (p. 202), da er die Dauer angibt, über die hinweg das Unternehmen Nutzen aus einem Training generiert, obwohl dieses bereits abgeschlossen wurde. „Benefits accrue long after training, and can be projected over the life of the trainees in the job for which they were trained“ (Parry, 1998, p. 202). Laut Parry (1998) bestimmt die Anwendungshäufigkeit der im Training vermittelten Inhalte die Dauer der „payback period“. Fertigkeiten und Verhaltensweisen, die im Anschluss an die Maßnahme regelmäßig praktiziert werden, können sich mitunter über die gesamte Zeit, die der trainierte Mitarbeiter im Unternehmen verbleibt, auszahlen. Besteht hingegen nur sporadisch die Möglichkeit, neu Erlerntes anzuwenden, so verkürzt sich diese sog. „payback period“. Auch Morrow, Jarrett und Rupinski (1997) betonen den positiven Zusammenhang zwischen der Dauer (T) der Trainingseffekte und „on-the-job reinforcement“ (p. 100). Baldwin und Ford (1988) sowie Gülpen (1996) verweisen auf die Abhängigkeit der Effektdauer der Weiterbildungsmaßnahme von der Vergessensrate, welche wiederum stark von der Art der vermittelten Trainingsinhalte und vom Trainingsdesign beeinflusst wird. Schmidt et al. (1982) gehen von folgender Annahme aus: „The effects of organizational interventions probably decline gradually with time rather than disappearing abruptly“ (p. 339). Verschiedene Autoren (Cascio, 2000; Morrow et al.,

1997; Parry, 1998) räumen ein, dass die Schätzung dieses Parameters eine schwierige Aufgabe darstellt. „In practice, we have little knowledge about the maintenance of training effects (Cascio, 2000, p. 306)“. Cascio (2000) merkt jedoch gleichzeitig an: „The lesson to be learned from those results is that if we do not understand the maintenance of training, we do not really understand the effect of training on organizational performance“ (p. 306). Eine Möglichkeit zur Schätzung der Größe T stellt die Delphi-Methode dar, bei der Experten beauftragt werden, die Dauer der Trainingseffekte zu schätzen (Cascio, 2000). Die individuellen Einschätzungen werden anonymisiert und einem neutralen Dritten, dem sog. Mediator, zugetragen. Dessen Aufgabe besteht in der Zusammenfassung der individuellen Daten sowie deren Rückmeldung an die teilnehmenden Experten, bevor diese in die zweite Beurteilungsrunde gehen. Die Prozedur der individuellen Einschätzungen wird so lange wiederholt, bis die Experten einen Konsens erreicht haben.

3.3.3 Weiterentwicklungen der Formel zur Nutzenerfassung nach Schmidt, Hunter und Pearlman

In aktuellen Weiterentwicklungen der Formel von Schmidt et al. (1982) wird die Notwendigkeit der Integration weiterer Faktoren in die Nutzenberechnung betont, „to provide a more complete and precise utility definition“ (Boudreau, 1983, p. 551). Ausgewählte Ansätze zur Optimierung der Formel sowie zu ihrer verbesserten Anwendbarkeit in der Praxis sollen nachfolgend dargestellt werden.

3.3.3.1 Erweiterung um drei ökonomische Konzepte nach Boudreau

Boudreau (1983) betont, dass die bisherige Nutzenberechnung nach Schmidt et al. (1982) „fails to reflect certain economic concepts which are basic to organizational investment decisions“ (p. 553) und fordert, in die investitionstheoretische Betrachtung von PE die gleichen ökonomischen Erwägungen einzubeziehen wie bei traditionellen Investitionsentscheidungen. Der Autor erweitert die Formel um drei ökonomische Überlegungen: die Differenzierung zwischen fixen und variablen Kosten (1), die Berücksichtigung von Gewinnsteuern (2) sowie die Diskontierung der Trainingserträge über die Zeit (3). Die Erweiterung des Modells um diese Faktoren verfolgt das Ziel, die Investitionen in PE-Maßnahmen mit anderen Investitionsmöglichkeiten im Unternehmen sowie deren Berechnungsarten vergleichbar zu machen. „Investments in personnel programs must be evaluated similarly to other investment options so that their bottom-line implications can be readily compared“ (Boudreau, 1983, p. 553). Boudreau (1983) zeigt, dass die Vernachlässigung dieser drei ökonomischen Konzepte zu verzerrten, häufig erhöhten Nutzenschätzungen führt. Die Berücksichtigung von variablen Kosten, Steuern und Diskontsätzen führt laut Boudreau (1983) zu einer „substantially diffe-

rent, more realistic, and more complete description of the effect of personnel programs“ (p. 574). Die in Geldeinheiten bewerteten Trainingseffekte verringern sich zwar unter Verwendung der modifizierten Methode zur Nutzenbestimmung, sie weisen jedoch auch nach der modifizierten Nutzenberechnung noch einen beträchtlichen Nutzenzuwachs auf und liefern somit „more compelling evidence for the utility to be derived from personnel programs“ (Boudreau, 1983, p. 574). Laut Gerlich (1999) kann die Anwendung des weiterentwickelten Modells zur Nutzwertanalyse nach Boudreau dazu beitragen, dass „Bildungs- oder Personalverantwortliche in Budgetdiskussionen größeres Gehör finden, da sie sich einer ökonomischen Nomenklatur und Vorgehensweise bedienen“ (S. 64).

3.3.3.2 Berücksichtigung des betroffenen Tätigkeitsbereiches nach Gülpen

Gülpen (1996) merkt an, dass sich ein Training für gewöhnlich nicht auf das gesamte Arbeitsspektrum einer Person bezieht. Aus diesem Grund wird empfohlen, im Zuge der Nutzenberechnung nur „den Anteil des durch das Training betroffenen Tätigkeitsbereiches (A) zu berücksichtigen“ (Gülpen, 1996, S. 55). Eine Verbesserung der Arbeitsleistung kann nur in den direkt durch die Weiterbildungsmaßnahme betroffenen Tätigkeitsbereichen erwartet werden. In der ursprünglichen Formel von Schmidt et al. (1982) wird angenommen, dass ein Training den gesamten Produktivitätsbereich eines Mitarbeiters umfasst; diese Vermutung kann jedoch zu verzerrten Ergebnissen führen. Wird die Formel um den Parameter A erweitert, so lautet diese (Darstellung in Anlehnung an Cascio, 2000):

Gleichung 5

$$\Delta U = N \cdot T \cdot A \cdot d_i \cdot SD_y - N \cdot C$$

Betrifft eine Maßnahme einen Tätigkeitsbereich, der für einen trainierten Mitarbeiter nicht relevant ist, so erfolgt keine Produktivitätssteigerung und die Maßnahme leistet somit auch keinen positiven Wertbeitrag (Gülpen, 1996). Der Faktor A wäre in diesem Fall gleich Null und der Nutzen des Trainings damit negativ in Höhe der Gesamtkosten. Mögliche für das Unternehmen wichtige, jedoch nicht monetisierbare Effekte werden in diesem Modell vernachlässigt.

3.3.4 Kritik an der Utility Analysis

Die Utility Analysis beachtet zahlreiche Faktoren zur Schätzung des finanziellen Nutzens von PE-Maßnahmen. Boudreau und Ramstad (2003) würdigen die „useful logic and rigor“ (p. 200) der Methode, merken jedoch kritisch an: „The complexity and assumptions may reduce credibility and usefulness“ (p. 200). Diese Befürchtung unterstützen Studien von Latham und Whyte (1994) sowie Whyte und Latham (1997), die ergaben, dass erfahrene Manager weit

weniger Bereitschaft zeigen, ein hypothetisches Auswahlinstrument in ihr Unternehmen zu implementieren, wenn neben der schriftlichen Empfehlung des Verfahrens durch einen Experten auch eine Utility Analysis (die Manager erhalten eine schriftliche Präsentation der Utility Analysis, sehen ein Video, in dem ein Experte diese Methode empfiehlt und haben die Möglichkeit, im Anschluss an den Kurzfilm Fragen direkt an diesen Experten zu richten) herangezogen wird, um den finanziellen Nutzen des Verfahrens zu verdeutlichen. „These findings were surprising, and undercut a central assumption of utility analysis that managers find this type of information helpful in decision making“ (Whyte & Latham, 1997, p. 602). Die Ergebnisse widerlegen entgegen allen Erwartungen die Annahme, dass die Nutzwertanalyse ein Instrument darstellt, welches Manager in ihrer Entscheidungsfindung unterstützt. Grundannahmen der Nutzwertanalyse wurden durch diese Studien in Frage gestellt, was dazu führte, dass sie vielfach diskutiert sowie repliziert wurden und Anlass zu vielen weiterführenden Untersuchungen zum Thema Brauchbarkeit der Utility Analysis gaben.

Mögliche Gründe, die Whyte und Latham (1997) zur Erklärung ihrer Ergebnisse anführen, finden sich in ähnlicher Form in der Kritik wieder, die Boudreau und Ramstad (2003) an der Utility Analysis in ihrer derzeitigen Form üben. Die Autoren vermuten, dass Manager der Nutzwertanalyse mit großer Skepsis begegnen, da sie der Beitrag, den diese Methode zu leisten vermag, nicht überzeugt. Die Führungskräfte in der Studie von Whyte und Latham (1997) standen vor der Aufgabe, eine Entscheidung zu fällen und ggf. Entscheidungsträger des Unternehmens von deren Richtigkeit zu überzeugen. Die Ergebnisse der komplexen Utility Analysis waren den Managern zu abstrakt und unverständlich, lieferten somit keinen zusätzlichen Erkenntnisgewinn und boten daher keine ersichtliche Unterstützung in der Entscheidungsfindung. Die Methode erschien den Führungskräften aus diesen Gründen weder hilfreich noch sinnvoll und wurde abgelehnt (Whyte & Latham, 1997). Ein Großteil der aktuellen Literatur zur Utility Analysis beschäftigt sich mit diesem Aspekt und untersucht die der Methode entgegengebrachte Akzeptanz (Carson, Becker & Henderson, 1998; Cronshaw, 1997; Hazer & Highhouse, 1997).

Carson et al. (1998) replizierten die zitierte Studie von Latham und Whyte (1994) und stellten die Hypothese auf „that the manner in which the utility information was presented led to the counterintuitive results“ (p. 85). Ihr Untersuchungsdesign weist ein zusätzliches Szenario auf, in dem die Nutzwertanalyse kürzer und verständlicher erklärt wird, ansonsten sind die Bedingungen identisch. Die Autoren können die überraschenden Ergebnisse von Latham und Why-

te (1994) nicht replizieren, sondern geringe bis moderate positive Effekte nachweisen, wenn die schriftliche Empfehlung zur Anwendung des hypothetischen Verfahrens um eine Utility Analysis ergänzt wird. Die Werte der Versuchspersonen ihrer Studie liegen sowohl im Bereich der Akzeptanz als auch in dem des Verständnisses der Utility Analysis über denen der ursprünglichen Studie – die Hypothese der Autoren wird somit bestätigt.

Interessante Anmerkungen zu der Studie von Whyte und Latham (1997) stammen auch von Cronshaw (1997), der in dieser Studie die Rolle des Experten spielte, der auf dem Video zu sehen war und außerdem als Ansprechpartner vor Ort agierte und Fragen zur Utility Analysis beantwortete. Der Autor berichtet, im Video ausschließlich Argumente genannt zu haben, die die Nutzwertanalyse unterstützen; Kritik an der Methode hätte er absichtlich vermieden.

As a result, the subjects might well have interpreted my message as high-pressure or coercive communication that was designed to persuade them to accept the utility analysis, rather than a neutral message intended to inform than a neutral message intended to inform them of a potentially valuable decision-making tool. (Cronshaw, 1997, p. 612)

Cronshaw (1997) nimmt an, dass die negativen Reaktionen der Manager auf die Nutzwertanalyse darauf zurückzuführen sind, dass die Methode sehr eindringlich und zwingend kommuniziert wurde. Es sei anzunehmen, dass Manager eher bereit sind, ihren Entscheidungen die Utility Analysis zugrunde zu legen, wenn das Verfahren mit dem Ziel vorgestellt worden wäre, umfassende Informationen zu liefern anstatt zu überreden.

Mattson (2003) vergleicht in einer Studie drei Evaluationsmethoden, nämlich „utility analysis, critical outcome technique, and the anecdotal approach“ (p. 131), hinsichtlich der wahrgenommenen Nützlichkeit und Anwendbarkeit der damit gelieferten Informationen: „Attitudinal response toward using information is, in turn, a function of two beliefs: perceived ease of use of the information and perceived usefulness of the information“ (p. 129). Seine Ergebnisse widerlegen zum Teil die Ergebnisse der Studie von Whyte und Latham (1997). Mattson (2003) konnte zeigen, dass die Führungskräfte den mit Hilfe der Utility Analysis gewonnenen Informationen ein höheres Maß an Nützlichkeit beimaßen als den rein subjektiven Informationen, die über Gespräche mit Experten (anecdotal approach) gewonnen wurden. Dennoch konnte auch dieser Autor die Hypothese stützen, dass Entscheidungsträger die Anwendbarkeit der Nutzwertanalyse häufig aufgrund ihrer Komplexität anzweifeln und daher lieber auf weniger komplexe Methoden – in diesem Fall die Critical Outcome Technique – zurückgreifen.

Morrow et al. (1997) konnten belegen, dass das Verständnis sowie die Vorerfahrungen, die Manager der Nutzwertanalyse entgegenbringen, eine entscheidende Rolle bei ihrer Akzeptanz und Anwendung spielen. Eine Studie dieser Autoren konnte keine generelle Ablehnung der Utility Analysis durch Führungskräfte bestätigen, sie unterschied sich jedoch in einem entscheidenden Punkt von den Studien von Latham und Whyte (1994) sowie Whyte und Latham (1997): Die Entscheidungsträger ihrer Studie kannten die Methode und befürworteten deren Einsatz bereits im Vorfeld. „This point related to management approval of a utility analysis model *prior* [Hervorhebung durch Verf.] to its application and consideration in a decision-making context is particularly important in light of recent critiques of utility analysis“ (Morrow et al., 1997, p. 114).

Law (1995) kritisiert die Unsicherheit in der Interpretation der Ergebnisse der Nutzenschätzung, welche zu großen Anteilen auf die kritische Variable der Standardabweichung (SD_y) zurückzuführen ist. Es ist unklar, welche Aussagen auf Basis der Resultate der Utility Analysis getroffen werden können. Hunter et al. (1988) haben unterschiedliche Definitionen von ΔU zahlreicher Autoren (u. a. Bobko et al., 1983; Boudreau, 1983; Schmidt et al., 1979; Schmidt et al., 1982) analysiert und kommen zu dem Schluss, dass eine alleinige korrekte Definition von Utility nicht existiert. „For some purposes, one index of utility might be more useful and for other purposes another might be preferable. . . . The benefits from personnel programs can be expressed in many different ways to answer different questions and address different concerns“ (Hunter et al., 1988, p. 526). Hierbei wird deutlich, dass die Anwendung der Nutzwertanalyse eine gewisse Ambiguitätstoleranz und Entscheidungsbereitschaft ihrer Anwender erfordert. Es sollte situationsspezifisch entschieden werden, welche Parameter (z. B. Steuern, variable Kosten, relevante Tätigkeitsbereiche, etc.) in der Utility-Gleichung berücksichtigt werden.

Kritiker der Utility Analysis heben zudem wiederholt die Komplexität der Methode hervor, welche sowohl deren Verständnis als auch ihre Akzeptanz erschwert (Hazer & Highhouse, 1997; Whyte & Latham, 1997). Einen klaren Standpunkt zu dieser Kritik vertreten Cascio und Morris (1990):

Even if these techniques are not currently understood by the human resource professional, the task of the scholar is to explain the valid technique so as to enhance its application, not to avoid it because it is difficult to explain. (p. 414)

Auch Rauschenberger und Schmidt (1987) betonen die Notwendigkeit, ein breiteres Verständnis der Utility Analysis in der Praxis zu schaffen, da Entscheidungsträger in Unternehmen die Methode verstehen müssen, bevor sie sich bereit erklären, diese anzuwenden.

Neben den Schwierigkeiten, welche die Interpretation der Utility Analysis oftmals mit sich bringt, wird häufig auch an der Höhe der über die Nutzenschätzung erhaltenen Ergebnisse Kritik geübt. Unterschiedliche Autoren (Hunter et al., 1988; Jones & Wright, 1992; Mattson, 2003) vertreten die Ansicht, dass der erwartete „payoff“ (Boudreau, 1983, p. 553) von Weiterbildungsmaßnahmen durch die Utility Analysis häufig zu hoch angesetzt wird. Jones und Wright (1992) kritisieren, dass die Utility Analysis den Nutzen von HRD-Maßnahmen systematisch überschätzen würde. Nach Meinung der Autoren ist diese Verzerrung darauf zurückzuführen, dass der Kostenfaktor in der Nutzengleichung nicht ausreichend spezifiziert sei. „The result is that utility estimates are currently unbounded“ (Jones & Wright, 1992, p. 272). Die Autoren empfehlen, Transaktionskosten im sozialen Austausch, die im Zusammenhang mit Weiterbildungsmaßnahmen auftreten, detaillierter zu berücksichtigen, um zu einer exakteren Schätzung des Nutzens von PE-Maßnahmen zu gelangen. Die Ergebnisse würden somit realistischer und ihre Glaubwürdigkeit größer. Boudreau (1995) nimmt an, dass es Managern häufig gar nicht um eine exakte Nutzenerfassung von Weiterbildungsmaßnahmen gehe. Vielmehr stünden trainingsbedingte, beobachtbare Verhaltensänderungen, welche die hohen Ergebnisse der Utility Analysis rechtfertigen können, im Mittelpunkt ihres Interesses. Eine häufig gestellte Frage an Anwender der Nutzwertanalyse laute daher: „What behaviors will I observe that will support the high utility values you obtained?“ (Boudreau, 1995, p. 5)

Abschließend lässt sich festhalten, dass die Utility Analysis trotz all der aufgeführten Bedenken bezüglich der Aussagekraft und Anwendbarkeit dieser Methode einen Ansatz darstellt, der geeignet erscheint, den finanziellen Wertbeitrag einer PE-Maßnahme angemessen zu erfassen. Mattson (2003) bezeichnet die Nutzwertanalyse als ein „viable means of evaluating the potential effectiveness of an HRD intervention“ (p. 148) und merkt an, dass diese Methode wie kaum ein anderer Ansatz den Erhalt quantitativer Informationen erlaube. Morrow et al. (1997) halten fest, dass nicht allein die Utility Analysis, sondern sämtliche existierende Methoden zur Schätzung des Nutzens und der Kosten eines Trainings in Geldeinheiten Parameter enthalten, die nicht objektiv erhoben werden können und somit Fehler enthalten. Dennoch ist die Anwendung der Utility Analysis häufig sinnvoll, da sie viele unterschiedliche Faktoren berücksichtigt und somit erlaubt, Entscheidungen auf einer möglichst breiten Informationsba-

sis zu fällen. „As with other decisions, training resource allocation decisions need to be based upon good information; perfectly accurate information is *not* [Hervorhebung durch Verf.] critical“ (Morrow et al., 1997, p. 113). Auch Boudreau (1991) weist darauf hin, dass es die Hauptaufgabe der Utility Analysis sei, einen Leitfaden zu liefern, anhand dessen Investitionsalternativen hinsichtlich ihrer Nützlichkeit verglichen werden können. Wird die Utility Analysis in diesem Sinne als Entscheidungshilfe verstanden, so muss sie die Fähigkeit aufweisen „to describe, predict, explain, and improve decisions“ (Boudreau, 1991. p. 630) – hingegen ist es nicht ihr ultimatives Ziel, die Auswirkungen einer PE-Maßnahme völlig fehlerfrei zu messen. Natürlich muss auch die Nutzwertanalyse bei ihren Bemühungen, die Realität möglichst genau abzubilden, mit Vereinfachungen arbeiten, dennoch ist ihre Anwendung immer dann sinnvoll und vertretbar, wenn Entscheidungsprozesse mit ihrer Hilfe strukturierter ablaufen und Entscheidungen somit sorgfältiger und mit mehr Bedacht getroffen werden können, als dies ohne die Utility Analysis möglich wäre.

4 Auswirkungen verschiedener Transferbedingungen auf den Trainingserfolg

Sowohl bei der 3. Ebene im Kirkpatrick-Modell als auch bei der Berechnung des ROI nach Phillips und bei der Schätzung der Effektgröße im Rahmen der Utility Analysis kommt dem Transfer einer Trainingsleistung eine entscheidende Bedeutung zu. „Transfer is the core issue with respect to linking individual change to the requirements of the organizational system“ (Kozlowski & Salas, 1997, p. 255) Der Fokus betrieblicher Trainingsmaßnahmen liegt also auf einer Veränderung des Traineeverhaltens im Arbeitskontext (Mathieu & Martineau, 1997; Yamnill & McLean, 2001). Goldstein und Ford (2002) betonen genau diesen Handlungstransfer, d. h. die Anwendung des Gelernten im Arbeitskontext, als wichtigen Aspekt einer jeden Weiterbildungsmaßnahme und stellen die Bedeutung der „improved performance in another environment“ (p. 1) heraus. Das übergeordnete Ziel einer jeden Weiterbildungsmaßnahme, nämlich optimal ausgebildete Mitarbeiter im Unternehmen zu haben, welche sowohl derzeitigen als auch zukünftigen Anforderungen erfolgreich begegnen können, ist jedoch nur unter Bereitstellung geeigneter Transferbedingungen zu erreichen.

Ein Hauptproblem der Trainingserfolgskontrolle liegt für Childs und Bell (2001) in der quantitativen und qualitativen Messung der Leistungsverbesserung, welche auf das Training zurückgeführt werden kann. Wird im Zuge der Trainingsevaluation festgestellt, dass ein Training unzureichende Ergebnisse auf der Verhaltensebene liefert und die Zielvorgaben nicht

erreicht wurden, so gilt es festzustellen, welche Gründe dafür vorliegen könnten. Hemmende Faktoren müssen mit Hilfe der Evaluation aufgedeckt werden, damit diese im besten Fall aufgelöst und der Transfer eines Trainings gewährleistet werden kann. In einer Literaturübersicht fassen Baldwin und Ford (1988) – und darauf aufbauend Noe und Ford (1992) – drei Transferinflussgrößen zusammen: Trainingsdesign, Traineeigenschaften und Arbeitsplatzumgebung. Diese drei Transferbedingungen werden im Folgenden beschrieben und anschließend werden Maßnahmen aufgezeigt, wie sie im Evaluationsprozess Berücksichtigung finden sollten.

4.1 Trainingsdesign und Bedarfsanalyse

Das Trainingsdesign, das den Inhalt und den Aufbau eines Trainings umfasst, wird von Baldwin und Ford (1988) als wichtige Transferbedingung bezeichnet. „Training design refers to factors built into the training program to increase the chances that transfer of training will occur” (Noe, 2002, p. 153). Dieser Bereich beinhaltet auch alle Bedingungen, bei denen zur Verbesserung der Trainingsgestaltung auf angemessene Lernprinzipien zurückgegriffen wird, wie z. B. ausreichend Zeit zur Anwendung des Lernstoffs in Übungen, aktive Einbeziehung der Teilnehmer durch Aufnahme ihrer Fragen und Anmerkungen (Hummel, 2001), die genaue Simulation des zukünftigen Anwendungsbereichs im Training oder zeitlich verteiltes statt massiertes Lernen (Höft, 2001). Da laut Nork (1989) nur wenige Trainingssituationen den tatsächlichen Arbeitsbereich repräsentieren, ergeben sich Schwierigkeiten dabei, die außerhalb des Arbeitsbereiches stattgefundenen Lernprozesse auf Verhalten innerhalb der Arbeitssituation zu transferieren. Wilkening (2002) betont die Notwendigkeit der frühzeitigen Beurteilung externer Trainingskonzeptionen durch die Bildungsverantwortlichen, was zeigt, dass dem Evaluationsprozess bereits vor der Ausgestaltung des Trainingsdesigns eine hohe Bedeutung zuteil werden sollte. Dies gilt auch für die Bedarfsanalyse: „One of the first steps in developing a training evaluation plan is the identification of the training requirements for the system” (Childs & Bell, 2001, p. 478). Die Bedarfsanalyse liefert die Informationen, welche für die Ausgestaltung des Trainingsprogramms notwendig sind (Goldstein & Ford, 2002), um Arbeitssituationen möglichst realitätsnah in Trainingssituationen simulieren und Trainingsziele formulieren zu können.

4.2 Berücksichtigung von Traineeigenschaften

Verschiedene Autoren (z. B. Colquitt et al., 2000; Baldwin & Ford, 1988; Mathieu, Tannenbaum & Salas, 1992; Noe & Schmitt, 1986; Quiñones, 1995; Russ-Eft & Preskill, 2001; Salas & Cannon-Bowers, 2001) betonen zunehmend die Berücksichtigung von Traineeigenschaften

ten im Trainings- und Evaluationsprozess. Damit beginnt der Prozess der Evaluation bereits im Vorfeld einer PE-Maßnahme, so z. B. bei der Auswahl von Teilnehmern und ihrer Zuordnung zu angebotenen Trainingsmaßnahmen.

Goldstein und Ford (2002) verweisen auf zwei wichtige Voraussetzungen, die Trainingsteilnehmer mitbringen müssen: „Before trainees can benefit from any form of training, they must be ready to learn, that is, they must have the particular background experiences necessary for being successful in the training program, and they must be motivated to learn” (p. 110). Fehlen einem Trainee Grundkenntnisse („can do“-Faktoren) oder Lernmotivation („will do“-Faktoren), so sollte er gesondert auf das Trainingsprogramm vorbereitet oder es sollte ihm ggf. von einem Besuch der Maßnahme abgeraten werden (Goldstein & Ford, 2002; Tannenbaum & Yukl, 1992). Daher wird empfohlen, im Vorfeld jeder PE-Maßnahme zu evaluieren, ob die potenziellen Trainees zum einen alle nötigen Vorkenntnisse und zum anderen die erforderliche Motivation mitbringen, da diese Faktoren notwendige Voraussetzungen für eine erfolgsversprechende Trainingsteilnahme sind.

Kraiger (2003) definiert die „readiness for training“ (p. 173) folgendermaßen: Trainees müssen über die notwendigen kognitiven Fähigkeiten verfügen, die erforderlich sind, um die Trainingsinhalte zu verstehen; außerdem sollten sie ein umfassendes Wissen über ihre Arbeitstätigkeit und deren Funktion im Unternehmen als Ganzes haben, da sie nur dann in der Lage sein werden zu erkennen, „how the tasks, knowledge, and skills covered in training are relevant to that job“ (p. 173). Neben kognitiven Fähigkeiten spielen auch Persönlichkeitseigenschaften wie z. B. Ängstlichkeit, Kontrollerleben, Vertrauen in die eigene Selbstwirksamkeit und der Wunsch danach, hochgesteckte Ziele zu erreichen, eine wichtige Rolle (Colquitt et al., 2000).

Darüber hinaus wird der Lernmotivation eines Trainees eine große Bedeutung beigemessen, wenn es darum geht, Faktoren zu identifizieren, die den Transfer beeinflussen. Lernmotivation tritt laut Kraiger (2003) auf, wenn Trainees die Inhalte eines Trainings für wichtig halten und bereit dazu sind, „to exert effort in the learning environment“ (p. 173). Colquitt et al. (2000) konnten durch eine Metaanalyse bestätigen, dass die Lernmotivation eines Trainees sowohl einen großen Einfluss auf das Ausmaß der Lernleistung als auch auf die Aneignung und Anwendung neuer Verhaltensweisen hat. Weitere Studien konnten ebenfalls einen stark positiven Zusammenhang zwischen Lernmotivation und Wissenserwerb (Mathieu & Marti-

neau, 1997; Noe & Schmitt, 1986) bzw. zwischen Lernmotivation und gewünschter Verhaltensmodifikation (Noe, 1986) der Trainingsteilnehmer aufzeigen. Colquitt et al. (2000) haben die Ergebnisse von über 100 Studien zusammengetragen, die sich mit dem Zusammenhang zwischen der Lernmotivation der Trainingsteilnehmer und unterschiedlichen Indikatoren für Trainingserfolg (u. a. Lern- und Transferleistung) befassen. Lernmotivation wird hier als ein multidimensionales Konstrukt verstanden, dass sowohl von Traineeigenschaften als auch von organisationalen Rahmenbedingungen einer PE-Maßnahme beeinflusst wird. Zu den Organisationsbedingungen zählt z. B., ob die Trainees freiwillig oder unfreiwillig an einer Maßnahme teilnehmen (Baldwin & Magjuka, 1997). Noe und Schmitt (1986) definieren Lernmotivation als den ausdrücklichen Wunsch der Trainingsteilnehmer, sich die Trainingsinhalte anzueignen. Noe (1986) betont, dass Motivation im Trainingskontext die Aufgabe hat, Trainees für die Maßnahme zu begeistern, sie dazu zu ermutigen, die Inhalte zu lernen und neu erlerntes Wissen sowie angeeignete Fähigkeiten im Anschluss an das Training anzuwenden. Mit Hilfe von Korrelationen konnten außerdem moderate bis starke Zusammenhänge zwischen unterschiedlichen Persönlichkeitseigenschaften (u. a. Kontrollerleben, Selbstwirksamkeit, Ängstlichkeit) und Lernmotivation festgestellt werden (Colquitt et al., 2000).

4.3 Transferkritische Umgebungsfaktoren

Mehrere Autoren (z. B. Arthur, Bennett et al., 2003; Baldwin & Ford, 1988; Brethower, 2001; Gaudine & Saks, 2004; Goldstein & Ford, 2002; Holton, Chen & Naquin, 2003; Noe & Colquitt, 2002; Richman-Hirsch, 2001; Rouiller & Goldstein, 1993; Ruona et al., 2002; Salas & Cannon-Bowers, 2001; Tracey, Tannenbaum & Kavanagh, 1995) beschäftigen sich mit organisationalen Faktoren, die den Transfer beeinflussen. Diese organisationalen Einflussfaktoren wirken sowohl vor als auch nach dem eigentlichen Training auf die Trainees ein und können somit den Transfer positiv unterstützen, erschweren oder sogar ganz unterdrücken (Kraiger, 2003; Mathieu & Martineau, 1997). Auch hier greift daher eine Evaluation, die sich ausschließlich auf den Zeitraum der Trainingsdauer beschränkt zu kurz, da sie wichtige Einflussfaktoren unberücksichtigt lässt.

Als transferkritische Umgebungsfaktoren werden u. a. „supervisory and peer support“ (Baldwin & Ford, 1988, p. 64), „tools and equipment“ (Mathieu & Martineau, 1997, p. 209), „organizational climate“ (Tracey et al., 1995, p. 241) und „organizational justice“ (Kraiger, 2003, p. 174) genannt. Trainingsteilnehmern muss im Anschluss an das Training die Möglichkeit gegeben werden, Erlerntes über einen längeren Zeitraum im Arbeitskontext anzuwenden zu können (Baldwin & Ford, 1988; Kozlowski & Salas, 1997; Noe & Colquitt, 2002), damit sie

Erlerntes fest in ihr Verhaltenrepertoire aufnehmen können. Wichtig ist hierbei, dass die Zeitspanne zwischen dem theoretischen Lernen und der praktischen Anwendung nicht zu groß ist, da ansonsten laut Hummel (2001) mit einem „Versandungseffekt“ (S. 76) zu rechnen ist. Erfahren sie die Unterstützung seitens ihres Arbeitsumfeldes nicht, besteht die Gefahr einer Lücke zwischen „posttraining potential“ (Smith-Jentsch, Salas & Brannick, 2001, p. 280) und gezeigter Leistung am Arbeitsplatz. „Trainees’ willingness and ability to perform may not translate into performance improvement if the situation inhibits their efforts“ (Mathieu & Martineau, 1997, p. 209).

Werden die oben genannten Rahmenbedingungen und Einflussfaktoren nicht berücksichtigt, kann dies die Qualität einer Weiterbildungsmaßnahme und somit auch ihren Nutzen für die Organisation erheblich schmälern. Häufig genügt laut Hummel (2001) bereits das Ausbleiben einiger unterstützender oder das Auftreten einiger hemmender Lerntransferfaktoren, um den Transfer zu verhindern – eine Erklärung dafür, warum der Autor lediglich von einer zugrundeliegenden Lerntransferquote von ca. 10% spricht.

4.4 Konsequenzen für den Evaluationsprozess

Die aufgezeigte Wichtigkeit der Transfersicherung lässt die Notwendigkeit erkennen, die unterschiedlichen Transferfaktoren im Zuge der Trainingsevaluation nicht zu vernachlässigen. Aufgrund der starken Zusammenhänge zwischen Persönlichkeitseigenschaften und Lernmotivation sowie zwischen Lernmotivation und Trainingserfolg wird zunehmend empfohlen, die interindividuell unterschiedlichen Dispositionen der Trainees bereits vor der Planung einer Trainingsmaßnahme zu berücksichtigen (Quiñones, 1997). „It may become important to collect evaluation data before training even occurs to ensure that trainees are ready for training“ (Quiñones, 1997, p. 193). Werden Eigenschaften der Trainingsteilnehmer bereits im Vorfeld der Maßnahme evaluiert, so können dabei erhobene Ergebnisse häufig zum besseren Verständnis von erzielten bzw. ausgebliebenen Trainingserfolgen herangezogen werden: Z. B. kann geklärt werden, warum es nicht bei allen Teilnehmern einer Maßnahme im jeweils gleichen Ausmaß zu einer Wissens- bzw. Leistungsverbesserung gekommen ist. Persönlichkeitseigenschaften, Erwartungen und Einstellungen (z. B. mit Hilfe von Persönlichkeits- und Motivationsfragebogen) der zukünftigen Trainingsteilnehmer müssen evaluiert und identifiziert werden, um die Wahrscheinlichkeit zu erhöhen, dass die Teilnahme an der Maßnahme zum gewünschten Erfolg führt. Werden beispielsweise falsche Erwartungshaltungen seitens der Teilnehmer erhoben, so ist es notwendig, diese vor dem Training anzusprechen, Gründe dafür zu eruieren und die Bedenken der Trainees auszuräumen bzw. deren Erwartungen mit den an

das Training gestellten Ansprüchen abzugleichen. Geschieht dies nicht, so besteht die Gefahr, dass nicht mit einer Lern- bzw. Transferleistung des Trainees gerechnet werden kann.

Außerdem wurde aufgezeigt, dass auch organisationale Umfeldbedingungen bereits vor dem eigentlichen Training analysiert werden sollten. Trainingsevaluation sollte daher den Anspruch stellen, den Erfolg einer Maßnahme im „ongoing organisational process“ (p. 207) zu bewerten (Mathieu & Martineau, 1997). Nur so ist es möglich, hemmende Faktoren rechtzeitig zu erkennen und auszuräumen, bevor der Trainee mit gewonnenem Wissen, neuen Fähigkeiten und der Erwartung an den Arbeitsplatz zurückkehrt, diese anwenden zu können. Treffen neu erlernte Verhaltensweisen auf Ablehnung und negative Reaktionen, so ist es sehr wahrscheinlich, dass der Trainingsteilnehmer diese nicht weiter anwendet, sondern sein altes Verhalten wieder annimmt. Die Evaluation muss sich hier also nicht länger allein mit dem Trainee, sondern auch mit seinem Umfeld beschäftigen und eventuelle Einflussfaktoren berücksichtigen, um somit einen optimalen Transfer zu gewährleisten.

Die obigen Ausführungen zeigen unserer Meinung nach die Notwendigkeit auf, mit der Wirkungs- und Qualitätskontrolle einer Maßnahme nicht erst nach deren Abschluss zu beginnen, sondern bereits im Vorfeld der Maßnahme die Weichen für erfolgreiche Personalentwicklung zu stellen. Zusammenfassend lässt sich diese Forderung mit einem Zitat von Childs und Bell (2001) benennen: „Evaluation should be integrated throughout the life cycle of the training“ (p. 480).

Das sog. Tannenbaum-Rahmenmodell (Cannon-Bowers, Salas, Tannenbaum & Mathieu, 1995) integriert neben weiteren Variablen die von Baldwin und Ford (1988) aufgeführten drei Randbedingungen und stellt diese mit den vier Ebenen Kirkpatrick's in Zusammenhang. Anspruch dieses Modells ist es, verschiedenste Komponenten zu berücksichtigen, die Einfluss auf den Transfer- und Trainingserfolg ausüben können. „The proposed model focuses on characteristics of the organization and work environment and characteristics of the individual trainee as crucial input factors“ (Cannon-Bowers et al., 1995, p. 143). Einflussfaktoren wie z. B. Traineeigenschaften und -motivation, Trainingsdesign sowie Organisations- und Situationsmerkmale werden umfassend beschrieben und hinsichtlich ihrer Wirkung vor, während sowie nach dem Training differenziert. Das Tannenbaum-Rahmenmodell beachtet verschiedenste Aspekte des Trainingsgeschehens sowie den Erfolg beeinflussende Variablen und kann somit „Hilfestellung zur exakteren Definition von Evaluationszielen geben“ (Höft, 2001, S.

637). Cannon-Bowers et al. (1995) entwickelten somit ein Modell, welches die vier Ebenen von Kirkpatrick nicht länger isoliert betrachtet evaluiert, sondern kritische Umgebungsfaktoren explizit in den Evaluationsprozess miteinbezieht. Somit wird hier der Forderung nachgekommen, Faktoren in der Evaluation zu berücksichtigen, die im Vorfeld des Trainings und auch im Anschluss daran Einfluss auf die Lern- und Transferleistung der Trainees nehmen.

5 Evaluationsstudie

Die beschriebenen Evaluationsansätze sollen im Folgenden anhand einer konkreten Personalentwicklungsmaßnahme der EADS Deutschland GmbH Anwendung finden. Dabei werden zunächst Informationen zu den organisatorischen Rahmenbedingungen des Trainings gegeben, die Zielsetzung des Trainings erklärt und der Trainingsauswahlprozess beschrieben. Vor dem Hintergrund der in Kapitel 3 aufgeführten Theorien und der unten stehenden Beschreibung des Trainings werden anschließend Hypothesen formuliert, die im Rahmen der Studie mit Hilfe der eingesetzten und ausführlich dargestellten Erhebungsinstrumente überprüft werden sollen. Abschließend werden die Ergebnisse der Evaluation aufgeführt und diskutiert.

Ziel dieser Evaluationsstudie war es, speziell für ein bestimmtes Trainingsmodul der EADS Evaluationsinstrumente zu entwickeln, die vielschichtige Ergebnisse und breiten Erkenntnisgewinn liefern können. Dabei sollten die vier Ebenen des Kirkpatrick-Modells berücksichtigt, der monetäre Nutzen des Trainingsmoduls mit Hilfe der Utility Analysis erhoben und der Trainings-ROI ausgewiesen werden. Bei der Konstruktion der Instrumente standen zwei Ansprüche im Vordergrund: Zum einen sollten die verschiedenen Ebenen der theoretischen Trainingsevaluation berücksichtigt werden, zum anderen galt es, Instrumente zu entwickeln, die in ihrer Struktur auf die Evaluation weiterer Trainings der EADS übertragbar sind.

Das von uns evaluierte Trainingsmodul fand am 19.04.2004 in einem Firmengebäude der EADS in Friedrichshafen/ Immenstaad in der Zeit von 9:30 Uhr bis 17:00 Uhr statt und wurde von zwei externen Trainern der Firma Kölsch & Altmann durchgeführt.

5.1 Trainingsbeschreibung

Die Personalentwicklungsmaßnahme, die Gegenstand der vorliegenden Studie ist, wurde uns freundlicherweise von der EADS als Praxispartner für die Evaluation zur Verfügung gestellt. Im Rahmen der Diplomarbeit wurde ein Trainingsmodul evaluiert, das den Titel „Systematischer Software-Test“ trägt und in einer umfangreichen Trainingsreihe eingebettet ist. Im Folgenden wird diese Trainingsreihe vorgestellt, die damit angesprochene Zielgruppe beschrie-

ben und erläutert, wie die Teilnehmer für diese ausgewählt wurden. Anschließend wird das zur Evaluation bereitgestellte Modul vorgestellt, dabei wird u. a. auf seine Kostenstruktur, Zielsetzung und derzeitige Evaluation im Unternehmen eingegangen.

5.1.1 Software integriertes Förderprogramm für Teilprojektleiter (SWIFT)

Das Modul „Systematischer Software-Test“ stammt aus einer umfassenden, bereits mehrfach durchgeführten Trainingsreihe mit dem Titel „Software integriertes Förderprogramm für Teilprojektleiter“ (SWIFT). Zielgruppe dieser Trainingsreihe sind Software (SW)-Entwickler und Systemingenieure, die mindestens ein Jahr Erfahrung in der Entwicklung von Softwaresystemen mitbringen und das Potenzial zur Übernahme von Aufgaben eines SW-Projektleiters bzw. Aufgabenleiters aufweisen. Erfahrene Mitarbeiter der Entwicklungsabteilungen sollen sich in dieser Trainingsreihe mit Methoden und Werkzeugen des SW-Managements auseinandersetzen und SW-Management-Instrumente kennenlernen, die bei der EADS Anwendung finden. Darüber hinaus sollen Entwicklungstrends im SW-Bereich aufgezeigt und diskutiert werden. Die komplette Trainingsreihe SWIFT umfasst ca. 20 Arbeitstage und verteilt sich über die Dauer eines Jahres. Die Kosten pro Teilnehmer belaufen sich auf € 6000.

5.1.2 Auswahl der Trainingsteilnehmer

Die Vorauswahl zur Teilnahme an dieser Trainingsreihe erfolgt durch die direkten Vorgesetzten im Rahmen der quantitativen und qualitativen Personalplanung. Die EADS hat hierzu ein Kompetenzmanagementmodell entwickelt, welches vorsieht, in jährlichen Mitarbeitergesprächen mit der jeweils zuständigen Führungskraft die individuelle Ist-Qualifikation des Mitarbeiters zu überprüfen. Weicht diese Ist-Qualifikation von der Soll-Anforderung ab, so werden in dem Mitarbeitergespräch die erforderlichen individuellen Kompetenzmanagementmaßnahmen vereinbart. Definieren Mitarbeiter und Vorgesetzte zu diesem Zweck Personalentwicklungsmaßnahmen, so ist der Zeitrahmen zur Durchführung der geplanten Maßnahmen für beide Seiten verbindlich festzulegen. Die Trainingsreihe „SWIFT“ stellt eine dieser Qualifizierungsmaßnahmen dar, welche zur Schließung von Kompetenzlücken im Bereich des SW-Managements angeboten werden.

5.1.3 Kosten des Moduls

Das Trainingsmodul „Systematischer Software-Test“ erstreckt sich über die Dauer eines Tages. Die Kosten für das Trainingsmodul belaufen sich auf € 2000. Die Nebenkosten (u. a. Raumkosten, Ausstattung, Kosten für Administrationspersonal) betragen insgesamt € 1266. Damit belaufen sich die Trainingskosten für diesen einen Tag auf € 3266. Am 19.04.04 nahmen 11 Teilnehmer an dem Trainingsmodul teil, d. h. die Kosten für diesen Trainingstag be-

liefen sich auf € 297 pro Teilnehmer. Berücksichtigt man zudem die Ausfallkosten (inklusive Lohnnebenkosten), die für das Unternehmen entstehen, da die Mitarbeiter für den Trainingstag entlohnt werden, obwohl sie keine Leistung am Arbeitsplatz erbringen, so erhöhen sich die Gesamtkosten des Trainings auf € 6.599, 44. Die Gesamtkosten pro Teilnehmer steigen somit auf € 599,95 an.

5.1.4 Zielsetzung des Moduls

Ausgewiesenes Ziel des Moduls ist es, die Notwendigkeit des systematischen Testens aufzuzeigen und dessen Bedeutung entlang des SW-Entwicklungsprozesses zu erkennen. Der Stellenwert des systematischen Testens im Bewusstsein der SW-Entwickler soll erhöht werden. Ziel ist außerdem, dass die Teilnehmer die verschiedenen Methoden des systematischen SW-Testens (u. a. Vor- und Nachteile, Grenzen der jeweiligen Methode) kennenlernen. Die Trainingsteilnehmer werden dafür in die Testplanung eingeführt und erhalten einen Überblick über Teststufen und Testtools. Des Weiteren sollen die Teilnehmer im Rahmen des Lernprojektes die Methoden im praktischen Einsatz und durch die gezielte Anwendung verstehen lernen, Erfahrungen gewinnen und diese vertiefen. Diese Aspekte sollen den Teilnehmern dabei helfen, eine adäquate Teststrategie für das eigene Projekt ableiten zu können.

5.1.5 Derzeitige Evaluation

Zur Zeit wird der Erfolg dieses Moduls anhand eines Feedbackbogens am Ende des Trainingsmoduls gemessen, der Teilen unseres Zufriedenheitsfragebogens ähnelt. Zur Erfolgsmessung des Trainings wird also eine Reaktionsmessung unmittelbar nach dem Training vorgenommen – die in Kapitel 3.1.2 aufgezeigten Forschungsergebnisse verschiedener Autoren (vgl. Al-Athari & Zairi, 2003; Grünewald et al., 2003; Van Buren & Erskine, 2002), die belegen, dass in der Praxis häufig lediglich auf Ebene 1 des Kirkpatrick-Modells evaluiert wird, finden hier somit Bestätigung. Vor dem Hintergrund einer umfassenden Literaturrecherche und vorhergehender theoretischer Überlegungen findet aus unserer Sicht zur Zeit keine umfassende Evaluation des Trainingsmoduls „Systematischer Software-Test“ statt.

5.1.6 Hypothesen

Vor dem Hintergrund einer umfassenden Literaturrecherche erwarten wir im Rahmen der Evaluation des Trainingsmoduls „Systematischer Software-Test“ die Bestätigung folgender Hypothesen:

1. Es besteht kein signifikanter Zusammenhang zwischen der Zufriedenheit der Teilnehmer mit dem Training und ihrer Lernleistung.

2. Von allen Zufriedenheitsfaktoren weist der Themenkomplex „Nützlichkeit des Trainingsmoduls“ die höchste Korrelation mit der Lernleistung auf.
3. Es besteht ein signifikant positiver Zusammenhang zwischen Trainingsmotivation und Lernleistung.
4. Die Anzahl der richtigen Antworten im Post-Lerntest ist signifikant höher als die Anzahl der richtigen Antworten im Pre-Lerntest.
5. Der Nutzen des Trainings ist gemessen über die Utility Analysis positiv.

5.2 Methode

Die zur Trainingsevaluation und zur Überprüfung der Hypothesen entwickelten Erhebungsinstrumente wurden in einem iterativen Prozess in Zusammenarbeit mit dem Forschungsprojekt „Corporate Universities“ der Fachhochschule Nordostniedersachsen erstellt. Die Befragung fand in Fragebogenform anhand von offenen und geschlossenen Fragen statt. Es wurden drei Fragebogen an die Teilnehmer ausgeteilt, von denen einer unmittelbar vor dem Training (Fragebogen Lernebene 1), einer unmittelbar danach (Zufriedenheitsmessung) und ein weiterer 1-2 Tage nach dem Training (Fragebogen Lernebene 2) bearbeitet werden sollte. Zur Entwicklung der notwendigen Fragebogen wurden die Trainingsunterlagen hinzugezogen, Fachliteratur studiert (z.B. Fachzeitschriften, Lehrbücher) sowie Gespräche mit IT-Spezialisten geführt. Des Weiteren wurden HR- und Trainingsverantwortliche der Firma EADS zu Hintergrund und Zielen des Trainingsmoduls befragt. Die Zustimmung der Mitarbeitervertretung seitens der EADS (Betriebsrat) wurde eingeholt. Im Hinblick auf die verschiedenen zu untersuchenden Forschungsfragen wurden verschiedene Erhebungsinstrumente eingesetzt, welche nach der Beschreibung der Stichprobe vorgestellt werden. In allen Fragebogen wurde der Nutzen der Evaluation für die Teilnehmer herausgestellt: Dieser besteht darin, dass anhand der Evaluation verschiedener Aspekte des Trainings die untersuchte PE-Maßnahme in Zukunft noch besser auf die Bedürfnisse der Mitarbeiter abgestimmt werden kann; zudem kann das Vorgehen in der Evaluation auf weitere Programme übertragen werden. Den Teilnehmern wurde absolute Vertraulichkeit im Umgang mit ihren Daten zugesichert und es wurde betont, dass Rückschlüsse auf Einzelpersonen weder gewünscht noch möglich sein werden, da alle Daten anonymisiert in die Berechnungen eingehen.

5.2.1 Stichprobenbeschreibung

Von den 11 EADS-Mitarbeitern, die an dem oben genannten Trainingsmodul teilnahmen, beteiligten sich alle an der freiwilligen Evaluationsstudie. Alle 11 Trainees waren männlichen Geschlechts und zwischen 27 und 39 Jahren alt. Zum Zeitpunkt des Trainings waren die Trai-

ningsteilnehmer durchschnittlich knapp 6 Jahre bei der EADS beschäftigt, wobei sich die geringste Beschäftigungsdauer auf 2, die höchste auf 9 Jahre belief und ihr durchschnittliches Bruttojahresgehalt € 57.230 betrug. Alle 11 Trainees konnten ein abgeschlossenes Hochschulstudium vorweisen: 6 absolvierten ihr Studium in dem Bereich der Informatik, 4 in dem Bereich des Ingenieurwesens und 1 in dem Bereich der Naturwissenschaften. 6 Trainingsteilnehmer konnten bereits vor dem Training Erfahrungen als Software-Projektleiter sammeln und 8 Personen hatten im Vorfeld des Trainings bereits mit systematischen Software-Tests gearbeitet. Ein Training zu ähnlichen Inhalten hatten bereits 3 Teilnehmer besucht, wobei diese vorherigen Trainings bereits 1998 bzw. 2001 stattgefunden hatten.

5.2.2 Zufriedenheitsmessung

Unmittelbar nach dem Training wurden die Teilnehmer um eine subjektive Einschätzung ihrer Zufriedenheit mit verschiedenen Aspekten bezüglich des Trainings gebeten. Grundlage für die Konstruktion der Themenkomplexe des Zufriedenheitsfragebogens (vgl. Anhang A) stellen Ergebnisse der Faktorenanalyse nach Morgan und Casper (2000) dar. Diese wurden bereits in Kapitel 3.1.1 vorgestellt. Der Faktor 3, welcher in der Studie von Morgan und Casper (2000) als „Zufriedenheit mit Prozess der Leistungsüberprüfung“ beschrieben wird, wurde in unserer Untersuchung nicht berücksichtigt, da das Modul „Systematischer Software-Test“ keine Leistungsüberprüfung innerhalb des Trainings vorsieht. Statt dessen wurde der Faktor „wahrgenommene Trainingsschwierigkeit“ nach Warr und Bunce (1995) hinzugefügt, um die kognitive und emotionale Beanspruchung der Teilnehmer, die sich auf den Schwierigkeitsgrad des Trainings zurückführen lässt, erheben zu können. Außerdem wurden unter einem separaten Punkt (Themenkomplex 7) Fragen zusammengestellt, die Auskunft über den Gesamteindruck der Teilnehmer und ihre Trainingsmotivation geben können. Der ausgeteilte Fragebogen beinhaltete insgesamt 31 Fragen zu folgenden Themenkomplexen/ Faktoren:

1. Zufriedenheit mit dem Trainer (Item 1-5)
2. Zufriedenheit mit dem administrativen Trainingsmanagement (Item 6-8)
3. Zufriedenheit mit dem Schwierigkeitsgrad des Trainings (Item 9-11)
4. Nützlichkeit des Trainingsmoduls (Item 12-15)
5. Zufriedenheit mit dem Trainingsumfeld (Kursmaterial, Medieneinsatz, Ausstattung) (Item 16-19)
6. Zufriedenheit mit dem Aufbau des Kurses (Item 20-24)
7. Weitere Aspekte (Item 25-31, wobei die Items 25, 27, 29 und 31 den Gesamteindruck und die Items 26, 28 und 30 die Motivation der Teilnehmer erfassen)

Es wurde eine 5er-Skale verwendet, wobei bei den Themenkomplexen 1-6 der Grad der Zufriedenheit (1= sehr unzufrieden, 2= eher unzufrieden, 3= teils/teils, 4= eher zufrieden, 5= sehr zufrieden) und bei Themenkomplex 7 der Grad der Zustimmung (1= stimme nicht zu, 2= stimme eher nicht zu, 3= teils/teils, 4= stimme eher zu, 5= stimme zu) angegeben werden sollte. Zu beachten ist dabei, dass die Items 27 und 30 bei Fragenkomplex 7 im Sinne der Merkmalsausprägung umkodiert werden müssen. Den Teilnehmern wurde außerdem die Möglichkeit gegeben, Anmerkungen bzw. Verbesserungsvorschläge am Ende jedes Themenkomplexes niederzuschreiben. Am Ende des Fragebogens wurden zusätzlich demographische Daten erhoben. Neben Geschlecht, Alter, Ausbildungshintergrund und derzeitiger Tätigkeit bei EADS interessierten hier vor allem Vorerfahrungen, die Teilnehmer mit Trainings zu inhaltsverwandten Themen gemacht hatten.

5.2.3 Lernerfolgsmessung

Für die Messung des Lernerfolgs wurde der Wissensstand der Teilnehmer mit Hilfe eines „Paper-and-Pencil Tests“ (Kirkpatrick, 1996, p. 303) vor und nach dem Training gemessen. Der Pre-Test (vgl. Anhang B) wurde den Teilnehmern am Trainingstag vor Beginn des Trainings ausgehändigt und nach der Bearbeitung von den Evaluatoren eingesammelt. Der Post-Test (vgl. Anhang C) wurde den Teilnehmern in adressierten und frankierten Briefumschlägen nach Beendigung des Trainings mit der Bitte mitgegeben, diesen innerhalb der nächsten 2-3 Tage ausgefüllt und ohne Zuhilfenahme von zusätzlichen Informationen, beispielsweise in Form von Fachliteratur, Trainingsunterlagen oder dem Rat Dritter, an die Evaluatoren zurückzuschicken. Letztendlich wurden die ausgefüllten Fragebogen in einem Zeitrahmen von bis zu zwei Wochen nach dem Training zurückgesandt. Der Fragebogen eines Teilnehmers lag zum Zeitpunkt der Datenauswertung noch nicht vor, so dass sich die Ergebnisse der Lernerfolgsmessung lediglich auf 10 Teilnehmer beziehen. Pre- und Posttest waren identisch und wurden mit Hilfe der Trainingsunterlagen der Firma Kölsch & Altmann, einschlägiger Literatur (z.B. Balzert, 1998; Knöll, Slotos & Suk, 1996; Rätzmann, 2002; Sneed & Winter, 2002) sowie dem Rat eines IT-Spezialisten ausgearbeitet. Der Fragebogen umfasste insgesamt 26 Aufgaben, darunter Multiple-Choice Fragen, offene Fragen, Alternativfragen und Zuordnungsaufgaben.

5.2.4 Transfererfolgsmessung

Die Messung des Transfererfolgs sollte ursprünglich unter der Zuhilfenahme von Fallstudien, welche von den Teilnehmern bearbeitet werden sollten, stattfinden. Aufgrund von EADS-internen Umstrukturierungen und daraus resultierenden zeitlichen Verzögerungen bei

der Kontaktaufnahme mit dem externen Trainingsanbieter, wurde auf die Erstellung von Fallstudien im Rahmen der Diplomarbeit jedoch verzichtet. Statt dessen wurde – ähnlich wie bei Gülpen (1996) – ein Transferfragebogen (vgl. Anhang D) entwickelt. Nach Absprache mit Verantwortlichen des Forschungsprojektes „Corporate Universities“ wurde der ursprünglich angedachte Erhebungszeitpunkt für den Transferfragebogen aus Validitätsgründen von ca. drei Wochen auf ca. drei Monate nach dem Training verschoben. Somit ist die Wahrscheinlichkeit höher, dass die Trainees zum Erhebungszeitpunkt die Möglichkeit haben werden, dass neu Gelernte anzuwenden. Der Zeitpunkt dieser Transfererfolgsmessung liegt daher außerhalb des für die Diplomarbeit vorgesehenen Zeitrahmens; die Erhebung wird voraussichtlich eigenständig und unter Verwendung der von uns entwickelten Instrumente von HR-Verantwortlichen der EADS durchgeführt. Neben den Fragen 1 und 2, die sich auf die Beschäftigung mit den Trainingsunterlagen und die Teilnahme an weiteren Trainings nach dem Modul „Systematischer SW-Test“ beziehen, werden hauptsächlich Fragen zur Anwendung einzelner Trainingsinhalte im Arbeitskontext formuliert. Mit Hilfe von geschlossenen Fragen soll zum einen erfasst werden, welche konkreten Inhalte des Trainings die Teilnehmer nach dem Besuch des Moduls anwenden können und zum anderen, welche dieser Inhalte die Teilnehmer bereits vor dem Besuch des Trainings angewendet haben. Außerdem ist von Interesse, wie häufig diese Inhalte vor bzw. nach dem Training Anwendung fanden und ob Schwierigkeiten bei ihrer Anwendung auftraten. Es ergaben sich somit zwei Fragenkomplexe mit je acht Fragen, wobei jeweils drei der acht Fragen aus mehreren Unterfragen bestehen. Mit Hilfe dieser Daten wird es der EADS möglich sein, zu ermitteln, welche Trainingsinhalte tatsächlich auf den Arbeitsalltag der Trainees transferiert werden konnten.

5.2.5 Auswertungsmethoden

Zur Auswertung der Fragebogen wurden verschiedene Kenngrößen aus dem Bereich der deskriptiven Statistik herangezogen (u. a. Mittelwerte, Häufigkeiten, bivariate Korrelationen, Range). Mit Hilfe dieser Berechnungen konnte ein Großteil der demographischen Angaben der Trainees ausgewertet werden. Auch ließ sich bestimmen, wie zufrieden die Trainingsteilnehmer mit einzelnen Trainingsaspekten waren und wie viele Punkte sie im Durchschnitt bei den einzelnen Lerntests erzielten. Die Stärke der Zusammenhänge zwischen den ausgewählten Faktoren der Trainingszufriedenheit nach Morgan und Casper (2000) bzw. Warr und Bunce (1995) (vgl. Kap. 5.2.2) und der Lernleistung der Teilnehmer wurde mit Hilfe der Produkt-Moment-Korrelation nach Pearson überprüft, wobei der Signifikanztest zweiseitig erfolgte. Zudem wurden alle Items, die wir entsprechend der ausgewählten sechs Zufriedenheitsfaktoren entwickelt haben, auf ihre interne Konsistenz nach Cronbachs Alpha überprüft, um so die

Reliabilität unserer Zufriedenheitsskala sowie die Trennschärfe der einzelnen Items zu überprüfen. Aus dem Bereich der statistischen Testtheorie lieferte ein Mittelwertsvergleich mittels t-Test für abhängige Stichproben Aufschluss darüber, ob die Wissensstände der Teilnehmer, die anhand des Pre-/ Post-Lerntests erhoben wurden, signifikante Unterschiede aufwiesen, die sich nicht mit zufälligen Schwankungen erklären lassen. Diese Signifikanzüberprüfung erfolgte auf einem Niveau von 1%. Zur Bestimmung des Nutzens des untersuchten Trainings wurde die Utility Analysis nach der Formel von Schmidt et al. (1982) herangezogen (vgl. Kap. 3.3.1). Dabei wurde zur Schätzung der Standardabweichung der Leistung in Geldeinheiten (SD_y) die 40%-Regel verwendet, da diese laut Literatur eine hohe Akzeptanz unter Praktikern erfährt. Außerdem lag das zur Berechnung benötigte durchschnittliche Bruttojahresgehalt der Trainingsteilnehmer der EADS vor und bedeutet für das Unternehmen somit keinen zusätzlichen Aufwand.

5.3 Ergebnisse

Bei der Beschreibung der Ergebnisse werden zunächst in einer zusammenfassenden Darstellung die explorativ ermittelten Ergebnisse bezüglich der Zufriedenheitsmessung und der weiteren Aspekte (Faktor 7 im Fragebogen zur Zufriedenheitsmessung) vorgestellt. In einem zweiten Teil der Ergebnisdarstellung werden die in Kapitel 5.1.6 aufgestellten Hypothesen überprüft.

5.3.1 Zusammenfassende Darstellung

Die über die sechs Faktoren des Zufriedenheitsfragebogens (vgl. Anhang A) gemessene durchschnittliche Zufriedenheit der Teilnehmer mit dem Training lag bei 3,8 (vgl. Anhang F.1), was im Hinblick auf die verwendete 5-er Skala (1 = sehr unzufrieden, 5 = sehr zufrieden) bedeutet, dass die Teilnehmer im Großen und Ganzen mit dem Training eher zufrieden waren. Die Reliabilitätsanalyse (vgl. Anhang F.3) konnte aufzeigen, dass die Items der sechs Themenkomplexe eine interne Konsistenz von Cronbachs Alpha = .91 aufweisen. Diese hohen Item-Interkorrelationen zeigen, dass unser Fragebogen das Konstrukt der Zufriedenheit sehr genau misst. Betrachtet man die sechs Faktoren der Zufriedenheit gesondert, indem jeweils die verschiedenen Items eines Faktors zu einem Summenitem zusammengefasst werden (vgl. Kap. 5.2.2, Übersicht über die Faktoren und die dazugehörigen Items), so wird ersichtlich, dass die Teilnehmer den Trainern (Faktor 1) mit einem mittleren Wert von 4,16 den höchsten Zufriedenheitswert entgegenbrachten (siehe. Abb. 4). Bei der Beantwortung der Items 1, 2 und 3 des Faktors „Zufriedenheit mit dem Trainer“ (Faktor 1) wählten alle 11 Teilnehmer entweder die Antwortalternative sehr zufrieden oder eher zufrieden; das Engagement

(Item 1) und die Kompetenz (Item 3) der Trainer sowie die Verständlichkeit, mit der die Inhalte präsentiert wurden (Item 2), fanden also bei den Teilnehmern besonders großen Zuspruch (vgl. Anhang F.2).

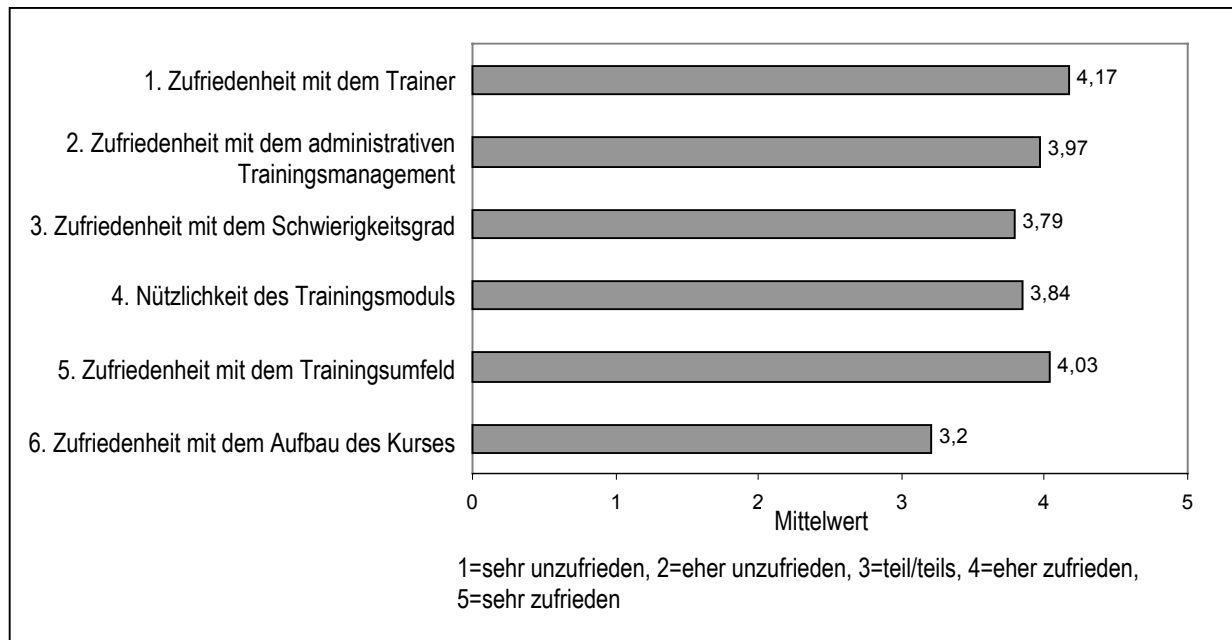


Abbildung 4: Differenzierte Betrachtung der Zufriedenheit der Trainingsteilnehmer über die sechs Zufriedenheitsfaktoren.

Wie Abbildung 4 hingegen ebenfalls zu entnehmen ist, waren die Trainees mit dem Aufbau des Kurses (Faktor 6) mit einem mittleren Wert von 3,2 am wenigsten zufrieden (vgl. Anhang F.1). Im Rahmen dieses Themenkomplexes waren 3 Teilnehmer (27,3%) mit der zeitlichen Länge des Trainings (Item 20) eher unzufrieden, einer (9,1%) sogar sehr unzufrieden (vgl. Anhang F.2). Jeweils 4 Teilnehmer (36,4%) gaben an, mit der Geschwindigkeit, mit der die Trainingsinhalte präsentiert wurden (Item 21) sowie mit der Anzahl praktischer Übungen im Training (Item 23) eher unzufrieden gewesen zu sein. Auch am Schwierigkeitsgrad des Trainings (Faktor 3), dessen mittlerer Zufriedenheitswert 3,79 beträgt, wurde von den Teilnehmern Kritik geübt; insgesamt 5 Teilnehmer (45,5%) waren mit der Menge der zu bearbeitenden Inhalte (Item 11) eher bzw. sogar sehr unzufrieden.

Der zusätzlich unter dem Themenkomplex „Weitere Aspekte“ (Faktor 7 im Fragebogen zur Zufriedenheitsmessung, vgl. Anhang A) über 4 Items erhobene Gesamteindruck der Teilnehmer vom Training weist einen mittleren Wert von 4 (Skala von 1 bis 5, vgl. Kap. 5.2.2) auf (vgl. Anhang F.1). 7 Teilnehmer (63,6%) stimmten der Aussage ganz oder zumindest eher zu, dass das Training voll und ganz ihren Erwartungen entsprochen habe (Item 25). 10 Teilnehmer (90,9%) waren der Meinung, das Training habe sich gelohnt (Item 27) und stimmten die-

ser Aussage zu bzw. eher zu; nur 1 Teilnehmer stimmte einer solchen Aussage eher nicht zu (9,1%) (vgl. Anhang F.2). Dem Item „Ich werde das Training auf jeden Fall KollegInnen weiterempfehlen“ (Item 31) stimmten 8 Trainingsteilnehmer eher zu (72,7%) - die übrigen 3 Personen (27,3%) brachten dieser Aussage nur zum Teil Zustimmung entgegen. Die ebenfalls im Themenkomplex 7 erhobene Trainingsmotivation der Teilnehmer kann mit einem mittleren Wert von 3,5 (Skala von 1 bis 5, vgl. Kap. 5.2.2) als mittelmäßig beschrieben werden (vgl. Anhang F.1). 5 Teilnehmer (45,5%) haben sich persönlich dafür eingesetzt, an dem Trainingsmodul „Systematischer Software-Test“ teilzunehmen (Item 28) und 10 Teilnehmer (90,9%) hatten laut eigener Angabe ein sehr hohes Eigeninteresse an den Inhalten des Trainings (Item 26) (vgl. Anhang F.2). Lediglich 3 Teilnehmer (27,3%) werden laut eigener Angabe eher nicht an den Trainingszielen festhalten, wenn sich zeigt, dass diese im Arbeitskontext schwer zu erreichen sind (Item 30).

5.3.2 Überprüfung der Hypothesen

Im Folgenden werden die in Kapitel 5.1.6 formulierten Hypothesen auf ihre Richtigkeit untersucht. Bei der Überprüfung der Annahmen, die sich auf Daten der Lernfragebogen beziehen, muss beachtet werden, dass nur Daten von 10 der 11 Trainingsteilnehmer berücksichtigt werden konnten, da ein Teilnehmer den Post-Lernfragebogen nicht zurückgesendet hat und dessen Werte somit nicht in die Ergebnisauswertung einfließen konnten. Zur Überprüfung der Hypothese 1 wurden die Variablen Zufriedenheit (gemessen über die 6 Zufriedenheitsfaktoren) und Lernleistung (Punktedifferenz zwischen Pre- und Post-Lernfragebogen) korreliert (vgl. Anhang F.4). Dabei konnte kein signifikanter Zusammenhang zwischen Zufriedenheit der Teilnehmer und ihrer Lernleistung festgestellt werden; die Hypothese 1 konnte somit bestätigt werden. Hingegen konnte die Hypothese 2 durch die Ergebnisse unserer Studie nicht belegt werden; der Themenkomplex „Nützlichkeit des Trainingsmoduls“ korrelierte ebenso wie die übrigen 5 Zufriedenheitsfaktoren nicht signifikant mit der Lernleistung. Auch über den von uns erwarteten signifikant positiven Zusammenhang zwischen Trainingsmotivation und Lernleistung (Hypothese 3) konnten die Daten dieser Evaluationsstudie keinen Nachweis führen. Die Ergebnisse eines Mittelwertsvergleiches mittels t-Test für gepaarte Stichproben konnten aufzeigen, dass, wie in Hypothese 4 angenommen, ein signifikanter Unterschied zwischen der mittleren Anzahl richtiger Antworten im Post-Lernfragebogen und der Anzahl richtiger Antworten im Pre-Lernfragebogen besteht (vgl. Anhang F.5). Da sich der Unterschied zwischen den Mittelwerten 21,1 (Post-Lernfragebogen) und 13,9 (Pre-Lernfragebogen) als signifikant auf dem Niveau von 1% erwies, kann mit sehr hoher Wahrscheinlichkeit davon ausgegangen werden, dass dieser Effekt nicht zufallsbedingt ist. Alle Trainingsteilnehmer

schnitten im zweiten Lernfragebogen im Vergleich zum ersten Lernfragebogen besser ab (siehe Abb. 5).

Die minimale Anzahl richtiger Antworten lag im Pretest-Learning bei 6, im Posttest bei 9; die maximale Anzahl richtiger Antworten lag im Pretest bei 20 und im Posttest bei 27. Die höchste Lernleistung wies ein Teilnehmer auf, der im Posttest 12 Fragen mehr richtig beantworten konnte, als im Pretest. Der Teilnehmer mit der geringsten Leistungssteigerung konnte im

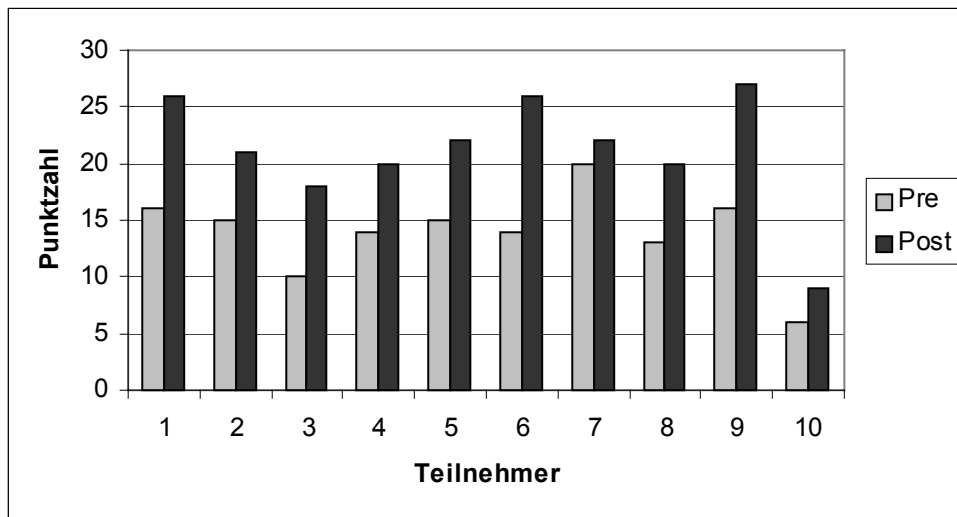


Abbildung 5: Erreichte Punktzahl in den Fragebogen zur Lernebene. Vergleich von Pre- und Post-Lernfragebogen.

Anmerkung: Da ein Teilnehmer keinen Post-Lernfragebogen zurückgeschickt hat, liegt dieser Auswertung ein N von 10 zugrunde.

Posttest nur 2 Fragen mehr richtig beantworten

Zur Überprüfung der Hypothese 5 wurde der Nutzwert nach der Formel von Schmidt et al. (1982) berechnet. Dazu wurden Angaben zu Trainingskosten, durchschnittlichem Bruttojahresgehalt der Trainingsteilnehmer, Lohnnebenkosten und voraussichtlicher Dauer der Trainingseffekte verwendet. Diese Angaben wurden uns von der EADS zur Verfügung gestellt (vgl. Anhang E). Zur Berechnung der Effektgröße wurde der mittlere Wissensstand vor und nach dem Training, ermittelt über die Differenz der Ergebnisse aus Post- und Pre-Lernfragebogen, herangezogen.

Berechnung der Gesamtkosten

Kosten für den Trainingstag € 3.266

Durchschnittliches Jahresbruttogehalt pro Teilnehmer (TN) € 57.230

Durchschnittlicher Tagessatz pro TN bei 230 Arbeitstagen	$\text{€ } 57.230 : 230 = \text{€ } 248,83$
Durchschnittliche Lohnnebenkosten pro TN p. a.	$\text{€ } 12.469$
Durchschnittliche Lohnnebenkosten pro TN pro Tag bei 230 Arbeitstagen	$\text{€ } 12.469 : 230 = \text{€ } 54,21$
Gesamtkosten für den Trainingstag mit 11 Teilnehmern	$\text{€ } 3.266 + 11 \cdot \text{€ } 248,83 + 11 \cdot \text{€ } 54,21$ $= \text{€ } 6.599,44$
Berechnung der Effektgröße	$d = \frac{\overline{X}_t - \overline{X}_u}{SD_y}$ $= \frac{21,1 - 13,9}{3,755} \approx 1,9$
Berechnung der SD _y	$\text{€ } 57.230 \cdot 0,4 = \text{€ } 22.892,08$
Parameter (einzusetzen in Gleichung 2, Kap. 3.3.1)	$N = 11$ $T = 5$ $d_t = 1,9$ $SD_y = \text{€ } 22.892,08$ $C = \text{€ } 6.599,44$
Berechnung des Trainingsnutzens	$\Delta U = N \cdot T \cdot d_t \cdot SD_y - C$ $= 11 \cdot 5 \cdot 1,9 \cdot \text{€ } 22.892,08 - \text{€ } 6.599,44$ $= \text{€ } 2.385.622,92$

Somit wurde unter Anwendung der Gleichung für die Utility Analysis des Trainings „Systematischer Software-Test“ ein positiver Nutzwert von € 2.385.622,92 nach Kostendeckung berechnet. Die Evaluation kann somit aufzeigen, dass der Nutzen, den das Trainingsmodul über die kommenden Jahre bringt, weit über dessen Kosten liegt.

5.4 Diskussion der Studie

Die Ergebnisse zeigen auf, dass das Training durch kleine Modifizierungen noch höhere Werte der Zufriedenheit bei den Teilnehmern erzielen könnte. Dazu würde es schon genügen, den Umfang zu bearbeitender Inhalte etwas zu verringern und den Trainingsteilnehmern die Möglichkeit zu geben, einen größeren Anteil der neu erlernten Verfahren bereits im Training probeweise anwenden zu können. Dies würde die Zufriedenheit der Trainees wahrscheinlich erhöhen und gleichzeitig dafür sorgen, dass die Trainingsinhalte stärker verinnerlicht und zugleich besser in das Verhaltensrepertoire der Teilnehmer implementiert werden können. Diese Ergebnisse stützen auch persönliche Anmerkungen und Kommentare einiger Trainingsteilnehmer, in denen diese z. B. darauf hinwiesen, dass das Training eigentlich um einen

Tag auf zwei Tage verlängert werden müsse, weil die aktuelle Konzeption eine zu große Menge an Inhalten in zu kurzer Zeit bearbeite.

Da die Mehrheit der Teilnehmer ganz oder zumindest eher der Aussage zustimmte, dass das Training voll und ganz ihren Erwartungen entsprach, fast alle der Meinung waren, dass sich die Teilnahme gelohnt habe und die meisten nach eigener Aussage das Training ihren Kollegen weiterempfehlen werden, kann grundsätzlich von einem positiven Gesamteindruck der Teilnehmer bezüglich des Trainings gesprochen werden. Dies ist ein wichtiger Aspekt, wenn man bedenkt, dass auch in Zukunft weitere EADS-Mitarbeiter an diesem Programm teilnehmen werden. Schon im Vorfeld kommender und im Rahmen der Trainingsreihe SWIFT stattfindender Trainingsmodule kann dies zu einer positiven Erwartungshaltung bei den Beteiligten führen.

Laut Reliabilitätsanalyse kann das Konstrukt der Zufriedenheit noch genauer gemessen werden, wenn aus dem von uns verwendeten Fragebogen zur Zufriedenheitsmessung die Items 23 und 24 entfernt werden. Diese Änderungen hätten zur Folge, dass Cronbachs Alpha von derzeit .9062 auf .9218 bzw. .9122 steigen würde. Aus statistischer Sicht hätte auch die Entfernung der Items 3 und 9 eine verbesserte Messung der Zufriedenheit zur Folge, allerdings bewirkt diese Modifikation nur eine sehr geringe Erhöhung von Cronbachs Alpha. Vorteilhaft wäre, dass die Teilnehmer bei der Eliminierung der genannten Items weniger Fragen zu beantworten hätten und daher dem Evaluationsprozess gegenüber vielleicht positiver eingestellt wären. Nachteilig dürfte sich jedoch auswirken, dass dem Evaluator somit wichtige Erkenntnisse verloren gehen würden, die von inhaltlicher Relevanz sind. Daher sollten die Items beibehalten werden, auch wenn das Verbleiben der Items im Zufriedenheitsfragebogen dessen interne Konsistenz verringert.

Die Bestätigung der Hypothese 1 deckt sich mit den in Kapitel 3.1.3.2 beschriebenen früheren Forschungsergebnissen anderer Autoren. Es konnte bestätigt werden, dass eine reine Zufriedenheitsmessung kein Gradmesser für eventuell auftretende Verbesserungen auf der Learning-Ebene des Kirkpatrick-Modells ist. Die Evaluation, wie sie bisher von EADS durchgeführt wird, sollte daher in Zukunft um weitere Aspekte ergänzt werden. Erst durch eine Evaluation auf höheren Ebenen und die Durchführung einer Utility Analysis lassen sich Rückschlüsse auf den Nutzen von PE-Maßnahmen ziehen. Hypothese 2, die Zusammenhänge zwischen der Nützlichkeitseinschätzung eines Trainings durch die Teilnehmer und deren Lern-

leistung vermutet, konnte entgegen der Ergebnisse von Alliger et al. (1997) nicht bestätigt werden, da die untersuchten Korrelationen keine signifikanten Werte aufwiesen. Auch die Ergebnisse zur Untersuchung des Zusammenhangs zwischen Trainingsmotivation und Lernleistung (Hypothese 3) fielen nicht hypothesenkonform aus. Die in Kapitel 4.2 vorgestellten Ergebnisse der Metaanalyse von Colquitt et al. (2000) fanden somit in dieser Studie keine Bestätigung. Kritisch anzumerken ist dabei vor dem Hintergrund einer statistisch problematischen Evaluationsstichprobe die Aussagekraft der Ergebnisse. Es stellt sich die Frage, ob eine Stichprobe von 11 bzw. 10 Personen tatsächlich Aussagen über ein Training ermöglicht, welches bereits seit Jahren Anwendung findet. Der Grund für die vielen nicht signifikanten Ergebnisse bei der Überprüfung der Hypothesen wird sicherlich in unserer sehr kleinen Stichprobe zu suchen sein. Hypothesen, die anhand dieser Stichprobe verworfen werden mussten, sollten daher nicht als endgültig widerlegt betrachtet werden. Gleiche Vorsicht in der Interpretation gilt für die bestätigten Hypothesen. Wir nehmen an, dass die Aussagekraft vieler Ergebnisse erhöht werden kann, indem eine erneute Evaluation des Trainingsmoduls „Systematischer Software-Test“ in einer der folgenden SWIFT-Trainingsreihen anhand derselben Erhebungsinstrumente durchgeführt und die beiden existierenden Datensätze im Anschluss daran zusammengeführt werden. Erst auf Basis einer größeren Stichprobe können wirklich aussagekräftige Ergebnisse vorgestellt werden.

Die in Hypothese 4 vermutete und durch einen Mittelwertsvergleich bestätigte Annahme, dass die Teilnahme am Training zu einem signifikanten Wissenszuwachs der Teilnehmer führt, liefert einen ersten Nachweis über die Wertsteigerung und den Nutzenzuwachs, den die Durchführung der Maßnahme „Systematischer Software-Test“ im Unternehmen bewirkt. Der unmittelbare Erfolg des Trainings auf der Learning-Ebene wird dadurch belegt, dass alle Teilnehmer in dem zweiten Lernfragebogen eine höhere Punktzahl aufwiesen als in dem ersten. Da die Inhalte der Fragebogen sehr auf die Trainingsunterlagen abgestimmt waren, konnte dieser Wissenszuwachs erwartet werden. Zu beachten ist, dass der Zeitrahmen der Fragebogenbearbeitung zwischen zwei Tagen und zwei Wochen nach dem Training variierte und der Zusicherung der Teilnehmer Glauben geschenkt werden muss, den Post-Lernfragebogen ohne Zuhilfenahme von Drittmitteln bearbeitet zu haben. Dies kann unserer Meinung nach allerdings angenommen werden, da keiner der bei uns eingegangenen Fragebogen eine unglaubliche Steigerung der Leistung aufwies. In einer Follow-up-Untersuchung ca. 3 Monate nach dem Training könnte der Post-Lernfragebogen ein weiteres Mal zum Einsatz kommen, um längerfristige Effekte auf der kognitiven Ebene zu ermitteln.

Hypothese 5 konnte bestätigt werden; der Nutzen des Trainings gemessen mit Hilfe der Utility Analysis ist positiv. Da die zur Bestimmung der Effektgröße hinzugezogene Lernleistung der Teilnehmer anhand von Fragebogen erhoben wurde, konnte diese Messung als objektiv und reliabel angesehen werden. Die geschätzte Effektgröße musste daher keinen weiteren Korrekturen unterzogen werden, wie es z. B. bei Vorgesetzteneinschätzungen, die eher subjektiven Charakter aufweisen, notwendig gewesen wäre. Der beispielhaft über die Methode der Utility Analysis berechnete Nutzen des Software-Trainings (Hypothese 5) erscheint sehr hoch. Dieses Ergebnis kann u. U. damit erklärt werden, dass die in unserer Formel verwendete Effektgröße nicht konsequent im Sinne von Schmidt et al. (1982) berechnet wurde. Im Falle der vorliegenden Evaluationsstudie wurden Lerndifferenzen anstelle von Differenzen in der Arbeitsleistung zur Berechnung der Effektgröße herangezogen (vgl. Kap. 5.3.2). Eine bessere Vorgehensweise wäre die Verwendung der mittleren Anwendungshäufigkeit der Trainingsinhalte vor und nach dem Training. Dazu benötigte Daten liegen vor, sobald die Teilnehmer den Transferfragebogen (vgl. Anhang D) bearbeitet haben werden. Da diese Daten jedoch erst 3 Monate nach unserer Diplomarbeit erhoben werden können, haben wir die Berechnung des Trainingsnutzens hier beispielhaft unter Zuhilfenahme der Wissensdifferenz vorgenommen. Wir nehmen allerdings an, dass die somit möglichen Verzerrungen des Ergebnisses minimal sind und gehen davon aus, dass der sehr hohe Nutzwert auch dann zustande käme, wenn die Effektgröße über Differenzen der Arbeitsleistung berechnet werden würde, da abgesehen von der Effektgröße alle übrigen der eingesetzten Parameter exakt nach den Vorstellungen von Schmidt et al. (1982) berechnet wurden. Vermutlich kann durch das hohe Ergebnis die Kritik vieler Autoren an der Utility Analysis bestätigt werden, welche nämlich die Ansicht vertreten, dass diese Methode den Nutzen von Weiterbildungsprogrammen systematisch überschätze (vgl. Kap. 3.3.4).

Für die Berechnung des Return on Investment dieser Trainingsmaßnahme hätten bereits im Vorfeld messbare Ziele formuliert werden müssen. Ein Großteil der jetzigen Ziele bezieht sich auf eine Erweiterung des Wissens- und Kenntnisstandes der Teilnehmer. Es geht hierbei hauptsächlich um das Kennenlernen von Methoden und Werkzeugen des systematischen Testens und damit um das Erreichen von Zielen auf der Lernebene durch die Trainingsteilnahme. Wie in Kapitel 3.2 aufgezeigt, erfordert die Berechnung des ROI quantifizierbare Daten auf der Verhaltens- oder Ergebnisebene. Im vorliegenden Fall könnte dies z. B. die Reduktion des Zeitaufwandes für die Durchführung eines kompletten Testzyklusses oder die Erhöhung der

Sensitivität zum Erkennen von Fehlern sein. Benötigt ein Mitarbeiter nun im Vergleich zu früheren Testzyklen weniger Zeit zur Durchführung eines gesamten Testzyklusses, geht dieses Zeitersparnis unter Zuhilfenahme der Mitarbeitergehälter als monetäre Größe in die Berechnung des ROI ein. Multipliziert mit dem Wert zur konservativen Schätzung und dem Isolationsfaktor ergibt sich der Gesamtgewinn des Trainings. Da eine Kostenaufstellung dieser Maßnahme bereits von der EADS vorgenommen wurde (vgl. Anhang E), wäre die Berechnung des ROI somit möglich. Beispielhaft soll eine Berechnung unter Berücksichtigung der angegebenen Trainingskosten und unter der Annahme fiktiver Zahlen durchgeführt werden. Fiktiv ist die Anzahl der durchgeführten Testzyklen pro Jahr und die jeweilige Dauer eines Testzyklusses vor bzw. nach der Maßnahme, da uns für diese beiden Aspekte keine Angaben zur Verfügung stehen.

Berechnung der Gesamtkosten

Kosten für den Trainingstag	€ 3.266
Durchschnittliches Jahresbruttogehalt pro TN	€ 57.230
Durchschnittlicher Tagessatz pro TN bei 230 Arbeitstagen	$€ 57.230 : 230 = € 248,83$
Durchschnittliche Lohnnebenkosten pro TN p. a.	€ 12.469
Durchschnittliche Lohnnebenkosten pro TN pro Tag bei 230 Arbeitstagen	$€ 12.469 : 230 = € 54,21$
Gesamtkosten für einen Arbeitstag pro TN	$€ 248,83 + € 54,21 = € 303,04$
Gesamtkosten für den Trainingstag mit 11 Teilnehmern	$€ 3.266 + 11 \cdot € 248,83 + 11 \cdot € 54,21 = € 6.599,44$

Berechnung des Gesamtgewinns

Anzahl der Mitarbeiter, die an dem Training teilnehmen	11
Durchgeführte Testzyklen pro Mitarbeiter p. a	7
Dauer der Durchführung eines Testzyklusses vor dem Training	6 Tage
Dauer der Durchführung eines Testzyklusses nach dem Training	5,5 Tage
Gesamtgewinn des Trainings	$(6 - 5,5) \cdot 11 \cdot 7 \cdot € 303,04 = € 11.667,32$
Parameter (einzusetzen in Gleichung 1, Kap. 3.2.2)	Gesamtgewinn = € 11.667,04 Gesamtkosten = € 6.599,44

ROI (%) des Trainings

$$\frac{€11.667,04 - €6.599,44}{€6.599,44} \cdot 100 \approx 77\%$$

Ein ROI von 77% bedeutet, dass es für jeden in das Trainingsprogramm investierten € einen Ertrag von € 0,77 Netto-Gewinn nach Kostendeckung gibt. Auf die Vor- und Nachteile der Berechnung und Nutzung des ROI einer Trainingsmaßnahme wurde in Kapitel 3.2.3 hingewiesen. Zu beachten gilt, dass es laut Phillips (1997) keinen allgemein akzeptierten Mindestwert für den ROI eines Trainings gibt. Nach Angaben des Autors wird in einigen Organisationen ein Mindestwert von 25% eingefordert, was über dem geforderten Return für andere Investitionsmaßnahmen liegt. Dies wird damit begründet, dass die Berechnung des ROI im Personalentwicklungsbereich nach wie vor neu ist und meist auf z. T. sehr subjektiven Schätzungen basiert.

In Kapitel 6.3.1 wird aufgezeigt, dass eine hohe Akzeptanz des Evaluationsvorhabens nur dann zu erwarten ist, wenn dieses im organisationalen Kontext hierarchisch hoch angesiedelt ist und offen kommuniziert wird. Unser Evaluationsvorhaben erfuhr starke Unterstützung von Führungskräften des Top-Managements der EADS, wie z. B. dem Vice President HR und dem Head of HR Policy & Controlling. Die Evaluation war demzufolge sehr gut in die Personalpolitik und -prozesse des Unternehmens implementiert. Die Mitarbeitervertretungen wurden früh informiert und in den Prozess eingebunden. Auch das Transferklima scheint aufgrund der jährlichen Mitarbeitergespräche, in denen erforderliche PE-Maßnahmen gemeinsam mit den direkten Vorgesetzten abgestimmt werden (vgl. Kap. 5.1.2), sehr positiv zu sein. Da die direkten Vorgesetzten unmittelbar in den Prozess der PE ihrer Mitarbeiter eingebunden sind, ist mit einem starken „supervisory support“ (Baldwin & Ford, 1988, p. 64) zu rechnen – ein Faktor, der die Transferleistung in erheblichem Maße beeinflusst (vgl. Kap. 4.3). Es kann davon ausgegangen werden, dass die direkten Vorgesetzten im Anschluss an das Training dafür Sorge tragen, dass ihre Mitarbeiter die Möglichkeit erhalten werden, Trainingsinhalte anzuwenden, da das Schließen von Kompetenzlücken am Arbeitsplatz als Ziel der Qualifizierungsmaßnahme SWIFT definiert wurde. Mit dem geplanten Einsatz des Transferfragebogens kann dies überprüft werden.

Wie bereits in Kapitel 5.2.4 angesprochen, wird der zusätzliche Einsatz von Fallstudien bei der Messung des Transfererfolgs von unserer Seite empfohlen, um zu erheben, ob Teilnehmer nach dem Training Teststrategien und Lösungsansätze auf konkrete Problemstellungen an-

wenden können. Die Erstellung der Fallstudien sollte in enger Zusammenarbeit mit dem Trainingsanbieter Kölsch & Altmann geschehen. Des Weiteren sollten bei der Messung des Transfers die in Kapitel 4 aufgeführten transferkritischen Faktoren beachtet und ggf. im Evaluationsprozess berücksichtigt werden. Dies hieße dann, dass der Prozess der Evaluation schon vor dem eigentlichen Training beginnen würde.

Ferner ist zusätzlich zu den genutzten Erhebungsinstrumenten der Einsatz von Fragebogen zu überdenken, mit deren Hilfe die Lern- und Verhaltensverbesserung der Teilnehmer durch deren Vorgesetzte eingeschätzt werden kann. Diese Fragebogen könnten den bereits entwickelten Instrumenten inhaltlich sehr ähneln. Dadurch wäre es möglich, einen Abgleich von Selbst- und Fremdeinschätzung der Lern- und Transferleistung vorzunehmen und die Evaluation auf eine objektivere Basis zu stellen. Dabei muss allerdings beachtet werden, dass dieses Vorgehen mit einem erheblichen Mehraufwand verbunden ist.

Mit der Unterstützung der EADS konnte beispielhaft aufgezeigt werden, wie die umfassende Evaluation einer Maßnahme aussehen kann. In ähnlicher Weise kann mit anderen Maßnahmen verfahren werden. Die durchgeführte Evaluation sollte als Pilotstudie verstanden werden, die aufzeigen konnte, wie die Erfolgsmessung eines Trainings praktisch durchgeführt werden kann und welche Schritte dazu unternommen werden müssen.

6 Ausblick

Die Evaluation und das Controlling von PE-Programmen wird eine Zukunftsaufgabe darstellen, der sich PE-Abteilungen verstärkt annehmen müssen, da zunehmend die Forderung des Nachweises über die wertschaffenden Auswirkungen von HRD-Aktivitäten gestellt wird. Es gilt, ein verstärktes Bewusstsein für die Notwendigkeit der Evaluation als Teil der strategischen Personalentwicklung zu schaffen, Vorurteile gegenüber den Methoden abzubauen, das Wissen über die Durchführung der Evaluation in Unternehmen zu erweitern und Organisationen bei der Implementierung theoretischer Modelle in die Praxis zu unterstützen. Außerdem müssen aktuelle Weiterentwicklungen im Bereich von Trainingsmethoden, -instrumenten und -design in der Evaluation berücksichtigt und in den Prozess der Erfolgskontrolle einer Trainingsmaßnahme aufgenommen werden.

6.1 Anwendbarkeit und Akzeptanz verschiedener Evaluationsansätze

Von wissenschaftlicher Seite wird an Unternehmen vermehrt die Forderung gestellt, Evaluationen auf der Ergebnisebene durchzuführen – es muss jedoch beachtet werden, dass hierzu

ein vorheriges Verständnis und eine damit einhergehende Akzeptanz der zur Verfügung stehenden Modelle unter den Anwendern erforderlich ist. „Attitudes about the usefulness and ease of use of HRD results lead to intentions, which lead to behaviors“ (Cascio, 2003, p. 155). In der Praxis scheinen Evaluationstätigkeiten häufig abgelehnt zu werden, da die existierenden Modelle nicht den Anforderungen der Unternehmen entsprechen: Das Verhältnis von Nutzen der theoriegeleiteten Evaluation und Kosten ihrer praktischen Anwendung muss vertretbar sein, die Methodik der Modelle soll leicht verständlich und die aus der Evaluation hervorgehenden Ergebnisse sollen nachvollziehbar sein. Wird der zusätzliche Aufwand für eine Evaluation betrieben, so sollten die Resultate auch den Wertbeitrag von PE herausstellen, um einen Vergleich zu anderen Investitionsoptionen zu ermöglichen. Die unterschiedlichen Methoden, die zur Evaluierung einer PE-Maßnahme vorgestellt wurden (vgl. Kap. 3), können jeweils Teilaspekte, kaum jedoch sämtliche dieser Forderungen erfüllen.

Mattson (2003) konnte belegen, dass Manager denjenigen Evaluationsberichten, die mit monetären Kennzahlen arbeiten und den Wertbeitrag einer Maßnahme ausweisen, am positivsten gegenüber stehen. Evaluationsmethoden, die hingegen mit Zufriedenheitsmessungen der Trainingsteilnehmer, deren Lernerfolgen oder Verhaltensänderungen arbeiten, stoßen auf geringere Akzeptanz. Entscheidungsträger lassen sich folglich leichter von der Notwendigkeit der Evaluation überzeugen, wenn diese die Möglichkeit bietet, den Wertbeitrag einer PE-Maßnahme in objektiven Zahlen, d. h. auf der Results-Ebene auszudrücken. Angesichts dieser praktischen Erfahrungen stellt sich die Frage, warum die in Kapitel 3.1.2 gezeigten Unternehmensbefragungen in der Praxis dennoch ergaben, dass lediglich 7% der befragten amerikanischen Unternehmen Evaluationen auf der Results-Ebene durchführen. Es scheint neben dem aufgezeigten Widerspruch zwischen Theorie und Praxis auch einen Widerspruch zu geben zwischen Evaluationsergebnissen, die vielversprechend hinsichtlich ihrer Akzeptanz sind und denen, die relativ einfach aufgrund ihrer verständlichen Methodik gewonnen werden können. Bei Evaluationen auf den Ebenen 1-3 erscheint die Anwendung relativ einfach und unkompliziert, den Ergebnissen fehlt jedoch die nötige Akzeptanz. Zufriedenheits- und Lernzuwacherhebungen stellen kostengünstige und verständliche Evaluationsansätze dar; da jedoch eine Darlegung des Nutzens eines Programms in Geldeinheiten durch diese Herangehensweisen nicht möglich ist, verwehren Manager deren Ergebnissen eine hohe Akzeptanz. Evaluationen, die auf der Ebene 4 und anhand von Nutzwertanalysen durchgeführt werden, liefern akzeptable Ergebnisse, werden jedoch aufgrund ihrer komplizierten Anwendung und damit einhergehenden Unverständlichkeit häufig abgelehnt.

Speziell für den Bereich der Utility Analysis raten Cabrera und Raju (2001) daher, diese Methode verständlicher zu kommunizieren, die Parameter der Formel auf die Bedürfnisse der Unternehmen abzustimmen und Manager stärker in den Prozess der Implementierung der Methode zu integrieren. Unternehmern muss die Chance gegeben werden, Methoden anzuwenden, die zu gewünschten Ergebnissen führen und zugleich Praktikabilität gewährleisten. Boudreau und Ramstad (2003) empfehlen außerdem, in Unternehmen verstärkt publik zu machen, dass die Utility Analysis eine der wenigen Methoden darstellt, die es schafft, den Wertbeitrag von HRD-Maßnahmen und somit deren Beitrag zur Erreichung monetärer Unternehmensziele zu erfassen. Würde ein Bewusstsein für den Nutzen der Utility Analysis geschaffen, so könne die Akzeptanz der Methode erhöht werden. Bisher wird der Zusammenhang zwischen positiven Utility Analysis Ergebnissen und „organizational performance“ (Boudreau, 1995, p. 3) selten untersucht bzw. herausgestellt. Boudreau (1995) nimmt an, dass sich die Nutzwertanalyse als hilfreich erweisen könne „in creating a bridge between scientific findings, organizational strategic choices, and managerial decisions“ (p. 22). Eine zukünftige Herausforderung für Vertreter der Utility Analysis besteht also darin, „to articulate the links between HRD interventions and organizational success“ (Cascio, 2003, p. 157).

Die Evaluationsaktivitäten der Unternehmen können vermutlich gesteigert werden, indem detaillierte Leitfäden zur Anwendung der unterschiedlichen Modelle zur Verfügung gestellt, Vor- und Nachteile der einzelnen Modelle aufgezeigt und kontextabhängige Anwendungsempfehlungen ausgesprochen werden. Zudem sollten Evaluationsmethoden zukünftig verstärkt in den organisationalen Kontext eingebettet werden, um zu verdeutlichen, wie sie bei der Erreichung von Unternehmenszielen unterstützend wirken können. Dies würde ein breiteres Verständnis der Methoden schaffen und somit deren Akzeptanz in der Praxis erhöhen.

6.2 Trends bei der Gestaltung zukünftiger PE-Programme

Zwei wesentliche Aspekte werden die Gestaltung zukünftiger Trainingsmaßnahmen charakterisieren. Zum einen wird zunehmend gefordert, Trainingsmaßnahmen an der Unternehmensstrategie auszurichten und zentrale Geschäftsprozesse bereits bei der Planung der Maßnahme zu berücksichtigen. Zum anderen wird verstärkt auf den Einsatz von technologiebasierten Lehr- und Lernmethoden wie z. B. programmierten Unterricht und E-Learning gesetzt.

6.2.1 Implementierung des Trainings in die Organisation

Salas und Cannon-Bowers (2001) verweisen auf folgende unzureichende Sichtweise von Trainingsmaßnahmen: „For many decades, training researchers have ignored the fact that training cannot be isolated from the system it supports” (p. 491). Dieser Umstand wurde laut Kozlowski und Salas (1997) bisher mehrheitlich vernachlässigt. Von mehreren Autoren (z. B. Baldwin & Magjuka, 1997; Dipboye, 1997; Goldstein, 1993; Goldstein & Ford, 2002; Noe, 2002; Noe & Ford, 1992; Quiñones & Ehrenstein, 1997; Tannenbaum & Yukl, 1992) wird verstärkt gefordert, Trainingsmaßnahmen als Teil der gesamten Organisationsentwicklung zu sehen und Trainingsziele dementsprechend an der Unternehmensstrategie auszurichten. PE-Maßnahmen dürfen nicht länger isoliert betrachtet werden, sondern müssen als „part of the ongoing dynamic and changeable processes in work organizations“ (Ford, 1997, p. 8) angesehen werden. Quiñones (1997) betont: „Training does not operate in a vacuum” (p. 193).

Da Trainingsprogramme laut Quiñones und Ehrenstein (1997) als ein „subsystem of the larger organizational entity“ (p. 5) betrachtet werden müssen, kann das Trainingsdesign, die Durchführung und die Effektivität einer Maßnahme durch organisationale Faktoren wie z. B. die Unternehmensziele, die Personalpolitik und finanzielle Ressourcen beeinflusst werden. Werden Organisationsfaktoren nicht berücksichtigt und Trainingsprogramme nicht an der Organisation ausgerichtet, hat dies Auswirkungen auf die Trainingseffektivität (Kozlowski & Salas, 1997). Auf der anderen Seite haben aber auch Trainingsprogramme Einfluss auf andere Bereiche der Organisation, wie etwa das Vergütungssystem, die Karriereplanung und das Mitarbeiterrecruitment (Quiñones & Ehrenstein, 1997). Dies bemerken ebenfalls Goldstein und Ford (2002): „More effective training programs can also affect all other systems in the work organization“ (p. 20). Koppeln Unternehmen z. B. kompetenzbasierte Vergütungsmodelle und strategische Karriereplanung an Trainingserfolgskontrollen, so können sie sich einen Wettbewerbsfaktor schaffen, der gerade in Zeiten mangelnder qualifizierter Nachwuchskräfte nicht unterschätzt werden darf. Ausgewiesene PE-Programme können Unternehmen für potenzielle Mitarbeiter attraktiv machen. Die Berücksichtigung der wechselseitigen Beziehungen zwischen Weiterbildung, Personal- und Unternehmenspolitik bietet also Potenzial, welches dadurch genutzt wird, indem Trainingsmaßnahmen nicht länger isoliert, sondern im Kontext der Organisation betrachtet werden.

6.2.2 Zunehmender Einsatz von Technologien

Immer mehr Organisationen setzen verstärkt neue Technologien wie z. B. webbasierte und softwaregestützte Anwendungen ein, um ihre traditionellen Trainingsprogramme zu verbes-

ern oder zu ersetzen (Goldstein & Ford, 2002; Noe, 2002; Quiñones, 1997; Salas & Cannon-Bowers, 2001). „However, the mere use of the latest training gadget will not ensure an adequate return on the training investment dollar“ (Quiñones & Ehrenstein, 1997, p. 11). Laut Kraiger und Jung (1997) unterscheiden sich technologieunterstützte Trainingsmaßnahmen u. a. durch folgende Aspekte von traditionellen Maßnahmen: Zum einen wird neben dem Erlernen der Trainingsinhalte ebenfalls das eigenständige Erlernen der Instruktionen vom Trainee erwartet. Zum anderen müssen die Teilnehmer einen sicheren Umgang mit dem System aufweisen und zusätzlich Wissen, welches in unterschiedlichster Weise präsentiert wird, aufnehmen können. Es ist Aufgabe der Trainees, den Prozess der Wissensaufnahme zu steuern und sich das nötige Wissen aus der Vielfalt der computergestützt dargebotenen Informationen zu beschaffen.

6.3 Berücksichtigung der aufgezeigten Trends durch die Evaluation

Wie oben aufgezeigt, werden sich Personalentwicklungsmaßnahmen zukünftig nicht nur in ihren Inhalten, sondern auch in ihrer Darbietungsform sowie bezüglich ihrer organisationalen und unternehmensbezogenen Implementierung verändern. Es gilt, diese Veränderungen auch im Bereich der Evaluation wahrzunehmen sowie Methoden und Instrumente ggf. daran auszurichten. Einige Ansätze und Überlegungen zu diesen Entwicklungen, die sich bei der Bewältigung dieser Herausforderungen als hilfreich erweisen könnten, werden im Folgenden beschrieben.

6.3.1 Integratives Modell zur formativen Evaluation

Die beiden oben genannten Trends bei der Gestaltung zukünftiger PE-Programme stellen laut Brown und Gerhardt (2002) neue Ansprüche an die Evaluation einer Trainingsmaßnahme. Diesen neuen Ansprüchen versuchen die Autoren mit Hilfe eines integrativen Modells zur formativen Evaluation gerecht zu werden. In diesem Modell werden die Vorzüge früherer formativer Evaluationsmodelle (Dick, Carey & Carey, 2000; Weston, McAlpine & Bordona-ro, 1995) vereint und gleichzeitig die geforderte strategisch-organisationale auf der einen und die technologische Perspektive auf der anderen Seite in den Trainingsevaluationsprozess miteinbezogen.

„Evaluation should include procedures that ensure alignment of a training activity with the organization’s strategy“ (Brown & Gerhardt, 2002, p. 960). Brown und Gerhardt (2002) fordern eine Einbindung der „training stakeholders“ (p. 961) bereits in frühen Stadien des Evaluationsvorhabens; dazu gehören u. a. die Vorgesetzten der Trainingsteilnehmer, die Fachabtei-

lungen und die Unternehmensleitung. Im Rahmen der formativen Evaluation steht nicht länger allein die Einschätzung der Wirksamkeit eines Trainings im Vordergrund, sondern die Empfehlungen und Verbesserungsvorschläge verschiedener Interessensgruppen. Ziel hierbei ist, die Trainingsmaßnahme zu optimieren (Verbesserungsfunktion der Evaluation). Dieses Umdenken kann die eingangs der Arbeit erwähnten Implementierungsschwierigkeiten von Evaluationssystemen reduzieren, die Ängste der Beteiligten verringern, Commitment bei diesen für die Maßnahme herstellen und dient somit der ebenfalls zu Beginn angesprochenen Kommunikationsfunktion. Wenn sowohl Vorgesetzte, Trainer als auch Teilnehmer in die Entscheidungs- und Verbesserungsprozesse involviert sind, identifizieren sie sich leichter mit dem Programm und fühlen sich somit für die Ergebnisse mitverantwortlich (Brown & Gerhardt, 2002). „This commitment must be visible, tangible, and consistently communicated throughout the organization” (Russ-Eft & Preskill, 2001, p. 412). Dies kann laut Brown und Gerhardt (2002) zu einer erwünschten Erhöhung der Evaluationsaktivität und kontinuierlichen Verbesserung der Weiterbildungsprogramme im Unternehmen führen. Außerdem kann durch den Einbezug von Entscheidungsträgern die hierarchische Stellung und die mangelnde Kompetenzausstattung des Bildungscontrollings verbessert werden, die bisher ein entscheidendes Hemmnis für die Akzeptanz von Evaluationstätigkeiten im Unternehmen waren (Hummel, 2001).

Um dem Aspekt der zunehmenden Technologisierung gerecht zu werden, sollte der Web-, Multimedia- bzw. Technologieeinsatz und die Mensch-Computer-Interaktion ebenfalls Bestandteil des Evaluationsprozesses sein (Brown & Gerhardt, 2002). Für die Evaluation dieser Trainingsprogramme ergeben sich folgende Konsequenzen: Auf der einen Seite sollte bei der Messung des Wissenszuwachses der Trainingsteilnehmer – ähnlich wie von Kraiger et al. (1993) vorgeschlagen – die Mehrdimensionalität von Lernen berücksichtigt werden. „It is of interest to determine how various knowledge presentation formats (text, graphics, animation, simulation, video) affect the acquisition and organization of knowledge” (Salas, Cannon-Bowers & Kozlowski, 1997, p. 364). Auf der anderen Seite sollte neben der bisher üblichen Evaluation der gelernten Trainingsinhalte auch evaluiert werden, wie sicher sich die Teilnehmer im Umgang mit dem Medium fühlen. Weiterhin besteht hier die Möglichkeit, eine unmittelbare Erfolgskontrolle vorzunehmen und Trainingsinhalte auf die individuellen Fähigkeiten und Bedürfnisse eines Teilnehmers schon während des Trainingsprozesses auszurichten; eine Art adaptiven Lehrens wäre somit denkbar.

6.3.2 Folgewirkungen der Evaluation

Aus der Forderung nach einer systemischen Betrachtungsweise und in Anbetracht der interdependenten Beziehung zwischen Training und Organisation ergeben sich zwei Aufgaben für die Personalentwicklung: Die Berücksichtigung vieler verschiedener und bereichsübergreifender Aspekte während der Programmausgestaltung und die Beachtung möglicher „cross-level effects“ (Quiñones & Ehrenstein, 1997, p. 5), d. h. die Beachtung von Auswirkungen des Trainings auf PE-fremde Organisationseinheiten. „Training is one operating system within an organization, and training should be evaluated as a system in support of other systems“ (Carnevale & Schulz, 1998, p. 239). Um die Frage nach der Effektivität von komplexen Personalentwicklungsprogrammen beantworten zu können, müssen nach Backhaus und Brauckhage (2004) die vielfältigen Wirkfaktoren von Interventionen und deren Interaktionen mit erhoben werden. Nach Ansicht der genannten Autoren reicht es nicht, die Frage nach Aufwand und Nutzen von Evaluation auf die kurzfristigen Effekte und den unmittelbaren Output von PE-Maßnahmen zu fokussieren, vielmehr ließe sich der Nutzen von Evaluation „erst im Zusammenhang mit Folgeaktivitäten in der Organisation angemessen würdigen“ (Backhaus & Brauckhage, 2004, S. 31). Wichtig ist, dass aus den Evaluationsergebnissen Konsequenzen abgeleitet werden, die im Anschluss auch tatsächlich umgesetzt werden, um der Entscheidungsfunktion von Evaluation gerecht zu werden. So könnten sich Entscheidungsträger eines Unternehmens vornehmen, turnusmäßig zu einer bestimmten Frist die 20% der Trainings einzustellen, die als am wenigsten leistungs- bzw. wertsteigernd identifiziert worden sind. Schneidet eine PE-Maßnahme in der Evaluation unzureichend ab, so ist es Aufgabe der HR- bzw. Trainingsverantwortlichen, diese bis zu der gesetzten Frist zu überarbeiten und somit die Verantwortung für die Fortsetzung des Programms zu übernehmen. Dies würde Engagement und Veränderungswillen der Verantwortlichen fördern und so aufgrund des kontinuierlichen Qualitätsmanagementprozesses geprüfte und hochwertige PE zur Folge haben.

Auch kann die Trainingsevaluation eines Unternehmens, welche durch Beiträge in einschlägigen Fachzeitschriften, Messeauftritte und Vorträge bei Fachtagungen publik gemacht werden kann, das Unternehmensimage am Markt verbessern und potenziellen Bewerbern erfolgreich das Bild vermitteln, dass Weiterbildung einen hohen Stellenwert einnimmt. Öffentlichkeitswirksame Evaluation kann somit auch als strategischer Erfolgsfaktor verstanden werden.

6.3.3 Evaluation von Softskill-Trainings

Zunehmende Bedeutung wird der Evaluation von Maßnahmen beigemessen, die auf die Weiterentwicklung von sog. Softskills ausgerichtet sind. Die wachsende Anzahl von Programmen

in diesem Bereich kann als eine Folge des fortschreitenden Wandels der Arbeitswelt gesehen werden, durch den auf dem Weg zur Wissensgesellschaft hohe Qualifikationen des Einzelnen sowie soziale Kompetenzen wie Führungsmotivation, Teamfähigkeit und planerische Fähigkeiten mehr und mehr an Bedeutung gewinnen. Konnte z. B. im Rahmen der technischen Fertigung die Ursache-Wirkungs-Kette von Training relativ leicht durch das Auszählen der produzierten Teile vor und nach dem Training aufgezeigt werden, gestaltet sich die Evaluation bei Weiterbildungsmaßnahmen zur Verbesserung der sozialen Kompetenz weitaus schwieriger, da diese nicht (mehr) so einfach zu quantifizieren sind (vgl. Kap. 3.1.1, Beschreibung der Results-Ebene). Dies sollte die Personalabteilungen der Unternehmen jedoch nicht davon abhalten, sich der Herausforderung zu stellen, nach geeigneten Evaluationsmethoden zur Überprüfung der zunehmenden Anzahl von Trainings im Softskill-Bereich zu suchen. Als ein Ansatz kann hier die Utility Analysis verstanden werden. Gelingt es z. B. anhand von Ratingskalen, Unterschiede zwischen gezeigtem Verhalten aufgrund eines Trainings messbar zu machen, kann der Nutzen eines Softskill-Trainings erhoben werden. Beispielsweise könnten Mitarbeiter Verbesserungen in Verhaltensweisen und sozialen Kompetenzen ihrer Führungskräfte identifizieren. Haben die Führungskräfte an einem Training zum Führen von Mitarbeitergesprächen teilgenommen, bewerten die Mitarbeiter mit Hilfe eines Fragebogens nach einer Zeitspanne von ca. drei Monaten aufgetretene Veränderungen in deren Verhalten und äußern sich zu verschiedenen Aspekten eines nach dem Training stattgefundenen Mitarbeitergesprächs. Veränderungen der „Softskills“ können somit messbar gemacht werden und in die Nutzengleichung der Utility Analysis mit aufgenommen werden. Sind derartige Methoden erst einmal implementiert, stellen sie für das Unternehmen einen Wettbewerbsfaktor dar, der sich in vielerlei Hinsicht auszahlen wird: PE-Maßnahmen zur Verbesserung „weicher“ Qualifikationen können nun legitimiert werden. Dadurch gelingt es, zukünftige Investitionen in derartige Trainings zu sichern, was sowohl dem Wandel der Arbeitswelt Rechnung tragen als auch die Außendarstellung und die Attraktivität des Unternehmens verbessern kann.

6.3.4 Trainingsstandards

In Anbetracht des demographischen Wandels und der damit einhergehenden Auswirkungen auf die Altersstruktur der Belegschaft ist u. a. zu berücksichtigen, dass in Zukunft eine größer werdende Anzahl an (älteren) Mitarbeitern umgeschult und innerhalb von Organisationen neu orientiert werden muss, da qualifizierte Nachwuchskräfte fehlen. Auch vor dem Hintergrund der aktuellen Diskussion um längere Lebensarbeitszeiten kommt dem Ausdruck „lebenslanges Lernen“ eine neue Bedeutung in der Personalentwicklung zu. Unternehmen werden vor der Aufgabe stehen, das Wissen und die Fähigkeiten ihrer Mitarbeiter bis zu einem weitaus

höheren Lebensalter als bisher zu erweitern und zu aktualisieren, um wettbewerbsfähig zu bleiben und den kommenden Anforderungen gerecht zu werden. Kein Unternehmen wird es sich leisten können, einem Großteil der Arbeitnehmer Weiterbildungsmöglichkeiten und sich selbst somit Innovation und Wettbewerbsvorteile vorzuenthalten. Diese Entwicklungen werden dazu führen, dass die Nachfrage nach Weiterbildung steigen und daher auch auf dem Markt externer Trainingsanbieter eine Angebotszunahme eintreten wird. Personalverantwortliche werden vor der Aufgabe stehen, aus der Fülle von Angeboten geeignete Trainings auszuwählen. Anhand welcher Kriterien soll jedoch die Qualität der einzelnen Weiterbildungsangebote richtig eingeschätzt werden? Wenig Transparenz herrscht zur Zeit auf dem Markt der Weiterbildungsanbieter und bezüglich deren Qualität. Mandl (2003) stellt verschiedene Forderungen an die Qualitätssicherung von Weiterbildungsmaßnahmen, so z. B. vergleichende Tests und konsequente Zertifizierung von Weiterbildungsangeboten und Weiterbildungsanbietern sowie Überlegungen zur Gründung einer „Stiftung Bildungstest“ (S. 296), die in ihren Zielen und Strukturen der Stiftung Warentest ähneln könnte. Nur so gelänge es, mehr Transparenz in der Angebots- und Anbieterqualität zu erzielen. Werden Qualitätskriterien für Weiterbildung gefordert, so stellt sich auch die Frage, wer die Legitimation besitzt, derartige Standards zu setzen und die einzelnen Anbieter anhand dieser systematisch zu beurteilen. Für Plinke (2003) ist dies weniger Aufgabe des Staates oder einer öffentlich-rechtlichen Konstruktion, vielmehr sollte darauf vertraut werden, dass der freie Markt über einen starken Imagewettbewerb diese notwendige Auswahl von guten und schlechten Weiterbildungsanbietern und deren Maßnahmen vornimmt. Er fordert sogar ein Konkurrenzprinzip in der Evaluierung der Weiterbildungsprogramme und plädiert für die Einführung konkurrierender Rankings.

6.3.5 Wo sollte Evaluation ansetzen?

Ist in einem ersten Schritt das Bewusstsein für die Notwendigkeit der Evaluation von PE-Programmen geschaffen und im Unternehmen kommuniziert worden, so stellt sich die Frage, an welcher Stelle die Evaluation angesetzt werden kann. Unserer Meinung nach sollte die Evaluation bereits in Planungsvorhaben Berücksichtigung finden, die sich mit der Konzeption eines Trainings beschäftigen. Wird ein Training von externen Anbietern eingekauft, so bietet es sich an, die Umsetzung der Evaluation von Anfang an mit in Auftrag zu geben. Nur so können optimale Instrumente entwickelt werden, von denen erste bereits während des Trainings Anwendung finden können – eine enge Zusammenarbeit von Unternehmen und Trainingsanbietern vorausgesetzt. Auch Arzberger und Mayer (2004) fordern, dass Trainingsanbieter einen Nachweis über die Wirksamkeit ihrer Maßnahmen erbringen und die Evaluation

als festen Bestandteil der Maßnahme verstehen. In einer Studie von Morrow et al. (1997) belaufen sich die Kosten der Trainingsevaluation auf lediglich ca. 0,2% des Trainingsbudgets. Diese Angaben sind nicht ohne weiteres auf andere Organisationen zu übertragen; so spricht Phillips (1997) z. B. davon, dass der Prozess der ROI-Berechnung nicht mehr als zusätzliche 3% - 5% der Trainingskosten ausmache. Auch wenn umfassendere Evaluationen höhere Ausgaben beanspruchen, kann grundsätzlich jedoch argumentiert werden, dass eine relativ geringe zusätzliche Investition dazu führen kann, umfangreiche Fehlinvestitionen zu verhindern. Die in Kapitel 4 vorgenommene Übersicht über die Faktoren, die den Transfer von Trainingsinhalten positiv wie negativ beeinflussen können, verstärkt die Notwendigkeit, so früh wie möglich mit der Evaluation als zentralem Bestandteil des PE-Prozesses zu beginnen.

Die Forderung, Weiterbildungsmaßnahmen auf ihren Erfolg hin zu überprüfen, steht momentan im Fokus nahezu jeder Diskussion und Fachtagung (Gloger, 2004; Mattson, 2003). Auch der Umfang an Theorien und Methoden zum Thema Trainingsevaluation nimmt kontinuierlich und beträchtlich zu (Salas & Kosarzycki, 2003), was als weiteres Indiz für die Aktualität des Themas verstanden werden kann. Mit einem aufgezeigten positiven Trainingseffekt, messbar gemacht und in Zahlen ausgedrückt, gelingt die Vergleichbarkeit mit anderen Unternehmensinvestitionen. Erst durch die Anwendung der aufgezeigten Modelle und Konzepte und deren nachvollziehbarer Kommunikation bei Mitarbeitern, deren Vertretungen, Entscheidungsträgern und Trainingsverantwortlichen gelingt es, wie zu Beginn dieser Arbeit gefordert, Investitionen in das Humanvermögen zu rechtfertigen und Personalentwicklung als zukunftssträchtige Investition darzustellen. Es kann dargelegt werden, ob ein Trainingsprogramm zur Verbesserung der Leistung eines Mitarbeiters und darüber hinaus zur Steigerung des Unternehmenswertes beitragen kann. Der weit verbreiteten Annahme, Weiterbildung verursache lediglich Kosten, könne jedoch keine messbaren Ergebnisse liefern, kann mit der Überprüfung ihrer Wirksamkeit eindrucksvoll begegnet werden. Wir hoffen, mit dieser Arbeit einen Beitrag dazu leisten zu können, dass Evaluation verstärkt als integrativer Bestandteil des Trainingsprozesses verstanden wird und dieser Einsicht bereit bei der Planung und Konzeption einer Weiterbildungsmaßnahme Rechnung getragen wird.

Literaturverzeichnis

- Al-Athari, A. & Zairi, M. (2002). Training evaluation: An empirical study in Kuwait. *Journal of European Industrial Training*, 26(5), 241-251.
- Aldisert, L. M. (2002). *Valuing people: How human capital can be your strongest asset*. Chicago: Dearborn Trade.
- Alliger, G. M. & Janak, E. A. (1989). Kirkpatrick's level of training criteria: Thirty years later. *Personnel Psychology*, 42, 331-342.
- Alliger, G. M., Tannenbaum, S. I., Bennet, W., Traver, H. & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50, 341-358
- Arthur, W., Jr., Bennet, W., Jr., Edens, P. S. & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, 88(2), 234-245.
- Arthur, W., Jr., Tubre, T. C., Paul, D. S. & Edens, P. S. (2003). Teaching effectiveness: The relationship between reaction and learning criteria. *Educational Psychology*, 23(3), 275-285.
- Arzberger, H. & Mayer, T. (2004). Evaluation als integraler Bestandteil der HR-Arbeit: Integrierte Evaluation beteiligt die Akteure der Personal- und Organisationsentwicklung an der Auswertung ihres Handelns. *Personalführung*, 3, 32-40.
- Backhaus, C. & Brauckhage, M. (2004). Evaluation eines Management-Entwicklungsprogramms: Evaluation als gezielte Intervention im Rahmen eines Trainings in der Automobilindustrie. *Personalführung*, 3, 24-31.
- Baldwin, T. T. & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41, 63-105.
- Baldwin, T. T. & Magjuka, R. J. (1997). Training as an organizational episode: Pretraining influences on trainee motivation. In J. K. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 99-127). Mahwah, NJ: Lawrence Erlbaum.
- Balzert, H. (1998). *Lehrbuch der Software-Technik: Software-Management, Software-Qualitätssicherung, Unternehmensmodellierung*. Heidelberg: Spektrum.
- Bassi, L. J. & Van Buren, M. E. (1998). The ASTD state of the industry report. Retrieved January 4, 2001, from http://www.astd.org/virtual_community/comm_trends/state_of_industry_td0198_cms.htm
- Bell, J. D. & Kerr, D. L. (1987). Measuring training results: Key to managerial commitment. *Training and Development Journal*, 41(1), 70-73.
- Bobko, P., Karren, R. & Parkington, J. J. (1983). Estimation of standard deviation in utility analysis: An empirical test. *Journal of Applied Psychology*, 68(1), 170-176.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3., überarb. Aufl.). Berlin: Springer.

- Boudreau, J. W. (1983). Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology*, 36, 551-576.
- Boudreau, J. W. (1991). Utility analysis for decisions in human resource management. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 621-745). Palo Alto, CA: Consulting Psychologists Press.
- Boudreau, J. W. (1995, August). Future utility analysis research: Continue, but expand the cognitive and strategic focus. In Center for Advanced Human Resource Studies/ Cornell University (Chair), *Should the research on utility theory continue? Why and why not?* Symposium conducted at the national meeting of the Academy of Management, Vancouver, British Columbia, Canada.
- Boudreau, J. W. & Ramstad, P. M. (2003). Strategic industrial and organizational psychology and the role of utility analysis models. In I. B. Weiner (Series Ed.), W. C. Borman, D. R. Ilgen & R. J. Klimoski (Vol. Eds.), *Handbook of psychology: Vol. 12. Industrial and organizational psychology* (pp. 193-221). Hoboken, NJ: John Wiley & Sons.
- Bramley, P. (1996). *Evaluating Training*. London: Institute of Personnel and Development.
- Brethower, D. M. (2001). Invited reaction: Posttraining interventions to enhance transfer. *Human Resource Development Quarterly* 12(2), 121-125.
- Brinkerhoff, R. O. (1987). *Achieving results from training*. San Francisco: Jossey-Bass.
- Brinkerhoff, R. O. (1998). An integrated evaluation model for HRD. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 78-83). Alexandria, VA: American Society for Training and Development.
- Brodgen, H. E. (1949). When testing pays off. *Personnel Psychology*, 2, 171-183.
- Brown, K. G. & Gerhard, M. (2002). Formative evaluation: An integrative practice model and case study. *Personnel Psychology*, 55, 951-983.
- Cabrera, E. F. & Raju, N. S. (2001). Utility analysis: Current trends and future directions. *International Journal of Selection and Assessment*, 9(1,2), 92-102.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., III & Weick, K. R., Jr. (1970). *Managerial behavior, performance, and effectiveness*. New York: McGraw-Hill.
- Campion, M. A. & Campion, J. E. (1987). Evaluation of an interviewee skills training program in a natural field experiment. *Personnel Psychology*, 40, 675-691.
- Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I. & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: A model and initial empirical investigation. *Military Psychology*, 7(3), 141-164.
- Carnevale, A. P. & Schulz, E. R. (1998). Return on investment: Accounting for training. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 224-252). Alexandria, VA: American Society for Training and Development.
- Carson, K. P., Becker, J. S. & Henderson, J. A. (1998). Is utility really futile? A failure to replicate and an extension. *Journal of Applied Psychology*, 83(1), 84-96.
- Cascio, W. F. (1987). *Applied psychology in personnel management*. Englewood Cliffs, NJ: Prentice Hall.

- Cascio, W. F. (1989). Using utility analysis to assess training outcomes. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp. 63-88). San Francisco, CA: Jossey-Bass.
- Cascio, W. F. (2000). *Costing human resources: The financial impact of behavior in organizations* (4th ed.). Cincinnati, OH: South-Western College.
- Cascio, W. F. (2003). Invited reaction: The effects of alternative reports of human resource development results on managerial support. *Human Resource Development Quarterly*, 14(2), 153-158.
- Cascio, W. F. & Morris, J. R. (1990). A critical reanalysis of Hunter, Schmidt, and Coggin's (1988) "Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing the utility of personnel programs". *Journal of Applied Psychology*, 75(4), 414-417.
- Cascio, W. F. & Ramos, R. A. (1986). Development and application of a new method for assessing job performance in behavioral/ economic terms. *Journal of Applied Psychology*, 71(1), 20-28.
- Childs, J. M. & Bell, H. H. (2001). Training systems evaluation. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of human factors testing and evaluation* (2nd ed., pp. 473-509). Mahwah, NJ: Lawrence Erlbaum.
- Claßen, M. (2003). Human Asset Management in deutschen Unternehmen. In K. Schwuchow & J. Gutmann (Hrsg.), *Jahrbuch Personalentwicklung und Weiterbildung: Praxis und Perspektiven 2004* (S. 20-27). München: Luchterhand.
- Colquitt, J. A., LePine, J. A. & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path of 20 years of research. *Journal of Applied Psychology*, 85(5), 678-707.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana: University of Illinois Press.
- Cronshaw, S. F. (1997). Lo! The stimulus speaks: The insider's view on Whyte and Latham's "The futility of utility analysis". *Personnel Psychology*, 50, 611-615.
- Day, E. A., Arthur, W., Jr. & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology*, 86(5), 1022-1033.
- Davidove, E. A. & Schroeder, P. A. (1998). Demonstrating ROI of training. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 205-206). Alexandria, VA: American Society for Training and Development.
- Delahaye, B. L. (2000). *Human resource development: Principles and practice*. Milton, Queensland, Australia: John Wiley & Sons Australia.
- Dick, W., Carey, L. & Carey, J. O. (2000). *The systematic design of instruction* (5th ed.). Boston: Pearson Allyn & Bacon.
- Dipboye, R. L. (1997). Organizational barriers to implementing a rational model of training. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Application of psychological research* (pp. 31-60). Washington, DC: American Psychological Association.
- Fitz-enz, J. A. (1988). Proving the value of training. *Personnel*, 65(3), 17-21.

- Ford, J. K. (1997). Advances in training research and practice: An historical perspective. In J. K. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 1-16). Mahwah, NJ: Lawrence Erlbaum.
- Galvin, T. (2003). 2003 Industry Report. *Training*, 40(9), 21-38.
- Gaudine, A. P. & Saks, A. M. (2004). A longitudinal quasi-experiment on the effects of post-training transfer interventions. *Human Resource Development Quarterly*, 15(1), 57-76.
- Gerlich, P. (1999). *Controlling von Bildung, Evaluation oder Bildungs-Controlling?*. München: Rainer Hampp.
- Gibb, S. (2002). *Learning and Development: Processes, practices and perspectives at work*. London: Palgrave Macmillan.
- Glass, G. V. & Ellet, F. S. (1980). Evaluation research. *Annual Review of Psychology*, 31, 211-228.
- Gloger, A. (2004). Weiterbildungscontrolling: Vom Nutzen des Trainings. *ManagerSeminare*, 73(2), 18-26.
- Goldsmith, T. E. & Kraiger, K. (1997). Applications of structural knowledge assessment to training evaluation. In J. K. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. S. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 73-96). Mahwah, NJ: Lawrence Erlbaum.
- Goldstein, I. L. (1991). Training in work organizations. In M. D. Dunnette & L. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed., Vol. 2, pp. 507-619). Palo Alto, CA: Consulting Psychologists Press.
- Goldstein, I. L. (1993). *Training in organizations: Needs assessment, development and evaluation* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Goldstein, I. L. & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth.
- Greer, O. L. & Cascio, W. F. (1987). Is cost accounting the answer? Comparison of two behaviorally based methods for estimating the standard deviation of job performance in dollars with a cost-accounting-based approach. *Journal of Applied Psychology*, 72(4), 488-595.
- Grünewald, U., Moraal, D. & Schönfeld, G. (2003). *Betriebliche Weiterbildung in Deutschland und Europa*. Bielefeld: Bertelsmann.
- Gülpen, B. (1996). *Evaluation betrieblicher Verhaltenstrainings unter besonderer Berücksichtigung des Nutzens*. München: Rainer Hampp.
- Hager, W. & Hasselhorn, M. (2000). Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen: Standards und Kriterien* (S. 41-82). Göttingen: Huber.
- Hamblin, A. C. (1974). *Evaluation and control of training*. London: McGraw-Hill.

- Hazer, J. T. & Highhouse, S. (1997). Factors influencing manager's reactions to utility analysis: Effects of SDy method, information frame, and focal intervention. *Journal of Applied Psychology*, 82(1), 104-112.
- Hesketh, B. (1997). Dilemmas in training for transfer and retention. *Applied Psychology: An International Review*, 46(4), 317-339.
- Hesketh, B. & Frese, M. (2002). Simulation and training in work settings. In P. B. Baltes & N. J. Smelser (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 14097-14101). Oxford, England: Elsevier Science.
- Höft, S. (2001). Erfolgsüberprüfung personalpsychologischer Arbeit. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 617-651). Göttingen: Hogrefe.
- Holton, E. F., III (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1), 5-21.
- Holton, E. F., III, Chen, H.-C. & Naquin, S. S. (2003). An examination of learning transfer system characteristics across organizational settings. *Human Resource Development Quarterly*, 14(4), 459-482.
- Huint, P. & Saks, A. M. (2003). Translating training science into practice: A study of manager's reactions to posttraining transfer interventions. *Human Resource Development Quarterly*, 14(2), 181-198.
- Hummel, T. R. (2001). *Erfolgreiches Bildungscontrolling: Praxis und Perspektiven* (2., überarb. Aufl.). Heidelberg: Sauer.
- Hunter, J. E., Schmidt, F. L. & Coggin, T. D. (1988). Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing the utility of personnel programs. *Journal of Applied Psychology*, 73(3), 522-528.
- Informationsdienst des Instituts der Deutschen Wirtschaft Köln. (2004, Januar). *Konjunkturprognosen: Die Wirtschaft nimmt Anlauf*. Abgerufen am 03.02.2004, von <http://www.iwkoeln.de/default.aspx?p=content&i=17189>
- Jones, G. R. & Wright, P. M. (1992). An economic approach to conceptualizing the utility of human resource management practices. In G. R. Ferris & K. M. Rowland (Eds.), *Research in personnel and human resources management* (pp. 271-299). Greenwich, CT: Jai Press.
- King, L. M., Hunter, J. E. & Schmidt, F. L. (1980). Halo in a multidimensional forced choice evaluation scale. *Journal of Applied Psychology*, 65(4), 507-516.
- Kirkpatrick, D. L. (1959a). Techniques for evaluation training programs. *Journal of the American Society of Training Directors*, 13(11), 3-9.
- Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2 - learning. *Journal of the American Society of Training Directors*, 13(12), 21-26.
- Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3 - behavior. *Journal of the American Society of Training Directors*, 14(1), 13-18.
- Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 4 - results. *Journal of the American Society of Training Directors*, 14(2), 28-32.

- Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler.
- Kirkpatrick, D. L. (1996). Evaluation. In R. L. Craig (Ed.), *The ASTD training and development handbook: A guide to human resource development* (4th ed., pp. 294-312). New York: McGraw-Hill.
- Kirkpatrick, D. L. (1998a). *Evaluating training programs: The four levels* (2nd ed.). San Francisco: Berrett-Koehler.
- Kirkpatrick, D. L. (1998b). Great ideas revisited. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 3-8). Alexandria, VA: American Society for Training and Development.
- Knöll, H.-D., Slotos, T. & Suk, W. (1996). *Entwicklung und Qualitätssicherung von Anwendungssoftware: Konzepte, Meßverfahren, Standards*. Heidelberg: Spektrum.
- Kozlowski, S. W. J. & Salas, E. (1997). An organizational systems approach for the implementation and transfer of training. In J. K. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 247-287). Mahwah, NJ: Lawrence Erlbaum.
- Kraiger, K. (2003). Perspectives on training and development. In I. B. Weiner (Series Ed.), W. C. Borman, D. R. Ilgen & R. J. Klimoski (Vol. Eds.), *Handbook of psychology: Vol. 12. Industrial and organizational psychology* (pp. 171-192). Hoboken, NJ: John Wiley & Sons.
- Kraiger, K., Ford, J. K. & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78(2), 311-328.
- Kraiger, K. & Jung, K. M. (1997). Linking training objectives to evaluation criteria. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Application of psychological research* (pp. 151-175). Washington, DC: American Psychological Association.
- Latham, G. & Whyte, G. (1994). The futility of utility analysis. *Personnel Psychology*, 47, 31-46.
- Law, K. S. (1995). Estimating the dollar value contribution of human resource intervention programs: Some comments on the Brodgen utility equation. *Australian Journal of Management*, 20(2), 197-206.
- Mandl, H. (2003). Auf dem Weg zu einer neuen Weiterbildungskultur. In N. Killius, J. Kluge & L. Reisch (Hrsg.), *Die Bildung der Zukunft* (S. 294-301). Frankfurt am Main: Suhrkamp.
- Mathieu, J. E. & Martineau, J. W. (1997). Individual and situational influences on training motivation. In J. K. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 193-221). Mahwah, NJ: Lawrence Erlbaum.
- Mathieu, J. E., Tannenbaum, S. I. & Salas, E. (1992). Influences of individual and situational characteristics on measures on training effectiveness. *Academy of Management Journal*, 35(4), 828-847.

- Mattson, B. W. (2003). The effects of alternative reports of human resource development results on managerial support. *Human Resource Development Quarterly*, 14(2), 127-151.
- Mincer, J. (1962). On-the-job-training: Costs, returns, and some implications. *Journal of Political Economy*, 70(Suppl. 2), 50-79.
- Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen: Standards und Kriterien: Ein Handbuch* (S. 102-128). Göttingen: Hans Huber.
- Morgan, R. B. & Casper, W. (2000). Examining the factor structure of participant reactions to training: A multi-dimensional approach. *Human Resource Development Quarterly*, 11(3), 301-317.
- Morrow, C. C., Jarrett, M. Q. & Rupinski, M. T. (1997). An investigation of the effect and economic utility of corporate-wide training. *Personnel Psychology*, 50, 91-119.
- Newstrom, J. W. (1978). Catch-22: The problems of incomplete evaluation of training. *Training and Development Journal*, 32(11), 22-24.
- Noe, R. A. (1986). Trainee's attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review*, 11(4), 736-749.
- Noe, R. A. (2002). *Employee training and development* (2nd ed.). New York: McGraw-Hill.
- Noe, R. A. & Colquitt, J. A. (2002). Planning for training impact: Principles of training effectiveness. In K. Kraiger (Ed.), *Creating, implementing, and maintaining effective training and development: State-of-the-art lessons for practice* (pp. 53-79). San Francisco: Jossey-Bass.
- Noe, R. A. & Ford, J. K. (1992). Emerging issues and new directions for training research. *Research on Personnel and Human Resources Management*, 10, 345-384.
- Noe, R. A., Hollenbeck, J. R., Gerhart, B. & Wright, P. M. (2004). *Fundamentals of human resource management*. New York: McGraw-Hill/Irwin.
- Noe, R. A. & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology*, 39, 497-523.
- Nork, M. E. (1989). *Management Training: Evaluation - Probleme - Lösungsansätze*. München: Rainer Hampp.
- Parry, S. B. (1997). *Evaluating the impact of training: A collection of tools and techniques*. Alexandria, VA: American Society for Training and Development.
- Parry, S. B. (1998). Measuring training's ROI. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 200-204). Alexandria, VA: American Society for Training and Development.
- Pershing, J. A. & Pershing, J. L. (2001). Ineffective reaction evaluation. *Human Resource Development Quarterly*, 12(1), 73-90.
- Pfau, B. N. & Cohen, S. A. (2003). Aligning human capital practices and employee behavior with shareholder value. *Consulting Psychology Journal: Practice and Research*, 55(3), 169-178.

- Phillips, J. J. (1996). Measuring the results of training. In R. L. Craig (Ed.), *The ASTD training and development handbook: A guide to human resource development* (4th ed., pp. 313-341). New York: McGraw-Hill.
- Phillips, J. J. (1997). *The return on investment in training and performance improvement programs*. Houston, TX: Gulf.
- Phillips, J. J. (1998a). How much is the training worth. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 220-223). Alexandria, VA: American Society for Training and Development.
- Phillips, J. J. (1998b). ROI: The search for best practice. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 210-215). Alexandria, VA: American Society for Training and Development.
- Phillips, J. J. (1998c). Was it the training? In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 216-219). Alexandria, VA: American Society for Training and Development.
- Phillips, P. P. & Phillips, J. J. (2001). Symposium on the evaluation of training. *International Journal of Training and Development*, 5(4), 240-247.
- Plinke, W. (2003). Management education oder Weiterbildung? In N. Killius, J. Kluge & L. Reisch (Hrsg.), *Die Bildung der Zukunft* (S. 302-305). Frankfurt am Main: Suhrkamp.
- Quiñones, M. A. (1995). Pretraining context effects: Training assignment as feedback. *Journal of Applied Psychology*, 80(2), 226-238.
- Quiñones, M. A. (1997). Contextual influences on training effectiveness. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Application of psychological research* (pp. 177-199). Washington, DC: American Psychological Association.
- Quiñones, M. A. & Ehrenstein, A. (1997). Introduction: Psychological perspectives on training in organizations. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Application of psychological research* (pp. 1-10). Washington, DC: American Psychological Association.
- Rae, L. (2002). *Assessing the value of your training: The evaluation process from training needs to the report to the board*. Burlington, VT: Gower.
- Raju, N. S., Burke, M. J. & Normand, J. (1990). A new approach to utility analysis. *Journal of Applied Psychology*, 75(1), 3-12.
- Ralphs, L. T. & Stephan, E. (1986). HDR in the Fortune 500. *Training and Development Journal*, 40(10), 69-76.
- Rätzmann, M. (2002). *Software-Testing*. Bonn: Galileo Press.
- Rauschenberger, J. M. & Schmidt, F. L. (1987) Measuring the economic impact of human resource programs. *Journal of Business and Psychology*, 2(1), 50-59.
- Remdisch, S. & Utsch, A. (2004). Evaluation als Beitrag zur Entwicklung von Qualität: Evaluationsverfahren und ihre Methoden im Kontext der Personal- und Organisationsentwicklung. *Personalführung*, 3, 19-23.

- Richman-Hirsch, W. L. (2001). Posttraining interventions to enhance transfer: The moderating effects of work environments. *Human Resource Development Quarterly*, 12(2), 105-120.
- Rouiller, J. Z. & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. *Human Resource Development Quarterly*, 4(4), 377-390.
- Ruona, W. E. A., Leimbach, M., Holton, E. F., III & Bates, R. (2002). The relationship between learner utility reaction and predicted learning transfer among trainees. *International Journal of Training and Development*, 6(4), 218-228.
- Russ-Eft, D. F. & Preskill, H. (2001). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change*. Cambridge, MA: Perseus.
- Sackett, P. R. & Mullen, E. J. (1993). Beyond formal experimental design: Towards an expanded view of the training evaluation process. *Personnel Psychology*, 46, 613-627.
- Salas, E. & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, 52, 471-499.
- Salas, E., Cannon-Bowers, J. A. & Kozlowski, S. W. J. (1997). The science and practice of training – current trends and emerging themes. In J. K. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 357-367). Mahwah, NJ: Lawrence Erlbaum.
- Salas, E. & Kosarzycki, M. P. (2003). Why don't organizations pay attention to (and use) findings from the science of training. *Human Resource Development Quarterly*, 14(4), 487-491.
- Schmidt, F. L. & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology*, 68(3), 407-414.
- Schmidt, F. L., Hunter, J. E., McKenzie, R. C. & Muldrow, T. W. (1979). Impact of valid selection procedures on work force productivity. *Journal of Applied Psychology*, 64(6), 609-626.
- Schmidt, F. L., Hunter, J. E. & Pearlman, K. (1982). Assessing the economic impact of personnel programs on work-force productivity. *Personnel Psychology*, 35, 333-347.
- Schmidt, U. (2004). Aus- und Weiterbildung im Arbeitsfeld Evaluation: Kompetenzanforderungen an Evaluatoren und das Angebot an Studiengängen und berufsbegleitenden Fortbildungen. *Personalführung*, 3, 42-47.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Scriven, M. (1991). *Evaluation thesaurus*. Thousand Oaks, CA: Sage.
- Shapiro, L. T. (1995). *Training effectiveness handbook: A high-results system for design, delivery, and evaluation*. New York: McGraw-Hill.
- Shelton, S. & Alliger, G. (1998). Who's afraid of level 4 evaluation? A practical approach. In D. L. Kirkpatrick (Ed.), *Another look at evaluating training programs* (pp. 171-174). Alexandria, VA: American Society for Training and Development.

- Siemers, S. H. (1995). Klassische Modelle zur Kosten-Nutzen-Analyse von Personalauswahlverfahren. In T. J. Gerpott & S. H. Siemers (Hrsg.), *Controlling von Personalprogrammen* (S. 115-138). Stuttgart: Schäfer-Poeschel.
- Smith-Jentsch, K. A., Salas, E. & Brannick, M. T. (2001). To transfer or not to transfer? Investigating the combined effects of trainee characteristics, team leader support, and team climate. *Journal of Applied Psychology*, 86(2), 279-292.
- Sneed, H. M. & Winter, M. (2002). *Testen objektorientierter Software: Das Praxishandbuch für den Test objektorientierter Client/Server-Systeme*. München: Carl Hanser.
- Sonntag, K. H. (2002). Personalentwicklung und Training: Stand der psychologischen Forschung und Gestaltung. *Zeitschrift für Personalpsychologie*, 1(2), 59-79.
- Stewart, J. (1999). *Employee development practice*. London: Financial Times Pitman.
- Tan, J. A., Hall, R. J. & Boyce, C. (2003). The role of employee reactions in predicting training effectiveness. *Human Resource Development Quarterly*, 14(4), 397-411.
- Tannenbaum, S. I. & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology*, 43, 399-441.
- Thierau, H. (1991). *Analyse und empirische Überprüfung wissenschaftlicher Evaluationskonzepte in der betrieblichen Weiterbildung: Dargestellt am Beispiel der Schulung von Führungskräften in Personalbeurteilung*. Unveröffentlichte Dissertation, Ruhr-Universität, Bochum.
- Thierau-Brunner, H., Stangel-Meseke, M. & Wottawa, H. (1999). Evaluation von Personalentwicklungsmaßnahmen. In K. H. Sonntag (Hrsg.), *Personalentwicklung in Organisationen: Psychologische Grundlagen, Methoden und Strategien* (2., überarb. und erw. Aufl., S. 261-286). Göttingen: Hogrefe.
- Tracey, J. B., Tannenbaum, S. I. & Kavanagh, M. J. (1995). Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology*, 80(2), 239-252.
- Van Buren, M. E. & Erskine, W. (2002). *State of the industry: ASTD's annual review of trends in employer-provided training in the United States*. Alexandria, VA: American Society for Training and Development.
- Wang, G. G., Dou, Z. & Li, N. (2002). A systems approach to measuring return on investment for HRD interventions. *Human Resource Development Quarterly*, 13(2), 203-224.
- Warr, P., Allan, C. & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, 72, 351-376.
- Warr, P., Bird, M. & Rackham, N. (1978). *Evaluation of management training* (5th ed.). Westmead, England: Grower.
- Warr, P. & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology*, 48, 347-375.
- Weiß, R. (2002). Betriebliche Weiterbildung: Investitionen in Köpfe und Know-how. Abgerufen am 03.02.2004, von <http://www.iwkoeln.de/default.aspx?p=content&i=16392>

- Weiß, R. (2003). Betriebliche Weiterbildung 2003: Ergebnisse einer IW-Erhebung. Abgerufen am 02.02.2004, von http://www.liquide.de/content/infobrief/weiterbildung_2001.pdf
- Weston, C., McAlpine, L. & Bordonaro, T. (1995). A model for understanding formative evaluation in instructional design. *Educational Technology Research and Development*, 43(3), 29-48.
- Wexley, K. N. & Latham, G. P. (2002). *Developing and training human resources in organizations*. Upper Saddle River, NJ: Prentice Hall.
- Wilkening, O. S. (2002). Bildungs-Controlling – Erfolgssteuerungssystem der Personalentwickler und Wissensmanager. In H.-C. Riekhof (Hrsg.), *Strategien der Personalentwicklung* (5. Aufl., S. 209-237). Wiesbaden: Gabler.
- Whyte, G. & Latham, G. (1997). The futility of utility analysis revisited: When even an expert fails. *Personnel Psychology*, 50, 601-610.
- Wolff, B. (2003). Effiziente Kosten-Nutzen-Abwägung im HR-Bereich. *Personalwirtschaft*, 11, 12-15.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2., vollst. überarb. Aufl.). Bern: Huber.
- Yamhill, S. & McLean, G. N. (2001). Theories supporting transfer of training. *Human Resource Development Quarterly*, 12(2), 195-208.

Verzeichnis der Anhänge

Anhang A	Zufriedenheitsmessung.....	XXI
Anhang B	Fragebogen Lernebene (1).....	XXV
Anhang C	Fragebogen Lernebene (2).....	XXIX
Anhang D	Transfererfolgsmessung.....	XXXIII
Anhang E	Eckdaten zur Berechnung des Nutzwertes und des ROI.....	XXXIX
Anhang F	Statistische Auswertung.....	XLI
F.1	Deskriptive Statistiken.....	XLI
F.2	Häufigkeitstabellen.....	XLII
F.3	Reliabilitätsanalyse.....	XLVI
F.4	Korrelationen.....	XLVII
F.5	T-Test.....	XLIX

**Fragebogen zur Zufriedenheitsmessung
SWIFT-Modul „Systematischer Software-Test“**

19. April 2004

Sehr geehrte Trainingsteilnehmerin, sehr geehrter Trainingsteilnehmer,

Sie haben heute an dem SWIFT-Modul „Systematischer Software-Test“ teilgenommen. Wie bereits angekündigt, möchten wir dieses Modul im Rahmen unserer Diplomarbeit, welche sich mit der Erfolgsmessung von Weiterbildungsprogrammen beschäftigt, untersuchen. Wir hoffen, mit dieser Arbeit einen Beitrag leisten zu können, dass die Überprüfung von Weiterbildungsmaßnahmen zukünftig verstärkt Beachtung findet und Personalentwicklungsprogramme somit noch besser auf den Bedarf von Mitarbeitern abgestimmt werden können.

Mit dem vorliegenden Fragebogen werden verschiedene Aspekte Ihrer Zufriedenheit mit dem heutigen Trainingsmodul erhoben. Wir bitten Sie um die Vervollständigung bzw. Bewertung verschiedener Aussagen, um Ihre Einstellung und Meinung zu dem heutigen Training erfassen zu können. Die Ergebnisse dieses Fragebogens sollen auch zur Qualitätssicherung und kontinuierlichen Verbesserung des Trainingsmoduls beitragen.

Bitte vervollständigen bzw. bewerten Sie jede einzelne Aussage, indem Sie den Grad Ihrer Zufriedenheit bzw. Zustimmung mit einem „X“ in der jeweiligen Tabellenspalte ausdrücken. Außerdem haben Sie die Möglichkeit, weitere Anmerkungen und Verbesserungsvorschläge zu den einzelnen Themenkomplexen jeweils im Anschluss an die Fragen anzufügen.

Ihre Daten werden selbstverständlich vertraulich behandelt und nur für Auswertungen im Rahmen der Diplomarbeit genutzt. Um Ihre Anonymität zu sichern, die von Ihnen bearbeiteten Fragebögen aber dennoch einander zuordnen zu können, bitten wir Sie erneut, Ihren persönlichen Code zu erstellen. Rückschlüsse auf Ihre Person sollen und werden anhand dieser Angaben nicht möglich sein. Bitte erstellen Sie Ihren Code hier:

Letzter Buchstabe Ihres Geburtsortes
Erster Buchstabe Ihres Geburtsmonats
Letzter Buchstabe des Vornamens Ihrer Mutter
Erster Buchstabe des Mädchennamens Ihrer Mutter

Die Bearbeitung des Fragebogens wird ca. 10 Minuten dauern. Für Ihre Unterstützung und Mitarbeit bedanken wir uns ganz herzlich!

Katharina Niehoff und Marius Werth

1. Zufriedenheit mit dem Trainer

	sehr zufrieden	eher zufrieden	teils/ teils	eher unzufrieden	sehr unzufrieden
1 Mit dem gezeigten Engagement des Trainers war ich...					
2 Mit der Verständlichkeit, mit der die Inhalte vom Trainer präsentiert wurden, war ich...					
3 Mit der Kompetenz des Trainers bezüglich systematischer Software-Tests war ich...					
4 Mit der Fähigkeit des Trainers, mich für die Thematik zu begeistern, war ich...					
5 Mit dem vom Trainer gezeigten Interesse am Lernerfolg der Teilnehmer war ich...					

Anmerkungen/ Vorschläge:

2. Zufriedenheit mit dem administrativen Trainingsmanagement

	sehr zufrieden	eher zufrieden	teils/ teils	eher unzufrieden	sehr unzufrieden
6 Mit der Zeitspanne, die zwischen Bekanntgabe des Trainingstermins und dem tatsächlichen Trainingstag bestand, war ich...					
7 Mit den Informationen, die ich im Vorfeld des Trainings über Inhalt und Ablauf erhalten habe, war ich...					
8 Mit der Unterstützung, die ich von meinem Vorgesetzten/ meiner Abteilung bezüglich meiner Trainingsteilnahme erhalten habe, war ich...					

Anmerkungen/ Vorschläge:

	sehr zufrieden	eher zufrieden	teils/ teils	eher unzufrieden	sehr unzufrieden
9 Mit der Verständlichkeit der Trainingsinhalte war ich...					
10 Mit den an mich gestellten Anforderungen im Training war ich...					
11 Mit der Menge der zu bearbeitenden Inhalte während des Trainings war ich...					

Anmerkungen/ Vorschläge:

	sehr zufrieden	eher zufrieden	teils/ teils	eher unzufrieden	sehr unzufrieden
12 Mit der Nützlichkeit der Trainingsinhalte im Hinblick auf meine Arbeitstätigkeit bin ich...					
13 Mit der Relevanz der besprochenen Inhalte für meine Arbeit als Software-ProjektleiterIn bin ich...					
14 Mit dem Ausmaß, in dem mich das Training darauf vorbereitet hat, meine Arbeitstätigkeiten effektiver durchzuführen, bin ich...					
15 Mit der Übertragbarkeit der Trainingsinhalte auf die Praxis bin ich...					

Anmerkungen/ Vorschläge:

	sehr zufrieden	eher zufrieden	teils/ teils	eher unzufrieden	sehr unzufrieden
16 Mit dem Einsatz der Medien (Beamer, FlipChart, Overhead Projektor, etc.) war ich...					
17 Mit den ausgehändigten Trainingsunterlagen bin ich...					
18 Mit den Trainingsräumlichkeiten war ich...					
19 Mit der Qualität der eingesetzten Lernhilfen (z. B. PCs) war ich...					

Anmerkungen/ Vorschläge:

	sehr zufrieden	eher zufrieden	teils/ teils	eher unzufrieden	sehr unzufrieden
20 Mit der Länge des Trainings war ich...					
21 Mit der Geschwindigkeit, mit der die Inhalte präsentiert wurden, war ich...					
22 Mit dem logischen Aufbau des Trainings war ich...					
23 Mit der Anzahl der praktischen Übungen war ich...					
24 Mit der Abwechslung der eingesetzten Methoden (Präsentation, Gruppenübung, Einzelarbeit, Rollenspiel, etc.) war ich...					

Anmerkungen/ Vorschläge:

	stimme zu	stimme eher zu	teils/ teils	stimme eher nicht zu	stimme nicht zu
25 Das Training entsprach voll und ganz meinen Erwartungen.					
26 Mein eigenes Interesse an den Inhalten des heutigen Trainings ist sehr hoch.					
27 Der Besuch dieses Trainings lohnt sich nicht.					
28 Ich habe mich persönlich dafür eingesetzt, an diesem Training teilzunehmen.					
29 Die Atmosphäre während des Trainings war sehr positiv.					
30 Ich werde nicht an den Trainingszielen festhalten, wenn sich zeigt, dass sie nur schwer zu erreichen sind.					
31 Ich werde das Training auf jeden Fall KollegInnen weiterempfehlen.					

Sonstige Anmerkungen:

Abschließend möchten wir Sie bitten, uns noch einige Informationen zu Ihrer Person zu geben. Rückschlüsse sollen und werden anhand dieser Angaben nicht möglich sein, da Ihre Daten selbstverständlich anonymisiert in die Auswertung der Untersuchung eingehen.

Demographische Daten

1. Geschlecht : männlich weiblich

2. Alter: _____ Jahre

3. Ausbildung: Abgeschlossene Berufsausbildung zum: _____

Abgeschlossenes Hochschulstudium in dem Bereich:

Wirtschaftswissenschaften

Naturwissenschaften

Ingenieurwesen

Informatik

Sozialwissenschaften

Sonstiges: _____

4. Seit wie vielen Jahren/ Monaten sind Sie bei EADS bereits beschäftigt?

seit _____ Jahren/ Monaten

5. Welche Tätigkeit üben Sie derzeit im Unternehmen aus?

6. Haben Sie bereits Erfahrung als Software-ProjektleiterIn?

Ja Nein

7. Haben Sie bereits mit systematischen Software-Tests gearbeitet?

Ja Nein

8. Haben Sie bereits vor dem heutigen Tag ein Training zu ähnlichen Inhalten besucht?

Ja Nein

Wenn Sie bei Frage 8 „Ja“ angekreuzt haben, beantworten Sie bitte die beiden nachfolgenden Fragen:

8a) Welchen Titel trug das von Ihnen im Vorfeld besuchte Training?

8b) Wann fand es statt? _____ (Monat/ Jahr)

Fragebogen – Lernebene (1) SWIFT-Modul „Systematischer Software-Test“

Sehr geehrte Trainingsteilnehmerin, sehr geehrter Trainingsteilnehmer,

Sie werden am kommenden Montag, den 19. April 2004 an dem SWIFT-Modul „Systematischer Software-Test“ teilnehmen. Wir, Katharina Niehoff und Marius Werth, möchten dieses Modul im Rahmen unserer Diplomarbeit, welche sich mit der Erfolgsmessung von Weiterbildungsprogrammen beschäftigt, untersuchen. EADS unterstützt als Praxispartner unsere Arbeit und stellt dieses Trainingsmodul für die Untersuchung zur Verfügung. Ziel ist nicht die Überprüfung Ihrer persönlichen Leistung, sondern die Entwicklung von Instrumenten zur Wirksamkeitskontrolle zukünftiger Trainings. Unser Vorgehen ist mit Ihrem Betriebsrat und dem Personalmanagement abgestimmt.

Mit dem vorliegenden Fragebogen werden verschiedene Aspekte Ihrer bisherigen Erfahrung im Umgang mit systematischen Software-Tests abgefragt und Ihr derzeitiges Wissen zum Thema erhoben. Der Fragebogen ist so konzipiert, dass Sie im Vorfeld Ihrer Trainingsteilnahme die unten stehenden Fragen nur zum Teil beantworten können – lassen Sie sich davon nicht entmutigen! Sie werden nach Abschluss des Trainings erneut gebeten, einen ähnlichen Fragebogen auszufüllen.

Wir bitten Sie, möglichst alle Fragen/ Aussagen zu bearbeiten. Bei einem Großteil können Sie durch Ankreuzen des jeweiligen Kästchens deutlich machen, ob Sie die Aussage für richtig oder falsch halten. Bei anderen müssen Sie zwischen Antwortalternativen auswählen. Hierbei kann es sein, dass mehrere Alternativen angekreuzt werden müssen, da sie alle richtige Antworten darstellen. Bei einigen Fragen bitten wir Sie, Ihre Antworten handschriftlich auf die dafür vorgesehenen Linien zu schreiben.

Ihre Daten werden selbstverständlich vertraulich behandelt und nur für Auswertungen im Rahmen der Diplomarbeit genutzt. Um Ihre Anonymität zu sichern, die von Ihnen bearbeiteten Fragebögen aber dennoch einander zuordnen zu können, bitten wir Sie, Ihren persönlichen Code zu erstellen. Rückschlüsse auf Ihre Person sollen und werden anhand dieser Angaben nicht möglich sein. Bitte erstellen Sie Ihren Code hier:

Letzter Buchstabe Ihres Geburtsortes

Erster Buchstabe Ihres Geburtsmonats

Letzter Buchstabe des Vornamens Ihrer Mutter

Erster Buchstabe des Mädchennamens Ihrer Mutter

Die Bearbeitung des Fragebogens wird ca. 15 Minuten dauern. Wir möchten uns bereits an dieser Stelle auch im Namen von Herrn Michael Suchan (HRO) und Herrn Dr. Wolfram Sauer (HRP) für Ihre Mitarbeit bedanken.

Nach Abschluss unserer Diplomarbeit haben Sie die Möglichkeit, die Ergebnisse der Untersuchung einzusehen. Bei Interesse können Sie sich gerne an uns oder Herrn Dr. Wolfram Sauer (HRP) wenden. Selbstverständlich stehen wir Ihnen auch für Rückfragen jederzeit zur Verfügung.

Katharina Niehoff
0171-677 54 85
katharina_niehoff@yahoo.de

Marius Werth
0174-76 95 018
mar.werth@gmx.de

1. Die Durchführung von Tests ermöglicht Aussagen bezüglich der Korrektheit eines Programms.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
2. Durch Testen kann sowohl die Anwesenheit, als auch die Abwesenheit von Fehlern bewiesen werden.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
3. Bitte ordnen Sie die Verfahren der statischen oder dynamischen Prüfung zu und kennzeichnen Sie diese mit einem „S“ (für statisch) oder einem „D“ (für dynamisch) in dem daneben stehenden Kästchen.	<input type="checkbox"/> Audits <input type="checkbox"/> Datenflussanalyse <input type="checkbox"/> Reviews <input type="checkbox"/> Black-Box Test <input type="checkbox"/> Mathematische Programmverifikation <input type="checkbox"/> Äquivalenzklassenmethode
4. Was ist ein Testzyklus und aus welchen Phasen besteht er? _____ _____ _____	
5. Was können Tests aufdecken?	<input type="checkbox"/> Falsche Operationsreihenfolgen <input type="checkbox"/> Falsche Spezifikationen <input type="checkbox"/> Falsche Berechnungen <input type="checkbox"/> Falsche Parameter
6. Der White-Box Test wird auch funktionaler Testansatz genannt.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
7. Von einem White-Box Test spricht man, wenn der Tester sich mit der Implementierung des Testgegenstands auseinandersetzt und seinen Test auf der Konstruktion des Objekts aufbaut.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
8. Der Black-Box Test setzt voraus, dass die innere Struktur des Testobjekts bekannt ist.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
9. Black-Box Fehler sind schwieriger zu beheben als White-Box Fehler.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
10. Bitte vervollständigen Sie die Teststufen.	1. _____ 2. _____ 3. Integrationstest 4. _____ 5. _____
11. Was bezeichnet das Requirements Management?	<input type="checkbox"/> Die systematische Erfassung und Umsetzung von Anforderungen. <input type="checkbox"/> Die Verwaltung der Anforderungen und ihrer Beziehungen. <input type="checkbox"/> Die Vorgehensweise, um die Anforderungen in einem iterativen Prozess zu ermitteln.

12. Bitte nennen Sie drei Nachweismethoden, die im Rahmen der Testplanung verwendet werden.	1. _____ 2. _____ 3. _____
13. Welche Aussagen treffen auf den Unit-Test zu?	<input type="checkbox"/> Man testet die Unit isoliert. <input type="checkbox"/> Man benutzt ausschließlich den „White-Box Test“. <input type="checkbox"/> Man testet auf alle in der Unit möglichen Exceptions. <input type="checkbox"/> Man testet keine Szenarios.
14. Der Komponententest überprüft auch externe Verknüpfungen, da fremde Komponenten Bestandteil des Tests sind.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
15. Was sind Methoden zum dynamischen Testen von Units und Komponenten?	<input type="checkbox"/> Codeanalyse <input type="checkbox"/> Performance-Test <input type="checkbox"/> White-Box Test <input type="checkbox"/> Szenario-Test
16. Die Basis für den kontrollflussorientierten White-Box Test ist ein Pfad, der eine alternierende Folge von Knoten und Kanten aufweist, in einem Startknoten beginnt und in einem Endknoten terminiert.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
17. Bitte nennen Sie drei Fehlerarten, die während des Unit- und Komponententests auftreten können.	1. _____ 2. _____ 3. _____
18. Beim nicht-inkrementellen Integrationstest ist die Lokalisierung von Fehlern, die erst beim Test des gesamten Programms auftreten, schwierig.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
19. Beim nicht-inkrementellen Integrationstest sind Treiber oder Stubs erforderlich.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
20. Für was stehen die Abkürzungen SIL und HIL? SIL = _____ HIL = _____	
21. Bitte vervollständigen Sie den folgenden Satz: Tracing von Inter-Komponenten-Kommunikation während des Szenario-Tests erfordert “design for _____“.	
22. Bitte nennen Sie drei Fehlerarten, die während des Integrationstests auftreten können:	1. _____ 2. _____ 3. _____
23. Ein Systemtest, bei dem die spezifizierte Last erhöht und zugleich die Ressourcen verknüpft werden, heißt _____ .	
24. Bitte nennen Sie neben Zuverlässigkeit und Benutzerfreundlichkeit zwei weitere nicht-funktionale Anforderungen (Requirements), die an den Systemtest gestellt werden.	1. _____ 2. _____

25. Ein Regressionstest tritt während oder nach einer Wartungsphase auf, wenn Änderungen am System vorgenommen werden.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
26. Testtools können feststellen, ...	<input type="checkbox"/> ... welcher Code auch tatsächlich ausgeführt wird. <input type="checkbox"/> ... welcher Code am meisten Zeit verbraucht. <input type="checkbox"/> ... welche Aktionen der Benutzer ausgeführt hat.

Für Ihre Unterstützung und Mitarbeit bedanken wir uns ganz herzlich!

Katharina Niehoff und Marius Werth

Fragebogen – Lernebene (2) SWIFT-Modul „Systematischer Software-Test“

Sehr geehrte Trainingsteilnehmerin, sehr geehrter Trainingsteilnehmer,

Sie haben am Montag, den 19. April 2004 an dem SWIFT-Modul „Systematischer Software-Test“ teilgenommen.

Mit dem vorliegenden Fragebogen möchten wir erneut verschiedene Aspekte zum Thema Software-Tests erheben. Es ist der gleiche Fragebogen, den Sie bereits zu Beginn des Trainings ausgefüllt haben. Bitte nehmen Sie zur Beantwortung der Fragen nicht Ihre Trainingsunterlagen zur Hilfe, sondern beantworten Sie die Fragen aus Ihrer Erinnerung an das Training heraus.

Wieder bitten wir Sie, möglichst alle Fragen/ Aussagen zu bearbeiten. Bei einem Großteil können Sie durch Ankreuzen des jeweiligen Kästchens deutlich machen, ob Sie die Aussage für richtig oder falsch halten. Bei anderen müssen Sie zwischen Antwortalternativen auswählen. Hierbei kann es sein, dass mehrere Alternativen angekreuzt werden müssen, da sie alle richtige Antworten darstellen. Bei einigen Fragen bitten wir Sie, Ihre Antworten handschriftlich auf die dafür vorgesehenen Linien zu schreiben.

Wir möchten Sie erneut darauf hinweisen, dass Ihre Daten selbstverständlich vertraulich behandelt und nur für Auswertungen im Rahmen der Diplomarbeit genutzt werden. Um Ihre Anonymität zu sichern, die von Ihnen bearbeiteten Fragebögen aber dennoch einander zuordnen zu können, bitten wir Sie wieder, Ihren persönlichen Code zu erstellen. Rückschlüsse auf Ihre Person sollen und werden anhand dieser Angaben nicht möglich sein. Bitte erstellen Sie Ihren Code hier:

Letzter Buchstabe Ihres Geburtsortes

Erster Buchstabe Ihres Geburtsmonats

Letzter Buchstabe des Vornamens Ihrer Mutter

Erster Buchstabe des Mädchennamens Ihrer Mutter

Die Bearbeitung des Fragebogens wird ca. 15 Minuten dauern. Wir möchten uns bereits an dieser Stelle wieder einmal für Ihre Mitarbeit und Unterstützung bedanken!

Selbstverständlich stehen wir Ihnen für Rückfragen jederzeit gerne zur Verfügung.

Katharina Niehoff

0171-677 54 85

katharina_niehoff@yahoo.de

Marius Werth

0174-76 95 018

mar.werth@gmx.de

1. Die Durchführung von Tests ermöglicht Aussagen bezüglich der Korrektheit eines Programms.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
2. Durch Testen kann sowohl die Anwesenheit, als auch die Abwesenheit von Fehlern bewiesen werden.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
3. Bitte ordnen Sie die Verfahren der statischen oder dynamischen Prüfung zu und kennzeichnen Sie diese mit einem „S“ (für statisch) oder einem „D“ (für dynamisch) in dem daneben stehenden Kästchen.	<input type="checkbox"/> Audits <input type="checkbox"/> Datenflussanalyse <input type="checkbox"/> Reviews <input type="checkbox"/> Black-Box Test <input type="checkbox"/> Mathematische Programmverifikation <input type="checkbox"/> Äquivalenzklassenmethode
4. Was ist ein Testzyklus und aus welchen Phasen besteht er? _____ _____ _____	
5. Was können Tests aufdecken?	<input type="checkbox"/> Falsche Operationsreihenfolgen <input type="checkbox"/> Falsche Spezifikationen <input type="checkbox"/> Falsche Berechnungen <input type="checkbox"/> Falsche Parameter
6. Der White-Box Test wird auch funktionaler Testansatz genannt.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
7. Von einem White-Box Test spricht man, wenn der Tester sich mit der Implementierung des Testgegenstands auseinandersetzt und seinen Test auf der Konstruktion des Objekts aufbaut.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
8. Der Black-Box Test setzt voraus, dass die innere Struktur des Testobjekts bekannt ist.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
9. Black-Box Fehler sind schwieriger zu beheben als White-Box Fehler.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
10. Bitte vervollständigen Sie die Teststufen.	1. _____ 2. _____ 3. Integrationstest 4. _____ 5. _____
11. Was bezeichnet das Requirements Management?	<input type="checkbox"/> Die systematische Erfassung und Umsetzung von Anforderungen. <input type="checkbox"/> Die Verwaltung der Anforderungen und ihrer Beziehungen. <input type="checkbox"/> Die Vorgehensweise, um die Anforderungen in einem iterativen Prozess zu ermitteln.

12. Bitte nennen Sie drei Nachweismethoden, die im Rahmen der Testplanung verwendet werden.	1. _____ 2. _____ 3. _____
13. Welche Aussagen treffen auf den Unit-Test zu?	<input type="checkbox"/> Man testet die Unit isoliert. <input type="checkbox"/> Man benutzt ausschließlich den „White-Box Test“. <input type="checkbox"/> Man testet auf alle in der Unit möglichen Exceptions. <input type="checkbox"/> Man testet keine Szenarios.
14. Der Komponententest überprüft auch externe Verknüpfungen, da fremde Komponenten Bestandteil des Tests sind.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
15. Was sind Methoden zum dynamischen Testen von Units und Komponenten?	<input type="checkbox"/> Codeanalyse <input type="checkbox"/> Performance-Test <input type="checkbox"/> White-Box Test <input type="checkbox"/> Szenario-Test
16. Die Basis für den kontrollflussorientierten White-Box Test ist ein Pfad, der eine alternierende Folge von Knoten und Kanten aufweist, in einem Startknoten beginnt und in einem Endknoten terminiert.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
17. Bitte nennen Sie drei Fehlerarten, die während des Unit- und Komponententests auftreten können.	1. _____ 2. _____ 3. _____
18. Beim nicht-inkrementellen Integrationstest ist die Lokalisierung von Fehlern, die erst beim Test des gesamten Programms auftreten, schwierig.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
19. Beim nicht-inkrementellen Integrationstest sind Treiber oder Stubs erforderlich.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
20. Für was stehen die Abkürzungen SIL und HIL? SIL = _____ HIL = _____	
21. Bitte vervollständigen Sie den folgenden Satz: Tracing von Inter-Komponenten-Kommunikation während des Szenario-Tests erfordert “design for _____“.	
22. Bitte nennen Sie drei Fehlerarten, die während des Integrationstests auftreten können:	1. _____ 2. _____ 3. _____
23. Ein Systemtest, bei dem die spezifizierte Last erhöht und zugleich die Ressourcen verknappt werden, heißt _____.	
24. Bitte nennen Sie neben Zuverlässigkeit und Benutzerfreundlichkeit zwei weitere nicht-funktionale Anforderungen (Requirements), die an den Systemtest gestellt werden.	1. _____ 2. _____

25. Ein Regressionstest tritt während oder nach einer Wartungsphase auf, wenn Änderungen am System vorgenommen werden.	<input type="checkbox"/> Richtig <input type="checkbox"/> Falsch
26. Testtools können feststellen, ...	<input type="checkbox"/> ... welcher Code auch tatsächlich ausgeführt wird. <input type="checkbox"/> ... welcher Code am meisten Zeit verbraucht. <input type="checkbox"/> ... welche Aktionen der Benutzer ausgeführt hat.

Für Ihre Unterstützung und Mitarbeit bedanken wir uns ganz herzlich!

Katharina Niehoff und Marius Werth

Transfererfolgsmessung SWIFT-Modul „Systematischer Software-Test“

Sehr geehrte Trainingsteilnehmerin, sehr geehrter Trainingsteilnehmer,

für Ihre bisherige tatkräftige Unterstützung der im Zuge der Neuorganisation der Division DS etablierten HR-Controlling Funktion danken wir Ihnen sehr herzlich. Nachdem Sie am 19. April 2004 an dem SWIFT-Modul „Systematischer Software-Test“ teilgenommen haben, bitten wir Sie hiermit ein letztes Mal, einen Fragebogen zu bearbeiten. Bei diesem Fragebogen geht es u. a. darum, ob Sie die im Training erlernten Inhalte und Methoden in Ihrem Arbeitskontext anwenden konnten bzw. diese bereits vor dem Training angewendet haben. Die genauen Instruktionen und Fragestellungen finden Sie vor den jeweiligen Fragen.

Mit dem vorliegenden Fragebogen werden somit verschiedene Aspekte Ihrer Erfahrung mit der Anwendbarkeit von Inhalten und Techniken des systematischen Testens erhoben. Wir bitten Sie, möglichst alle Fragen zu bearbeiten und genau ein „X“ in das jeweilige Kästchen zu setzen. Wichtig ist außerdem, dass es keine richtigen oder falschen Antworten gibt, da ausschließlich Ihre Erfahrung im Umgang mit Inhalten des Trainings von Interesse ist.

Wie Sie bereits wissen, werden Ihre Daten selbstverständlich vertraulich behandelt und nur für Auswertungen im Rahmen der Effizienzüberprüfung des Weiterbildungsprogramms „Systematischer Software-Test“ genutzt. Um Ihre Anonymität zu sichern, die von Ihnen bearbeiteten Fragebögen aber dennoch einander zuordnen zu können, bitten wir Sie ein letztes Mal, Ihren persönlichen Code zu erstellen. Rückschlüsse auf Ihre Person sollen und werden anhand dieser Angaben nicht möglich sein. Bitte erstellen Sie Ihren Code hier:

Letzter Buchstabe Ihres Geburtsortes

Erster Buchstabe Ihres Geburtsmonats

Letzter Buchstabe des Vornamens Ihrer Mutter

Erster Buchstabe des Mädchennamens Ihrer Mutter

Die Bearbeitung des Fragebogens wird ca. 20 Minuten dauern.

Für eventuelle Rückfragen steht Ihnen Herr/ Frau XY unter der Telefonnummer 0000 oder der email-Adresse name@sysde.eads.net zur Verfügung. Wie bereits angekündigt, haben Sie nach Abschluss dieser Untersuchung die Möglichkeit, die Ergebnisse einzusehen. Bei Interesse wenden Sie sich bitte an Herrn Dr. Wolfram Sauer (HRP).

Mit freundlichen Grüßen,

Frau/ Herr XY

1. Haben Sie sich nach dem Training „Systematischer Software-Test“ noch einmal mit den Trainingsunterlagen beschäftigt? Ja Nein

2. Haben Sie seit dem Training „Systematischer Software-Test“ an weiteren Trainings zu ähnlichen Inhalten teilgenommen? Ja Nein

2.a) Falls Sie soeben „Ja“ angekreuzt haben, wie lautete der genaue Titel des Trainings und wann fand es statt?

Titel: _____

Monat/ Jahr: _____

Bitte geben Sie an, welche der folgenden konkreten Inhalte des Trainings „Systematischer Software-Test“ Sie **nach dem Besuch des Trainings** bereits persönlich angewendet haben:

3. Haben Sie nach dem Training einen Unit-/ Komponententest durchgeführt? Ja Nein

a) Wie oft haben Sie die folgenden Testmethoden nach dem Training im Rahmen des Unit-/ Komponententests angewendet?

	sehr oft	oft	ab und zu	selten	nie
Codeanalyse					
White-Box Test					
Grey-Box Test					
Black-Box Test					

b) Wie oft bereitete Ihnen die Anwendung der folgenden Testmethoden nach dem Training im Rahmen des Unit-/ Komponententests Probleme?

	sehr oft	oft	ab und zu	selten	nie
Codeanalyse					
White-Box Test					
Grey-Box Test					
Black-Box Test					

c) Wie oft sind die folgenden Arten von Fehlern nach dem Training im Rahmen des Unit-/ Komponententests aufgetreten?

	sehr oft	oft	ab und zu	selten	nie
SW-Spezifikationsfehler					
Schnittstellenfehler					
Initialisierungsfehler					
Speicherfehler					
Programmierfehler					
Andere Fehler					

4. Haben Sie nach dem Training einen **Integrationstest** durchgeführt?

- Ja
 Nein

a) Wie oft haben Sie die folgenden Tests/ Testmethoden nach dem Training im Rahmen des Integrationstests angewendet?

	sehr oft	oft	ab und zu	selten	nie
Nicht-inkrementeller Integrationstest					
Inkrementeller Integrationstest					
Szenario-Test					
Performance Test					
Feedback-Loops					

b) Wie oft bereitete Ihnen die Anwendung der folgenden Tests/ Testmethoden nach dem Training im Rahmen des Integrationstests Probleme?

	sehr oft	oft	ab und zu	selten	nie
Nicht-inkrementeller Integrationstest					
Inkrementeller Integrationstest					
Szenario-Test					
Performance Test					
Feedback-Loops					

c) Wie oft sind die folgenden Arten von Fehlern nach dem Training während des Integrationstests aufgetreten?

	sehr oft	oft	ab und zu	selten	nie
SW-Spezifikationsfehler					
Schnittstellenfehler					
Timing-Fehler					
Multithreading-Fehler					
Initialisierungs-/ Shutdown-Fehler					
Speicherfehler					
Fehler in der Umgebung					
Fehler in der Produktion					
Andere Fehler					

5. Haben Sie nach dem Training einen **Systemtest** durchgeführt?

- Ja
 Nein

a) Wie oft haben Sie die folgenden Testmethoden nach dem Training im Rahmen des Systemtests angewendet?

	sehr oft	oft	ab und zu	selten	nie
Lasttest					
Stress-Test					
Performance-Test					

b) Wie oft bereitete Ihnen die Anwendung der folgenden Tests nach dem Training im Rahmen des Systemtests Probleme?

	sehr oft	oft	ab und zu	selten	nie
Lasttest					
Stress-Test					
Performance-Test					

6. Wie oft waren Sie nach dem Training in die Planung eines Software-Tests einbezogen?
7. Wie oft waren Sie nach dem Training mit der Entwicklung eines Test-Designs betraut?
8. Wie oft konnten Sie nach dem Training an der Durchführung eines Tests mitwirken?
9. Wie oft haben Sie nach dem Training Aufgaben im Bereich der Trainingsevaluation übernommen?

	sehr oft	oft	ab und zu	selten	nie

10. Wie wurde die Umsetzung der Trainingsinhalte in Ihrem Arbeitsumfeld durch direkte Vorgesetzte nach dem Training gefördert?

- stark
- eher stark
- weder noch
- eher schwach
- schwach

11. Wie wurde die Umsetzung der Trainingsinhalte in Ihrem Arbeitsumfeld durch Kollegen nach dem Training gefördert?

- stark
- eher stark
- weder noch
- eher schwach
- schwach

Bitte geben Sie an, welche der folgenden konkreten Inhalte des Trainings „Systematischer Software-Test“ Sie **bereits vor dem Besuch des Trainings** persönlich angewendet haben:

12. Haben Sie vor dem Training einen **Unit-/ Komponententest** durchgeführt?

- Ja
- Nein

- a) Wie oft haben Sie die folgenden Testmethoden vor dem Training im Rahmen des Unit-/ Komponententests angewendet?

	sehr oft	oft	ab und zu	selten	nie
Codeanalyse					
White-Box Test					
Grey-Box Test					
Black-Box Test					

- b) Wie oft bereitete Ihnen die Anwendung der folgenden Testmethoden vor dem Training im Rahmen des Unit-/ Komponententests Probleme?

	sehr oft	oft	ab und zu	selten	nie
Codeanalyse					
White-Box Test					
Grey-Box Test					
Black-Box Test					

c) Wie oft sind die folgenden Arten von Fehlern vor dem Training während des Unit-/ Komponententests aufgetreten?

	sehr oft	oft	ab und zu	selten	nie
SW-Spezifikationsfehler					
Schnittstellenfehler					
Initialisierungsfehler					
Speicherfehler					
Programmierfehler					
Andere Fehler					

13. Haben Sie vor dem Training einen **Integrationstest** durchgeführt? Ja Nein

a) Wie oft haben Sie die folgenden Tests/ Testmethoden vor dem Training im Rahmen des Integrations- tests angewendet?

	sehr oft	oft	ab und zu	selten	nie
Nicht-inkrementeller Integrationstest					
Inkrementeller Integrationstest					
Szenario-Test					
Performance Test					
Feedback-Loops					

b) Wie oft bereitete Ihnen die Anwendung der folgenden Tests/ Testmethoden vor dem Training im Rah- men des Integrationstests Probleme?

	sehr oft	oft	ab und zu	selten	nie
Nicht-inkrementeller Integrationstest					
Inkrementeller Integrationstest					
Szenario-Test					
Performance Test					
Feedback-Loops					

c) Wie oft sind die folgenden Arten von Fehlern vor dem Training während des Integrationstests aufgetre- ten?

	sehr oft	oft	ab und zu	selten	nie
SW-Spezifikationsfehler					
Schnittstellenfehler					
Timing-Fehler					
Multithreading-Fehler					
Initialisierungs-/ Shutdown-Fehler					
Speicherfehler					
Fehler in der Umgebung					
Fehler in der Produktion					
Andere Fehler					

14. Haben Sie vor dem Training einen Systemtest durchgeführt?

- Ja
 Nein

a) Wie oft haben Sie die folgenden Testmethoden vor dem Training im Rahmen des Systemtests angewendet?

	sehr oft	oft	ab und zu	selten	nie
Lasttest					
Stress-Test					
Performance-Test					

b) Wie oft bereitete Ihnen die Anwendung der folgenden Tests vor dem Training Probleme?

	sehr oft	oft	ab und zu	selten	nie
Lasttest					
Stress-Test					
Performance-Test					

15. Wie oft waren Sie vor dem Training in die Planung eines Software-Tests einbezogen?
 16. Wie oft waren Sie vor dem Training mit der Entwicklung eines Test-Designs betraut?
 17. Wie oft konnten Sie vor dem Training an der Durchführung eines Tests mitwirken?
 18. Wie oft haben Sie vor dem Training Aufgaben im Bereich der Trainingsevaluation übernommen?

	sehr oft	oft	ab und zu	selten	nie

Abschließend möchten wir uns noch einmal für Ihr Engagement und Ihre Unterstützung bedanken!

Notwendige Eckdaten (Kosten, Effektdauer) zur Berechnung des Nutzwertes und des ROI des Trainingsmoduls „Systematischer Software-Test“

Kostenklassifikation	
Aufwendungen	Angabe in Euro
Durchschnittliches Bruttojahresgehalt der Trainingsteilnehmer	€ 57.230
Lohnnebenkosten der Trainingsteilnehmer	€ 12.469
Gesamtkosten für das Trainingsmodul „Systematischer Software-Test“ (dazu gehören u. a. Entwicklungs- und Materialkosten, Tagessätze des externen Trainingspersonals von Kölsch & Altmann, Reisekosten und Unterkunft für Trainingspersonal, sonstige Aufwendungen für Trainingspersonal)	€ 2.000
Raumkosten (Mietkosten für Raum Nonnenhorn im GästeCasino)	€ 166
Ausstattung der Trainingsräumlichkeiten (Nutzungskosten für Overhead, Beamer, Flipchart, etc.)	€ 120
Verpflegung während des Trainings für Trainer und Teilnehmer (Mittagessen, Getränke etc.)	€ 180
Reisekosten und ggf. Unterkunft Teilnehmer	Werden von TN bezahlt
Kosten für Administrationspersonal (Angabe unter Berücksichtigung von Zeitaufwand, Bruttojahresgehalt und Lohnnebenkosten)	€ 800

Bestimmung der Dauer und Stärke der Trainingseffekte

Zur Bestimmung der „Payback-period“, die einen Schätzwert liefert, wie lange Trainingseffekte anhalten und zur Bestimmung der Effektgröße sind folgende Informationen erforderlich:

- Über wie viele Jahre werden die Trainingsteilnehmer voraussichtlich verbesserte Leistung zeigen, die auf das Modul zurückgeführt werden können?
5 - 10 Jahre, da es sich um ein Grundlagenseminar handelt
- Auf wieviel Prozent steigt die Gesamtleistung eines Mitarbeiters nach dem Besuch des Trainingsmoduls ihrer Meinung nach an, wenn die Leistung vor dem Besuch des Moduls auf 100% angesetzt wird?
Gesamtleistung steigt auf ca. 105% - 110% an.
- Wie lange sind Mitarbeiter durchschnittlich bei EADS beschäftigt?
8 Jahre
- Können die neu erlernten Modulinhalt im Arbeitskontext der Teilprojektleiter angewendet werden?
Ja Nein

Zielsetzung des Moduls „Systematischer Software-Test“

Bei der Durchführung des Moduls stehen folgende Ziele im Vordergrund:

- » Die Bedeutung (die Notwendigkeit) des systematischen Software-Tests entlang des SW-Entwicklungsprozesses erkennen. Die verschiedenen Methoden des systematischen SW-Tests kennenlernen (inklusive Vor- und Nachteile, sowie die Grenzen der jeweiligen Methode).
- » Werkzeuge für den systematischen SW-Test kennenlernen.
- » Methoden (ggf. Werkzeuge) im Rahmen des Lernprojektes im praktischen Einsatz kennenlernen, Erfahrungen gewinnen bzw. durch kollegiales Lernteam Erfahrungen vertiefen.
- » Aus den verschiedenen Methoden eine Teststrategie für das eigene Projekt ableiten.
- »
- »

Sollten die Ziele zum jetzigen Zeitpunkt nicht benennbar sein, werden wir diese aus den Trainingsinhalten ableiten, sobald uns diese zur Verfügung stehen.

Anhang F

Anhand der im Folgenden dargestellten Tabellen wird ersichtlich, welche Berechnungen zur statistischen Auswertung der Fragebogen herangezogen wurden.

F.1 Deskriptive Statistiken

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Alter	11	27	39	33,09	3,618
Beschäftigungsdauer bei EADS in Jahren	11	2,00	9,00	5,6545	2,33939
Gültige Werte (Listenweise)	11				

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Mittelwert Zufriedenheit über 6 Skalen	11	3,19	4,55	3,8321	,48019
Gültige Werte (Listenweise)	11				

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Mittelwert Zufriedenheit mit Trainer	11	3,60	5,00	4,1636	,51239
Mittelwert Zufriedenheit mit administrativem Trainingsmanagement	11	3,00	5,00	3,9697	,67420
Mittelwert Zufriedenheit mit Schwierigkeitsgrad des Trainings	11	2,67	5,00	3,7879	,71915
Mittelwert Zufriedenheit mit Nützlichkeit des Moduls	11	2,50	5,00	3,8409	,62523
Mittelwert Zufriedenheit mit Trainingsumfeld	11	2,50	5,00	4,0303	,72291
Mittelwert Zufriedenheit mit Aufbau des Kurses	11	2,00	4,60	3,2000	,70993
Gültige Werte (Listenweise)	11				

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Mittelwert Gesamteindruck (25,27,29,31)	11	3,50	4,50	4,0000	,35355
Gültige Werte (Listenweise)	11				

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Mittelwert Motivation (26,28,30)	11	2,67	4,67	3,5455	,70353
Gültige Werte (Listenweise)	11				

Statistics

		Anzahl richtiger Antworten im Pre-Lerntest	Anzahl richtiger Antworten im Post-Lerntest
N	Valid	10	10
	Missing	0	0
Mean		13,90	21,10
Median		14,50	21,50
Std. Deviation		3,755	5,195
Minimum		6	9
Maximum		20	27

F.2 Häufigkeitstabellen

Hochschulstudium

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig Naturwissenschaften	1	9,1	9,1	9,1
Ingenieurwesen	4	36,4	36,4	45,5
Informatik	6	54,5	54,5	100,0
Gesamt	11	100,0	100,0	

Erfahrung als Software-ProjektleiterIn

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ja	6	54,5	54,5	54,5
nein	5	45,5	45,5	100,0
Gesamt	11	100,0	100,0	

Bereits mit systematischen SW-Test gearbeitet

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ja	8	72,7	72,7	72,7
nein	3	27,3	27,3	100,0
Gesamt	11	100,0	100,0	

Bereits Training zu ähnlichen Inhalten besucht

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ja	3	27,3	27,3	27,3
nein	8	72,7	72,7	100,0
Gesamt	11	100,0	100,0	

1 Engagement des Trainers

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig eher zufrieden	7	63,6	63,6	63,6
sehr zufrieden	4	36,4	36,4	100,0
Gesamt	11	100,0	100,0	

2 Verständlichkeit der Inhalte

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig eher zufrieden	6	54,5	54,5	54,5
sehr zufrieden	5	45,5	45,5	100,0
Gesamt	11	100,0	100,0	

3 Kompetenz des Trainers bezüglich systematischer SW-Test

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig eher zufrieden	6	54,5	54,5	54,5
sehr zufrieden	5	45,5	45,5	100,0
Gesamt	11	100,0	100,0	

20 Länge des Trainings

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig sehr unzufrieden	1	9,1	9,1	9,1
eher unzufrieden	3	27,3	27,3	36,4
teils/ teils	3	27,3	27,3	63,6
eher zufrieden	3	27,3	27,3	90,9
sehr zufrieden	1	9,1	9,1	100,0
Gesamt	11	100,0	100,0	

21 Geschwindigkeit mit der Inhalte präsentiert wurden

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher unzufrieden	4	36,4	36,4	36,4
	teils/ teils	3	27,3	27,3	63,6
	eher zufrieden	3	27,3	27,3	90,9
	sehr zufrieden	1	9,1	9,1	100,0
	Gesamt	11	100,0	100,0	

23 Anzahl praktischer Übungen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	sehr unzufrieden	1	9,1	9,1	9,1
	eher unzufrieden	4	36,4	36,4	45,5
	teils/ teils	2	18,2	18,2	63,6
	eher zufrieden	3	27,3	27,3	90,9
	sehr zufrieden	1	9,1	9,1	100,0
	Gesamt	11	100,0	100,0	

11 Menge der zu bearbeitenden Inhalte

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	sehr unzufrieden	1	9,1	9,1	9,1
	eher unzufrieden	4	36,4	36,4	45,5
	teils/ teils	3	27,3	27,3	72,7
	eher zufrieden	1	9,1	9,1	81,8
	sehr zufrieden	2	18,2	18,2	100,0
	Gesamt	11	100,0	100,0	

25 Training entsprach voll und ganz Erwartungen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	stimme eher nicht zu	1	9,1	9,1	9,1
	teils/ teils	3	27,3	27,3	36,4
	stimme eher zu	6	54,5	54,5	90,9
	stimme zu	1	9,1	9,1	100,0
	Gesamt	11	100,0	100,0	

27 Besuch des Trainings lohnt sich

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	stimme eher nicht zu	1	9,1	9,1	9,1
	stimme eher zu	4	36,4	36,4	45,5
	stimme zu	6	54,5	54,5	100,0
	Gesamt	11	100,0	100,0	

31 Werde Training Kollegen weiterempfehlen

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig teils/ teils	3	27,3	27,3	27,3
stimme eher zu	8	72,7	72,7	100,0
Gesamt	11	100,0	100,0	

26 Eigenes Interesse an Trainingsinhalten

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig teils/ teils	1	9,1	9,1	9,1
stimme eher zu	8	72,7	72,7	81,8
stimme zu	2	18,2	18,2	100,0
Gesamt	11	100,0	100,0	

28 persönlich für Teilnahme eingesetzt

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig stimme eher nicht zu	3	27,3	27,3	27,3
teils/ teils	3	27,3	27,3	54,5
stimme eher zu	1	9,1	9,1	63,6
stimme zu	4	36,4	36,4	100,0
Gesamt	11	100,0	100,0	

30 Werde nicht an Zielen festhalten, wenn schwer zu erreichen

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig stimme eher nicht zu	3	27,3	27,3	27,3
teils/ teils	6	54,5	54,5	81,8
stimme eher zu	1	9,1	9,1	90,9
stimme zu	1	9,1	9,1	100,0
Gesamt	11	100,0	100,0	

Anzahl richtiger Antworten im Pre-Lerntest

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 6	1	10,0	10,0	10,0
10	1	10,0	10,0	20,0
13	1	10,0	10,0	30,0
14	2	20,0	20,0	50,0
15	2	20,0	20,0	70,0
16	2	20,0	20,0	90,0
20	1	10,0	10,0	100,0
Total	10	100,0	100,0	

Anzahl richtiger Antworten im Post-Lerntest

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 9	1	10,0	10,0	10,0
18	1	10,0	10,0	20,0
20	2	20,0	20,0	40,0
21	1	10,0	10,0	50,0
22	2	20,0	20,0	70,0
26	2	20,0	20,0	90,0
27	1	10,0	10,0	100,0
Total	10	100,0	100,0	

F.3 Reliabilitätsanalyse

***** Method 1 (space saver) will be used for this analysis *****

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
TR_ENGA	91,2857	130,2381	,8977	,8983
TR_VER	91,4286	131,9524	,7523	,9002
TR_KOMP	91,1429	139,1429	,1820	,9073
TR_BEG	92,0000	121,0000	,8220	,8945
TR_LERN	91,8571	123,8095	,7489	,8967
TM_ZEIT	91,0000	137,3333	,4515	,9048
TM_INFO	92,2857	128,5714	,5401	,9018
TM_UNT	91,7143	118,2381	,8236	,8939
SG_VER	91,4286	139,6190	,1244	,9082
SG_ANF	91,7143	128,5714	,4852	,9034
SG_MENGE	92,2857	122,9048	,6875	,8980
NU_INHA	91,7143	130,5714	,6613	,9002
NU_RELEV	91,7143	137,9048	,3862	,9054
NU_EFFEK	91,8571	133,8095	,5490	,9026
NU_PRAX	92,2857	132,9048	,3386	,9065
TU_MEDIE	91,8571	123,1429	,6634	,8987
TU_TU	91,8571	119,8095	,8044	,8946
TU_RAUM	91,7143	122,2381	,9263	,8931
TU_HILF	92,0000	124,6667	,6563	,8989
AUF_LÄN	92,4286	127,2857	,6012	,9004
AUF_GES	92,2857	131,5714	,3999	,9051
AUF_LOG	91,7143	137,2381	,4624	,9047
AUF_ÜBU	92,5714	146,9524	-,2489	,9218
AUF_MET	92,5714	141,6190	-,0397	,9122

Reliability Coefficients

N of Cases = 7,0

N of Items = 24

Alpha = ,9062

F.4 Korrelationen

Correlations

		Mittelwert Zufriedenheit über 6 Skalen	Lernleistung
Mittelwert Zufriedenheit über 6 Skalen	Pearson Correlation	1	-,169
	Sig. (2-tailed)	,	,642
	N	10	10
Lernleistung	Pearson Correlation	-,169	1
	Sig. (2-tailed)	,642	,
	N	10	10

Correlations

		Lernleistung	Mittelwert Motivation (26,28,30)
Lernleistung	Pearson Correlation	1	,254
	Sig. (2-tailed)	,	,480
	N	10	10
Mittelwert Motivation (26,28,30)	Pearson Correlation	,254	1
	Sig. (2-tailed)	,480	,
	N	10	10

Correlations

		Lernleistung	Mittelwert Zufriedenheit mit Trainer	Mittelwert Zufriedenheit mit administrativem Trainingsmanagement	Mittelwert Zufriedenheit mit Schwierigkeitsgrad des Trainings	Mittelwert Zufriedenheit mit Nützlichkeit des Moduls	Mittelwert Zufriedenheit mit Trainingsumfeld	Mittelwert Zufriedenheit mit Aufbau des Kurses
Lernleistung	Pearson Correlation	1	-,422	-,024	,311	-,407	-,410	,137
	Sig. (2-tailed)	,	,225	,947	,382	,243	,239	,706
	N	10	10	10	10	10	10	10
Mittelwert Zufriedenheit mit Trainer	Pearson Correlation	-,422	1	,665*	,545	,631	,858**	,429
	Sig. (2-tailed)	,225	,	,036	,103	,050	,001	,216
	N	10	10	10	10	10	10	10
Mittelwert Zufriedenheit mit administrativem Trainingsmanagement	Pearson Correlation	-,024	,665*	1	,706*	,379	,553	,628
	Sig. (2-tailed)	,947	,036	,	,022	,280	,097	,052
	N	10	10	10	10	10	10	10
Mittelwert Zufriedenheit mit Schwierigkeitsgrad des Trainings	Pearson Correlation	,311	,545	,706*	1	,275	,406	,385
	Sig. (2-tailed)	,382	,103	,022	,	,442	,244	,272
	N	10	10	10	10	10	10	10
Mittelwert Zufriedenheit mit Nützlichkeit des Moduls	Pearson Correlation	-,407	,631	,379	,275	1	,388	,089
	Sig. (2-tailed)	,243	,050	,280	,442	,	,268	,806
	N	10	10	10	10	10	10	10
Mittelwert Zufriedenheit mit Trainingsumfeld	Pearson Correlation	-,410	,858**	,553	,406	,388	1	,216
	Sig. (2-tailed)	,239	,001	,097	,244	,268	,	,550
	N	10	10	10	10	10	10	10
Mittelwert Zufriedenheit mit Aufbau des Kurses	Pearson Correlation	,137	,429	,628	,385	,089	,216	1
	Sig. (2-tailed)	,706	,216	,052	,272	,806	,550	,
	N	10	10	10	10	10	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

F.5 T-Test

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Anzahl richtiger Antworten im Pre-Lerntest	13,90	10	3,755	1,187
Anzahl richtiger Antworten im Post-Lerntest	21,10	10	5,195	1,643

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 Anzahl richtiger Antworten im Pre-Lerntest & Anzahl richtiger Antworten im Post-Lerntest	10	,787	,007

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	99% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Anzahl richtiger Antworten im Pre-Lerntest - Anzahl richtiger Antworten im Post-Lerntest	-7,20	3,225	1,020	-10,51	-3,89	-7,060	9	,000

Eidesstattliche Erklärung

Hiermit versichern wir, dass wir die entsprechend der beigelegten Notiz gekennzeichneten Teile der vorliegenden Arbeit mit dem Thema „Evaluation einer betrieblichen Personalentwicklungsmaßnahme: Theoretische Fundierung, praktische Anwendung und Möglichkeiten zur Implementierung“ selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt haben. Die Textpassagen, die anderen Werken dem Wortlaut nach oder sinngemäß entnommen sind, haben wir in jedem einzelnen Fall durch die Angabe der Quelle als solche kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Lüneburg, den 14. Mai 2004

Katharina Niehoff

Marius Werth