

Freie Universität  Berlin

Freie Universität Berlin  
Arbeitsbereich Erziehungswissenschaft und Psychologie  
Sommersemester 2012

# **Wirksamkeit von Weiterbildung**

## *Evaluation der Ausbildung zum Innovationspromotor*

**Masterarbeit im Projekt GI:VE - „Grundlagen nachhaltiger  
Innovationsfähigkeit: Vertrauenskultur und evolutionäre Wissensproduktion“  
am Institut für Psychologie der Humboldt-Universität zu Berlin**

Erstgutachter: Professor Dr. Wolfgang Scholl  
Zweitgutachter: Professor Dr. Harm Kuper

vorgelegt am 26. Juli 2012 von:

**Sandra Tirre**

Masterstudiengang Erziehungswissenschaft: Forschung und Entwicklung  
in sozialen und pädagogischen Organisationen

## Zusammenfassung

Ausgehend vom steigenden Bedarf in der Unternehmens- und Trainingspraxis nach einerseits validen und andererseits praktikabel umsetzbaren Konzepten zur Trainingsevaluation, widmet sich die vorliegende Untersuchung gezielt der Fragestellung, wie ein solcher *ganzheitlicher* Evaluationsansatz aussehen kann. Insbesondere für verhaltensorientierte Trainingsprogramme stellt die präzise Erfassung von den zumeist *weichen Kriterien* des Erfolges - wie Einstellungs- oder Verhaltensänderungen - im Rahmen von Evaluationsvorhaben eine große Herausforderung dar. Der Fokus der vorliegenden Untersuchung lag daher auf der Entwicklung eines Evaluationsansatzes, der sich insbesondere zur Überprüfung der Wirksamkeit der verhaltensorientierten Trainingsformate eignet. Die Beantwortung der Forschungsfrage erfolgte auf zwei Ebenen. In einem ersten Schritt wurde auf konzeptioneller Ebene ein Ansatz entwickelt, der sowohl den Anforderungen der Wissenschaft als auch denen der Praxis genügen sollte. Den Kern des Ansatzes bildete das Vier-Ebenen-Modell von Kirkpatrick (1959), das die Überprüfung der Wirksamkeit eines Trainings entlang der vier Ebenen Reaktion, Lernen, Verhalten und Resultate vorsieht. Ein weiterer Schwerpunkt in der Konzeptentwicklung lag auf der Kopplung von ergebnis- und prozessbezogenen Evaluationsmodellen. Neben den Output-Kriterien wurden daher Einflussvariablen auf den Trainings- und Transferprozess (Baldwin & Ford, 1988) einbezogen. In einem zweiten Schritt wurde der Nutzen und Erfolg des Ansatzes im Rahmen der Evaluation des verhaltensorientierten Trainingsprogramms *Ausbildung zum Innovationspromotor* empirisch überprüft. Im Rahmen eines Längsschnittdesigns wurde die Erhebung der Erfolgskriterien mittels maßgeschneiderter und publizierter Verfahren zu mehreren Zeitpunkten umgesetzt. Der entwickelte Ansatz hat sich dabei grundsätzlich bewährt. Es konnte ein differenziertes Bild über die Wirksamkeit der Ausbildung gezeichnet werden. Das Anliegen, dabei sowohl wissenschaftlichen als auch praktikablen Anforderungen zu genügen, wurde teilweise erfüllt. Auf Grundlage der Ergebnisse konnten jedoch bereits konkrete und vor allem handhabbare Möglichkeiten zur Weiterentwicklung des Ansatzes entwickelt werden. Ebenso ist es gelungen, aus den Befunden erste Empfehlungen für die Trainingspraxis abzuleiten, die ein praktikables und wissenschaftlich anspruchsvolles Vorgehen im Rahmen von Evaluationen verbinden.

## Vorbemerkung

Die vorliegende Arbeit ist im Rahmen des Forschungsprojektes GI:VE - „Grundlagen nachhaltiger Innovationsfähigkeit: Vertrauenskultur und evolutionäre Wissensproduktion“ entstanden. In dem vom BMBF<sup>1</sup> und ESF<sup>2</sup> geförderten Verbundprojekt von Wissenschaft und Praxis arbeiten Vertreter<sup>3</sup> der Humboldt-Universität zu Berlin und des Beratungsinstituts artop GmbH gemeinsam daran, die Ausgestaltung einer betrieblichen Vertrauenskultur und der nachhaltigen Innovationsfähigkeit von Unternehmen voranzutreiben (GI:VE, 2010).

Die Ausbildung zum Innovationspromotor als **Untersuchungsgegenstand der vorliegenden Arbeit** ist als eine wichtige Transfermaßnahme aus dem Projekt hervorgegangen. Von April 2011 bis Januar 2012 wurde die Ausbildung erstmalig durchgeführt und begleitend evaluiert. Die systematische Evaluation der Wirksamkeit der Ausbildung führte ich in enger Kooperation mit einer weiteren Projektdiplomandin, Frau Grit Rudinger, durch. Sie begleitete die ersten fünf Module der Ausbildung von April bis August 2011 als Evaluatorin und Protokollantin und war an der Entwicklung des Evaluationskonzeptes und der Erhebungsinstrumente maßgeblich beteiligt. Im Verlauf der Arbeit finden sich entsprechende Verweise dazu. Frau Rudinger stellte mir die erhobenen Daten der ersten fünf Module der Ausbildung zur Verfügung, sodass im Rahmen meiner Arbeit die Auseinandersetzung mit dem vollständigen Datensatz möglich wurde.

An dieser Stelle möchte ich Frau Rudinger für die gute Zusammenarbeit, ihre Hilfsbereitschaft und ihre Unterstützung danken.

---

<sup>1</sup> Bundesministerium für Bildung und Forschung

<sup>2</sup> Europäischer Sozialfond

<sup>3</sup> In dieser Arbeit werden soweit wie möglich geschlechtsneutrale Formulierungen verwendet. Wo dies nicht gelingt, wird zur sprachlichen Vereinfachung und besseren Lesbarkeit die allgemeine männliche Form verwendet. Die Form bezieht in der gesamten Arbeit immer beide Geschlechter mit ein.

## **Danksagung**

Eine Abschlussarbeit zu erstellen ist eine anspruchsvolle Lebensphase, die ich ohne die tolle Unterstützung meines familiären und privaten Umfelds nicht so gut gemeistert hätte.

Mein besonderer Dank gilt meinen Eltern, die zu jeder Zeit uneingeschränkt für mich da waren. Danke für eure Anteilnahme und euer Vertrauen!

Ich danke meinen lieben Freunden für ihre Unterstützung, ihr Verständnis und den festen Glauben an mich und mein Können.

Besonders danken möchte ich Frieder, Hannah, Rike, Nora und Jonas.

Wesentlichen Anteil an den Ergebnissen dieser Arbeit hat der Projektmanager des GI:VE Projektes Professor Dr. Sebastian Kunert. Ich möchte ihm an dieser Stelle für seine große Unterstützung und seine vielen interessanten Denkanstöße danken.

Außerdem danke ich dem gesamten GI:VE Team, den Teilnehmern der Ausbildung zum Innovationspromotor und meinen beiden Betreuern Professor Dr. Wolfgang Scholl und Professor Dr. Harm Kuper.

# Inhalt

I Einleitung	1
1. Entdeckungszusammenhang und Problemstellung	1
2. Zielsetzung und Anlage der Arbeit	3
3. Aufbau der Arbeit	5
II Theorie	6
1. Die Praxis von Trainings als Gegenstand von Forschung	6
1.1 Kommunikations- und Verhaltenstrainings: Lernfelder und -methoden	6
1.2 Trainingsprogramme zur Ausbildung professioneller Rollen	7
2. Evaluation von Trainings	10
2.1 (Trainings-)Evaluation: Eine begriffliche Annäherung	10
2.2 Strategien der Evaluation	12
2.3 Design- und Planungsfragen in der (Trainings-)Evaluation	13
2.3.1 Störfaktoren	14
2.3.2 Kontrollmöglichkeiten	15
2.4 Modelle der Trainings- und Transferevaluation	18
2.4.1 Ebenen und Inhalte von Trainingsergebnissen	19
2.4.1.1 Ebene Reaktion	19
2.4.1.2 Ebene Lernen	21
2.4.1.3 Ebene Verhalten	23
2.4.1.4 Ebene Resultate	26
2.4.1.5 Diskussion	27
2.4.2 Bedingungen des Transfers von Trainingsinhalten	30
III Empirischer Teil	32
1. Fragestellung und Anlage der Studie	32
1.1 Ausgangslage	32
1.2 Forschungsfrage der Arbeit	35
2. Methoden	37
2.1 Die Trainingsmaßnahme	37
2.1.1 Projekthintergrund und Zielstellung	37
2.1.2 Aufbau und Gestaltung	38
2.1.3 Begründung der Auswahl des Evaluationsgegenstandes	40

2.2 Entwicklung eines <i>ganzheitlichen</i> Evaluationsansatzes	41
2.2.1 Evaluationsmodell	41
2.2.2 Erhebungsmethoden	44
2.2.2.1 Ebene Reaktion	44
2.2.2.2 Ebene Lernen	45
2.2.2.3 Ebene Verhalten	48
2.2.2.4 Ebene Resultate	50
2.2.2.5 Transferbedingungen	51
2.2.3 Evaluationsdesign	53
2.3 Untersuchungsfragen und Hypothesen	58
2.3.1 Ebene Reaktion	58
2.3.1.1 Überprüfung der Wirksamkeit	58
2.3.1.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	58
2.3.2 Ebene Lernen	58
2.3.2.1 Überprüfung der Wirksamkeit	58
2.3.2.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	59
2.3.3 Ebene Verhalten	60
2.3.3.1 Überprüfung der Wirksamkeit	60
2.3.3.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	60
2.3.4 Ebene Resultate	61
2.3.5 Transferbedingungen	61
2.3.5.1 Überprüfung der Wirksamkeit	61
2.3.5.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	62
2.4 Beschreibung der Stichproben	63
2.5 Auswertungsstrategie	63
3. Ergebnisse der Untersuchung	67
3.1 Überprüfung der Voraussetzung zur Anwendung statistischer Verfahren	67
3.2 Ebene Reaktion	68
3.2.1 Überprüfung der Wirksamkeit	68
3.2.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	71
3.3 Ebene Lernen	73
3.3.1 Überprüfung der Wirksamkeit	73
3.3.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	75
3.4 Ebene Verhalten	77

3.4.1 Überprüfung der Wirksamkeit	77
3.4.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	78
3.5 Ebene Resultate	79
3.6 Transferbedingungen	81
3.6.1 Überprüfung der Wirksamkeit	81
3.6.2 Überprüfung von Indikatoren für die Eignung des Ansatzes	86
4. Diskussion	87
4.1 Die Eignung des Ansatzes im Lichte der Befunde	87
4.1.1 Ebene Reaktion	87
4.1.2 Ebene Lernen	89
4.1.3 Ebene Verhalten	91
4.1.4 Transferbedingungen	92
4.1.5 Gesamtmodell	93
4.2. Implikationen für die Forschung	94
4.3. Implikationen für die Trainingspraxis	96
4.3.1 Evaluationsmodell	97
4.3.2 Erhebungsmethoden	97
4.3.3 Evaluationsdesign	98
4.4 Methodische Einschränkungen der Studie	99
4.5 Fazit und Ausblick	99
Literatur	101
Eidesstattliche Erklärung	110
Anhang	111

## **Abbildungs- und Tabellenverzeichnis**

### **Abbildungen**

Abbildung 1: Schematische Veranschaulichung der Transferarten (Bergmann, 1999)	24
Abbildung 2: Ein Modell zur Beschreibung des Transferprozesses (nach Baldwin & Ford, 1988)	31
Abbildung 3: Konzeptuelles Modell des GLTSI (Kauffeld, 2008)	51
Abbildung 4: Mittelwerte und Standardabweichungen der Gesamtzufriedenheit Module 1-10	68

### **Tabellen**

Tabelle 1: Erhebungs- und Versuchsplan	57
Tabelle 2: Ergebnisse der qualitativen Analyse der offenen Kommentare des Zufriedenheitsfragebogens	70
Tabelle 3: Reliabilitäten der Skalen des Zufriedenheitsfragebogens zu den 10 Messzeitpunkten	72
Tabelle 4: Ursprünglicher Rücklauf der Daten auf der Ebene Reaktion	72
Tabelle 5: Ergebnisse der qualitativen Auswertung von Interviewfrage 1 „Maßnahmen der Innovationspromotion“	80
Tabelle 6: Ergebnisse der qualitativen Auswertung von Interviewfrage 2 „Kennzahlen der Innovationspromotion“	80
Tabelle 7: Ergebnisse der deskriptiven Analysen für die Skalen des GLTSI	81
Tabelle 8: Deskriptive Analyse des Teilnehmerkreises (in Anlehnung an Rudinger, 2012)	112



# I Einleitung

## 1. Entdeckungszusammenhang und Problemstellung

Bei der Evaluation von Trainingsmaßnahmen hat man es immer mit den Mühen und Mängeln einer Feldstudie zu tun: Man trifft bspw. auf *kleine und schwindende Stichproben, Ängste und Widerstände bei Teilnehmern, geringe Rücklaufquoten, begrenzte Möglichkeiten zur Kontrolle äußerer und forschungsmethodischer Störeinflüsse*.

Es gilt also Wege zu finden, mit den Unwägbarkeiten und spezifischen Bedingungen in der Trainingspraxis umzugehen und gleichzeitig wissenschaftlichen Anforderungen zu genügen. Lange Zeit wurde die unzureichende Wissenschaftlichkeit von Trainingsevaluationen bemängelt: „By and large, the training and development literature is voluminous, nonempirical, non-theoretical, poorly written and dull“, kritisiert Campbell (1971) den Stand der Forschung im Bereich von Trainingsmaßnahmen vor gut 40 Jahren (p. 565). Die Trainingsevaluation diente zu dieser Zeit vorrangig als Mittel zum Zweck, um den Erfolg eines neu entwickelten Trainingsansatzes zu belegen. Auf Grundlage der Ergebnisse wurde der Ansatz modifiziert und kam dann so lange zur Anwendung bis ein Neuer erschien. Der Kreislauf von Neukonzeption, Evaluation, Modifikation und Verbreitung begann erneut (Merzenich-Hieker, 1996).

Seither gibt es Bemühungen die Trainingsevaluation auf ein wissenschaftlich akzeptables Niveau zu heben. Nicht nur aus wissenschaftlicher Perspektive, sondern vor allem auch aus Sicht der Unternehmen ist dies von erheblichem Interesse:

*„Organisations invest considerable resources and effort into training their workforce. It is reasonable to pose questions such as:[...]What is the value of training to the 'bottom line'? How can organizations know that training is effective?“ (Salas, Cannon-Bowers & Kozlowski, 1997, p. 357).*

Laut der aktuellen Weiterbildungserhebung des Institutes der deutschen Wirtschaft Köln investieren deutsche Unternehmen jährlich rund 28,6 Milliarden Euro in Weiterbildungsmaßnahmen (Lenske & Werner, 2009). Steigende Arbeitsanforderungen und komplexer werdende Prozesse machen dabei vor allem Trainingsformate notwendig, die die Ausbildung überfachlicher Qualifikationen und Kompetenzen fördern (Hochholdinger, Rowold & Schaper, 2008a). Die Hoffnung wird zunehmend in **Kommunikations- und Verhaltenstrainings** gesetzt, die als arbeitsnahe, selbstgesteuerte und kollektive Lernform, Verhaltensorientierung und Übungseffekt statt Wissensvermittlung in den Vordergrund stellen (Schaper & Sonntag, 2007).

Die empirische Überprüfung der sogenannten *weichen* Kriterien des Erfolges - wie Einstellungs- und Verhaltensänderungen oder die Verbesserung der Kommunikationsfähigkeit - erhöht die bereits angeführten Anforderungen einer Trainingsevaluation noch einmal im Besonderen für die verhaltensorientierten Formate: *Wie lassen sich die vielschichtigen Wirkungen des Trainings präzise operationalisieren? Welche Methoden eignen sich zur Erhebung möglicher Erfolgskriterien? Welche Evaluationsdesigns lassen die eindeutige Interpretation der Ergebnisse zu?*

Diese und andere ungeklärte Fragen der Evaluationsforschung führen dazu, dass sich neben der Hoffnung vor allem auch Zweifel über die Wirksamkeit von Trainings einstellen: In seiner Publikation „Die Weiterbildungslüge“ erklärt der ehemals selbst in der Weiterbildungsbranche tätige Trainer Richard Gris (Pseudonym) Trainings für teuer, aber wirkungslos (Gris, 2008). Seine Streitschrift rangiert monatelang in den Bestsellerlisten der Wirtschaftsliteratur und findet Eingang in die Medien: Ist Weiterbildung ein „Erlebnis ohne Ergebnis“ fragt die FAZ (2009) und ruft den großen „Trainingsschwindel“ aus (FAZ, 2008). Zu einem ähnlich ernüchternden Urteil kommt DIE WELT schon vier Jahre früher: „Berufliche Weiterbildung in Deutschland krankt an der Umsetzung“ (DIE WELT, 2004).

Der Bedarf nach systematischer, empirisch gestützter Evaluation von Trainings, die Belege erbringt, dass der Nutzen der Maßnahmen die immensen Kosten auch tatsächlich rechtfertigt, ist also größer denn je (Seidel, 2012; Hochholdinger u.a., 2008b; Kauffeld, 2010). Während Untersuchungen zu Trainingsevaluationen lange Zeit eher rar waren, gibt es im Bereich der Arbeits- und Organisationspsychologie Bemühungen die Forschung zur Trainingsevaluation voranzutreiben.

In der ***Trainings- und Transferforschung*** werden dazu einerseits allgemeine methodische Strategien aus der Evaluation in anderen Disziplinen auf den Trainingskontext übertragen (Hochholdinger u.a., 2008b). Andererseits werden inhaltliche Modelle und spezifische methodische Strategien zur Evaluation von Trainings (weiter-)entwickelt (ebd.). Modelle der Trainingsevaluation sind also nicht mehr nur Mittel zum Zweck, sondern stehen selbst im Mittelpunkt von Forschung. Anhand von exemplarischen Trainingsevaluationen erfolgt die Überprüfung der theoretischen Annahmen. Die Erkenntnisse sollen Aufschluss über die Wirkungen und Einflussvariablen im Trainingsprozess bringen (z.B. Weissweiler, 2008; Scharpf, 1999; Karg, 2006) und der Weiterentwicklung von Evaluationsmodellen dienen (z.B. Rowold, 2008).

Die Überprüfung der Wirksamkeit einzelner Trainings ist dabei lediglich ein Nebenprodukt der Forschung. **Die Frage, wie sich die Erkenntnisse in ein praktikables Konzept zur Trainingsevaluation gießen lassen, steht nicht im Fokus.** Evaluationsstudien, die demgegenüber die Überprüfung der Wirksamkeit eines konkreten Trainingsansatzes verfolgen, legen zwar auch zunehmend aufwendige Designs und standardisierte Verfahren zugrunde. **In der Diskussion der Ergebnisse interessiert hierbei aber letztlich der Erfolg und Nutzen des überprüften Trainings und weniger der Erfolg und Nutzen des Evaluationskonzeptes** (z.B. Saborowski & Muellerbuchhof, 2010; Lang & Otto, 2010).

So kann festgehalten werden, dass sich der Anspruch an Wissenschaftlichkeit für Trainingsevaluationen zwar zunehmend durchsetzt. So lange die Ergebnisse aber nicht in Evaluationskonzepte transferiert werden, die auch in der Unternehmens- bzw. Trainingspraxis praktikabel und ökonomisch einsetzbar sind, bleiben die erfreulichen Entwicklungen allerdings auf den Elfenbeinturm der Forschung begrenzt.

## **2. Zielsetzung und Anlage der Arbeit**

Die vorliegende Arbeit nimmt den aufgedeckten Bedarf zum Ansatzpunkt und zielt auf die Entwicklung eines Evaluationsansatzes, der die Überprüfung von Trainingswirksamkeit ermöglichen soll. Im Fokus steht dabei die Auseinandersetzung mit folgender Forschungsfrage:

*Wie kann ein ganzheitlicher Evaluationsansatz zur Überprüfung von Trainingswirksamkeit aussehen, der eine ausgewogene Balance zwischen wissenschaftlichem Anspruch und praktikablen Anforderungen im Feld herstellt?*

In einem ersten Schritt erfolgt die **konzeptionelle Auseinandersetzung** mit der Fragestellung. Auf Basis aktueller Befunde aus der Trainings- und Transferforschung wird ein Evaluationsansatz entwickelt, der beiden Anspruchsgruppen gerecht werden soll. In dem Rahmen erfolgen Überlegungen zu einem geeigneten Evaluationsmodell, adäquater Evaluationskriterien und Erhebungsmethoden. Überdies müssen Designfragen geklärt werden.

In einem zweiten Schritt soll der Nutzen des Ansatzes **empirisch überprüft** werden. Dazu wird der Ansatz in Bezug auf eine ausgewählte Trainingsmaßnahme konkretisiert und dient einmal exemplarisch der Überprüfung der Wirksamkeit.

Die *Ausbildung zum Innovationspromotor* liefert die empirische Plattform für die Untersuchung. Das Ziel der empirischen Untersuchung besteht zwar darin, die Wirksamkeit der Ausbildung nachzuweisen. In Frage steht aber nicht der Erfolg der Ausbildung, sondern der Erfolg und Nutzen des entwickelten Evaluationsansatzes.

Eine umfassende empirische Modellüberprüfung ist im vorliegenden Untersuchungsrahmen allerdings nicht möglich. In der empirischen Studie sollen aber systematisch Anhaltspunkte dafür gewonnen werden, inwieweit der entwickelte Evaluationsansatz einerseits *wissenschaftlichen Ansprüchen genügt* und sich andererseits *unter Feldbedingungen als praktikabel erweist*.

Daran anknüpfend werden Empfehlungen und Konsequenzen für die Evaluationspraxis im Weiterbildungsbereich abgeleitet und darüber hinaus Möglichkeiten und Grenzen der Evaluation von Trainingswirksamkeit diskutiert.

### 3. Aufbau der Arbeit

Die Grundlage für die Auseinandersetzung mit dem Feld der Trainingsevaluation bildet im Theorieteil der Arbeit (**Teil II**) die Annäherung an den Gegenstandsbereich *Training* (**Kap. 1**). Es wird erörtert, was unter *Training* in dieser Arbeit verstanden wird und welche Gestaltungsvarianten im Mittelpunkt stehen. Darauf aufbauend werden die Besonderheiten bei modular aufgebauten Trainingsprogrammen zur Ausbildung professioneller Rollen thematisiert. In diesem Rahmen lässt sich die Ausbildung zum Innovationspromotor als späterer Evaluationsgegenstand sinnvoll verorten.

Im Abschnitt *Evaluation von Trainings* (**Kap. 2**) geht es darum, für die Besonderheiten des Forschungsfeldes zu sensibilisieren und Methoden der Wirksamkeitsüberprüfung von Trainings zu erläutern. Insbesondere bei der Wahl eines geeigneten Evaluationsdesigns eröffnet sich das Spannungsfeld zwischen wissenschaftlichen Ideallösungen und dem, was in der Praxis umsetzbar ist. Die Auseinandersetzung mit Designfragen nimmt daher einen besonderen Stellenwert im Theorieteil ein.

Anknüpfend an die allgemeinen Evaluationsmethoden werden spezifische Konzepte zur Evaluation von Trainings und der Überprüfung ihrer Wirksamkeit dargestellt. Bei der Operationalisierung und Erhebung von Variablen zur Wirksamkeitsüberprüfung wird die Kluft zwischen wissenschaftlich anzustrebenden und in der Praxis umsetzbaren Lösungen erneut akut. Auf der Auseinandersetzung mit dieser Thematik liegt daher ebenfalls ein Schwerpunkt der vorliegenden Arbeit.

Den empirischen Teil der Arbeit (**Teil III**) eröffnet das Kapitel zur übergeordneten *Fragestellung und Anlage der Studie* (**Kap. 1**). Im *Methodenteil* der Arbeit (**Kap. 2**) steht die Entwicklung eines ganzheitlichen Evaluationsansatzes zur Überprüfung von Trainingswirksamkeit im Fokus, der eine Balance zwischen wissenschaftlichem Anspruch und praktikablen und ökonomischen Anforderungen im Feld herstellt. Dazu werden in einem ersten Schritt konzeptionelle Überlegungen zum Evaluationsmodell, möglicher Erhebungsmethoden und einem adäquaten Evaluationsdesign vorgenommen, die in einem zweiten Schritt in Bezug auf das Training der exemplarischen Evaluation konkretisiert werden. Abschließend werden spezifische Untersuchungsfragen und Hypothesen für die Evaluationsstudie abgeleitet. Diese werden im *Ergebnisteil* (**Kap. 3**) überprüft und anschließend im Hinblick auf die übergeordnete Forschungsfrage der Arbeit *diskutiert* (**Kap. 4**).

## **II Theorie**

### **1. Die Praxis von Trainings als Gegenstand von Forschung**

In diesem Abschnitt der Arbeit soll es darum gehen, den Gegenstandsbereich für die spätere Evaluation zu beleuchten. Zunächst gilt es die Merkmale und Besonderheiten von Kommunikations- und Verhaltenstrainings und darauf aufbauend von umfassenden Trainingsprogrammen herauszuarbeiten. Damit soll eine erste Idee davon vermittelt werden, welche Unwägbarkeiten und Bedingungen sich bei der Evaluation von Trainings(-programmen) ergeben. Im anschließenden Kapitel zur Trainingsevaluation wird diese Auseinandersetzung vertieft.

#### **1.1 Kommunikations- und Verhaltenstrainings: Lernfelder und -methoden**

Die Lernform, um die es im Rahmen dieser Arbeit gehen soll, wird als sogenanntes Kommunikations- und Verhaltenstraining von dem Fachtraining abgegrenzt. Im Unterschied zu den fachlich orientierten Formaten, die lange Zeit den Trainingsmarkt dominierten, zielen Kommunikations- und Verhaltenstrainings stärker auf die Förderung überfachlicher Qualifikationen und Kompetenzen (Hochholdinger u.a., 2008a). Statt der Vermittlung fachlicher Inhalte stehen der Aufbau und/oder die Verbesserung längerfristig verfügbarer Verhaltensdispositionen im Sinne von Fertigkeiten, Fähigkeiten und Einstellungen im Fokus (Hager & Hasselhorn, 2000). Damit sollen Kommunikations- und Verhaltenstrainings auf einen Qualifizierungsbedarf antworten, der sich in Folge einer sich wandelnden Arbeitswelt ergibt: *Globalisierungstrends, eine dramatische Beschleunigung von Innovationsprozessen, sich verändernden Organisationsstrukturen sowie weiterentwickelte und neue Techniken* und weitere veränderte Außenfaktoren der Wirtschaft, spiegeln sich in steigenden und komplexer werdenden Arbeitsanforderungen für die Mitarbeiter (Meier, 2007).

Um den aufgemachten Bedarf der Ausbildung überfachlicher Qualifikationen und Kompetenzen zu erfüllen, werden im Rahmen von kommunikations- und verhaltensbezogenen Trainings spezifische Lehr- und Lernmethoden notwendig, die sich von der klassischerweise deduktiven Ausrichtung von Fachtrainings klar unterscheiden. Für Kommunikations- und Verhaltenstrainings ist eine handlungs- und teilnehmerorientierte Gestaltung kennzeichnend. Phasen, in denen theoretischer Input vermittelt und/oder interaktiv mit den Teilnehmern erarbeitet wird, wechseln sich mit Phasen des aktiven Übens ab (Bachmann, Runkel & Scholl, 2010).

Dahinter steht das Ziel, die erlernten Inhalte immer auch praktisch erlebbar zu machen und das eigene Verhalten und Erleben in unterschiedliche Situationen zu erfahren (ebd.). An solche Übungsphasen knüpfen in der Regel Phasen der Nachreflexion an, in denen bestimmte Aspekte noch einmal diskutiert und weiterentwickelt werden. Als klassische Trainingsmethoden können neben Rollen- und Planspielen, Gruppendiskussionen, Fallstudien sowie Übungen zur Interaktion und Kooperation, zur Wahrnehmung oder zur Selbstreflexion angeführt werden (Malmendier, 2003).

In Kommunikations- und Verhaltenstrainings wird die Trainingsgruppe selbst als Lernfeld betrachtet. Ein Wechsel zwischen Großgruppenformaten und der Arbeit in Kleingruppen soll eine optimale Nutzung dieser Ressource ermöglichen (Bachmann u.a., 2010). Es werden gruppenspezifische Prozesse in Gang gesetzt, die **soziales Lernen** ermöglichen sollen (*dazu zählen bspw. Lernmechanismen wie das Lernen durch Feedback, Lernen am Modell, Lernen durch Reflexion*).

Übliche Themen im Bereich der Kommunikations- und Verhaltenstrainings sind *Führung, Verkauf, Präsentation, Moderation, Umgang mit Konflikten, Selbst- und Zeitmanagement, Kommunikation etc.*<sup>1</sup> (Graf, 2010).

## 1.2. Trainingsprogramme zur Ausbildung professioneller Rollen

Daneben etablieren sich zunehmend umfassendere Trainingsprogramme am Markt, die über die übliche Zielstellung von Kommunikationstrainings noch hinausgehen. Statt der Förderung einzelner Kompetenzbereiche und Qualifikationsfelder geht es um die Ausbildung **professioneller Rollen**. Dazu zählen bspw. Trainer, Coaches, und Prozessberater, all jene Beraterrollen, auf die Unternehmen in Zeiten dynamischer Veränderungsprozesse und komplexer werdender Anforderungen zunehmend angewiesen sind. Die **Ausbildung zum Innovationspromotor** als Evaluationsgegenstand im Rahmen der empirischen Studie zielt auf die Ausbildung einer professionellen Beraterrolle für Unternehmen und lässt sich daher in diesem Kontext verorten. Die ausführliche Vorstellung der Ausbildung erfolgt im Methodenteil der Arbeit.

Die Trainingsprogramme umfassen mehrere Module, die als Kommunikations- und Verhaltenstrainings gestaltet sind und in der Regel in einer festen Ausbildungsgruppe durchgeführt werden (*die übliche Gruppengröße liegt bei 15 Teilnehmern*).

---

<sup>1</sup> Zur Abgrenzung lassen sich als klassische Themen aus dem Bereich der Fachtrainings wie *Anwendung und Programmierung von Softwaretools, Telekommunikation und Internet, Betriebswirtschaft und Controlling, Marketing und Vertrieb, Personalwirtschaft, Sprachen etc.* nennen (Graf, 2010).

Die Ausführungen zu den Lehr- und Lernmethoden können daher auch für die Beraterausbildungen gelten. Die Lern- und Entwicklungsfelder im Rahmen der Ausbildung von Beraterrollen gehen über die Inhalte der bisher thematisierten Trainingsformate noch hinaus und beziehen sich laut Bachmann und Kollegen (2010)

*„(...) auf die Veränderung von Einstellungen und Verhalten von Menschen zur Ausgestaltung professioneller Rollen, auf die Steuerung von Kommunikations- und Interaktionsprozessen in Gruppen, Teams und Organisationen sowie auf grundlegende Prozesse der menschlichen Wahrnehmung, des Lernens und der Entwicklung“ (S. 210).*

Die Autoren unterscheiden weiter fünf Kernkompetenzen<sup>2</sup>, die im Rahmen der Trainingsprogramme gezielt ausgebaut werden sollen:

Die **Fachkompetenz** umfasst das Wissen, welches für ein Verstehen und Begreifen organisationaler Zusammenhänge und Prozesse von Bedeutung ist. Dabei können Modelle, Theorien, Ansätze und Fakten aus den Disziplinen der Psychologie, der Betriebswirtschaft und der Organisationssoziologie zur Anwendung kommen (ebd.). Die sogenannte Feldkompetenz wird in der Literatur üblicherweise der Fachkompetenz zugeordnet (z.B. Heyse & Erpenbeck, 2004).

Für Berater, die als *Externe* für Organisationen arbeiten, bietet es sich allerdings an, die **Feldkompetenz** im Sinne eines spezifischen Wissens über Herkunft und Branche des Klientensystems separat aufzuführen (Bachmann u.a., 2010).

Zentral für Berater ist zudem die Entwicklung einer umfassenden **Methodenkompetenz**. Diese umfasst neben der Kenntnis verschiedener Methoden vor allem das Vermögen, diese auch zielgruppen- und situationsadäquat einsetzen zu können (ebd.).

Das Lernfeld der **Sozialkompetenz** umfasst vor allem die Klarheit über die eigene Rolle und das Vermögen diese bewusst auszugestalten. Überdies werden Phänomene aufgezählt, die typischerweise unter *sozialer Kompetenz* verstanden werden: *Perspektivübernahme, Kommunikationskompetenz, Beziehungsmanagement u.a.* (ebd.).

Die **persönliche Kompetenz** umfasst schließlich individuelle Fähigkeiten und Persönlichkeitseigenschaften, welche für die Ausübung der jeweiligen Beraterrolle relevant sind. Dazu zählen *intellektuelle Fähigkeiten, Selbstbewusstsein, die Fähigkeit zur Selbstreflexion und zum Selbstmanagement u.a.*

---

<sup>2</sup> Das Konstrukt der *Kompetenz* ist in der wissenschaftlichen Auseinandersetzung umstritten und es existiert eine Vielzahl an Forschungsansätzen und divergierenden Begriffsverständnissen zu dem Feld. Im Rahmen der vorliegenden Untersuchung kann eine differenzierte Betrachtung des Diskurses nicht erfolgen. Zur kritischen Diskussion des Kompetenzbegriffes siehe z.B. Bethscheider, Höhns & Münchhausen (2010); Edelmann & Tippelt (2004).



Aufgrund der komplexen Zielstellungen und vielfältigen Wirkungen ergeben sich bei der Evaluation von kommunikations- und verhaltensorientierten Trainingsprogrammen verschiedene Herausforderungen.

Bachmann, Runkel und Scholl (2010) postulieren, dass eine empirische Prüfung des Erfolges für beschriebene Trainingsprogramme kaum zufriedenstellend möglich sei. Dafür wird das Zusammenwirken zahlreicher persönlicher und beruflicher Einflussfaktoren und Ziele als zu vielschichtig und komplex betrachtet. Lediglich die Abfrage der Zufriedenheit der Teilnehmer mit dem Training wird als mögliche Evaluationsform für Trainingsprogramme vorgeschlagen.

In der nachfolgenden Auseinandersetzung mit dem Themenfeld der Trainingsevaluation sollen dennoch Ideen entwickelt werden, welche Wege und Möglichkeiten es geben kann, die Wirksamkeit von verhaltensorientierten Trainings(-programmen) systematisch zu prüfen.

## 2. Evaluation von Trainings

In diesem Kapitel der Arbeit wird das Themenfeld der Trainingsevaluation sowohl theoriegeleitet als auch anwendungsorientiert beschrieben und diskutiert. Dabei werden einerseits Methoden und Ansätze zum allgemeinen Evaluationsvorgehen erläutert, die auch in der Trainingsevaluation Anwendung finden. Andererseits kann in der Trainingsevaluation zunehmend auf eigene Modelle und methodische Strategien zurückgegriffen werden, die ebenfalls ausführlich beleuchtet werden.

### 2.1 (Trainings-)Evaluation: Eine begriffliche Annäherung

Evaluation spielt in verschiedenen gesellschaftlichen Zusammenhängen - im Bildungs-, Sozial-, oder Gesundheitswesen, der Wirtschaft oder in politischen Kontexten - eine Rolle und ist damit auch ganz automatisch Gegenstand interdisziplinärer Anwendungsforschung. Je nach Anwendungsfeld werden mit dem Terminus Maßnahmen der Erfolgs-, Wirkungs-, oder Qualitätskontrolle, der Effizienzmessung oder der Kosten-Nutzen-Analyse etc. assoziiert, mit denen wiederum eine Fülle an mehr oder minder systematischen Vorgehensweisen und Methoden verknüpft ist (Gollwitzer & Jäger, 2009). In den einschlägigen Publikationen wird der inflationäre Gebrauch des Begriffes bemängelt und konstatiert, dass dieser eine eindeutige Definition erschwere (Gollwitzer & Jäger, 2009; Balzer, 2005; Wottowa & Thierau, 2003). Die nähere Betrachtung zeigt, dass sich trotz divergierender Facetten aber eine breite Übereinstimmung im grundsätzlichen Begriffsverständnis durchsetzen konnte.

Ausgehend von der engeren Wortbedeutung wird *Evaluation* von lat. *valere* = *bewerten* als der Prozess der Bewertung und Beurteilung des Wertes einer Sache, bzw. von Produkten, Programmen oder Prozessen definiert (Gollwitzer & Jäger, 2009; Wottowa & Thierau, 2003; Weissweiler, 2008; Hanke, 2005, Suchmann, 1967). In den Vordergrund wird häufig die Evaluation von Programmen gestellt und daher in einigen Publikationen der Begriff *Programmevaluation* genutzt, um den Gegenstandsbe- reich explizit mitauszuweisen (Kuper, 2011; Rossi, Freeman & Hofmann, 1988).

Andere Autoren bemühen sich darum, die Unterscheidung zwischen einem alltäglichen und einem wissenschaftlichen Gebrauch des Begriffes *Evaluation* zu markieren und ein datengestütztes Vorgehen mit Hilfe systematischer Verfahren ausschließlich der *Evaluationsforschung* zuzuschreiben oder dafür den Zusatz *wissenschaftlich gestützte* Evaluation zu verwenden (Gollwitzer & Jäger, 2009; Wottowa & Thierau,

2003; Suchmann, 1967). Die beschriebenen begrifflichen Unterscheidungen sind jedoch nicht Konsens und es überwiegt eine synonyme Verwendung der Termini, wie sie auch im Rahmen dieser Arbeit gelten soll: „(...) we use the terms evaluation, program evaluation, and evaluation research interchangeably (Rossi, Lipsey & Freeman, 2004, p. 6).

Die wissenschaftliche Orientierung und damit Anwendung sozialwissenschaftlicher Forschungsmethoden stellt dabei ein zentrales Bestimmungsstück von Evaluationen dar (Solga, 2011a). Es hat sich durchgesetzt, weitere allgemeine Kennzeichen von Evaluationen zu definieren, statt eine einheitliche Evaluationsformel festlegen zu wollen (Rossi u.a., 1988; Wottawa & Thierau, 2003; Stockmann, 2007; Thierau-Brunner, Wottowa & Stangel-Meseke, 2006):

Evaluation wird als ziel- und zweckorientierter Prozess zur Beurteilung von Konzeption, Ausgestaltung, Umsetzung und Nutzen sozialer Interventionsprogramme beschrieben. Die Evaluation kann der Überprüfung und/oder Verbesserung praktischer Maßnahmen und als Planungs- und Entscheidungshilfe über deren Fortbestand dienen. Mit der Kontroll-, Optimierungs-, und Entscheidungsfunktion sind bereits drei zentrale Funktionen von Evaluationen angesprochen worden. Daneben kann eine Evaluation die Funktion haben, wissenschaftliche Erkenntnisse über die Eigenschaften und Wirkungen von Interventionen zu sammeln (*Erkenntnisfunktion*) oder die Entwicklung und Durchführung einer Intervention nach außen zu legitimieren (*Legitimationsfunktion*).

Die zentralen Elemente der allgemeinen Definitionsbemühungen werden in die Trainingsevaluation übertragen. Im Fokus stehen fachliche und überfachliche Trainings, aber auch weitere Maßnahmen der Personalentwicklung, die systematisch *bewertet* und *beurteilt* werden (Hochholdinger u.a., 2008b)<sup>3</sup>. Dabei können das Trainingskonzept, das Trainingsdesign, die Umsetzung des Trainings oder die Wirksamkeit näher betrachtet werden. Es wird bspw. untersucht, ob Inhalt, Organisation, Ablauf, Räumlichkeiten, Trainer, Materialien etc. für Lernen und die Umsetzung des Gelernten in der Praxis förderlich waren und die Identifikation von Stärken und Schwächen des Trainings vorgenommen (Hanke, 2005).

---

<sup>3</sup> Daher wird das Themenfeld der Trainingsevaluation auch häufig unter dem Thema *Evaluation von Personalentwicklungsmaßnahmen* mit abgehandelt (z.B. Solga, 2011a).

Die Beurteilungskriterien lassen sich in Aufwand, Effektivität und Angemessenheit gliedern und werden mit wissenschaftlichen Verfahren der Datenerhebung und -auswertung erfasst (Bortz & Döring, 2006). Auch im Rahmen von Trainingsevaluationen bilden Überlegungen zur strategischen Zielsetzung und Funktion des Vorhabens den Ausgangspunkt der Planung. Je nachdem, ob eine Evaluationsstudie bspw. die Bewertung eines Trainings ermöglichen oder als Entscheidungsgrundlage über dessen Fortbestand dienen soll, werden andere Evaluationsstrategien, -modelle, -designs und -methoden gewählt. Nachfolgend wird es zunächst darum gehen, in Anlehnung an die beschriebenen Funktionen, adäquate Strategien für Evaluationsvorhaben zu diskutieren.

## **2.2 Strategien der Evaluation**

Die bekannteste Unterscheidung von Evaluationsstrategien, die auch für die Trainingsevaluation Relevanz besitzt, ist die der *formativen* und *summativen* Evaluation (Scriven, 1967, 1991). Im Rahmen einer formativen Evaluation wird eine Maßnahme begleitend zur Durchführung bewertet. Damit fließen Intervention und Evaluation gewissermaßen zusammen und es ist erlaubt, Anpassungen im laufenden Prozess vorzunehmen (Solga, 2011a).

In der summativen Evaluation werden hingegen keine Eingriffe in laufende Prozesse vorgenommen. Die Wirksamkeit und der Erfolg einer Maßnahme werden ausschließlich im Anschluss an die Durchführungsphase bewertet (Kontrollfunktion). Die Ergebnisse sollen als Entscheidungsgrundlage dienen und Aussagen zur Eignung der Maßnahme ermöglichen sowie Ansatzpunkte zur Weiterentwicklung liefern (Optimierungsfunktion) (Gollwitzer & Jäger, 2009).

In aktuellen Trainingsstudien überwiegt die summative Herangehensweise zur Überprüfung von Trainingswirksamkeit. Dabei steht die sorgfältige und systematische Anwendung von Messinstrumenten und -methoden im Vordergrund. Es werden experimentelle Designs mit quantitativen Erhebungsmethoden präferiert, die ein hypothesengeleitetes Vorgehen und Messungen des Lern- oder Transfererfolgs ermöglichen (Hochholdinger u.a., 2008b).

Nachfolgend werden die methodischen Aspekte der Wirksamkeitsanalyse weiterführend diskutiert: Dabei stehen Überlegungen zu Gütekriterien, möglichen Störvariablen und Kontrollmöglichkeiten über die Wahl von Evaluationsdesigns im Fokus.

### 2.3 Design- und Planungsfragen in der (Trainings-)Evaluation

Evaluationsvorhaben, die die Überprüfung der *Wirksamkeit* von Interventionen zum Ziel haben, gehen zunächst der Frage nach, ob die mit der Maßnahme intendierten Veränderungen (*Wirkungen*) tatsächlich eingetreten sind. Wirksamkeitsanalyse ist also immer *Veränderungsmessung*. Dabei ist in der Regel der Grad der Veränderung von Interesse (Hager & Hasselhorn, 2000). Ein prototypisches Vorgehen umfasst die Messung von Kriterien für Wirksamkeit zu zwei Messzeitpunkten (vor und nach der Intervention). Die Ergebnisse von Prä- und Postmessung können anschließend verglichen und die Differenz als Ausmaß und die Richtung der Veränderung interpretiert werden (*Vortest-Nachtest-Differenz*) (Solga, 2011a; Hager & Hasselhorn, 2000). Daran anknüpfend stellt sich die Frage nach der *Gültigkeit* der so registrierten Veränderungen bezogen auf zwei Aspekte: Lassen sich die gemessenen Veränderungen tatsächlich auf die Intervention und nicht etwa auf externe Einflüsse zurückführen (*interne Validität*) (Gollwitzer & Jäger, 2009)? Können die Ergebnisse über die spezifischen Untersuchungsbedingungen hinaus, also „(...) nicht nur für die untersuchten Personen, sondern für die Gesamtheit aller von der Maßnahme betroffenen Personen gelten (...)“ (Bortz & Döring 2006, S. 104)? Die zentrale Herausforderung einer Evaluation liegt in der Wahl eines Untersuchungsdesigns, welches einerseits die Anzahl plausibler Alternativerklärungen für die Ergebnisse minimiert und gleichzeitig eine hohe Natürlichkeit der Untersuchungsbedingungen und Repräsentativität der untersuchten Stichprobe gewährleistet (ebd.). Es ist daher ratsam im Rahmen einer Wirksamkeitsevaluation vorab zu überlegen, welche Alternativwirkungen zusätzlich zu den maßnahmenspezifischen Wirkungen auftreten können. Neben zwar nicht intendierten, dennoch durch die Maßnahme erreichten Neben- und Folgewirkungen, sind hierbei eine Reihe von *Störfaktoren* zu beachten. Störeinflüsse und maßnahmenspezifische Wirkungen (*Nettoeffekt*) zusammengenommen, werden auch als *Bruttowirkung* im Sinne der gesamten zu beobachtenden Veränderung bezeichnet (Rossi u.a., 1999). Von der beschriebenen Validität des Untersuchungsdesigns ist die *Validität der Messinstrumente* zu unterscheiden. Diese betrifft die Frage, inwieweit es gelingt, das Phänomen zu erfassen, welches erfasst werden soll (Patry & Perrez, 2000). Zudem müssen eingesetzte Messverfahren zwei weiteren sogenannten psychometrischen Gütekriterien genügen: So muss sichergestellt werden, dass die resultierenden Kennwerte unabhängig von irrelevanten Randbedingungen der Messung, wie dem Evaluator, der Auswertung und der Interpretation der Daten sind (*Objektivität*)

(Gollwitzer & Jäger, 2009). Die **Reliabilität** bezieht sich auf die Genauigkeit der Merkmalerfassung und zielt auf die Minimierung von Messfehlern. Liegen eine mangelnde instrumentelle Reliabilität oder eine mangelnde instrumentelle Validität vor, ist die interne Validität einer Untersuchung gefährdet (ebd.). Im Folgenden werden weitere mögliche Störfaktoren für Evaluationen dargestellt, um daran anknüpfend Kontrollmöglichkeiten in entsprechenden Untersuchungsdesigns zu diskutieren.

### 2.3.1 Störfaktoren

Im Rahmen von Evaluationsstudien sind prinzipiell eine Fülle an Störfaktoren denkbar, die sich anknüpfend an Rossi und Kollegen (2004) noch einmal in **störende Umgebungseinflüsse** (äußere konfundierende Faktoren) und **forschungsmethodische Störeinflüsse** (Designeffekte) untergliedern lassen. Nachfolgend werden die meist genannten Störfaktoren dieser beiden Gruppen erläutert, die zunächst eine Gefährdung der internen Validität einer Evaluation darstellen<sup>4</sup>: Unabhängig vom Untersuchungsgeschehen können sich die Versuchspersonen selbst verändern (*älter, erfahrener, weniger aufmerksam werden*) und in Folge dessen ihr Verhalten ändern (**Reifung**) (ebd.). Als weiterer störender Umgebungseinfluss sind **Zeiteinflüsse** zu nennen. So kann ein Ereignis, das zeitlich parallel zur Trainingsmaßnahme stattfindet, sich auf das Verhalten der Trainingsteilnehmer auswirken (Merzenich-Hieker, 1996). Bei Längsschnittstudien kann zudem der Ausfall von Personengruppen über die Zeit (**Mortalität**) auftreten (ebd.). Als möglicher **Designeffekt** muss bei Wirksamkeitsanalysen der Einfluss der Messung selbst beachtet werden, da bei jeder Testung mit Messwiederholung die Gefahr besteht, dass sich das Antwortverhalten aufgrund der vorherigen Messung ändert (**Testübung**) (Solga, 2011a). Als weiterer forschungsmethodischer Störfaktor muss die sogenannte **Regression zur Mitte** bedacht werden. Dabei kommt es vor, dass extreme Prätestwerte die Tendenz haben, sich bei einer wiederholten Messung zur Mitte der Merkmalsverteilung hin zu verändern (Bortz & Döring, 2006). Auch bei der Auswahl der Stichprobe können Störfaktoren zu Ergebnissen einer Evaluation führen, die nichts mit der Trainingsmaßnahme zu tun haben. So können Unterschiede in den Effekten durch personelle Unterschiede zwischen den Gruppen bedingt sein (**Selektion**). Insbesondere bei Designs, in denen die Untersuchungsteilnehmer nicht per Zufall ausgewählt werden, können Selektionseffekte

---

<sup>4</sup> Prinzipiell wirkt sich die Gefährdung der internen Validität indirekt auch immer auf die externe Validität aus. In der Literatur werden Störfaktoren dennoch häufig danach unterschieden, ob sie die interne oder die externe Validität einschränken. Allerdings finden sich hierbei unterschiedliche Zuordnungen (z. B. Solga, 2011; Bortz & Döring, 2006). Daher dient die Aufteilung lediglich als Orientierungshilfe.

eine Rolle spielen (ebd.). So haben sich bspw. nur die besonders motivierten Mitarbeiter für das Training angemeldet oder diejenigen, die nach Möglichkeiten suchen Abstand vom beruflichen Umfeld zu gewinnen. Die externe Validität kann von Störfaktoren in Frage gestellt werden, die dem Phänomen der **experimentellen Reaktivität** - das Verhalten der Teilnehmer resultiert aus den konkreten Bedingungen des Untersuchungssettings - zugeordnet werden (Solga, 2011a). Hierunter fällt bspw. der sogenannte **Hawthorne-Effekt**. Dieser besagt, dass die empirisch nachgewiesene Wirksamkeit einer Intervention nicht auf die Maßnahme selbst,

*„(...)sondern lediglich darauf zurückzuführen ist, dass mit den Personen überhaupt irgendwas durchgeführt wurde bzw. dass eine Messung stattgefunden hat. Allein die Tatsache, dass etwas passiert, erhöht die Motivation der Teilnehmer, gute Leistungen zu zeigen, positive Werte abzugeben (...) usw.“ (Gollwitzer & Jäger 2009, S. 148).*

Je nachdem inwieweit man mit dem zu erfassenden Merkmal eine starke normative Valenz messen will, kann es sein, dass die Befragten zu **sozial erwünschtem Antwortverhalten** neigen, um bspw. nicht als dumm oder faul zu gelten (Gollwitzer & Jäger, 2009). In Evaluationsvorhaben liegen oftmals sehr spezifische Bedingungen vor, was die Replikation der Ergebnisse erschwert. Zudem scheint die Sicherung der internen Validität Herausforderung genug zu sein, weshalb die externe Validität häufig vernachlässigt wird (Höft, 2006). Solga (2011) spricht sich explizit dafür aus, Fragen der externen Validität immer mitzubedenken. „Andernfalls werden später Programmerrträge erwartet, die die Maßnahme unter regulären Anwendungsbedingungen nicht leisten kann“ (S. 351).

### **2.3.2 Kontrollmöglichkeiten**

Um also beurteilen zu können, ob die Ergebnisse einer (Trainings-)Evaluation auf die Wirkungen der Maßnahme und nicht auf beschriebene andere Einflüsse zurückzuführen sind, gilt es Untersuchungsbedingungen zu konstruieren, die eine möglichst isolierte Betrachtung der Trainingseffekte zulassen (Thierau-Brunner u.a., 2006). Verschiedene Untersuchungsdesigns variieren dazu mit der Anzahl und der Art der Trainings- und Kontrollgruppen, der Anzahl der Messungen und mit Zeitintervallen, in denen die Kriterien gemessen werden (ebd.). Zunächst stellt sich die Frage, ob es möglich ist, eine **Kontrollgruppe** zur Trainingsgruppe zu konstruieren, die zur Trainingsgruppe in jeder Hinsicht (*Ausbildung, Alter, Geschlecht, Funktion, Hierarchieebene etc.*) äquivalent ist, aber nicht die gleiche bzw. keine Intervention erhält. In **experimentellen Designs** soll die Ähnlichkeit der Untersuchungsgruppen durch die

zufällige Zuteilung (**Randomisierung**) gewährleistet werden (Hochholdinger u.a., 2008b). Damit sollen optimale Voraussetzungen für kausale Schlussfolgerungen geschaffen werden. In experimentellen Designs, die zudem eine Prä- und Postmessung vorsehen, lassen sich Störeffekte kontrollieren, bei denen davon ausgegangen werden kann, dass sie in beiden Gruppen gleichermaßen auftreten. Erhält die Kontrollgruppe kein Treatment, können Umgebungseinflüsse wie Reifung und Zeiteinflüsse sowie Designeffekte wie Messungseffekte und statistische Regressionseffekte kontrolliert werden (ebd.). Durch die zufällige Zuweisung zu den Untersuchungsbedingungen sollen zudem Selektionseffekte minimiert werden.

*„Ein echter Effekt, der kausal auf das Training zurückzuführen ist, wäre empirisch nachgewiesen, wenn die Unterschiede zwischen Prä- und Post-Test in der Trainingsgruppe größer sind als in der nicht-trainierten Kontrollgruppe“ (Gollwitzer & Jäger, 2009, S. 169).*

Auch in der Kontrollgruppe können also bedingt durch externe Einflussfaktoren Haupteffekte auftreten. Im Fokus der Wirksamkeitsanalyse steht die **Differenz** zwischen den Gruppen, die als besagter Nettoeffekt interpretiert werden kann (Solga, 2011a). Wird zusätzlich (oder stattdessen) eine behandelte Kontrollgruppe eingesetzt, die eine Placebo-Intervention (*vorgeblich systematische Durchführung einer tatsächlich unwirksamen Maßnahme*) erhält, können zusätzlich Effekte der experimentellen Reaktivität wie die **soziale Erwünschtheit** und der **Hawthorne Effekt** kontrolliert werden (Gollwitzer & Jäger, 2009). Solga (2011a) betont, dass diesem Vorgehen jedoch sowohl ökonomische als auch ethische Grenzen gesetzt sind.

Wenngleich experimentelle Designs die größte Wirksamkeit zur Minimierung von Störfaktoren und damit in der Regel eine hohe interne Validität aufweisen, sind diese im Bereich der Trainingsevaluation nur schwer umzusetzen. Sie erfordern oft künstliche Arrangements, die schwer realisierbar sind und zudem die externe Validität einschränken können (Merzenich-Hieker, 1996). Im Bereich der Trainingsevaluation überwiegen **quasi-experimentelle Designs**, in denen die Teilnehmenden nicht zufällig den Untersuchungsbedingungen zugeordnet werden (Hochholdinger u.a., 2008b). Ergeben sich hierbei Gruppenunterschiede zwischen Untersuchungs- und Kontrollgruppe, können diese nicht eindeutig auf Wirkungen der Intervention zurückgeführt werden. Störfaktoren lassen sich in quasi-experimentellen Designs nur schwer kontrollieren und die interne Validität ist daher deutlich geringer als in experimentellen Designs (Bortz & Döring, 2006). Im Bereich der Trainingsevaluation finden sich daneben auch häufig sogenannte **Einzelgruppen-Designs**, in denen ganz auf eine



Kontrollgruppe verzichtet wird und die daher noch weniger Möglichkeiten bieten, Alternativerklärungen für Effekte auszuschließen (Merzenich-Hieker, 1996). Für die Kontrolle von Störfaktoren sind quasi-experimentelle Designs (*zu denen in dieser Arbeit in Anlehnung an Merzenich-Hieker (1996) auch Einzelgruppen-Designs gezählt werden*) stärker als experimentelle Designs auf die Variation der Anzahl der Messungen angewiesen. In der einfachsten Form, der **Pretest-Posttest-Untersuchung**, werden Messungen zu genau zwei Zeitpunkten, einmal vor und einmal nach dem Training vorgenommen. Die Differenz zwischen den Messzeitpunkten wird dann auf die Wirkung der Intervention zurückgeführt (Gollwitzer & Jäger, 2009):

*„[Die] Messung der Stärke, Verbreitung oder Häufigkeit bestimmter Merkmale bei Programmbeginn dient als Vergleichsgrundlage zur Bestimmung der nach der Implementation bestimmter Maßnahmen auftretenden Effekte“ (Rossi u.a. 1988, S. 220).*

Damit ist eine Prätestung in quasi-experimentellen Designs unabdingbar, um eine solide Vergleichsgrundlage zu generieren. Die sogenannten Längsschnittstudien (in Abgrenzung zu Querschnittuntersuchungen mit einmaliger Messung) lassen sich danach untergliedern, ob die Messung an immer der gleichen Stichprobe wiederholt wird (**Paneluntersuchung**), oder ob zu jedem Zeitpunkt andere Untersuchungspersonen befragt werden (**Trendstudie**) (Häder, 2006). Im Bereich der Trainingsevaluation bietet sich die Paneluntersuchung an, wenn es darum geht, die Wirksamkeit einer Maßnahme zu prüfen und entsprechend Veränderungen an den Trainingsteilnehmern festzustellen. Designs, die keine Voruntersuchung vorsehen und in denen die Ausgangssituation über retrospektive Fragen ermittelt werden soll (**Retrospektive Vorher-Nachher-Untersuchung**), sind hingegen extrem anfällig für Störeinflüsse (Merzenich-Hieker, 1996). Bortz und Döring (2006) sprechen sich stattdessen dafür aus, die Anzahl der Messzeitpunkte in Veränderungsmessungen noch weiter zu erhöhen. Generell gehen sie davon aus, dass sich der Einfluss eines wenig reliablen Messinstrumentes mit zunehmender Anzahl der Messzeitpunkte kompensieren lässt. In sogenannten **Zeitreihendesigns** werden an der Trainingsgruppe zu mehreren Zeitpunkten vor und nach der Intervention Messungen vorgenommen (ebd.). Dabei wird die Trainingsgruppe selbst als reflexive Kontrollgruppe genutzt: „Zeigen sich im Kurvenverlauf der Messzeitpunkte in enger Beziehung zum Training Sprünge oder Diskontinuitäten, so wird dies als Indiz für Veränderungen interpretiert“ (Merzenich-Hieker 1996, S. 31). Typische Störeinflüsse in Längsschnittdesigns, wie Testübung oder Reifung können dabei kontrolliert werden (Gollwitzer & Jäger, 2009).

Allerdings sind für die Umsetzung von Zeitreihendesigns sehr viele Messzeitpunkte erforderlich, was sich in der Trainingsevaluation kaum sinnvoll realisieren lässt. Aber schon die Erhöhung von zwei auf drei Messzeitpunkte kann laut Bortz und Döring (2006) einen enormen Reliabilitätszugewinn bedeuten. Sie berufen sich dabei auf die Studienergebnisse von Willett (1980) der berichtet, dass schon das Hinzufügen eines dritten Messzeitpunktes die Reliabilität um 250% und mehr erhöhen kann. Auch für die Verteilung von Messzeitpunkten werden unterschiedliche Variationen diskutiert, die allerdings vor allem in Designs mit zahlreichen Messzeitpunkten zum tragen kommen (vgl. hierzu ausführlich Bortz & Döring, 2006, S. 554). Im Bereich der Trainingsevaluation kann es zur Prüfung der zeitlichen Stabilität von Effekten insbesondere sinnvoll sein, mehrere Messungen in unterschiedlichen Zeitabständen nach der Intervention vorzunehmen (Solga, 2011a). Idealtypische Designs verbinden die verschiedenen Kontrollmöglichkeiten und setzen mehrere Messzeitpunkte in ausreichend großen Zeitabständen und mehrere randomisierten Kontrollgruppen ein.<sup>5</sup>

In der Evaluationsforschung kommen neben den wissenschaftlichen Ansprüchen auch Anforderungen und Bedingungen im Anwendungsfeld zu tragen, die möglichen Evaluationsdesigns oftmals enge Grenzen setzen.

## **2.4 Modelle der Trainings- und Transferevaluation**

Anknüpfend an die vorangehende Darstellung methodischer Herausforderungen der Wirksamkeitsanalyse, gilt es im nachfolgenden Teil der Arbeit stärker inhaltlich zu klären, welche Wirksamkeitskriterien in der Trainingsevaluation Berücksichtigung finden. Diese Kriterien können entweder prozess- oder ergebnisorientiert ausfallen und werden in einem Evaluationsmodell spezifiziert.

Darüber hinaus legt das Evaluationsmodell die Instrumente zur Erhebung dieser Kriterien sowie den forschungsmethodischen Aufbau und Ablauf der Untersuchung fest (Solga, 2011a). Im Bereich der Trainingsevaluation lassen sich Evaluationsmodelle danach systematisieren, inwieweit sie den Erfolg einer Maßnahme an Kriterien bzw. Ergebnisvariablen festmachen oder inwieweit sie den Prozess betrachten und stärker die Wirkung verschiedener Variablen auf den Trainings- und Transfererfolg benennen (Hochholdinger u.a., 2008b).

---

<sup>5</sup> Als idealtypischer Untersuchungsplan wird oftmals der Salomon-4-Gruppen-Plan diskutiert, in dem sich zwei Interventionsgruppen und zwei Kontrollgruppen gegenüber stehen. Es handelt sich weiter um ein Prätest-Posttest-Design, wobei mögliche Störeinflüsse des Posttests, wie die Wechselwirkung mit der Bedingung oder die Teilstützung kontrolliert werden: Nur zwei der vier Gruppen nehmen an einem Prätest teil (eine Gruppe, die eine Intervention erhält und eine Kontrollgruppe) (Gollwitzer & Jäger, 2009).

Nachfolgend werden zwei populäre Modelle der Trainingsevaluation diskutiert, die sich in ihrer Ausrichtung in beschriebener Weise unterscheiden. Im Rahmen der eigenen Evaluationsstudie sollen schließlich beide Perspektiven integriert werden. Der Fokus liegt aber auf der Betrachtung ergebnisbezogener Faktoren, sodass die Auseinandersetzung mit den Ebenen und Inhalten von Trainingsergebnissen einen weitaus größeren Stellenwert einnimmt als die Beschreibung der Transferbedingungen.

#### **2.4.1 Ebenen und Inhalte von Trainingsergebnissen**

Eines der bekanntesten Modelle zur Evaluation der Wirksamkeit von Trainingsprogrammen welches bereits in den 1960er Jahren entwickelt wurde, ist das *Vier-Ebenen-Modell* von Kirkpatrick (1959). Er unterscheidet darin vier Ebenen (*Reaktion, Lernen, Verhalten, Resultate*) auf denen Wirksamkeitsbelege erbracht werden sollen. Für jede Ebene werden spezifische Evaluationskriterien und Erhebungsmöglichkeiten vorgeschlagen, die nachfolgend differenziert betrachtet werden. Im ursprünglichen Modell stehen die Ebenen in einem hierarchisch-kausalen Abhängigkeitsverhältnis: Es wird angenommen, dass der Informationsgehalt je Ebene zunimmt und der Erfolg der jeweils vorgegangenen Ebene keine hinreichende, aber eine notwendige Bedingung für den Erfolg auf der nächst höheren Ebene darstellt. Dementsprechend wird auch von positiven Korrelationen zwischen den Ebenen ausgegangen (Kirkpatrick & Kirkpatrick, 2006).

Die beschriebenen Grundannahmen werden bei der Darstellung des Modells kritisch zu diskutieren sein. Dabei werden je Ebene auch die Befunde aus aktuellen Studien angeführt und in die Diskussion mit einbezogen. Eine übergreifende Diskussion des Modells im Gesamt bildet schließlich den Abschluss des Kapitels.

##### **2.4.1.1 Ebene Reaktion**

Nach Kirkpatrick (1967) ist *reaction* die erste Ebene, auf der er in seinem Evaluationsmodell eine Überprüfung der Wirksamkeit vorschlägt. Wörtlich abgeleitet geht es auf dieser Ebene um die Frage, wie die Teilnehmer eines Trainings auf dieses *reagieren* und damit um das subjektive Erleben der Teilnehmer (Solga, 2011a). Kirkpatrick selbst schlägt vor, die Effekte auf dieser Ebene mit der Zufriedenheit der Teilnehmer mit dem Training zu übersetzen (Kirkpatrick & Kirkpatrick, 2006). In der Evaluationspraxis findet sich häufig eine Untergliederung der Zufriedenheit nach der Zufriedenheit mit dem Trainer, den Inhalten, der Didaktik etc. (Kauffeld, 2010). Als Erhebungsinstrument für die Ebene Reaktion schlagen Kirkpatrick und Kirk-

patrick (2006) einen standardisierten Fragebogen vor, der eine Quantifizierung der Reaktionen ermöglicht. Um ein umfassendes Bild der Teilnehmerzufriedenheit zu bestimmen, sollte die Möglichkeit zu freien Kommentaren als Ergänzung gegeben werden. Die Autoren betonen in den Ausführungen zu dem Modell, dass die Reaktionsebene für die Beurteilung der Trainingswirksamkeit bedeutsam und vergleichsweise einfach zu erheben sei (ebd.). Der Annahme kausaler Bezüge der Ebenen folgend, postulieren sie weiter, dass positive Reaktionen eine wichtige Voraussetzung für den Trainingserfolg darstellen. Die Zustimmung auf dieser Ebene zeige das Interesse, die Motivation und die Aufmerksamkeit der Teilnehmer für das Training. Für diese Komponenten wird wiederum ein Einfluss auf die nachfolgende Ebene des Lernens angenommen. Weiter wird den Ergebnissen auf der Ebene Reaktion eine wichtige Feedbackfunktion zugeschrieben, die als Basis für Weiterentwicklungen und Verbesserungen eines Trainings dienen kann (ebd.).

In der fachwissenschaftlichen Auseinandersetzung sind die Bedeutung und der Einfluss der Reaktionsebene im Rahmen des Gesamtmodells umstritten. Insbesondere der postulierte Einfluss der Teilnehmerreaktionen auf die Lernebenen gibt Anstoß zur Kritik: In ihrer viel zitierten Metaanalyse kommen Alliger & Janak (1989) zu dem Ergebnis, dass die erhobenen Zufriedenheitsmaße nur sehr schwach mit Lernerfolgsmaßen korrelieren ( $r_{12} = .07$ ). Auch der Zusammenhang zwischen Zufriedenheitsmaßen und Indikatoren für Lerntransfer wird in der Studie zurückgewiesen ( $r_{13} = .05$ ) (ebd.). Die Ergebnisse werden in späteren Metanalysen bestätigt (Alliger, Tannenbaum, Bennet, Traver & Shotland, 1997; Colquitt, LePine & Noe, 2000). Weissweiler (2008) merkt an, dass die Annahme der kausalen Verbundenheit der Ebenen Reaktion und Lernen auch im Widerspruch zu den vorgeschlagenen Messzeitpunkten für die Ebenen steht: So schlägt Kirkpatrick für beide Ebenen die Erfassung direkt nach dem Training vor, womit eine kausale Abhängigkeit allein aus einem mangelnden zeitlichen Abstand nicht logisch erscheint. Es gibt also Gründe anzunehmen, „(...) dass Reaktionen nicht das Lernen begründen, und dies ist umso wahrscheinlicher, da Reaktionen nicht verhaltensbasiert, sondern mehr als ‚state‘ bzw. als momentane Einstellung gemessen werden“, resümiert Weissweiler (2008) die genannten Kritikpunkte (S. 21).

In Adaptionen und Weiterentwicklungen des Modells wird die Ebene Reaktion daher konsequenterweise unverbunden neben die Lernen-Verhalten-Ergebnis-Kette gestellt (Alliger u.a., 1997; Kraiger, 2002) oder gar nicht berücksichtigt (Holton, 1996).

#### 2.4.1.2 Ebene Lernen

Auf der Ebene Lernen geht es um Aussagen zur Aufnahme, Verarbeitung und Bewältigung der Lerninhalte und -prinzipien durch die Teilnehmer (Weissweiler, 2008). Dabei lassen sich nicht die Lernprozesse selbst erfassen. Vielmehr werden über den Zuwachs an Wissen und Kompetenzen oder über Änderungen von Einstellungen, Schlüsse über den Lernerfolg gezogen. Während einige Trainer erst dann von *Lernen* ausgehen, wenn sich Verhaltensänderungen zeigen, legen Kirkpatrick und Kirkpatrick (2006) dem Vier-Ebenen-Modell folgendes Verständnis von Lernen zugrunde: „In the four levels [...], learning has taken place when one or more of the following occurs: Attitudes are changed. Knowledge is increased: Skill is improved“ (p. 22). Je nach Ansatz und Ausrichtung eines Trainings kann unter *Lernerfolg* also etwas sehr Unterschiedliches verstanden werden. Umso wichtiger ist es, im Rahmen einer Trainingsevaluation konkret festzulegen, welche Lernziele mit dem Training primär verbunden sind und wie diese überprüft werden können.

Während sich mit Hilfe standardisierter Testverfahren vergleichsweise einfach prüfen lässt, ob ein gewünschter Wissenserwerb stattgefunden hat, lassen sich Kompetenzzuwachs und Einstellungsänderung schwerer erfassen. Hierbei werden vor allem Urteile von Teilnehmern, Trainern oder Beobachtern zum Lernfortschritt herangezogen (Schuler, 1998). Die Erhebungen auf der Lernebene gestalten sich also im Vergleich zu der Erhebung der Ebene Reaktion weitaus schwieriger. Kauffeld (2010) folgert, dass spezifische Trainingsprogramme mit oftmals komplexen und speziellen Zielstellungen erst die Entwicklung adäquater Erhebungsinstrumente fordern: „Für maßgeschneiderte Interventionen müssen daher spezifische Tests entwickelt werden“ (Kauffeld, 2010, S. 112). Damit ergeben sich für den Evaluator insbesondere methodische Herausforderungen, bei denen Kirkpatrick und Kirkpatrick (2006) durch die Formulierung einiger Leitlinien für die Erhebung auf der Lernebene Unterstützung leisten wollen: Zunächst wird der Einsatz einer Kontrollgruppe, die kein Training erhalten hat, nahe gelegt, um eine bessere Beweisführung der Effektivität des Trainings zu ermöglichen. Mit der Einschränkung „if practical“ verweisen Kirkpatrick und Kirkpatrick (2006) bereits auf mögliche Schwierigkeiten in der Evaluationspraxis, die dem Einsatz einer Kontrollgruppe entgegenstehen können. Die Autoren sprechen sich weiter für ein Design mit Prä- und Postmessung aus, das einen unmittelbaren Vergleich der Ergebnisse vor und nach der Trainingsmaßnahme ermöglichen soll. Um den Erwerb und den Ausbau von Wissen und Fähigkeiten zu

erfassen, wird zudem der Einsatz von standardisierten Leistungstests empfohlen. Für alle Lernerfolgsmaße wird überdies eine Erhebung in schriftlicher Form angeraten.

Unabhängig von diesen methodischen Überlegungen wird in der fachwissenschaftlichen Auseinandersetzung für bestimmte Lernerfolgskriterien ganz grundsätzlich angezweifelt, ob sich diese überhaupt empirisch erfassen lassen. So wird in Frage gestellt, ob sich die Einstellung eines Menschen als relativ stabiles Konstrukt, welches sich über langfristige Prozesse herausbildet und wenn überhaupt auch nur in langfristigen Prozessen verändert, direkt im Anschluss an ein Training - soweit der empfohlene Messzeitpunkt nach Kirkpatrick und Kirkpatrick (2006) - erfassen lässt (Bergmann, 1999; Nork, 1991). Auch ein Zuwachs an Kompetenzen, wie bspw. der Problemlösefähigkeit, lässt sich laut Häring (2003) nur schwer messen und muss sich nicht unmittelbar in quantitativen Ergebnissen niederschlagen.

Neben den viel diskutierten Problemen, die sich bei der Erfassung von Kriterien auf der Lernebene ergeben können, ziehen die Kritiker des Modells auch hier wieder die postulierte Abhängigkeit der Ebenen in Frage. Während Kirkpatrick und Kirkpatrick (2006) davon ausgehen, dass die Änderung in dem jeweils erwünschten Lernfeld eine notwendige, wenn auch nicht hinreichende Voraussetzung für eine Verhaltensänderung darstellt, finden Alliger und Kollegen (1997) empirisch lediglich einen schwachen Zusammenhang zwischen den Ebenen Lernen und Verhalten (*die durchschnittlichen Korrelationen der unterschiedlichen Lernerfolgsmaße mit Lerntransfer variieren von .08 bis .18*). Kirkpatrick und Kirkpatrick (2006) selbst räumen ein, dass der Übertragung des Gelernten in Verhaltensweisen eine Reihe von Faktoren im Weg stehen können, weshalb der scheinbar logische Rückschluss von unveränderten Ergebnissen auf der Verhaltensebene auf einen mangelnden Lernerfolg nicht zulässig sei. Was sich hinter dem Phänomen des Lerntransfers genau verbirgt, wird in den nachfolgenden Ausführungen zu der Ebene Verhalten zu klären sein. Welche Faktoren den Lerntransfer beeinflussen, findet in dem outputorientierten Modell Kirkpatricks jedoch keine Berücksichtigung. Die Beschreibung transferförderlicher bzw. -hemmender Bedingungen kommt erst bei der Auseinandersetzung mit dem prozessorientierten Evaluationsmodell zum Tragen.

### 2.4.1.3 Ebene Verhalten

*Was passiert, wenn die Teilnehmer das Lernfeld verlassen und zurück in den Arbeitsalltag und damit das Anwendungsfeld kehren?*, fragen Kirkpatrick und Kirkpatrick (2006) einleitend zur Erläuterung der Ebene Verhalten. Im Bereich verhaltensorientierter Trainings kann in der Anwendung des Gelernten im beruflichen Alltag ein wesentlicher Nachweis für die Wirksamkeit einer Weiterbildungsmaßnahme erbracht werden (Weissweiler, 2008). Der **Transfer** des Gelernten wird somit zur erfolgskritischen Größe von Trainingsmaßnahmen. In seinem Ausmaß bemisst sich für den Betrieb der Erfolg einer Weiterbildungsmaßnahme (Kauffeld, 2008).

Bevor mögliche Erhebungsformen und -instrumente für die Ebene Verhalten sinnvoll diskutiert werden können, gilt es eine nähere Betrachtung des abstrakten Konstrukts *Lerntransfer* vorzunehmen. Dabei finden sich in der fachwissenschaftlichen Auseinandersetzung zahlreiche Begriffsbestimmungen und Konkretisierungsversuche, die sich in Abstraktionsgrad und Schwerpunktsetzung unterscheiden. Solga (2011b) fasst die zentralen Bestimmungsstücke von Lerntransfer wie folgt zusammen:

*„Generell bedeutet Lerntransfer, dass Kenntnisse und Fertigkeiten, die in einer bestimmten Lernumgebung (Lernfeld) erworben wurden, auf Anwendungskontexte (Funktionsfelder) übertragen werden, deren Merkmale sich von denen der Lernumgebung mehr oder weniger stark unterscheiden“ (S. 342).*

Bezogen auf die betriebliche Weiterbildung kann die Trainingsmaßnahme als Lernfeld und das Unternehmen als Funktionsfeld bezeichnet werden. Die Übertragungsleistung impliziert dabei die **Generalisierung** des Gelernten auf andere Kontexte (Bergmann & Sonntag, 2006). Unterscheiden sich die Lerninhalte eines Trainings stark von den konkreten Anforderungen im Berufsfeld - dies ist für Kommunikations- und Verhaltenstrainings bzw. komplexe Ausbildungsprogramme zu erwarten - ist eine stärkere Abstraktionsleistung erforderlich, um das Gelernte sinnvoll anzuwenden (Solga, 2011b).

Neben der Generalisierung ist die **Aufrechterhaltung** der Transferleistung ein zweites zentrales Bestimmungsstück für Lerntransfer. So ist mit einem Training die Hoffnung verbunden, dass die gelernten Inhalte auch langfristig zur Anwendung kommen und nicht schon nach kurzer Zeit wieder in Vergessenheit geraten (Bergmann & Sonntag, 2006). Zur Systematisierung möglicher Verlaufsformen von Transferprozessen haben sich folgende Typologisierungen etabliert: Dem **positiven Lerntransfer**, der eine Verbesserung der Arbeitstätigkeit bedeutet, wird der sogenannte **negative**

**Lerntransfer** gegenüber gestellt (Solga, 2011b). Darunter werden Übertragungsversuche gefasst, die die Leistung im Anwendungsfeld eher beeinträchtigen statt fördern. Von einem **Nulltransfer** ist die Rede, wenn sich ein Training gar nicht auf das Funktionsfeld auswirkt (Weissweiler, 2008).

Ziel einer jeden Trainingsmaßnahme ist ein **positiver** Transfer. Dieser lässt sich noch einmal nach der Komplexität der Anforderungssituation unterteilen: Liegt für die Anwendungssituationen ein vergleichbares Anforderungsniveau vor, wird auch von **horizontalem Transfer** gesprochen (ebd.). Die im Lernfeld erworbenen Fähigkeiten und Kenntnisse werden im Arbeitsfeld angewandt. Wird das Gelernte darüber hinaus für weitere komplexere Lerngelegenheiten genutzt, liegt ein sogenannter **vertikaler Transfer** vor: „Vertikaler Transfer ist sukzessives Dazulernen durch Anwendung des Gelernten auf Herausforderungen eines komplexeren Anforderungsniveaus“ (Solga, 2011b). Abbildung 1 dient der Veranschaulichung der beschriebenen Transferarten:

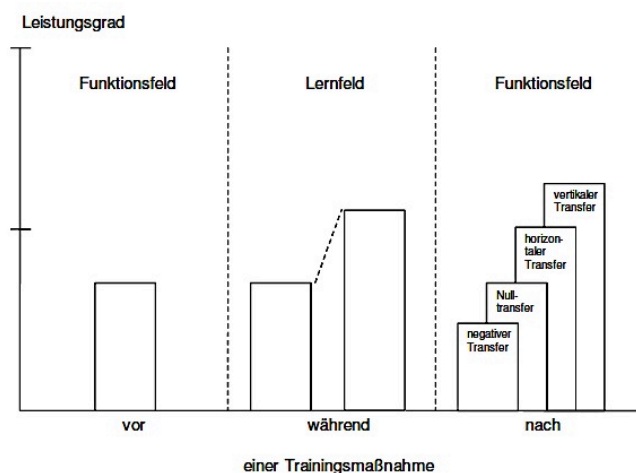


Abbildung 1: Schematische Veranschaulichung der Transferarten (B. Bergmann, 1999)

Den differenzierten Beschreibungen von Transferarten und möglichen Verlaufsformen von Lerntransfer wird bei der Überprüfung des Transfers bisher wenig Rechnung getragen. Im Modell von Kirkpatrick (1967) liegt der Fokus wie gewohnt auf dem Output und damit auf der Frage, ob ein Lerntransfer stattgefunden hat. Veränderungen im Arbeitsverhalten sollen als Kriterium zur Wirksamkeitsüberprüfung auf dieser Ebene dienen. Kirkpatrick und Kirkpatrick (2006) räumen ein, dass sich die Überprüfung von Effekten auf dieser Ebene im Vergleich zu den beiden vorhergegangenen Ebenen noch einmal deutlich komplexer und schwieriger gestaltet.



Zur Unterstützung formulieren sie auch für die Verhaltensebene einige Leitlinien, die sich insbesondere auf Designfragen beziehen: Analog zu den Erhebungsempfehlungen für die Lernebene wird der Einsatz einer Kontrollgruppe sowie eine Prä- und Postmessung der Effekte empfohlen. Der Zeitpunkt der Erhebung sollte hingegen - anders als bei der Lern- und Reaktionsebene - nicht unmittelbar nach dem Training liegen: „Allow time for behaviour change to take place“ (ebd., p. 53). Um den Transferprozess und -verlauf beurteilen zu können, wird zudem eine mehrmalige Wiederholung der Messung in größeren Zeitabständen empfohlen. Zum Mittel der Wahl zur Beurteilung von Verhaltensänderungen werden Beobachtungen erklärt (Kauffeld, 2006; Erpenbeck & Rosenstiel, 2007). Daneben ist der Einsatz von Interviews oder Fragebögen üblich (Weissweiler, 2008). Es bietet sich dabei an, Personen zur Einschätzung des Verhaltens heranzuziehen, die im Funktionsfeld unmittelbar davon betroffen sind: Kollegen, Mitarbeiter, Führungskräfte etc. Selbsteinschätzungsverfahren zur Erhebung der Transfererwartungen finden sich in der Praxis häufig, lassen aber keine Aussagen über tatsächliche Transferwirkungen zu (ebd.).

Unabhängig von den forschungsmethodischen Herausforderungen, die sich insbesondere für die Ebene Verhalten stellen, wird in der Trainingsforschung das sogenannte *Transferproblem* beklagt: Ergebnisse verschiedener Studien zeigen, dass das Gelernte oftmals gar nicht oder nur unzureichend in der Arbeit angewendet wird (Baldwin & Ford, 1988; Broad & Newstrom, 1992). In einer Metaanalyse zur Effektivität von Trainingsmaßnahmen (Arthur, Bennett, Edens & Bell, 2003) zeigen sich zwar insgesamt mittlere bis große durchschnittliche Effektstärken auf den vier Ebenen ( $d = 0.60$  bis  $d = 0.63$ ). Zwischen den Ebenen Lernen und Verhalten wurden aber erhebliche Effektverluste gefunden: „[T]he average decrease in effect size for learning and behavioral comparisons was 0.77, a fairly large decrease“ (ebd., p. 238).

Auch Taylor, Rust und Chan (2005) decken in einer Metaanalyse zum *Behavior Modelling Training*<sup>6</sup> eine Verminderung der Effektstärke von der Ebene Lernen zur Ebene Verhalten auf: Während die Effektstärken für Lernerfolg vergleichsweise groß ausfallen ( $d = 1.05$  für *Wissenserwerb* und  $d = 1.09$  für *den Erwerb von Fertigkeiten*), zeigen sich für den Lerntransfer deutlich geringere Effektstärken von  $d = 0.25$ .

---

<sup>6</sup> Die sogenannte Verhaltensmodellierung gilt als eine effektive Trainingsmethode. Sie geht theoretisch auf Banduras Theorie des Beobachtungslernens (1979) zurück und zielt auf den systematischen Aufbau neuer Verhaltensweisen (vgl. dazu ausführlich Demmerle, Schmidt, Hess, Solga & Ryschka, 2011, 289 ff.)

Eine andere interessante Perspektive auf das Lerntransferproblem beleuchten Taylor, Russ-Eft und Taylor (2009) in ihrer publizierten Metaanalyse zur Effektivität von Managementtrainings. Sie decken auf, dass die Stärke der Lerntransfereffekte unter Umständen stark von den einschätzenden Personen abhängt: Während die Lernenden selbst zu relativ optimistischen Einschätzungen kommen, was den Transfer des Gelernten betrifft ( $d = 0.61$ ), bewerten die Kollegen und insbesondere die Mitarbeiter der trainierten Führungskräfte den Lerntransfer als deutlich schwächer ( $d = 0.25$  und  $d = 0.12$ ). Die Werte sind aus forschungsmethodischer Sicht alarmierend und verdeutlichen die große Schwierigkeit, valide Kennzahlen für den Erfolg auf der Verhaltensebene zu erhalten. So sollte man genau prüfen, wen man für die Erhebungen des Lerntransfers sinnvollerweise befragt und wenn möglich mehrere Anspruchsgruppen zur Einschätzung heranziehen.

Da der Erfolg auf der vierten Ebene Resultate laut Kirkpatrick und Kirkpatrick (2006) maßgeblich von dem Transfererfolg abhängt, stimmen diese Werte eher pessimistisch, was zu erwartende Veränderungen auf der organisationalen Ebene betrifft.

#### **2.4.1.4 Ebene Resultate**

Gegenstand der vierten Ebene ist die Frage, inwieweit sich das geänderte Verhalten am Arbeitsplatz positiv auf die organisationale Leistung auswirkt. Dazu sollen konkrete Kennzahlen betrachtet werden, die für die jeweilige Organisation als Indikatoren des Unternehmenserfolges gelten (Solga, 2011a). Neben den mit dem Begriff *Kennzahlen* meist assoziierten *harten Faktoren*, wie eine **Kostenreduktion** oder eine **Steigerung der Produktivität**, stehen auf der Ebene der Resultate auch sogenannte *weiche Faktoren* im Fokus (Phillips & Schirmer, 2008). Darunter fallen bspw. die Konfliktrate oder die Mitarbeiterzufriedenheit, Fehlzeiten und Krankentage als Indikatoren des **Arbeits- und Organisationsklimas**. Auch werden Maße für die **Kundenzufriedenheit**, z.B. gemessen in einem Zufriedenheitsindex oder der Anzahl an Beschwerden, zur Messung des Unternehmenserfolges herangezogen. Daneben interessieren Indikatoren für die **Qualität der Leistung**, wie die Fehlerhäufigkeit, Nacharbeiten etc. (ebd.).

Kirkpatrick und Kirkpatrick (2006) bezeichnen diese Ebene - dem hierarchischen Aufbau des Modells folgend - als die bedeutendste Erhebungsebene des Modells.

Sie nennen zugleich verschiedene Gründe, warum die Messung auf dieser Ebene selten erfolgreich verläuft bzw. in der Regel gar nicht erst erfolgt: Eine große Schwierigkeit bestehe zunächst darin, festzulegen, mit welchen Kriterien der Unternehmenserfolg erfasst werden soll und wie sich diese überprüfen lassen. Während harte Kriterien meist objektiv messbar und in monetäre Werte übertragbar sind, müssen die weichen Kriterien als subjektive Faktoren erst in objektive transferiert werden (Solga, 2011a). Wurden Indikatoren des organisationalen Erfolges erfolgreich festgelegt, ergibt sich die Schwierigkeit, wie überprüft werden soll, dass die gemessenen Veränderungen auf das Training und nicht auf zahlreiche andere denkbare Entwicklungen und Einflussgrößen in der Organisation zurückgehen (Kirkpatrick & Kirkpatrick, 2006). Dieses Problem wird auch als **Kriterienkontamination** bezeichnet (Goldstein & Ford, 2002). „Ein Evaluationskriterium ist kontaminiert, wenn es Aspekte beinhaltet oder von Einflüssen abhängig ist, die sich der Veränderbarkeit durch das PE-Programm entziehen“ (Solga 2011a, S. 348). Daneben ist zu bedenken, dass im Rahmen von Trainingsmaßnahmen immer nur bestimmte Anteile der jeweiligen Arbeitsaufgabe trainiert werden können. Ein hundertprozentiger Zusammenhang zwischen Transfer und Arbeitsleistung ist daher nicht zu erwarten (Hochholdinger u.a., 2008b). Kirkpatrick und Kirkpatrick (2006) empfehlen auch für diese Ebene die Einhaltung gewisser Leitlinien für die Messung. Neben den ersten vier Regeln, welche denen auf der Verhaltensebene entsprechen (*Kontrollgruppe, Prä-/Postvergleich, Messwiederholungen, Messzeitpunkte in größerem Abstand zur Maßnahme*), schlagen sie eine Kosten-/Nutzen-Analyse vor, welche in den Bereich der Effizienzanalyse fällt. Ihre letzte Empfehlung bezieht das beschriebene Problem der Kriterienkontamination mit ein: „Be satisfied with evidence if proof is not possible“ (ebd., p. 64).

#### **2.4.1.5 Diskussion**

Kirkpatrick's Vier-Ebenen-Ansatz ist nach wie vor das am meisten verbreitete Modell zur Trainingsevaluation. Aufgrund seiner einfachen Systematik und Anwendbarkeit ist es in Wissenschaft und Unternehmenspraxis gleichermaßen populär. Der Umgang mit dem Modell und das Verständnis darüber gehen in den beiden Feldern allerdings stark auseinander.

In der Wissenschaft dominiert die kritische Auseinandersetzung mit dem Modell und seinen immanenten Annahmen. Dabei wird insbesondere auf drei bereits erwähnte

Metaanalysen Bezug genommen, die die ursprünglich postulierte Abhängigkeit der vier Ebenen in Frage stellen: So kommen Alliger & Janak (1989) anhand von 12 bis dato verfügbaren Studien zu dem Schluss, dass die vier Ebenen nur schwach korrelieren. Während die Zusammenhänge zwischen Zufriedenheitsmaßen und Lernerfolgs-, Transfer-, und Leistungsindikatoren wie bereits aufgeführt sehr gering ausfallen, sind die Korrelationen zwischen den anderen drei Ebenen zwar etwas höher ausgeprägt, aber immer noch als schwach zu beurteilen ( $r_{23} = .13$ ;  $r_{24} = .40$ ;  $r_{34} = .19$ ). Problematisch an der Studie ist allerdings neben der geringen Fallzahl der Umstand, dass die untersuchten Studien überwiegend lediglich zwei Kriterienebenen berücksichtigten. Untersuchungen in denen alle vier Ebenen zum Einsatz kamen, existierten bis dato kaum (Rowold, 2008). Die Zusammenhänge zur Resultate-Ebene, die vergleichsweise höher ausfallen, basieren lediglich auf einer einzelnen Studie und können daher nur unter Vorbehalt interpretiert werden. In ihrer Übersichtsarbeit von 1997 bestätigen Alliger und Kollegen die Ergebnisse jedoch auf Grundlage einer deutlich größeren Datenbasis (34 Studien). Drei Jahre später gelingt es Colquitt, Lepine und Noe (2000) in der bisher umfangreichsten Metaanalyse unter Einbezug von 106 empirischen Studien nachzuweisen, dass die Abhängigkeiten zwischen den vier Ebenen der Trainingsevaluation lediglich geringfügig ausgeprägt sind. Obwohl die Annahme einer linearen Abfolge der Ebenen damit empirisch zurückgewiesen wurde, zeigen sich in Weiterentwicklungen des Modells lediglich Konsequenzen für die Einbindung der Reaktionsebene. Diese wird - wie erwähnt - entweder separat betrachtet (z.B. Alliger & Janak, 1989; Kraiger, 2002) oder gar nicht berücksichtigt (z.B. Holton, 1996). Für die anderen drei Ebenen wird in vielen Modellen weiter von einer kausalen Verknüpfung ausgegangen. Cannon-Bowers, Salas, Tannenbaum und Mathieu (1995) beziehen die Abhängigkeit der Ebenen sogar auf alle vier Ebenen.

Während einige Autoren (z.B. Rowold, 2008; Alvarez, Salas & Garofano, 2004) daran anknüpfend eine Diskrepanz zwischen Empirie und Theorieentwicklung beklagen, muss der größere Bruch zwischen den wissenschaftlichen Erkenntnissen über das Modell und dem Umgang damit in der Evaluationspraxis in den Unternehmen festgestellt werden: Twitchell, Holton und Trott (2001) berichten, dass 73 % der von ihnen befragten US-amerikanischen Unternehmen Teilnehmerurteile erheben, um den Erfolg ihrer Weiterbildungsprogramme zu schätzen. Deutlich seltener werden die Kriterien der Ebenen Lernen, Verhalten und Resultate gemessen (47 %, 31 % und 21 % befragter Organisationen).

In einer anderen häufig zitierten US-amerikanischen Studie wird ein noch pessimistischeres Bild gezeichnet: Laut Buren und Erskine (2002) messen zwar 78% der Unternehmen Kriterien für den Zufriedenheitserfolg. Maße für den Lernerfolg werden hingegen nur noch von 32% der Unternehmen erhoben und für den Transfer- und Unternehmenserfolg interessieren sich nach der Studie weniger als 10% der Unternehmen (Transfererfolg = 9%; Unternehmenserfolg = 7%).

Wenngleich die Zufriedenheitsebene im Gesamtmodell durchaus wichtige Erkenntnisse liefern kann, ist sie als Hauptindikator für die Wirksamkeit einer Maßnahme nicht geeignet. So klar diese Erkenntnis in der fachwissenschaftlichen Auseinandersetzung verbreitet wird, so konsequent wird sie in der Evaluationspraxis in den Unternehmen ignoriert. Kritiker verleihen ihrem Missmut über die Evaluationspraxis in abfälligen Bezeichnungen der Ebene als *Happiness-Index*, *Happy-Sheets-Befragung* Ausdruck (Kauffeld, 2010). Auch in vielen wissenschaftlichen Studien wird auf die Erhebung einzelner Ebenen (in der Regel der vierten Ebene) verzichtet (z.B. Saborowski & Muellerbuchhof, 2010). Rowold (2008) fordert explizit dazu auf, Mehrebenen-Analysen in der Trainingsevaluation umzusetzen und Trainings konsequent auf allen vier Ebenen nach Kirkpatrick zu evaluieren. Nur so könne man zu differenzierten Ergebnissen über die Wirksamkeit von Trainings gelangen.

Die Annahme der hierarchischen Abhängigkeit der Ebenen wird in der wissenschaftlichen Auseinandersetzung ebenfalls kritisch gesehen, jedoch mehr aus einer wertorientierten denn empirischen Sicht. So würde durch die Annahme suggeriert, dass Unternehmenserfolg bedeutsamer ist als Lernerfolg oder die Zufriedenheit mit einem Training. Alliger & Janak (1989) weisen überdies darauf hin, dass es durchaus Maßnahmen gibt, die gar nicht auf die Veränderung ökonomischer Größen zielen:

„Some training may be largely rewarding, spirit-building, or perquisite in nature. For example, programs designed to instill company pride or rejuvenate employees may be realistically expected to have impacts at the reaction level only.“ (p. 332).

Ein weiterer Kritikpunkt an den vier Ebenen der Trainingsevaluation nach Kirkpatrick ist die lediglich grobe Eingrenzung davon, welche Kriterien es nun konkret auf den jeweiligen Ebenen zu überprüfen gilt. Während sich die Maße auf der Zufriedenheitsebene noch vergleichsweise einheitlich erheben lassen, sind für die Erhebung der Lern- und Verhaltensebene ganz unterschiedliche Kriterien denkbar und je nach Zielstellung der Evaluation angemessen. Ein Großteil der Trainingsprogramme die nach Kirkpatrick evaluiert werden, zielt auf die Entwicklung relativ eingegrenz-

ter Wissens- und Kompetenzbereiche. Diese lassen sich, wie auch von Kirkpatrick und Kirkpatrick (2006) empfohlen, oftmals mit Hilfe standardisierter Verfahren erheben (z.B. Rowold, 2008; Scharpf, 1999; Saborowski & Muellerbuchhof, 2010). Für Kommunikations- und Verhaltenstrainings oder modulare Ausbildungsprogramme, die komplexe Zielstellungen verfolgen, wird die Operationalisierung von Lernerfolgs- und Transferkriterien hingegen schwierig. Verdeutlicht man sich Kirkpatricks ursprünglichen Anspruch an sein Modell, dass nie als etwas anderes gedacht war als ein erster Zugang, eine globale Heuristik zur Erfassung von Kriterien zur Trainingsevaluation, ist dieser Umstand nicht weiter problematisch. „As such it has done well“ (Alliger & Janak, 1989, p. 340). Der hohe Abstraktionsgrad der Ebenen, der auf der einen Seite die Anwendbarkeit des Modells auf viele unterschiedliche Trainingsformate ermöglicht, eröffnet auf der anderen Seite aber einen Interpretationsspielraum, der aus wissenschaftlicher Sicht problematisch ist:

„[Kirkpatrick’s model] provides a vocabulary and rough taxonomy for criteria. At the same time, Kirkpatrick’s model, through its easily adopted vocabulary and a number of (often implicit) assumptions, can tend to misunderstandings and overgeneralizations“ (ebd., p. 331).

#### **2.4.2 Bedingungen des Transfers von Trainingsinhalten**

Im Rahmen des Evaluationsmodells von Kirkpatrick (1967) liegt der Fokus, wie beschrieben, auf der Überprüfung von Outputvariablen. Auf der Ebene Verhalten interessiert die Frage, ob ein Lerntransfer stattgefunden hat (Teil II, Kap 2.4.1.3). Zeigen sich dabei keine oder nur geringfügige Verhaltensänderungen, lässt das Modell keinen Rückschluss auf mögliche Gründe zu. Einen anderen Zugang zur Trainingsevaluation wählen daher Baldwin und Ford (1988). Auf der Grundlage einer ausführlichen Sichtung der damaligen Befunde zu dem Themenfeld gingen sie der Frage nach, welche Einflussvariablen auf Lerntransferprozesse wirken. Sie entwickelten ein Rahmenmodell zur Beschreibung des Transferprozesses, das drei Gruppen von Einflussvariablen unterscheidet: Die erste Gruppe umfasst teilnehmerbezogene Merkmale wie Persönlichkeitsfaktoren, Fähigkeiten und Fertigkeiten und Motivation (***Trainee-Charakteristika***). Die zweite Gruppe betrifft das Training selbst und bezieht Faktoren wie Lernprinzipien und -methoden oder den Trainingsinhalt ein (***Trainingsdesign***). Als dritte Gruppe werden Faktoren der ***Arbeitsumgebung*** unterschieden. Hierzu zählen bspw. die Unterstützung durch Vorgesetzte und Kollegen, Möglichkeiten der Anwendung etc. (Baldwin & Ford, 1988).

Auch im Rahmenmodell von Baldwin und Ford (1988) werden in Anlehnung an Kirkpatrick Outputvariablen berücksichtigt. Dabei werden *Trainings-Outputs* - Lerneffekte, die unmittelbar aus dem Training resultieren - von den darauf aufbauenden Transferleistungen (*Conditions of Transfer*) unterschieden (Kuper, 2011). Letztere umfassen - anknüpfend an die bereits vorgenommene definitorische Eingrenzung im Rahmen dieser Arbeit - die Generalisierung und die Erhaltung des Gelernten im Arbeitskontext (Bergmann & Sonntag, 2006).

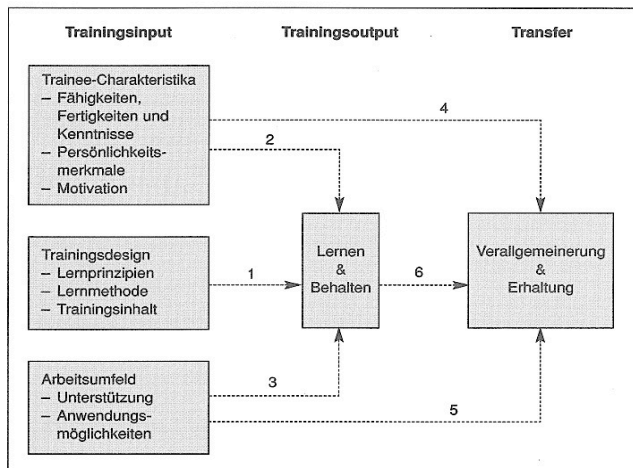


Abbildung 2: Ein Modell zur Beschreibung des Transferprozesses (nach Baldwin & Ford, 1988)

In Übereinstimmung mit der im Vier-Ebenen-Modell (Kirkpatrick, 1967) indizierten Annahme der Abhängigkeit der Ebenen wird ein direkter Effekt des Lernerfolges auf den Transfererfolg angenommen (Baldwin & Ford, 1988). Auch für die Faktoren der drei beschriebenen Wirkgruppen, die im Modell als **Inputvariablen** verstanden werden, werden Effekte auf die Lern- und Transfereffekte angenommen. Dabei wird für die Variablengruppe der Trainee-Charakteristika und der Arbeitsumgebung jeweils ein direkter Einfluss auf den Lernerfolg und den Transfererfolg angenommen, womit sich zusätzlich automatisch ein über den Lernerfolg vermittelter indirekter Effekt auf den Transfererfolg ergibt (ebd.). Das Trainingsdesign als dritte Gruppe von Einflussfaktoren wirkt in direkter Form ausschließlich auf den Lerntransfer ein. Auch hierfür wird zusätzlich ein indirekter Effekt auf den Transfererfolg angenommen (Abb. 2).<sup>7</sup> Kauffeld (2010) erklärt die Suche nach Einflussfaktoren des Transfererfolges zu einem existenziellen Schritt, „um Trainingsprogramme zu verbessern und strategische Entscheidungen zu treffen“ (S. 130). Damit stellt sich die Frage nach Möglichkeiten der systematischen Erfassung der Wirkfaktoren auf den Transfererfolg.

<sup>7</sup> Eine umfassende Diskussion dieser Einflussbedingungen findet sich in zahlreichen Publikationen der Transferforschung (z.B. Baldwin & Ford, 1988; Scharpf, 1999; Hochholding et al., 2008; Bergmann & Sonntag, 2006).

### **III Empirischer Teil**

#### **1. Fragestellung und Anlage der Studie**

##### **1.1 Ausgangslage**

Wie die Auseinandersetzung mit den theoretischen Grundlagen der Trainingsforschung zeigen konnte, bewegt sich die Evaluation von Trainings immer in einem Spannungsfeld zwischen wissenschaftlichem Anspruch einerseits und den spezifischen Bedingungen im Feld und der Forderung nach praktikablen Lösungen andererseits. Insbesondere in der Konstruktion eines adäquaten Evaluationsdesigns und in der Operationalisierung von Evaluationskriterien liegen die zentralen Herausforderungen einer Trainingsevaluation.

So sind aufwendigen Versuchsplanungen, mit denen die Auftretenswahrscheinlichkeit von Störfaktoren im Vorhinein verringert werden könnte, in der Praxis oftmals Grenzen gesetzt. Höft (2006) fragt zu Recht „wie kann bei vorgegebenen Arbeitsgruppen eine Randomisierung erfolgen?“ (S. 787), und spielt damit auf technische Hürden einer experimentellen Versuchsplanung an. Im Bereich der Trainingsevaluation ist eine zufällige Zuweisung zu Untersuchungs- und Kontrollbedingung selten möglich, da ein Großteil der Trainings für bestimmte Personengruppen entwickelt wird. Die Teilnahme wird dann entweder verpflichtend verordnet oder ist freiwillig, was die besonders interessierten und engagierten Personen anspricht. In beiden Szenarien ist von störenden Selektionseffekten auszugehen. Die Hoffnung, durch eine Randomisierung die Gleichverteilung aller Störvariablen zu erreichen, scheitert zudem oftmals an zu kleinen Stichproben (Gollwitzer & Jäger, 2009). Häufig beginnt die Problematik aber noch davor, bei dem Versuch überhaupt eine Kontrollgruppe zur Trainingsgruppe zu konstruieren, an der sich wenn möglich auch noch eine Placebo-Intervention durchführen lässt. Dies ist insbesondere für längerfristig angelegte Maßnahmen nur schwer realisierbar und wie bereits angedeutet auch ethisch bedenklich (Solga, 2011a). Inwieweit sich Kontrollgruppen, die keine Maßnahme enthalten, zur Kontrolle von Störfaktoren eignen, ist ebenso fraglich. Hingegen der postulierten Bedeutung des Transferthemas finden sich zudem kaum Untersuchungsdesigns, die mehr als zwei Messungen verwirklichen (Lang & Otto, 2010; Merzenich-Hieker, 1996). Gerade für die Überprüfung der Generalisierung und Aufrechterhaltung des Gelernten wäre eine Follow-Up-Messung einige Zeit nach der Maßnahme aber wünschenswert.



Bei der Operationalisierung der Wirksamkeitskriterien und der Entwicklung und Auswahl der Erhebungsinstrumente schlägt sich, wie bereits in der Diskussion von Kirkpatrick's Ebenen vermutet, die ungenaue Eingrenzung möglicher Erfolgskriterien nieder (Teil II, Kap. 2.4.1.5). Die Kriterien müssen einerseits hinreichend spezifisch für die jeweilige Maßnahme festgelegt werden (Kauffeld, 2010). Andererseits ist aus wissenschaftlicher Perspektive der Einsatz standardisierter und publizierter Verfahren wünschenswert, um die Gültigkeit der Ergebnisse zu sichern. Bei der Festlegung von Erfolgskriterien bei Trainingsmaßnahmen aus dem Bereich der Kommunikations- und Verhaltenstrainings kommt erschwerend hinzu, dass die *weichen* Kriterien des Erfolges wie Einstellungs- oder Verhaltensänderung sich nur schwer eindeutig operationalisieren und in quantitativer Form erfassen lassen.

Aufwendige Designs und Messmethoden stoßen in der Praxis zudem nicht selten auf Akzeptanzprobleme. So können standardisierte Maßnahmen der Erfolgsmessung bei Teilnehmern und Trainern Kontroll- und Versagensängste auslösen, was sich wiederum in Widerständen und Verweigerungen der Messungen oder in verzerrtem Antwortverhalten äußern kann. Auch kommt es vor, dass die Befürchtung von aufkommenden Ängsten und Widerständen von vornherein zu einer oberflächlichen Evaluation führt (Hanke, 2005). Sollte es dennoch gelingen ein Evaluationsdesign zu konstruieren, was als intern valide gelten kann, stellt sich weiter die Frage, ob und inwieweit sich Ergebnisse einer Evaluationsstudie überhaupt replizieren lassen (Höft, 2006). So lassen die spezifischen Interventions- und Evaluationsbedingungen in der Praxis meist nur singuläre Aussagen zu. Das mag für die den konkreten Auftraggeber genügen, ist unter den Prämissen Wissenschaftlichkeit und Qualitätssicherung und -entwicklung in der Evaluationsforschung jedoch zu problematisieren.

Die Deutsche Gesellschaft für Evaluation (DeGEval) (2008) formuliert vier Standards zur Qualitätssicherung in Evaluationen, in denen sich das Spannungsfeld zwischen wissenschaftlichem Anspruch und Anforderungen und Bedingungen im Anwendungsfeld gewissermaßen widerspiegelt. Auf der einen Seite sollen Evaluationsvorhaben dem Anspruch der *Genauigkeit* genügen: „Aus Evaluationsprozessen sollen akkurate und gültige Befunde zum Evaluationsgegenstand und zu den Evaluationsfragestellungen resultieren“ (Solga, 2011a, S. 334). Damit sind auch die beschriebenen forschungsmethodischen Anforderungen angesprochen. Auf der anderen Seite wird gefordert, dass die Planung der Evaluation realistisch, kostenbewusst und gut durchdacht ist und bei Betroffenen und Beteiligten eine hohe Akzeptanz erreicht

(Bortz & Döring, 2006). Die Gewährleistung der damit angesprochenen **Durchführbarkeit** kann zu Lasten der Einhaltung der Genauigkeitsstandards gehen. Auch der Standard der **Fairness**, der die Sicherstellung eines respektvollen und fairen Umgangs mit den Beteiligten und die Achtung ethischer und rechtlicher Standards formuliert (Bortz & Döring, 2006; Solga, 2011a), setzt aufwendigen Designplanungen enge Grenzen. Mit dem **Nützlichkeitsstandard** werden darüber hinaus die Ausrichtung des Evaluationsprozesses an den Interessen und Bedürfnissen der vorgesehenen Nutzer und die Berücksichtigung der vereinbarten Evaluationszwecke gefordert (Solga, 2011a). Auch hier lässt sich die Warnung entnehmen, Nutzen und Zweck des Vorhabens nicht zugunsten wissenschaftlicher Interessenlagen zu vernachlässigen.

Evaluationsvorhaben erfordern also immer einen Spagat zwischen verschiedenen Anspruchsgruppen. Der fachwissenschaftliche Diskurs lässt dabei unterschiedliche Prioritätensetzungen erkennen: Wottowa und Thierau (2003) postulieren die Notwendigkeit praktikabler Kompromisse für die Evaluationsforschung. Man dürfe nicht den Fehler machen, sich „sklavisch“ an idealen Designvorstellungen festzuhalten und Untersuchungen, die den methodischen Anforderungen nicht entsprechen, als unwissenschaftlich und irrelevant abzuwerten (ebd.). Sie verstehen die Designanforderungen als ein Ideal, dem es sich anzunähern gilt, das sich aber nie ganz erreichen lässt. Das macht auf der einen Seite Mut, sich den Anforderungen einer Evaluation zu stellen, lässt auf der anderen Seite aber viel Spielraum für Spekulationen darüber, wieviel Abweichung vom Ideal denn noch erlaubt ist. So schreiben sich andere Autoren eher die Verteidigung des wissenschaftlichen Anspruchs auf die Fahnen:

*„Die Komplexität der Praxissituation, die ein Abwägen des sinnvollen und machbaren Vorgehens notwendig macht, darf aber nicht als Begründung für methodisch mangelhafte Evaluationen dienen“ (Höft, 2006, S. 787).“*

Die Betrachtung aktueller Evaluationsstudien zeigt, dass zwar eine Weiterentwicklung zugunsten des wissenschaftlichen Anspruchs stattgefunden hat. Die Studien verfolgen jedoch entweder die Zielsetzung, weiterentwickelte Theoriemodelle zu prüfen (z.B. Rowold, 2008) oder den Nutzen und Erfolg konkreter Trainingsprogramme nachzuweisen (z.B. Saborowski & Muellerbuchhof, 2010). Studien, die danach fragen, wie ein geeigneter Ansatz zur Evaluation von Trainings aussehen kann, der die verschiedenen Anforderungen von Wissenschaft und Praxis gleichermaßen berücksichtigt, sind nach wie vor rar. Damit eröffnet sich ein Forschungsbedarf, der im Rahmen dieser Arbeit bearbeitet werden soll.

## 1.2 Forschungsfrage der Arbeit

Auf Basis der bis hierhin erarbeiteten Rahmenbedingungen, Methoden und theoretischen Konzepte soll nun ein Evaluationsansatz entwickelt werden, der den aufgedeckten Forschungsbedarf im Bereich der Trainingsevaluation als Ansatzpunkt nimmt und der Beantwortung folgender Forschungsfrage dienen soll:

***Wie kann ein ganzheitlicher Evaluationsansatz zur Überprüfung von Trainingswirksamkeit aussehen, der eine ausgewogene Balance zwischen wissenschaftlichem Anspruch und praktikablen Anforderungen im Feld herstellt?***

In einem ersten Schritt wird dazu auf *konzeptioneller Ebene* ein *ganzheitlicher* Evaluationsansatz entwickelt, der sowohl den Anforderungen der Wissenschaft als auch denen der Praxisseite hinreichend gerecht werden soll.

Der Zusatz *ganzheitlich* steht in diesem Fall für *divergierende Anforderungen beachtend*. Es wird nicht der Anspruch erhoben, den einen Königsweg zur Trainingsevaluation aufzuzeigen und den immanenten Konflikt widersprüchlicher Bestrebungen in Evaluationen etwa aufzulösen. Es geht vielmehr um einen bewussten Umgang mit den unterschiedlichen Herausforderungen einer Trainingsevaluation und um den Versuch, einen Mittelweg aufzuzeigen, der beide Perspektiven sinnvoll integriert.

In einem zweiten Schritt soll die Eignung des entwickelten Evaluationsansatzes *empirisch überprüft* werden. Dazu wird der Ansatz in Bezug auf eine ausgewählte Trainingsmaßnahme - *die Ausbildung zum Innovationspromotor* - konkretisiert. Eine umfassende Modellüberprüfung ist im vorliegenden Untersuchungsrahmen allerdings nicht möglich. Dennoch sollen gezielt empirische Anhaltspunkte für die Eignung, den Erfolg des Ansatzes gewonnen werden. Auf empirischer Ebene steht also die Beantwortung folgender Fragestellung im Fokus der Analysen:

***Eignet sich der entwickelte Evaluationsansatz zur Überprüfung von Trainingswirksamkeit?***

Die Gewinnung von Indikatoren für die Eignung findet auf zwei Ebenen statt:

Zum einen erfolgt die exemplarische Wirksamkeitsüberprüfung der Ausbildung. Auf Grundlage der Ergebnisse der Analysen sollen in der Diskussion Schlussfolgerungen über folgende Fragestellung gezogen werden:

***Gelingt es im Rahmen der Evaluation ein differenziertes Bild über die Wirksamkeit der Ausbildung zu zeichnen?***

Zur Beantwortung dieser Fragestellung muss im Rahmen der exemplarischen Evaluation folgende Untersuchungsfrage empirisch bearbeitet werden:

*Ist die Ausbildung zum Innovationspromotor wirksam?*

Zum anderen sollen auch auf empirischer Ebene gezielt Anhaltspunkte dafür gewonnen werden, inwieweit der Evaluationsansatz einerseits *wissenschaftlichen Ansprüchen genügt* und sich andererseits *unter Feldbedingungen als praktikabel erweist*:

***Inwieweit genügt der entwickelte Evaluationsansatz wissenschaftlichen Ansprüchen?***

***Inwieweit erweist sich der entwickelte Evaluationsansatz unter Feldbedingungen als praktikabel?***

## 2. Methoden

Bevor es darum gehen kann, die Entwicklung eines *ganzheitlichen* Evaluationsansatzes zu skizzieren und entsprechende Modell- und Designüberlegungen auszuführen, wird die Ausbildung zum Innovationspromotor als Trainingsmaßnahme, anhand derer der Evaluationsansatz empirisch geprüft werden soll, beleuchtet. Nur so können die Modell-, Design-, und Methodenüberlegungen in einem zweiten Schritt auf die spezifischen Bedingungen der zu evaluierenden Maßnahme bezogen werden.

### 2.1 Die Trainingsmaßnahme

Anknüpfend an die Ausführungen zu betrieblichen Trainings und der Ausbildung von Beratern, werden die Konzeption und die Zielstellung der Ausbildung zum Innovationspromotor und der Aufbau und die Gestaltung der Module skizziert.

#### 2.1.1 Projekthintergrund und Zielstellung

*Die Ausbildung zum Innovationspromotor* wurde als Qualifikationsansatz entwickelt, der interessierte Personen in Unternehmen befähigen soll, Innovationsprojekte systematisch zu unterstützen und voranzutreiben. Ausgehend von den neuesten Erkenntnissen der Innovationsforschung, in der die Bedeutung engagierter Individuen für den Erfolg von Innovationsvorhaben wiederholt betont wird (z.B. Gemünden, Salomo & Hölzle, 2007; Howell, Shea & Higgins 2005), wurde ein professionelles Rollenbild für solche Promotoren der Innovation entwickelt.<sup>8</sup> Zur Ausbildung der Promotorenrolle wurden die Teilnehmer auf der einen Seite zu Experten für Innovationen in ihrer Organisation ausgebildet. Auf der anderen Seite wurde der Erwerb notwendiger methoden-, sozial- und persönlicher Kompetenzen gefördert, um die Teilnehmer zu befähigen, im Rahmen von Innovationsprozessen als interne Berater, Unterstützer und Helfer wirksam zu werden. Damit sollten sie nachhaltig dazu beitragen - so das übergeordnete Ziel der Ausbildung - die Erfolgsaussichten von Innovationsprojekten deutlich zu steigern (Hüttner & Scholl, 2011).

Die Ausbildung richtete sich u.a. an Geschäftsführer, Innovationsbeauftragte, Planer und Organisationsentwickler aus Unternehmen, an Innovationen interessierte Betriebsräte und an andere motivierte Innovationstreiber aus Unternehmen.

---

<sup>8</sup> Das Rollenkonzept ist in Anlehnung an den Prozesspromotor in der Weiterentwicklung des Promotorenmodells (ursprünglich Witte, 1973) nach Hauschildt und Chakrabarti (1988) entstanden. Einen vertiefenden Einblick dazu bieten Hauschildt & Gemünden (1999) in ihrer Publikation „Promotoren: Champions der Innovation“. Einen komprimierten und aktuellen Überblick zur Promotorenforschung gibt Mansfeld (2010).

### 2.1.2 Aufbau und Gestaltung

Das Curriculum der Ausbildung zum Innovationspromotor umfasste zehn Module, die über den Zeitraum von zehn Monaten umgesetzt (160 Ausbildungsstunden) wurden. Die Module fanden immer an zwei aufeinanderfolgenden Wochentagen unter der Leitung von je zwei Ausbildern von artop oder der Humboldt-Universität statt. Durch die Verortung der Ausbildung in einem geförderten Projekt konnte der erste Durchgang für die Teilnehmer kostenlos angeboten werden. So wurden insbesondere interessierte Personen aus den Partnerunternehmen des Projektes zur Ausbildung zugelassen. Daneben konnten Personen aus anderen Unternehmen für die Teilnahme akquiriert werden.<sup>9</sup> Es wurde eine feste Ausbildungsgruppe à 15 Personen gebildet, die einen optimalen Lernrahmen ermöglichen sollte. Die zehn Module wurden als Kommunikations- und Verhaltenstrainings gestaltet und lassen sich in ihrer inhaltlichen Ausrichtung noch einmal in eine erste und zweite Ausbildungshälfte gliedern. Anders als in den beschriebenen Beraterrollen nahm die Erarbeitung theoretischer Inhalte einen großen Stellenwert in der ersten Hälfte der Ausbildung ein. Die Teilnehmer sollten verschiedene Facetten des Themenfeldes Innovation kennenlernen und ein gemeinsames Verständnis von Innovation erarbeiten.

Im Rahmen von **Modul zwei *Ideenmanagement*** lauteten die zentralen Fragestellungen bspw. „Was sind Innovationen? Was unterscheidet sie vom Alltagsgeschäft? Wie kommt überhaupt Neues bzw. neues Wissen in die Welt und wie kann es sich etablieren?“ (Hüttner & Scholl 2011, S. 78). Darauf aufbauend ging es im **dritten Modul** zu dem Thema ***Innovationsprozesse und ihre Analyse*** darum, den analytischen Blick auf Innovationsprozesse und ihre Entstehung in Unternehmen zu schulen und „(...) erfolgskritische Faktoren auf dem Weg einer Idee zur Innovation zu betrachten und zu bewerten“ (ebd., S. 79). In **Modul 4 *Rahmenbedingungen für Innovationen und Veränderungen*** wurde der Blick auf Veränderungsprozesse als möglicher Anlass, Inhalt oder die Folge von Innovationen gerichtet: „Was sind typische Probleme und Reaktionen? Wie kann man überhaupt gezielt in laufende soziale Systeme eingreifen? Wie tickt mein Unternehmen?“ (ebd., S. 79), lauteten nur einige Fragestellungen, die in diesem Kontext bearbeitet werden sollten. Das **fünfte Modul** mit dem Titel ***Interventionen zur Förderung von Innovationen*** bildete den Abschluss des Thementeils Innovation. Es zielte darauf ab, Möglichkeiten für ein *bewusstes Reagieren* zu beleuchten und konkrete Maßnahmen zur Gestaltung und Ver-

---

<sup>9</sup> Nähere Erläuterungen zur Zusammensetzung der Stichprobe erfolgen in Kapitel 2.4.

änderung bisheriger Abläufe und Strukturen kennenzulernen (Hüttner & Scholl, 2011). In der ersten Hälfte der Ausbildung stand die Förderung der Fachkompetenz klar im Fokus. Gleichzeitig wurden immer auch Methoden und Handwerkszeug mit erarbeitet, die in diesem Kontext relevant erschienen. Dabei spielten eher solche Methoden eine Rolle, die zur Analyse und Klärung dienen konnten.

Die Förderung methodischer Kompetenzen im Sinne der Intervention und Promotion stand stärker im Fokus der zweiten Hälfte der Ausbildung. Unter den Schlagworten ***Moderation und Gesprächsführung, Konfliktmanagement, Projektmanagement*** (Module 6-8) sollte das relevante Handwerkszeug und methodische Know How zur gezielten Promotion von Innovationsprozessen erarbeitet werden. ***Modul 9 Evaluation und Verankerung*** bot zwar auch Perspektiven auf konkrete Möglichkeiten zur Projekt - und Prozessevaluation, war aber wieder stärker thematisch an das Themenfeld der Innovation gekoppelt. Die Module 1 und 10 nahmen eine Sonderstellung ein und dienten vor allem der Reflexion und Rollenklärung und -festigung. ***Modul 1 Positionsbestimmung und Rollenklärung*** hatte zum Ziel eine Gruppe zu formieren, in der ein gemeinsames Lernen möglich wurde. Die individuellen Erwartungen und Zielstellungen der Teilnehmer wurden ausgetauscht und eine gemeinsame Blickrichtung für die Ausbildung erarbeitet. Es folgte ein inhaltlicher Einstieg in das Themenfeld der Innovation und Innovationspromotion sowie die Auseinandersetzung mit dem GI:VE Modell als Theorierahmen im Projekt. Fragen zu Haltung und Selbstverständnis als Innovationspromotor - als Grundstein der Ausbildung einer professionellen Rolle - wurden zum ersten Mal aufgeworfen (ebd.). ***Modul 10*** bildete den ***Abschluss*** der Ausbildung. Den Teilnehmern sollte der Raum geboten werden, ihre Rolle als Innovationspromotoren bzw. alle relevanten Inhalte und Lern- und Entwicklungsbereiche der Ausbildung systematisch zu reflektieren. Es wurde der Blick auf die Zeit nach der Ausbildung geworfen und Möglichkeiten des Transfers des Gelernten in den Unternehmenskontext diskutiert. Die Teilnehmer stellten sich gegenseitig Praxisprojekte vor, die sie begleitend zur Ausbildung durchgeführt hatten.

Die zentralen **Lern- und Entwicklungsfelder** der Ausbildung zum Innovationspromotor lassen sich den ausführlich skizzierten Kompetenzfeldern für die Trainingsprogramme zur Ausbildung von Beratern zuordnen (Teil II, Kap. 1.2). Lediglich die Förderung der Feldkompetenz spielte für die Innovationspromotoren keine gesonderte Rolle, da sie anders als Coaches oder Trainer in eine Organisation eingebunden sind und auch dort in ihrer neuen Rolle tätig werden sollten.

Auch die **Lehr- und Lernmethoden** der Ausbildung zum Innovationspromotor entsprachen denen, die für die Ausbildung von Beraterrollen bzw. für Kommunikations- und Verhaltenstrainings beschrieben wurden. In Rollenspielen, Gruppendiskussionen, Übungen zur Selbstreflexion u.a. wurden die erlernten Inhalte erlebbar gemacht.

### **2.1.3 Begründung der Auswahl des Evaluationsgegenstandes**

Die Ausbildung zum Innovationspromotor wurde aus verschiedenen Gründen für die vorliegende Evaluationsstudie ausgewählt: Zunächst fußt das Trainingskonzept - wie angedeutet - auf einer breiten Basis aus wissenschaftlicher Expertise und Erfahrungswissen. In die Konzeption der Ausbildung sind das Know How aller beteiligten Projektpartner aus Wissenschaft und Praxis und die ersten Projektergebnisse eingegangen. An der Humboldt Universität zu Berlin wird seit Jahren zu Themenfeldern der Innovationsforschung gearbeitet. Das Rollenbild des Innovationspromotors wurde wie beschrieben auf Basis der neusten Erkenntnissen der Innovationsforschung entwickelt und bildete das Grundgerüst der weiteren Ausbildungskonzeption. Die Inhalte der ersten Ausbildungshälfte zum Themenfeld der Innovation basieren ebenfalls auf den Ergebnissen jahrelanger Forschung in dem Feld (Scholl, 2010, 2004). Die Inhalte der zweiten Hälfte der Ausbildung sind klassische Themenfelder aus dem Bereich der Kommunikations- und Verhaltenstrainings, die von den artop-Beratern seit Jahren erfolgreich im Rahmen von Trainings bearbeitet werden. Die artop GmbH gilt als etablierter Anbieter auf dem Markt berufsbegleitender Ausbildungen und verfügt entsprechend über eine ausgewiesene Expertise auf dem Gebiet der Ausbildungsgestaltung.<sup>10</sup> Auf dieser Grundlage kann von einem erfolgreichen Verlauf der Ausbildung ausgegangen und daher der entwickelte Evaluationsansatz im Rahmen der Evaluation in Frage gestellt werden.

Darüber hinaus gab insbesondere der Zeitrahmen bzw. Aufbau der Maßnahme über mehrere Module den Ausschlag zur Auswahl der Ausbildung als exemplarischen Evaluationsgegenstand. Zum einen wurde eine engmaschige und tiefgehende evaluative Begleitung möglich. Zum anderen bot sich die Umsetzung eines Längsschnittdesigns in besonderer Weise an. Die komplexe Zielstellung der Ausbildung gesamt, aber auch der Lernziele für die einzelnen Module machten die Operationalisierung der Wirksamkeitskriterien und Entwicklung adäquater Erhebungsinstrumente zu einer besonderen Herausforderung.

---

<sup>10</sup> Einen Überblick über das Spektrum der bei artop angebotenen Ausbildungen gibt die Institutshomepage: [http://artop.de/3000\\_Ausbildungen/3000\\_Ausbildungen.html](http://artop.de/3000_Ausbildungen/3000_Ausbildungen.html).



## **2.2 Entwicklung eines *ganzheitlichen* Evaluationsansatzes**

Nachfolgend wird die Entwicklung eines *ganzheitlichen* Ansatzes zur Trainingsevaluation dargelegt, der sich auf die Entwicklung eines geeigneten Evaluationsmodells, eines Evaluationsdesigns und von Erhebungsmethoden bezieht. In jedem Abschnitt werden, wie bereits beschrieben, aus den grundsätzlichen Überlegungen zur Vereinbarkeit von wissenschaftlichem Anspruch und den Anforderungen nach praktikablen Lösungen im Feld Ableitungen für das Vorgehen im Rahmen der exemplarischen Trainingsevaluation anhand der Ausbildung zum Innovationspromotor getroffen.

### **2.2.1 Evaluationsmodell**

#### *Konzeptionelle Überlegungen*

Bei der Betrachtung aktueller Trainingsstudien zeigt sich der Trend, ergebnisbezogene und prozessbezogene Evaluationsmodelle zu verbinden. Dabei lässt sich eine Diskrepanz zwischen Theorieentwicklung und Umsetzung in der Evaluationspraxis feststellen. Während zunehmend komplexe integrierte Modelle entwickelt und als Königsweg der Trainings- und Transferevaluation postuliert werden (z.B. Hochholdinger u.a., 2008b), reicht die Evaluationspraxis über eine lose Koppelung der ergebnis- und prozessbezogenen Perspektive in der Regel nicht hinaus. Dabei wird der Fokus nach wie vor zumeist auf die Überprüfung von Kriterien der Trainingswirksamkeit nach Kirkpatrick gelegt (z.B. Saborowski & Muellerbuchhof, 2010; Rowold, 2008). Darüber hinaus wird versucht die Ergebnisvariablen mit einzelnen Prozessvariablen in Verbindung zu bringen. Die meisten Studien im Bereich der Trainings- und Transferevaluation fokussieren dabei auf einen der drei zentralen von Baldwin und Ford (1988) formulierten Einflussbereiche auf Lernen und Transfer. So untersucht Scharpf (1999) im Rahmen der Evaluation eines Rhetoriktrainings bspw. die Bedeutung spezifischer Lernermerkmale für den Trainingserfolg und den Praxistransfer. Auch Rowold (2008) untersucht im Rahmen einer Evaluationsstudie, inwieweit Trainee-Charakteristika, konkret die subjektiven Einstellungen der Teilnehmer, Ergebnisse auf den Evaluationsebenen nach Kirkpatrick beeinflussen. In der Trainingsevaluation setzt sich insgesamt der Trend durch, über die Frage „Wirkt das Training?“ hinaus zu gehen. Es wird gezielt nach Anhaltspunkten für mögliche Ursachen gesucht, sollten die Ergebnisse nicht optimal ausfallen und die Frage gestellt, welche Faktoren den Erfolg einer Maßnahme beeinflussen und wo Stellschrauben im Prozess liegen.

Das Anliegen der Studien besteht in der Regel darin, Erkenntnisse über die komplexen Wirkzusammenhänge im Trainings- und Transferprozess zu generieren. Die Frage, wie ein praktikabler Weg zur Verbindung der Modelle aussehen kann, rückt in den Hintergrund.

Bei der Entwicklung eines *ganzheitlichen* Evaluationsansatzes im Rahmen der Arbeit soll ein praktikabler Weg zur Kombination der ergebnisbezogenen und der prozessbezogenen Modelle gefunden werden. Aus wissenschaftlicher Perspektive ist diese Kombination der Ansätze insbesondere interessant, um Ansatzpunkte für kausale Erklärungen der Ergebnisse (vor allem im Fall mangelnder Wirksamkeit) zu generieren. Um die Praktikabilität des Evaluationsmodells zu gewährleisten, ist es ratsam, die Transferbedingungen als Randbedingungen für Wirksamkeit lediglich mit zu erheben und den Schwerpunkt der Erhebung nach wie vor auf die Outputs zu legen.

Trotz der angeführten Kritik an Kirkpatrick's Vier-Ebenen-Modell zur Trainingsevaluation (Teil II, Kap. 2.4.1.5) kann das Modell sowohl als hinreichend praktikabel gelten als auch wissenschaftlichen Anforderungen genügen, um den Kern des *ganzheitlichen* Evaluationsansatzes zu bilden. Während praktikable und ökonomische Anforderungen aufgrund der klaren und verständlichen Struktur und einfachen Systematik des Modells unumstritten erfüllt sind, müssen erst weitere wissenschaftliche Standards beachtet werden, damit das Modell als hinreichend wissenschaftlich gelten kann: Wie bereits in der Diskussion des Modells beleuchtet, hängt das wissenschaftliche Niveau der Umsetzung von Kirkpatrick's *Vier Ebenen* maßgeblich davon ab, wie diese operationalisiert und erhoben werden. So genügt das Modell nicht per se wissenschaftlichen Anforderungen. Werden aber für jede Ebene separat - unter Beachtung aktueller Forschungsergebnisse und der formulierten Leitlinien von Kirkpatrick - präzise Eingrenzungen von Erfolgskriterien vorgenommen, kann ein akzeptables wissenschaftliches Niveau erreicht werden. Eine ausführliche Beschreibung der für *geeignet* befundenen Erhebungsformen folgt im anschließenden Kapitel.

Aus wissenschaftlicher Perspektive ist es zudem von zentraler Bedeutung, eine Erhebung von Erfolgskriterien auf allen vier Ebenen des Modells anzustreben. Problematisch ist die Berufung auf Kirkpatrick in Evaluationsstudien vor allem dann, wenn wie bereits diskutiert, lediglich einzelne Ebenen erhoben werden und dennoch Aussagen über die Wirksamkeit einer Maßnahme abgeleitet werden (Teil II, Kap. 2.4.1.5).

Zwar können die Ergebnisse auf den einzelnen Ebenen auch für sich genommen relevante Informationen liefern. Erst im Zusammenspiel betrachtet lassen sich aber Folgerungen über die Wirksamkeit einer Maßnahme ziehen.

Dabei wird im Evaluationsansatz dieser Arbeit allerdings von der engen Kausalbeziehung der Ebenen abgesehen. Den bereits diskutierten Ergebnissen aktueller Studien folgend, gibt es keine validen Befunde für die Abhängigkeit der Ebenen (Teil II, Kap. 2.4.1.5). Ebenso wenig soll eine hierarchische Beziehung zwischen den Ebenen angenommen werden.

Die Fokussierung auf Outputvariablen entspricht einem summativen Evaluationsansatz. Die sorgfältige und systematische Anwendung von Messinstrumenten und -methoden steht damit im Vordergrund. Mit Hilfe von quantitativen Verfahren soll die Überprüfung von Hypothesen zum Lern- und Transfererfolg möglich werden.

#### *Ableitungen für die exemplarische Evaluation*

Im Rahmen der exemplarischen Evaluation der Ausbildung zum Innovationspromotor soll das beschriebene Evaluationsmodell unverändert zu Grunde gelegt werden. Es müssen keine weiteren Anpassungen aufgrund der Feldbedingungen vorgenommen werden, sodass die vorangehenden Ausführungen genauso auch für die konkrete Studie gelten können.

Auf Grundlage der Ergebnisse soll die Wirksamkeit der Maßnahme abschließend bewertet werden. Der Evaluation kommt damit vorrangig eine Kontroll-, bzw. Beurteilungsfunktion zu. Auf Basis der Ergebnisse können überdies Anhaltspunkte für die systematische Weiterentwicklung der Maßnahme generiert werden (Optimierungsfunktion).

## 2.2.2 Erhebungsmethoden

Die Definition präziser Evaluationskriterien, die praktikabel, erfassbar und valide sind, kann als erfolgskritische Größe der Wirksamkeitsevaluation gelten. Für die fünf entwickelten Evaluationskomplexe müssen daher Kriterien abgeleitet werden, die einerseits hinreichend standardisiert sind, um die Gültigkeit der Ergebnisse zu gewährleisten. Andererseits erfordern spezifische Zielsetzungen und Bedingungen im Feld oftmals maßgeschneiderte Methoden. Die Betrachtung aktueller Studien zeigt, dass ein Königsweg hier noch nicht gefunden ist und sich die Lösungswege insbesondere zwischen den Ebenen unterscheiden.

Nachfolgend werden zunächst auf konzeptioneller Ebene Wege zur Operationalisierung und Erhebung der Evaluationskriterien diskutiert, die einerseits wissenschaftlichen Anforderungen genügen und andererseits dem Kriterium der Praktikabilität entsprechen. Auf dieser Grundlage werden je Ebene die konkreten Erhebungsverfahren für die exemplarische Evaluation abgeleitet.<sup>11</sup>

### 2.2.2.1 Ebene Reaktion

#### Konzeptionelle Überlegungen

Die Ebene Reaktion wird in aktuellen Studien vergleichsweise einheitlich über standardisierte Fragebögen erhoben („Das Seminar hat Spaß gemacht“; „Insgesamt war ich mit dem Training sehr zufrieden“, z.B. Rowold, 2008; Weissweiler, 2008).

Das Erhebungsformat kann sowohl in der Entwicklung als auch im Einsatz und in der Umsetzung als praktikabel und ökonomisch gelten. Unter Einhaltung gewisser Standards bei der Fragebogenentwicklung kann auch der wissenschaftliche Anspruch gewahrt werden: So sollte jedem Item nur ein sachlicher Inhalt zugeordnet, mehrdeutige Begriffe und Verallgemeinerungen sollten hingegen vermieden werden (Bühner, 2006). Als Antwortformat bietet sich eine fünfstufige Ratingskala an, die die hinreichende Differenzierung der Antworten ermöglicht, ohne die Probanden mit zu vielen Antwortkategorien zu überfordern. Überdies sind die Leitlinien nach Kirkpatrick und Kirkpatrick (2006) bei der Erhebung der Ebene zu beachten (Teil II, Kap. 2.4.1.1).

Unter Berücksichtigung dieser Aspekte kann die Erhebung der Ebene Reaktion im Rahmen des *ganzheitlichen* Ansatzes sinnvoll über einen standardisierten Fragebogen erhoben werden. Das konkrete Verfahren wird nachfolgend dargestellt.

---

<sup>11</sup> Die in der Untersuchung verwendeten Erhebungsverfahren sind im Anhang einsehbar (Appendix B, S. 113 ff).

### Ableitung des Erhebungsverfahrens für die exemplarische Evaluation

In dem Beratungsinstitut artop GmbH existierte bereits ein Fragebogen, der jahrelang erfolgreich als Feedbackinstrument in verschiedenen Ausbildungsgängen eingesetzt wurde und sich zur Überprüfung der *Reaktion* bewährt hat. Ein Abgleich mit den Anforderungen, welche Kirkpatrick und Kirkpatrick (2006) an ein Erhebungsinstrument für die Messung der Zufriedenheit formulieren, fiel positiv aus. So bietet der Fragebogen eine Untergliederung in die Zufriedenheit mit dem Trainer (8 Items), den Inhalten (4 Items) und der Didaktik (5 Items). Eine fünfstufige Likert-Skala, auf der die Probanden von 1 *stimmt ganz und gar nicht* bis 5 *stimmt voll und ganz* Zutreffendes auswählen können, ermöglicht eine Quantifizierung der Ergebnisse. Zudem können die Probanden am Ende des Fragebogens offene Kommentare zu positiven und negativen Aspekten abgeben. In einer Überarbeitungsphase wurde lediglich die Formulierung der Items an die zu evaluierende Ausbildung angepasst (*Ich habe viele Anregungen für eine spätere Tätigkeit als Innovationspromotor/in erhalten*). Da alle Module von einem Trainertandem durchgeführt wurden, wurde eine Unterscheidung in Trainer 1 und Trainer 2 eingeführt, womit sich insgesamt 16 Items (2 x 8) für die Bewertungsdimension *Trainer* ergaben.<sup>12</sup>

#### **2.2.2.2 Ebene Lernen**

##### Konzeptionelle Überlegungen

Während sich die Erhebung der Ebene Reaktion also vergleichsweise einfach umsetzen lässt, erweist sich die Operationalisierung und Erhebung der Lernebene als deutlich schwieriger. In aktuellen Evaluationsstudien setzen sich dabei vor allem zwei Wege durch: Zum einen werden Wissenstests entwickelt, die die Inhalte des Trainings abbilden sollen. Da die Trainings in der Regel auf die Förderung relativ eingrenzter Kompetenzbereiche zielen (*Zeitmanagement, Rhetorik etc.*), werden zum anderen standardisierte Verfahren aus den theoretischen Konstrukten abgeleitet oder bereits bestehende und validierte Instrumente eingesetzt.

So erhebt Weissweiler (2008) den Lernerfolg in der Evaluation eines Zeitmanagementtrainings einerseits mittels eines Wissenstests, der aus offenen Fragen zu den behandelten Inhalten besteht. Andererseits setzt sie das bereits bestehende Inventar zur Erfassung von Lernstrategien im Studium ein (LIST, Wild & Schiefele, 1994).

---

<sup>12</sup> Die Anpassungen an dem Erhebungsinstrument auf der Ebene Reaktion hat die andere Projekt-Diplomandin Frau Grit Rudinger vorgenommen.

Rowold (2008) entwickelt zur Erhebung des Lernerfolges im Rahmen der Evaluation eines Trainings zur Förderung der Kundenorientierung für Call-Center Mitarbeiter ein standardisiertes Testverfahren, welches die im Training vermittelten Inhalte als multiple-choice-Fragen erfasst. Saborowski und Muellerbuchhof (2010) setzen im Rahmen der Evaluation eines Selbstmanagementtrainings hingegen auf den Einsatz standardisierter und publizierter Verfahren. Neben dem Fragebogen zum Lernen in der Arbeit (LIDA, Bergmann 2007) bedienen sie sich einer Skala aus dem Fragebogen zum Vorgehen in Problemsituationen (PRO-LÖ, Bergmann 2007).

Zwar ist der Einsatz publizierter Verfahren auf der Ebene Lernen durchaus wünschenswert, jedoch kann nicht für jedes Training davon ausgegangen werden, dass es sich bei den Lernbereichen um relativ eingegrenzte Kompetenzbereiche handelt, für die bereits Erhebungsverfahren existieren. Auch der klassische Wissenstest ist für Kommunikations- und Verhaltenstrainings ein eher ungeeignetes Format, da es in diesen Trainings nicht primär um die Vermittlung kompakter Wissensinhalte geht.

Im Rahmen eines *ganzheitlichen* Evaluationsansatzes, der sich auf verschiedene Trainingsformate übertragen lässt, müssen also neue Optionen für die Erhebung der Lernebene durchdacht werden. In Anlehnung an Kirkpatrick und Kirkpatrick (2006) soll ein praktikabler Weg im Rahmen dieser Arbeit darin bestehen, die Erfolgskriterien aus den *Lernzielen des Trainings* abzuleiten. Als Erhebungsverfahren soll wieder ein standardisierter Fragebogen als praktikables Format gewählt werden, welches unter Einhaltung der bereits beschriebenen Konstruktionsstandards auch wissenschaftlichen Anforderungen genügt.

Um die Erfolgskriterien auf der Lernebene differenziert zu erfassen, soll die Verarbeitung der Lerninhalte im Hinblick auf zwei Perspektiven eingeschätzt werden: Zum einen sollte erfasst werden, inwieweit die Teilnehmer mit der Definition und Bedeutung der Items vertraut sind, also über *deklaratives Wissen* im Sinne von Faktenwissen verfügen. Zum anderen interessiert die Frage, inwieweit die Teilnehmer mit der Umsetzung der Bereiche vertraut sind, also über *prozedurales Wissen* im Sinne von Handlungswissen verfügen.

### Ableitung des Erhebungsverfahrens für die exemplarische Evaluation

Anders als in den „klassischen“ Kommunikations- und Verhaltenstrainings, die auf die Entwicklung eines konkreten Kompetenzbereiches abzielen, ging es in der Ausbildung um die Entwicklung einer professionellen Rolle. Dafür waren ganz unterschiedliche Wissensinhalte und Kompetenzbereiche relevant, die in den unterschiedlichen Modulen behandelt wurden. Den konzeptionellen Überlegungen zur Erhebung der Lernebene folgend, wurde ein intuitiv-rationales Vorgehen für die Entwicklung der Items erwogen (Tränkle, 1983), das sich an den globalen Lernzielen je Modul orientierte.<sup>13</sup> Damit sollte gewährleistet werden, dass auf der einen Seite die abstrakten und umfassenden Zielstellungen je Modul hinreichend konkretisiert werden, ohne auf der anderen Seite zu Lasten des Informationsgehaltes zu gehen.

Unter Einbeziehung der im Rahmen des GI:VE Fragebogens erhobenen Konstrukte wurden für jedes Modul zunächst Lernziele festgelegt. In Absprache mit den Ausbildern wurden schließlich 4-5 Items je Modul ausgewählt, die die Lerninhalte am besten repräsentieren konnten. Die Items beziehen sich je nach Modul auf Inhalte aus dem Themenfeld Innovation („aktive Unterstützung von Innovationsprozessen“, „innovationsförderliche Arbeitsplatzgestaltung“) oder repräsentieren Methoden und Verhaltensweisen („Moderation“, „zweckvolle Konfliktbehandlung“, „unternehmerisches Handeln“). Die Formulierungen wurden dabei bewusst global und knapp gehalten. So stellen die Items zum Teil lediglich einzelne Begriffe wie „Mitarbeiterorientierung“ und „Gesprächsführung“ dar. Oftmals werden die Begriffe um ein erläuterndes Adjektiv ergänzt („nachhaltige Konfliktbewältigung“) oder kurze Aussagesätze als Items formuliert („angemessener Umgang mit Widerständen“, „Lernen in und mit der Organisation“). Der Gesamtfragebogen umfasste schließlich 36 Items.

Um die Erhebung des deklarativen und prozeduralen Wissens umzusetzen, wurden die Teilnehmer gebeten für jedes Item anzugeben, inwieweit sie das Konstrukt *kennen* und inwieweit sie es praktisch *umsetzen*. Das Antwortformat entsprach wie bereits erwähnt einer fünfstufigen Ratingskala mit Abstufen von *trifft überhaupt nicht* zu bis *trifft absolut zu*. Zu jedem Messzeitpunkt waren die Teilnehmer angehalten alle Items zu bewerten, unabhängig davon, ob diese bereits trainiert worden waren. Die Festlegung der Item-Reihenfolge erfolgte durch Randomisierung, sodass auch nicht auf den ersten Blick ersichtlich war, welches Item welches Modul repräsentiert.

---

<sup>13</sup> Das Erhebungsverfahren auf der Ebene Lernen wurde von Frau Grit Rudinger entwickelt.

### 2.2.2.3 Ebene Verhalten

#### Konzeptionelle Überlegungen

Die Betrachtung aktueller Trainingsevaluationen zeigt, dass auf der Verhaltensebene statt standardisierter Testverfahren üblicherweise spezifische Methoden entwickelt werden, die Aussagen über Verhaltensänderungen ermöglichen sollen (*Rollenspiele, Simulationsverfahren, Videoanalysen etc.*). So erstellt Weissweiler (2008) zur Erfassung des Zeitmanagementverhaltens eine Simulationsübung „Postkorb zur Erfassung von Zeitmanagementverhalten“, welche die Versuchspersonen etwa drei Monate nach dem Training bearbeiten. Rowold (2008) entwickelt zur Erhebung der Verhaltensebene ein Rollenspiel, welches mit jedem Teilnehmer direkt im Anschluss an das Training durchgeführt wird. Während die Ebenen Reaktion und Lernen in einem Gros der Evaluationen über Selbsteinschätzungen der Teilnehmer erhoben werden, wird auf der Verhaltensebene, wie auch von Kirkpatrick empfohlen auf Fremdurteile gesetzt. Hierbei werden z.B. Trainer, Kollegen oder gezielt geschulte Beobachter zur Einschätzung des Verhaltens herangezogen (Weissweiler, 2008; Rowold, 2008).

Für den *ganzheitlichen* Evaluationsansatz sind die beschriebenen Verfahren vor allem im Hinblick auf die Prämisse der Praktikabilität zu problematisieren. Sowohl bei der Entwicklung als auch bei der Durchführung und Auswertung derartiger Verfahren ist der zeitliche und auch personelle Aufwand (z.B. Beobachter) immens. Unabhängig davon ist es unter inhaltlichen Gesichtspunkten wünschenswert, die Instrumente der Lern- und Verhaltensebene stärker aufeinander zu beziehen und keine voneinander losgelösten Verfahren einzusetzen. Der Ansatzpunkt für die Erhebung der Verhaltensebene im Rahmen eines *ganzheitlichen* Evaluationsansatzes wird daher zunächst über die Lernebene gewählt. Aus den abstrakt gehaltenen Items zur Beschreibung der Lerninhalte können konkrete Verhaltensaussagen formuliert werden, welche dann im Rahmen eines Fragebogens eingeschätzt werden können. Dabei soll - wie in der Literatur empfohlen und auch in den meisten Evaluationsstudien umgesetzt - auf Fremdurteile gesetzt werden. Als geeignete Personen aus dem Funktionsfeld, sollen Kollegen für die Befragung herangezogen werden. So kann grundsätzlich davon ausgegangen werden, dass Veränderungen im Verhalten der Teilnehmer für die Kollegen in der täglichen Zusammenarbeit erfahrbar sind und implizit beobachtet werden. Zwar ist die Befragung von Kollegen mit einem höheren Aufwand verbunden als die Befragung der Trainingsteilnehmer selbst.



Aus wissenschaftlicher Perspektive haben die Ergebnisse aus einer Fremdbeurteilung, insbesondere auf der Verhaltensebene, aber eine höhere Güte als Selbsturteile. Im Hinblick auf die Praktikabilität der Erhebung auf der Verhaltensebene entsprechen sowohl das standardisierte Verfahren eines Fragebogens als auch die Personengruppe der Kollegen den Anforderungen. So ist davon auszugehen, dass die Kollegen über die Trainingsteilnehmer noch vergleichsweise unkompliziert für eine Befragung zu erreichen sind und eine grundsätzliche Bereitschaft besteht, für ihre Kollegen an einer Befragung auch teilzunehmen.

#### Ableitungen des Erhebungsverfahrens für die exemplarische Evaluation

Den konzeptionellen Überlegungen für die Erhebung der Verhaltensebene folgend, wurde als Grundlage für die Beschreibung von Verhaltensweisen der künftigen Innovationspromotoren der Lernfragebogen herangezogen. Während man davon ausgehen konnte, dass die Teilnehmer die abstrakten Items als Lerninhalte der Ausbildung wiedererkennen und ihren Gehalt erfassen können, bestand bei den Kollegen Erläuterungsbedarf. Für jedes Item auf der Lernebene wurde also eine konkrete Verhaltensbeschreibung vorgenommen.<sup>14</sup> Einzelne Items, die zu komplex waren, wurden nochmal in mehrere Items untergliedert („aktive Unterstützung von Innovationsprozessen“ → „ermuntert andere aktiv, ihre Ideen zu äußern“ + „unterstützt die Auswahl von Ideen“ + „setzt sich für eine erfolgreiche Umsetzung neuer Ideen ein“).<sup>15</sup> Die Items „strategische Personalentwicklung“ und „strategische Projektplanung“ wurden für den Verhaltensfragebogen hingegen gar nicht berücksichtigt, da davon ausgegangen werden musste, dass sie sich in der Fremdbeobachtung nicht sinnvoll erfassen lassen bzw. sich dem Beurteilungsradius der Kollegen entziehen. Um die Verständlichkeit der Items zu prüfen, wurde eine Prätestung an Kommilitonen und bekannten Personen aus Unternehmen vorgenommen. Der Fragebogen wurde überarbeitet und modifiziert. Die finale Version umfasste schließlich 34 Items. Da es aus organisatorischen Gründen nicht möglich war, die Kollegen bereits vor der Ausbildung zu befragen, sollte die Verhaltensänderung retrospektiv eingeschätzt werden. Für jedes Item wurden die Kollegen um ihre Einschätzung zu dem *gegenwärtigen Verhalten* und zu dem *Verhalten vor einem Jahr* gebeten. Das Antwortformat entsprach auch hier einer fünfstufigen Ratingskala mit Abstufungen von *trifft überhaupt nicht zu* bis *trifft absolut zu*.

---

<sup>14</sup> Eine Übersicht zu den Ausgangs-Items des Lernfragebogens und den daraus abgeleiteten Items des Fragebogens der Verhaltensebene findet sich im Anhang (Appendix B3.2, S. 126 ff.).

<sup>15</sup> Bei den Berechnungen wurden die drei Items aber wieder zu einem Item aggregiert.

#### 2.2.2.4 Ebene Resultate

##### Konzeptionelle Überlegungen

Die Erhebung der vierten Ebene im Rahmen von Trainingsevaluationen stellt nach wie vor eine der größten Herausforderungen dar. Wenngleich die große Bedeutung der Ebenen im Gesamtmodell einhellig postuliert wird (Kirkpatrick & Kirkpatrick, 2006; Rowold, 2008), finden sich kaum Studien, in denen die präzise Erfassung von Erfolgsfaktoren auf dieser Ebene gelingt. Das größte Problem liegt in der eindeutigen Rückführung möglicher Effekte auf die Wirkung eines Trainings. Alle vorgeschlagenen Kennzahlen für die Ebene (*Kostenreduktion, Steigerung der Produktivität, Konfliktrate, Mitarbeiterzufriedenheit, Fehlzeiten und Krankentage, Fehlerhäufigkeit, etc.*) (Teil II, Kap. 2.4.1.4) sind prinzipiell auch auf zahlreiche andere denkbare Entwicklungen und Einflussgrößen in der Organisation zurückzuführen. Bei der Evaluation von Kommunikations- und Verhaltenstrainings wird das Problem noch potenziert. Wie lassen sich die in der Regel *weichen Kriterien* des Erfolges, wie Einstellungs- und Verhaltensänderung in präzise Kriterien auf der vierten Ebene übersetzen? Am ehesten bieten sich die eher weichen Faktoren an, die in den Bereich des Arbeits- und Organisationsklimas oder der Qualität der Leistung fallen. Diese lassen sich aber nicht allgemeingültig bestimmen, sondern müssen an die zu überprüfende Maßnahme bzw. an die betroffenen Unternehmen angepasst werden. Eine empirische Überprüfung möglicher Wirkungen auf dieser Ebene war aus Gründen des Aufwands und datenschutzrechtlicher Einschränkungen im Rahmen der vorliegenden Untersuchung leider nicht möglich. Dennoch sollte die Untersuchung genutzt werden, um an der Lösungssuche für die systematische Erhebung dieser Ebene zu arbeiten.

##### Ableitungen für die exemplarische Evaluation

Die Expertise der Teilnehmer und ihr heterogener Unternehmensbackground sollten gezielt genutzt werden, um zu eruieren, worin mögliche Erfolgskriterien auf der Ebene Resultate liegen könnten. Im Rahmen eines telefonischen Interviews wurden die Teilnehmer gebeten sich mit folgenden Fragen auseinanderzusetzen:

1. *Welche Maßnahmen zur Innovationspromotion konntest du seit der Teilnahme an der Ausbildung in deinem Unternehmen umsetzen?*
2. *Gibt es in deinem Unternehmen Kennzahlen, die deine Arbeit als Innovations-Promotor reflektieren (bspw. im Rahmen der unter 1 genannten Maßnahme)?*

Frage 1 sollte dabei vor allem dazu dienen, die Ableitung von konkreten Kennzahlen aus der zunächst mal vielfältigen Arbeit eines Innovationspromotors zu erleichtern.

### 2.2.2.5 Transferbedingungen

#### Konzeptionelle Überlegungen

Für die Erhebung der Transferbedingungen soll im Rahmen der Arbeit die deutsche Version des Lerntransfer-System-Inventar (GLTSI) (Kauffeld, 2008) als validiertes Verfahren eingesetzt werden. So ist es insgesamt ratsam wenigstens in einem Erhebungskomplex auf ein validiertes Verfahren zurückzugreifen.

Nachfolgend wird die Struktur des Instrumentes knapp skizziert: In Anlehnung an das Rahmenmodell zur Beschreibung des Transferprozesses von Baldwin und Ford (1988), entwickelten Holton, Bates & Ruona (2000) das Lerntransfer-System-Inventar zur Erhebung der Einflussvariablen auf Transfererfolg. Das Instrument bildet die drei Wirkkomplexe Trainee-Charakteristika, Trainingsdesign und Arbeitsumgebung differenziert nach 16 Subdimensionen ab. Dabei werden 11 **spezifische Faktoren** berücksichtigt, die sich unmittelbar auf ein zu evaluierendes Training beziehen und 5 **generelle Faktoren**, welche den Lerntransfer unabhängig von der spezifischen Maßnahme beeinflussen (Holton u.a. 2000). Kauffeld und Kollegen (2008) haben das Verfahren ins Deutsche übersetzt und mittels explorativer Faktorenanalysen die Struktur des Original-LTSI (Bates, Kauffeld, Holton, 2007) in der deutschen Version bestätigt. In der Validierung des Verfahrens wurden einzelne Items des Original Instrumentes ersetzt, um die psychometrischen Eigenschaften des Verfahrens zu verbessern. Es liegen weitere 16 überprüfte Versionen des Instrumentes in 14 verschiedenen Sprachen vor, womit das LTSI als global validiertes Verfahren gelten kann (Kauffeld, 2010). Abbildung 3 zeigt die zugrundeliegende Struktur des Verfahrens.

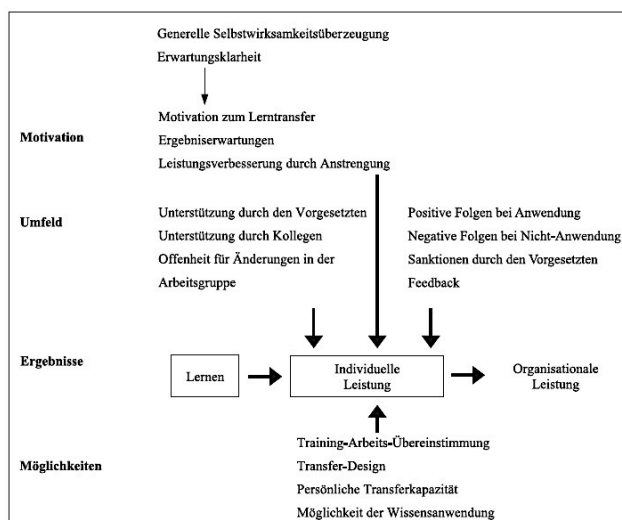


Abbildung 3: Konzeptuelles Modell des GLTSI (Kauffeld, 2008)

Das Verfahren besteht aus 67 Aussagen, für die die Teilnehmer gebeten werden, auf einer Skala von 1 (*stimme überhaupt nicht zu*) bis 5 (*stimme völlig zu*) ihre Einschätzung abzugeben. Die Reliabilität der Skalen wird als gut bis zufriedenstellend beschrieben (Kauffeld, 2008). Das Inventar soll in seiner Gesamtheit dazu dienen, Informationen zu liefern, „wie Bedingungen zu gestalten sind, um Trainingsmaßnahmen wirksamer und nutzbarer zu machen“ (Kauffeld, 2008, S. 52). Während der Einsatz des Instrumentes grundsätzlich verschiedene Funktionen erfüllen kann, soll er im Rahmen des *ganzheitlichen* Evaluationsansatzes die Ergebnisse auf den Evaluationsebenen nach Kirkpatrick sinnvoll ergänzen und daher als **Diagnoseinstrument** möglicher bestehender Transferprobleme zum Einsatz kommen.

#### Ableitungen des Erhebungsverfahrens für die exemplarische Evaluation

Im Rahmen der eigenen Erhebung wurden ausschließlich die spezifischen Faktoren des GLTSI erhoben, die sich auf die konkret zu evaluierende Trainingsmaßnahme beziehen. Aus den 11 Skalen wurden noch einmal 9 Skalen ausgewählt, die sich für die vorliegende Untersuchung eignen. Nicht berücksichtigt wurden die Skalen *Positive Folgen bei Anwendung* und *Negative Folgen bei Nicht-Anwendung*, da es sich um sehr spezifische Skalen handelt, die für die Teilnehmer der Untersuchung (*überwiegend Mitarbeiter in leitenden Positionen*) nicht passend erschienen. Um Umfang und Aufwand der Erhebung gering zu halten, wurde zudem auf die Erhebung der generellen Erfahrungen mit Trainings verzichtet. Die vorherigen Erfahrungen der Teilnehmer mit Trainings beeinflussen zwar ebenfalls den Lerntransfer, schienen aber für die vorliegende Untersuchung nicht so bedeutsam zu sein.

### 2.2.3 Evaluationsdesign

#### Konzeptionelle Überlegungen

Als Konsequenz der theoretischen Auseinandersetzung mit den Designfragen im Rahmen von Evaluationen steht die Erkenntnis, dass sich die Wahl eines geeigneten Untersuchungsdesigns in besonderer Weise in einem Spannungsfeld zwischen wissenschaftlichem Anspruch und den spezifischen Bedingungen einer Feldstudie bewegt. Es gilt im Rahmen dieser Arbeit ein Evaluationsdesign zu konstruieren, welches einerseits die Anzahl plausibler Alternativerklärungen für die Ergebnisse minimiert und gleichzeitig eine hohe Natürlichkeit der Untersuchungsbedingungen sicherstellt. Die Betrachtung von aktuellen Evaluationsstudien zeigt, dass sich die Autoren gezielt darum bemühen, wissenschaftlichen Ansprüchen gerecht zu werden.

Zwar überwiegen nach wie vor quasi-experimentelle Designs, die sich nur bedingt zur Kontrolle von Störfaktoren eignen. Es werden aber überwiegend sowohl mehrere Messzeitpunkte als auch Kontrollgruppen einbezogen. So wenden Saborowski und Muellerbuchhof (2010) in ihrer Studie zur Überprüfung der Wirksamkeit eines Selbstmanagementtrainings ein multivariates Versuchs-Warte-Kontrollgruppen-Design mit Prä- und Post-Messung an. Auch Lang und Otto (2010) evaluieren die Trainingsmaßnahme zur Verbesserung strategischer Verhandlungsplanung im Rahmen eines Prätest-Posttest-Kontrollgruppendedesigns. Die Kontrollgruppe erhielt dabei gar keine Trainingsmaßnahme. Auch Merzenich-Hieker (1996) wählt in ihrem Prätest-Posttest-Kontrollgruppen-Design Personen für die Kontrollgruppe aus, die entweder kein Training erhielten oder bereits früher an dem untersuchten Training teilgenommen hatten. Die Anzahl der Messzeitpunkte geht in den meisten Designs über zwei (Prä-/Post-Testung) nicht hinaus. Dabei liegt der Schwerpunkt der Untersuchungen oftmals auf der Frage nach den Transferwirkungen, was einen dritten Zeitpunkt eigentlich unumgänglich macht (z.B. Weissweiler, 2008). Andere Studien verzichten hingegen ganz auf die Variation der Messzeitpunkte und setzen lediglich Querschnittsdesigns um (z.B. Rowold, 2008). Die Anzahl der Studien, die Längsschnittsdesigns mit mehr als zwei Messzeitpunkten realisieren, ist insgesamt rar.

Damit kann ein Forschungsdesiderat für aktuelle Trainingsevaluationen konstatiert werden, das im Rahmen eines *ganzheitlichen* Evaluationsansatzes bearbeitet werden soll. Es soll ein Längsschnittsdesign verwirklicht werden, das mehrere Messzeitpunkte vor, während und nach der Trainingsmaßnahme vorsieht und damit Aussagen zum

Lern- und Transfererfolg des Trainings erst möglich macht. Um Aussagen über Veränderungen an den Trainingsteilnehmern festzustellen, wird im Sinne einer Paneluntersuchung die gleiche Stichprobe, nämlich die Trainingsgruppe, zu verschiedenen Messzeitpunkten untersucht. Durch die Umsetzung mehrerer Messzeitpunkte sollen gleichzeitig die unzureichenden Möglichkeiten, eine adäquate Kontrollgruppe zur Trainingsgruppe zu bilden, ausgeglichen werden. Zwar werden - wie beschrieben - in aktuellen Studien überwiegend Kontrollgruppen eingesetzt. Inwieweit es für kommunikations- und verhaltensorientierte Trainings aber überhaupt sinnvoll möglich ist, eine Vergleichsgruppe zur Trainingsgruppe zu bilden, die bis auf die Trainingsmaßnahme, über vergleichbare Bedingungen und Merkmale verfügt, bleibt fraglich. Erhält die Vergleichsgruppe überdies gar keine Intervention und wird zudem nicht zufällig zugewiesen, können Unterschiede zwischen den Gruppen nur sehr bedingt auf die Wirkungen der Maßnahme zurückgeführt werden. Dies ist aus wissenschaftlicher Perspektive bedenklich. Zudem müsste während der Zeit der Intervention immer wieder geprüft werden, ob die Vergleichbarkeit noch gegeben ist, was wiederum aus Zeit- und Kostengründen nicht gewährleistet werden kann. Eine Warte- oder Placebo-Kontrollgruppe wird ebenfalls eher aus Aufwandsgründen und ethischen Bedenken abgelehnt (Teil III, Kap. 1.).

Unter Abwägung der beschriebenen Diskussionspunkte im Hinblick auf die Wissenschaftlichkeit und Praktikabilität möglicher Evaluationsdesigns wird also ein sogenanntes Einzel-Gruppen-Design konstruiert, das für die Kontrolle von Störfaktoren stärker als experimentelle Designs auf die Variation der Anzahl der Messungen angewiesen ist und daher als Längsschnittuntersuchung angelegt wird. Statt der Differenz zwischen Trainings- und Kontrollgruppe rückt damit die Differenz zwischen den Messzeitpunkten in den Mittelpunkt der Betrachtung. Diese wird auf die Wirkung der Intervention zurückgeführt. Einschränkend muss aber bereits an dieser Stelle darauf verwiesen werden, dass das Design keine Möglichkeit zur eindeutigen Interpretation der Ergebnisse bietet. So kann der Einfluss von Störfaktoren nur sehr bedingt kontrolliert werden. Zwar ist auch die Umsetzung mehrerer Messzeitpunkte vergleichsweise aufwendig. Geht man jedoch davon aus, dass zu den verschiedenen Messzeitpunkten jeweils das gleiche Erhebungsverfahren eingesetzt wird, kann das Vorgehen dennoch als hinreichend praktikabel gelten. Es wird in Kauf genommen, dass das Untersuchungsdesign im Rahmen der *ganzheitlichen* Trainingsevaluation damit von den Empfehlungen von Kirkpatrick und Kirkpatrick (2006) abrückt.

### Ableitungen für die exemplarische Evaluation

Den konzeptionellen Überlegungen zu den Designfragen folgend, soll im Rahmen der exemplarischen Evaluation ein Längsschnittdesign ohne Kontrollgruppe umgesetzt werden. Aus inhaltlichen Gesichtspunkten, aber auch aus Gründen der Umsetzbarkeit, unterscheidet sich die Anzahl der Messungen auf den vier Ebenen. Nachfolgend wird immer auch die Durchführung der Messungen je Ebene erläutert.

Auf der **Ebene Reaktion** erfolgte eine Messung nach jedem Modul (= 10 Messzeitpunkte). Da sich die Module inhaltlich unterscheiden und auch von verschiedenen Ausbildertandems durchgeführt wurden, ist die Bewertung nach jedem Modul unumgänglich, um ein differenziertes Bild über die Zufriedenheit mit der Ausbildung zu erhalten. So kann bezogen auf jedes Modul Änderungsbedarf identifiziert werden. Am Ende jeder Sitzung wurde dazu der Fragebogen ausgeteilt und auch zur Formulierung offener Kommentare angeregt. Die Beantwortung des Fragebogens dauerte zwischen 5 und 10 Minuten. Die Herausforderung für den Evaluator bestand darin, trotz der vorherrschenden Aufbruchsstimmung gegen Sitzungsende, eine Stimmung zu erzeugen, die die Teilnehmenden in die notwendige Ernsthaftigkeit und Ruhe für die Beantwortung der Fragen versetzte. Damit sollte einer oberflächlichen Beantwortung der Fragen im Vorhinein begegnet werden. Neben einer Papier-Version wurde eine onlinegestützte Version des Fragebogens in dem Onlineportal Unipark erstellt, um einen möglichst hohen Rücklauf zu erzielen.<sup>16</sup>

Auf der **Ebene Lernen** ging es anders als auf der Ebene Reaktion explizit darum, einen Zuwachs an deklarativem und prozeduralem Wissen im Verlauf der Ausbildung festzustellen. Der Abgleich einer Prä- und Posttestung war damit zentral. Neben einer Prätestung zur Erhebung des Ausgangsniveaus, erfolgte eine Zwischenerhebung nach dem fünften und eine weitere Messung einige Zeit nach dem zehnten Modul. Aus organisatorischen Gründen wurde die Prätestung erst im Anschluss an das erste Modul durchgeführt. Da hier aber Kennenlernen und Einfinden im Vordergrund standen, wird die erste Messung trotzdem im Sinne einer Prämessung behandelt. Für den Lernfragebogen wurde ebenfalls eine onlinegestützte Version erstellt.<sup>17</sup> Nach Modul 1 und 5 erhielten die Teilnehmer eine E-Mail mit der Aufforderung an der Befragung teilzunehmen und knappen Hinweisen zu Zweck und Vorgehen der Erhebung. Die Beantwortung der Befragung nahm etwa 20 Minuten in Anspruch.

---

<sup>16</sup> Die onlinegestützte Version des Verfahrens wurde von Frau Grit Rudinger erstellt.

<sup>17</sup> Die onlinegestützte Version des Verfahrens erstellte ebenfalls Frau Grit Rudinger.

Auch darüber wurden die Teilnehmer vorab informiert. So sollte gewährleistet werden, dass sie nicht unter Zeitdruck und in einer unpassenden oder hektischen Situation den Fragebogen ausfüllten.

Die Kollegenbefragung auf der ***Ebene Verhalten*** fand eine Woche nach Abschluss der Ausbildung statt. Auch hierfür wurde eine Onlineversion des Fragebogens auf der Plattform Unipark erstellt. Um die Anonymität zu gewährleisten, wurden die Kollegen nicht direkt kontaktiert. Die Teilnehmer bekamen eine vorgefasste E-Mail an die Kollegen mit dem Link zum Befragungsportal und sollten diese nach Eintragung ihres Codes an die auserwählten Kollegen weitersenden. So sollte den Teilnehmern ein Gefühl der Kontrolle und Sicherheit gegeben werden. Andererseits ist die Gefahr, dass Teilnehmer den Fragebogen einfach selbst ausfüllen oder in anderer Weise Einfluss auf das Antwortverhalten der Kollegen nehmen, nicht gänzlich ausgeschlossen. Die Möglichkeiten der positiven Beeinflussung der Rücklaufquote sind für den Evaluator in diesem Setting ebenfalls begrenzt. Auch für diese Erhebung wäre die Umsetzung eines Prä-/Posttest-Designs wünschenswert gewesen. Dieses ließ sich aber aus Aufwandsgründen nicht umsetzen. Stattdessen wurde, wie bereits erörtert, ein retrospektives Vorher-Nachher-Design gewählt, das die Ausgangssituation über retrospektive Fragen ermittelt.

***Ebene Resultate:*** Die telefonischen Interviews wurden kurz nach Ausbildungsabschluss geführt und nahmen jeweils etwa 10 Minuten in Anspruch.

Die ***Transferbedingungen*** wurden ebenfalls ausschließlich in einer Posttestung erhoben. Im Rahmen der Abschlusserhebung - eine Woche nach Ende der Ausbildung - wurden neben dem Lernfragebogen auch die ausgewählten Items des GLTSI zur Beantwortung online gestellt.

#### Übergreifende Anmerkungen

Die Untersuchung erstreckte sich insgesamt über zehn Monate von April 2011 bis Januar 2012. In dieser Zeit fanden die zehn Module der Ausbildung in Abständen von etwa vier Wochen statt, also etwa ein Modul pro Monat. Die Daten auf der Ebene Reaktion für die Module 1 bis 5 wurden von Frau Grit Rudinger erhoben. Zudem führte sie die Prämessung und die Zwischenerhebung auf der Ebene Lernen durch. Alle weiteren Messungen der Untersuchung habe ich selbst durchgeführt. Die Module 6 bis 10 begleitete ich als Evaluatorin und Protokollantin der Ausbildung.



Um dem Auftreten von Störeinflüssen im Vorhinein zu begegnen, wurde bei der Untersuchungsdurchführung auf die Einhaltung wissenschaftlicher Standards geachtet. Die Teilnehmer wurden im ersten Modul über die begleitende Evaluation informiert und um ihre Mitarbeit gebeten. Dabei wurden die Zielstellung, der Rahmen und das Vorgehen des Vorhabens erläutert und darauf geachtet, Transparenz im Umgang mit den Daten und Offenheit für Nachfragen zu signalisieren. Jeder Erhebung wurde eine ausführliche Instruktion vorangestellt, in der Zielsetzung und Vorgehen der Befragung noch einmal skizziert wurden. Es wurde betont, dass es bei der Erhebung darum gehe die Wirksamkeit der Ausbildung und nicht etwa die *Leistung* der Teilnehmer zu überprüfen. Damit sollten Ängste, Widerstände, die aus Unsicherheit und mangelndem Vertrauen in die Untersuchung resultieren und bspw. zu sozial erwünschtem Antwortverhalten führen, im Vorhinein minimiert werden.

Wie im Theorieteil beschrieben, kann es zur Prüfung der zeitlichen Stabilität von Effekten insbesondere sinnvoll sein, mehrere Messungen in unterschiedlichen Zeitabständen nach der Intervention vorzunehmen. Die Ausbildung ist aufgrund ihres enormen Umfangs ein Sonderfall. Es ist davon auszugehen und wünschenswert, dass die Teilnehmer, anders als bei einem punktuellen Training, nicht erst nach Abschluss der Maßnahme Änderungen im Verhalten zeigen. Der Messzeitpunkt für die Ebene Verhalten wurde daher als geeignet für die Transfermessung befunden. Kirkpatrick und Kirkpatrick (2006) Empfehlung „Allow time for behaviour change to take place“ (p. 53), wurde erfüllt. Darüber hinaus wäre es natürlich wünschenswert gewesen, weitere Nachher-Messungen einige Zeit nach der Ausbildung umzusetzen. Dies konnte aus Aufwandsgründen aber nicht realisiert werden. Die nachfolgende Abbildung ermöglicht einen Überblick aller Kriterien, Instrumente und Erhebungszeitpunkte für jede Ebene.

<b>Kriterium /Einflussvariable</b>	<b>Instrument</b>	<b>Quelle</b>	<b>Messzeitpunkt</b>
Reaktion	Zufriedenheitsfragebogen	Teilnehmer	nach jedem Modul
Lernen	FB „deklaratives /prozedurales Wissen“	Teilnehmer	Prä- und Posttest + Zwischenevaluation
Verhalten	Fragebogen „Verhalten“	Kollegen	Posttest
Resultate	Interview	Teilnehmer	Posttest
Transferbedingungen	GLTSI	Teilnehmer	Posttest

Tabelle 1: Erhebungs- und Versuchsplan

## **2.3 Untersuchungsfragen und Hypothesen**

In diesem Kapitel werden die konkreten Untersuchungsfragen und Hypothesen aufgestellt, die zur Beantwortung der Forschungsfrage dieser Arbeit dienen sollen. Für jeden Erhebungskomplex werden Untersuchungsfragen und Hypothesen für die beiden aufgestellten Untersuchungsebenen formuliert.

### **2.3.1 Ebene Reaktion**

#### **2.3.1.1 Überprüfung der Wirksamkeit**

Auf der Ebene Reaktion wird die Zufriedenheit der Teilnehmer mit der Ausbildung untersucht. Es wird davon ausgegangen, dass eine hohe Zustimmung auf dieser Ebene die Ergebnisse der nachfolgenden Ebenen begünstigt. Auf dieser Ebene bietet sich ein exploratives Vorgehen an. Folgende Untersuchungsfragen stehen im Fokus:

**Frage 1: Wie ist die Zufriedenheit der Teilnehmer mit der Ausbildung ausgeprägt?**

**Frage 2: Wie variiert die Zufriedenheit zwischen den einzelnen Modulen?**

#### **2.3.1.2 Überprüfung von Indikatoren für die Eignung des Ansatzes**

##### *Wissenschaftliche Anforderungen*

Aus wissenschaftlicher Perspektive stellt sich die Frage, inwieweit das Verfahren den psychometrischen Gütekriterien entspricht. Die Überprüfung der Güte des Verfahrens kann sich aufgrund der Möglichkeiten im Rahmen der Untersuchung allerdings lediglich auf die Überprüfung der Reliabilität beziehen.

**Frage 1: Wie ist die Reliabilität des Erhebungsverfahrens einzuschätzen?**

##### *Praktikabilität*

Um das Erhebungsverfahren hinsichtlich der Praktikabilität zu beurteilen, sollen die Rücklaufquoten als quantitative Daten einbezogen werden:

**Frage 2: Wie hoch ist der Rücklauf der Daten?**

### **2.3.2 Ebene Lernen**

#### **2.3.2.1 Überprüfung der Wirksamkeit**

Als Indikator für den Lernerfolg der Ausbildung soll überprüft werden, inwieweit die Ausbildung zu einem Zuwachs an deklarativem und prozeduralem Wissen geführt hat.

Es stellen sich folgende Untersuchungsfragen:

**Frage 1: Bewirkt die Ausbildung zum Innovationspromotor einen Zuwachs an deklarativem Wissen?**

**Frage 2: Bewirkt die Ausbildung zum Innovationspromotor einen Zuwachs an prozeduralen Wissen?**

Das Design erlaubt folgende Hypothesen bezüglich des Wissenszuwachses:

*Hypothese 1: Zu mindestens zwei Messzeitpunkten unterscheiden sich die Mittelwerte für das deklarative Wissen signifikant.*

*Hypothese 2: Zu mindestens zwei Messzeitpunkten unterscheiden sich die Mittelwerte für das prozedurale Wissen signifikant.*

Der simultane Vergleich der Ergebnisse zu den drei Messzeitpunkten lässt zunächst keine weitere Spezifikation der Hypothesen zu. Im Rahmen von Post-Hoc Analysen können aber Einzelvergleiche vorgenommen werden, die eine präzisere Beantwortung der Untersuchungsfragen zulassen.

### **2.3.2.2 Überprüfung von Indikatoren für die Eignung des Ansatzes**

#### Wissenschaftliche Anforderungen

Aus wissenschaftlicher Perspektive stellt sich auch für das Instrument der Lernebene die Frage der Güte. Die zugrunde liegende Logik des Verfahrens auf der Lernebene macht eine herkömmliche Reliabilitätsschätzung allerdings nicht sinnvoll möglich. So wurden zur Konstruktion der Skalen die globalen Lernziele je Modul formuliert. Dabei handelt es sich aber nicht unbedingt um die verschiedenen Facetten eines Konstruktes, da die Module zum Teil sehr umfassende Themengebiete abdecken. So sind bspw. die Items Selbstmanagement und Selbstreflexion als Lernziele des ersten Moduls unabhängige Lernbereiche. Um dennoch Hinweise auf die Güte des Verfahrens zu erlangen, wurde der Umstand genutzt, dass die Teilnehmer zu jedem der drei Zeitpunkte bereits ihr Wissen für die Lerninhalte aller zehn Module einschätzen, obwohl diese zu Zeitpunkt 2 nur zu Hälfte trainiert worden sind.

Wenn die Skalen hinreichend zwischen den Lerninhalten der Ausbildung differenzieren, müssten die Mittelwerte für die zu t2 bereits trainierten Inhalte signifikant höher ausfallen.

*Hypothese 1: Für das deklarative Wissen zeigt sich von t1 zu t2 lediglich für die Items der bis dato trainierten Module ein signifikanter Zuwachs.*

*Hypothese 2: Für das prozedurale Wissen zeigt sich von t1 zu t2 lediglich für die Items der bis dato trainierten Module ein signifikanter Zuwachs.*

Um die Hypothesen zu überprüfen, werden zwei Cluster gebildet, denen die Items nach dem Kriterium trainiert/nicht trainiert zugeordnet werden. Cluster 1 umfasst alle Items, die zu t2 bereits trainiert worden sind (M 1-5). Cluster 2 umfasst, die bis dato noch nicht trainierten Items (M 6-9). Es werden folgende Annahmen überprüft.

*H1a: Der Mittelwert des deklarativen Wissens ist für Cluster 1 zu t2 höher als zu t1.*

*H1b: Der Mittelwert des deklarativen Wissens unterscheidet sich für Cluster 2 nicht zu den Messzeitpunkten t1 und t2.*

*H2a: Der Mittelwert des prozeduralen Wissens ist für Cluster 1 zu t2 höher als zu t1.*

*H2b: Der Mittelwert des prozeduralen Wissens unterscheidet sich für Cluster 2 nicht zu den Messzeitpunkten t1 und t2.*

### Praktikabilität

Um das Erhebungsverfahren hinsichtlich der Praktikabilität zu beurteilen sollen auch für diese Ebene die Rücklaufquoten als quantitative Daten einbezogen werden:

**Frage 1: Wie hoch ist der Rücklauf der Daten?**

## **2.3.3 Ebene Verhalten**

### **2.3.3.1 Überprüfung der Wirksamkeit**

Auf der Verhaltensebene soll überprüft werden, ob der Transfer des Gelernten in das berufliche Handeln gelingt. Folgende Untersuchungsfrage steht im Fokus:

**Frage 1: Bewirkt Teilnahme an der Ausbildung zum Innovationspromotor eine Veränderung im Arbeitsverhalten der Teilnehmer?**

Diese Veränderung soll sich in signifikant höheren Mittelwerten für das gegenwärtige Verhalten im Vergleich zu dem Verhalten vor einem Jahr zeigen. Das Design erlaubt folgende Hypothese bezüglich der Verhaltensänderung:

*Hypothese 1: Der Mittelwert für das gegenwärtige Verhalten ist signifikant höher als der Mittelwert für das Verhalten vor einem Jahr.*

### **2.3.3.2 Überprüfung von Indikatoren für die Eignung des Ansatzes**

#### Wissenschaftliche Anforderungen

Aufgrund des Einzelgruppen-Designs mit nur einem Messzeitpunkt auf der Verhaltensebene bieten sich kaum Kontrollmöglichkeiten zur Überprüfung dieser Hypothese und der Eliminierung potenzieller Störfaktoren an. Gerade im Transfer des Gelernten liegt aber die erfolgskritische Größe für den Erfolg einer Trainingsmaßnahme. Um die Ergebnisse auf der Verhaltensebene zu validieren, soll die Einschätzung des prozeduralen Wissens durch die Teilnehmer zu t3 als Selbsteinschätzung des eigenen

Verhaltens behandelt werden und ein Abgleich mit den Ergebnissen aus der Kollegenbefragung erfolgen. Der Abgleich zwischen Selbst- und Fremdeinschätzung wird von folgender Hypothese geleitet:

*Hypothese 1: Die Selbst- und Fremdeinschätzung des Verhaltens stimmen überein.*

Um die Hypothese zu überprüfen, werden die Konstrukte wie folgt operationalisiert:

*Hypothese 1a: Der Mittelwert für das gegenwärtige Verhalten unterscheidet sich nicht signifikant von der Einschätzung des prozeduralen Wissens durch die Teilnehmer zu t3.*

### Praktikabilität

Als Indikator für die Praktikabilität des Verfahrens wird - wie auch schon auf den vorherigen Ebenen - die Rücklaufquote betrachtet:

**Frage 2: Wie hoch ist der Rücklauf der Daten?**

### **2.3.4 Ebene Resultate**

Das Vorgehen auf der Ebene Resultate weicht, wie bei der Methodenbeschreibung erläutert, von dem üblichen Vorgehen der Untersuchung ab. Statt der Überprüfung von Erfolgskriterien im Rahmen der Wirksamkeitsanalyse und Überprüfung von Indikatoren zur Eignung dieses Vorgehens, wird explorativ untersucht, welche Kennzahlen sich zur Erfassung des Ausbildungserfolges auf der Unternehmensebene eignen könnten. Damit sollen gezielt Anhaltspunkte für die künftige Integration der Ebene im Rahmen des ganzheitlichen Evaluationsansatzes generiert werden:

**Frage 1: Welche Kennzahlen eignen sich zur Erfassung des Ausbildungserfolges auf der Unternehmensebene?**

Die Eignung der Kriterien macht sich wieder an den beiden Perspektiven fest. Die Kriterien sollen wissenschaftlichen und praktikablen Anforderungen genügen.

### **2.3.5 Transferbedingungen**

#### **2.3.5.1 Überprüfung der Wirksamkeit**

Das Rahmenmodell von Baldwin und Ford (1988) unterscheidet Bedingungsfaktoren des Lern- und Funktionsfelds sowie der Person, die sich fördernd oder hemmend auf den Transfer auswirken können. Vor diesem Hintergrund sollen die Ausprägungen der verschiedenen Variablen zunächst explorativ untersucht werden:

**Frage 1: Wie sind die Transferbedingungen bei den Teilnehmern ausgeprägt?**

Da das Verfahren im Rahmen der Arbeit gezielt zur Diagnose möglicher Barrieren für den Transfer eingesetzt werden soll, werden die Ausprägungen der Transferbedingungen in einem zweiten Schritt gezielt mit den Ergebnissen der Evaluationsebenen nach Kirkpatrick in Verbindung gebracht:

**Frage 2: Wie hängen die Transfervariablen und die Ergebnisse auf den unterschiedlichen Output-Ebenen zusammen?**

### **2.3.5.2 Überprüfung von Indikatoren für die Eignung des Ansatzes**

#### *Wissenschaftliche Anforderungen*

Wenngleich die Güte des Verfahrens bereits empirisch geprüft wurde (Kauffeld, 2008), wird im Rahmen der Untersuchung erneut die Reliabilitätsschätzung für die eingesetzten Skalen des Verfahrens vorgenommen.

**Frage 1: Wie ist die Reliabilität des Erhebungsverfahrens einzuschätzen?**

#### *Praktikabilität*

Um das Erhebungsverfahren hinsichtlich der Praktikabilität zu beurteilen, sollen die Rücklaufquoten als quantitative Daten einbezogen werden:

**Frage 2: Wie hoch ist der Rücklauf der Daten?**

## **2.4 Beschreibung der Stichproben**

### Untersuchungsgruppe 1

Die Stichprobe für die Untersuchung umfasst die 15 Ausbildungsteilnehmer. Diese wurden etwa zur Hälfte direkt oder indirekt über das GI:VE-Projekt akquiriert. Die Teilnahme an der Ausbildung war freiwillig. Die Geschlechterverteilung kann mit 8 Frauen (53 %) und 7 Männern als ausgeglichen bezeichnet werden. Die Altersspanne der Teilnehmer reicht von 25 bis 51 Jahren. Der Altersdurchschnitt der Ausbildungsgruppe liegt bei 38 Jahren. Bezüglich der Berufserfahrung und Funktionen sowie der Branchen und Unternehmensgrößen der Herkunftsunternehmen kann eine große Heterogenität konstatiert werden. So liegt die durchschnittliche Berufserfahrung der Teilnehmer bei 12,7 Jahren, wobei 1 Jahr Erfahrung das Minimum und 30 Jahre das Maximum an Berufserfahrung ausmachen. Eine Tabelle, der alle relevanten Informationen zu Bildungsgrad, aktueller Position im Unternehmen, Unternehmensgröße und Branche entnommen werden können, befindet sich im Anhang (Appendix A, S. 112)

### Untersuchungsgruppe 2

Auf der Ebene Verhalten wurden zusätzlich die Kollegen der Trainingsteilnehmer befragt. Die Teilnehmer wurden gebeten, selbst auszuwählen, welche Person in ihrem Unternehmen für die Einschätzung ihres Arbeitsverhaltens in Frage kommt. Je Teilnehmer sollte mindestens ein Kollege an der Befragung teilnehmen. Für die Stichprobe konnten letztlich 7 Fälle berücksichtigt werden. Es liegen keine weiteren Daten zu den befragten Personen vor, da sich dies aus Gründen des Datenschutzes und der Gewährleistung der Anonymität nicht sinnvoll umsetzen ließ.

## **2.5 Auswertungsstrategie**

Die Eingabe und Auswertung der erhobenen Daten erfolgte mit Hilfe der Statistiksoftware PASW 19. Um bereits im Vorfeld der Analyse Auffälligkeiten in der Datenstruktur zu erkennen und Informationen über Lage und Verteilung der Werte zu erhalten, wurden die Daten explorativ untersucht. Eingabefehler und Ausreißer konnten identifiziert und bei der Auswertung berücksichtigt werden. Bei der Analyse der fehlenden Werte zeigten sich keine Auffälligkeiten, sodass auf allen Erhebungsebenen von einer zufälligen Verteilung dieser ausgegangen wurde. Da der Rücklauf der Daten noch ausführlich als Indikator der Praktikabilität betrachtet wird, soll an dieser Stelle nur kurz auf den Umgang mit fehlenden Werten eingegangen werden.

Auf der Ebene Reaktion wurden die fehlenden Werte mit einem aufwendigen manuellen Schätzverfahren imputiert, um die Stichprobengröße nicht weiter zu verringern.<sup>18</sup> Auf der Lernebene waren die Datensätze von zwei Personen lückenhaft. Da sich unter Ausschluss der Fälle zeigte, dass sich die Ergebnisse nicht verändern, wurden die beiden Fälle in die Berechnungen eingeschlossen. Die Datensätze auf der Ebene Verhalten und der Transferbedingungen waren jeweils vollständig.

### Beurteilung der Wirksamkeit

In der Analyse wurde zunächst die Wirksamkeit der Ausbildung exemplarisch überprüft. Zur Beantwortung der dazu aufgestellten Untersuchungsfragen und Hypothesen wurden zunächst deskriptive Analysen durchgeführt. Darüber hinaus stand die Anwendung inferenzstatistischer Verfahren im Mittelpunkt der Analysen. Diese wurden teilweise aus den deskriptiven Zusammenhängen abgeleitet oder unmittelbar zur Testung der formulierten Hypothesen herangezogen.

Die zentrale Analyseverfahren für die vorliegende Untersuchung ist die ***einfaktorielle Varianzanalyse mit Messwiederholung***. Das Verfahren eignet sich zur Testung von Unterschiedshypothesen für mehr als zwei Teilstichproben, die voneinander abhängig sind. Es werden also die Messwerte *derselben Personen* zu verschiedenen Zeitpunkten überprüft. Der große Vorteil des Verfahrens ist, dass in einem Test beliebig viele Messzeitpunkte simultan miteinander verglichen werden können. Damit werden Probleme der alpha-Fehler-Kumulierung und der Verringerung der Teststärke umgangen, die sich bei einem t-Test ergeben würden, wenn man mehr als zwei Mittelwerte vergleicht (Rasch, Friese, Hofmann & Naumann, 2006a). Die Varianzanalyse prüft zunächst ungerichtet, ob ein Haupteffekt vorliegt, sich also zwischen mindestens zwei Mittelwerten signifikante Unterschiede finden lassen. Als Prüfgröße gilt die F-Statistik. Mit Hilfe sogenannter Post-hoc-Analysen kann darüber hinaus bestimmt werden, zwischen welchen Messzeitpunkten die Signifikanzen bestehen und in welche Richtung der Unterschied geht. Im Rahmen der Arbeit wurde die Bonferroni-Korrektur als Verfahren zur Post-hoc-Analyse angewandt<sup>19</sup>. Neben der statistischen Bedeutsamkeit interessiert zudem die praktische Bedeutsamkeit gefundener Unterschiede, also die Effektstärke. Als Maß soll hierfür das partielle Eta-Quadrat

---

<sup>18</sup> Der Umgang mit den fehlenden Werten auf der Ebene Reaktion wurde von Frau Grit Rudinger entwickelt. Eine ausführliche Darstellung des Vorgehens findet sich in Rudinger (2012).

<sup>19</sup> Bei der Interpretation der Ergebnisse ist zu beachten, dass die Teststärke dieses Verfahrens eingeschränkt ist, da nicht alle Daten in die Einzelvergleiche eingehen. Weiterführende Informationen zu dem Verfahren finden sich z.B. bei Rasch und Kollegen (2006a).



( $\eta_p^2$ ) verwendet werden. Dabei muss einschränkend darauf verwiesen werden, dass Effektgrößen in messwiederholenden Untersuchungen nur unter Vorbehalt interpretiert werden dürfen, da ihre Ausprägungen stark von den speziellen Bedingungen einer Studie abhängen (Rasch u.a., 2006a). Für die Interpretation eines Effektes aus einer messwiederholten Varianzanalyse kann also auch nicht die übliche Konvention nach Cohen (1988) herangezogen werden, da die Effektgröße von den Korrelationen zwischen den Messzeitpunkten abhängig ist. Rasch und Kollegen (2006a) kommen zu dem Schluss, dass  $\eta_p^2$  ein ungünstiges Maß für die Effektstärke in messwiederholenden Untersuchungen ist. Aufgrund mangelnder Alternativen soll  $\eta_p^2$  trotz dieser Kritik in der vorliegenden Untersuchung als Effektstärkemaß, aber lediglich im Sinne einer Orientierungsgröße verwendet werden. Als Voraussetzung für die valide Anwendung der einfaktoriellen Varianzanalyse gelten neben dem Intervallskalenniveau der Daten, die Normalverteilung der abhängigen Variablen und die Sphärizität der Daten. Die Ergebnisse der Überprüfung der Voraussetzungen werden im Ergebnisteil berichtet.

Zur Überprüfung von Unterschiedshypothesen, bei denen genau zwei empirisch gefundene Mittelwerte auf ihren systematischen Unterschied hin überprüft werden sollen, wird der ***t-test für abhängige Stichproben*** als Verfahren zur Signifikanzprüfung herangezogen. Anders als beim t-test für unabhängige Stichproben geht es hier nicht primär um die Unterschiede zwischen den Mittelwerten, sondern um den Mittelwert der Differenzen (Rasch u.a., 2006c). Es wird gegen die Nullhypothese getestet, dass der Mittelwert der Differenzen 0 beträgt. Als Voraussetzungen für die Anwendung des Verfahrens gelten wieder die Normalverteilung und die Intervallskaliertheit der Daten sowie die Varianzhomogenität. Bei dem t-test für abhängige Stichproben werden zudem Korrelationen zwischen den beiden Messreihen überprüft. Diese sollten nicht negativ sein, da dies die Teststärke verringert. Für die Berechnung der Effektstärke für abhängige t-tests finden sich in der Literatur verschiedene Empfehlungen. Problematisch ist wie auch bei der messwiederholenden Varianzanalyse, dass der Einfluss der Stärke der Abhängigkeit beachtet werden muss. Field (2009) folgend soll in dieser Arbeit  $r$  als Effektgröße für den abhängigen t-test verwendet werden. Während Rasch und Kollegen (2006c) aufgrund der Abhängigkeit der Daten keine feste Konvention zur Interpretation der Effektgrößen empfehlen, soll in dieser Arbeit in Anlehnung an Field (2009) folgende Konventionen zur Interpretation gelten: kleiner Effekt:  $r = 0,10$ ; mittlerer Effekt:  $r = 0,30$ ; großer Effekt:  $r = 0,50$ .

Neben den beschriebenen Verfahren zur Testung von Unterschiedshypothesen geht es im Rahmen der Analysen im Rahmen der Arbeit auch darum, **Zusammenhänge** zwischen Variablen aufzuzeigen. Um den Grad eines Zusammenhangs zu bestimmen, werden die Kovarianzen und die darauf aufbauende **Produkt-Moment-Korrelationen nach Pearson** bestimmt (Rasch u.a., 2006c). Als Korrelationskoeffizient wird  $r$  angegeben. Da  $r$  an den Streuungen der beteiligten Variablen standardisiert ist, kann es direkt als Effektstärkemaß interpretiert werden. Nach Cohen (1988) soll folgende Konvention zur Interpretation der Effektstärke gelten: kleiner Effekt:  $r = 0,10$ ; mittlerer Effekt:  $r = 0,30$ ; großer Effekt:  $r = 0,50$ .

Für die beschriebenen parametrischen Verfahren wurde ein im Forschungskontext übliches Signifikanzniveau von  $\alpha = 0.5$  festgelegt. Das Signifikanzniveau drückt die Irrtumswahrscheinlichkeit aus, die man bei Annahme der Alternativhypothese maximal akzeptieren will (Rasch u.a., 2006b).

#### Beurteilung der Eignung

Für das Erhebungsverfahren auf der Zufriedenheitsebene soll die **Reliabilität** der Skalen als Indikator für die Güte des Instrumentes erfasst werden. Zur Reliabilitätschätzung sind prinzipiell verschiedene Methoden denkbar. Die zugrunde liegende Annahme bei Reliabilitätsschätzungen ist die Gleichwertigkeit von Messungen (**Äquivalenz**). Unter Messungen können dabei Tests, Testteile oder Items verstanden werden (Bühner, 2006). Da im vorliegenden Untersuchungszeitpunkt der Einsatz des Erhebungsverfahrens zu den zehn Messzeitpunkten nicht als äquivalent wiederholte Messung gelten kann (*so soll sich die Einschätzung der Zufriedenheit zwischen den Modulen gezielt unterscheiden, das gemessene Merkmal ist also nicht stabil*), soll die Reliabilität über die Methode der Testhalbierung (interne Konsistenz) geprüft werden. Als Verfahren zur Schätzung der internen Konsistenz wird Cronbachs Alpha als die übliche Standardmethode gewählt. Jedes Item wird dabei als Testteil betrachtet. Cronbachs Alpha prüft die mittlere Korrelation zwischen allen Items einer Skala (Bühner, 2006). Die Höhe von Cronbachs Alpha soll in Anlehnung an Bühner (2006) wie folgt interpretiert werden: Werte kleiner 0.80 zeigen eine niedrige, Werte zwischen 0.80 und 0.90 eine mittlere und Werte größer 0.90 eine hohe Reliabilität. Es finden sich auch Definitionen wonach bereits Werte größer 0.7 als befriedigend angesehen werden (Seidel, 2012).

### 3. Ergebnisse der Untersuchung

In diesem Abschnitt werden die Ergebnisse der Untersuchung systematisch dargestellt. Die Gliederung der Ergebnisse orientiert sich an den aufgestellten Untersuchungsfragen und Hypothesen für die fünf Evaluationskomplexe der Studie. Die Darstellung der Ergebnisse wird auf die notwendigen Informationen zur Beantwortung der Fragestellungen eingegrenzt.<sup>20</sup>

Für jeden Untersuchungskomplex wird zudem die Interpretation der Ergebnisse unmittelbar mit vorgenommen. Vorab werden die Ergebnisse der Überprüfung der Voraussetzungen zur Anwendung statistischer Verfahren skizziert.

#### 3.1 Überprüfung der Voraussetzungen zur Anwendung statistischer Verfahren

In diesem Abschnitt wird die Prüfung der Normalverteilungsannahme mit Hilfe des Kolmogoroff-Smirnov-Tests (Field, 2009) berichtet. Die Überprüfung der Sphärizitätsannahme - als weitere Voraussetzung für die Anwendung der messwiederholenden Varianzanalyse - erfolgt direkt bei der Anwendung des Verfahrens. Für alle vorliegenden quantitativen Daten wird aus Plausibilitätsgründen - wie im Forschungskontext üblich - Äquidistanz und damit Intervallskalenniveau angenommen (Rasch, 2006c).

Die Überprüfung der *Normalverteilungsannahme* zeigt für die verwendeten Daten keine größeren Abweichungen.<sup>21</sup> Für die Daten der Ebene Reaktion gibt der Kolmogoroff-Smirnov-Test lediglich für die Variable *Trainier* zu Messzeitpunkt 6 und 10 eine signifikante Abweichung von der Normalverteilung aus. Allerdings verhält sich sowohl die Varianzanalyse als auch der t-test gegenüber der Verletzung der Normalverteilungsannahme prinzipiell robust (Bühner, 2006).

Die erhobenen Daten auf den Ebenen Lernen und Verhalten zeigen keine signifikanten Abweichungen von der Normalverteilung. Die Ausprägungen der Skalen des LTSI sind laut Kolmogoroff-Smirnov-Test ebenfalls normalverteilt.

---

<sup>20</sup> Die vollständige Darstellung der in diesem Kapitel berichteten Analysen ist im Anhang einsehbar (Appendix C, S. 131 ff.)

<sup>21</sup> Die Darstellung der Ergebnisse der Überprüfung der Normalverteilung findet sich im Anhang (Appendix C1, S. 131ff.).

## 3.2 Ebene Reaktion

### 3.2.1 Überprüfung der Wirksamkeit

Um die Untersuchungsfragen nach der Zufriedenheit der Teilnehmer mit der Ausbildung und dem Verlauf der Bewertungen über alle Module hinweg auszuwerten, zeigen deskriptive Analysen zunächst ein Bild von der Verteilung der Antworten auf.<sup>22</sup>

#### Deskriptive Analysen

Die Betrachtung der Mittelwerte zeigt insgesamt eine hohe Zufriedenheit mit der Ausbildung zum Innovationspromotor und deren einzelner Module (Abb. 6). Die Gesamtzufriedenheit über alle Module hinweg liegt bei 4,4 (Skala 1-5). Besonders positiv werden die Module 4 (4,6), 6 (4,8), 7 (4,6) und 8 (4,6) bewertet, wobei das sechste Modul mit dem Thema *Gesprächsführung und Moderation* am besten abschneidet. Da die Standardabweichung für dieses Modul sehr klein ausfällt (0,2), ist davon auszugehen, dass der Mittelwert das Antwortverhalten der Teilnehmer gut repräsentiert. Die Module 3 (4,0), 5 (3,9) und 9 (4,0) werden im Vergleich etwas schlechter bewertet. Für diese Module fallen auch die Werte für die Standardabweichung höher aus, was auf eine kontroversere Einschätzung der Module hindeutet.

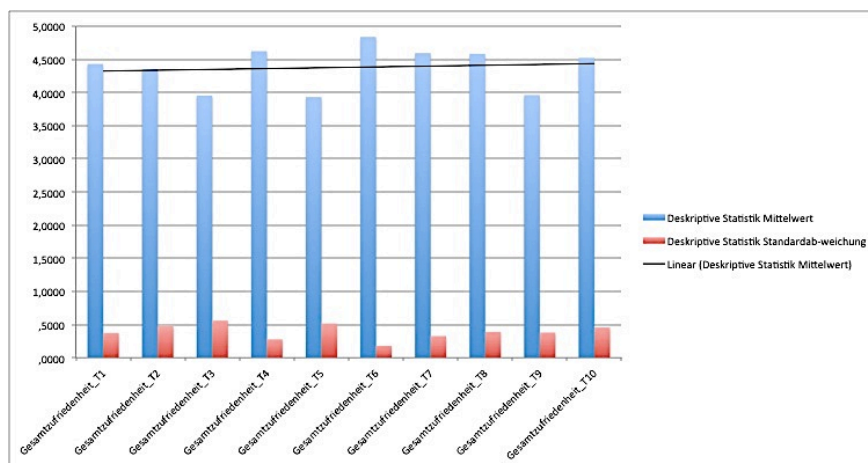


Abbildung 4 Mittelwerte und Standardabweichungen der Gesamtzufriedenheit Module 1-10

Um zu einer differenzierteren Einschätzung der Zufriedenheitswerte zu gelangen, wird die Verteilung für die drei Dimensionen separat betrachtet. Es zeigt sich, dass insbesondere die Dimension *Trainer* eine hohe Bewertung pro Modul erfährt und das auch in den Modulen, die ansonsten schlechter bewertet werden (M3, M5, M9). Der Mittelwert der Zufriedenheit mit den Trainern über alle Module hinweg liegt bei 4,6. Die Dimensionen *Didaktik* und *Inhalt* werden im Vergleich dazu etwas schlechter bewertet und liegen etwa gleich auf (4,3; 4,2).

<sup>22</sup> Die vollständigen Ergebnisse der deskriptiven Analysen sind im Anhang einsehbar (Appendix C2, S. 137ff.).

## Signifikanzprüfung

Anknüpfend an die Darstellung der Verteilung der Mittelwerte aus der Zufriedenheitsevaluation stellt sich die Frage, ob die gefundenen Unterschiede auch statistisch bedeutsam sind.<sup>23</sup>

Vergleicht man zunächst wieder die Gesamtzufriedenheit je Modul, lässt das Ergebnis der Varianzanalyse darauf schließen, dass zwischen mindestens zwei Modulen signifikante Unterschiede in der Zufriedenheitsbewertung bestehen. Unter Annahme der Sphärizität ergibt sich ein signifikanter Haupteffekt ( $F_{(2;126)} = 16.655$ ;  $p = .000$ ;  $\eta_p^2 = .54$ ). Die Effektgröße gibt aus, dass der Anteil der durch den messwiederholten Faktor aufgeklärten Variabilität, also hier die unterschiedlichen Module, 54% beträgt. Die Varianz in der Zufriedenheit mit der Ausbildung kann also zur Hälfte auf Unterschiede zwischen den Modulen zurückgeführt werden.

Um das Ergebnis zu spezifizieren, wurden Einzelvergleiche im Rahmen von Post Hoc Analysen vorgenommen. Betrachtet man zunächst die im Vergleich schlechter bewerteten Module 3, 5 und 9, zeigt sich, dass sich die Vermutungen der deskriptiven Analysen bestätigen. Die Bewertung für Modul 9 unterscheidet sich von allen anderen Modulen signifikant, außer von den Modulen, die ebenfalls schlechter bewertet werden (M3, M5). Der Unterschied in der Bewertung wird für Modul 5 ebenfalls für alle Module ausgenommen der Module 3 und 9 signifikant. Zusätzlich wird der Unterschied zu Modul 7 knapp nicht signifikant ( $p = .053$ ). Für Modul 3 wird hingegen zusätzlich der Unterschied zu Modul 10 knapp nicht signifikant ( $p = .077$ ). Betrachtet man nun die Module, die im Vergleich besonders positiv bewertet wurden, zeigt sich, dass sich lediglich Modul 6, als Modul welches dem Mittelwert nach am besten bewertet wurde, signifikant von einem Großteil der anderen Module unterscheidet. Nicht signifikant wird lediglich der Unterschied zu den beiden auch sehr gut bewerteten Modulen 7 und 8 und zu Modul 10.

Abschließend soll überprüft werden, ob die gefundenen Mittelwertsunterschiede zwischen den drei Bewertungsdimensionen signifikant sind ( $d_{\text{trainer-didaktik}} = 0,4$ ;  $d_{\text{trainer-inhalt}} = 0,3$ ;  $d_{\text{inhalt-didaktik}} = 0,1$ ). Für die Analyse werden die aggregierten Mittelwerte für die Dimensionen Inhalt, Trainer und Didaktik über alle Module hinweg verglichen. Unter Annahme der Sphärizität ergibt sich ein signifikanter Haupteffekt ( $F_{(2;28)} = 36.882$ ;  $p = .000$ ;  $\eta_p^2 = .725$ ).

---

<sup>23</sup> Die vollständigen Analysen der Signifikanzprüfung finden sich im Anhang (Appendix C2, S. 139).

Die Effektgröße zeigt an, dass etwa 73% der Variabilität der Gesamtzufriedenheit durch die Bewertungsdimension erklärt werden. Die Betrachtung der Einzelvergleiche zeigt, dass lediglich für die Dimension Trainer ein signifikanter Unterschied zu den Dimensionen Inhalt und Didaktik ausgewiesen wird (jeweils  $p = .000$ ), nicht aber zwischen den Dimensionen Inhalt und Didaktik ( $p = 1.000$ ). Die Bewertung der Trainer fällt also über alle Module hinweg signifikant positiver aus als die Bewertung für den Inhalt und die Didaktik.

### Auswertung der offenen Kommentare

Die qualitative Auswertung der offenen Kommentare zu positiven und negativen Aspekten soll die bisher dargestellten quantitativen Ergebnisse sinnvoll ergänzen. Daher soll die Auswertung im Hinblick auf zwei Perspektiven erfolgen:

1. Welche Aspekte der Ausbildung werden in den besonders gut bewerteten Modulen (4, 6, 7 und 8) positiv hervorgehoben?
2. Welche Aspekte werden in den schwächer bewerteten Modulen (3, 5 und 9) explizit als negativ genannt?

Nachfolgend werden die wesentlichen Ergebnisse tabellarisch gegenübergestellt.

Positive Kommentare (Module, 4, 6, 7, 8)	Negative Kommentare (Module 3, 5, 9)
<u>Modul 4</u> <ul style="list-style-type: none"> <li>• Verbindung von Theorie &amp; Praxis</li> <li>• Rollenspiel + Präsentation</li> <li>• Struktur der Themen, Inhalte und Übertragung auf Praxisbeispiele</li> <li>• Modelle Inhalt sehr gut greifbar</li> <li>• Möglichkeit der einfachen Übertragung in das eigene Unternehmen</li> <li>• Projekt- und Fallarbeit</li> <li>• Diskussion</li> </ul>	<u>Modul 3</u> <ul style="list-style-type: none"> <li>• Inhalte schwer greifbar</li> <li>• theoretische Untermauerung fehlte</li> <li>• unklare Aufgabenstellungen</li> <li>• zu wenig Hilfestellung bei den Übungen</li> <li>• Methode „Cockpit“ unklar</li> <li>• Auswertung der Übung zu umfangreich</li> </ul>
<u>Modul 6</u> <ul style="list-style-type: none"> <li>• Rollenspiele (4 Nennungen)</li> <li>• Wechsel Kleingruppen - Großgruppe (5 Nennungen)</li> <li>• Visualisierung</li> <li>• Praktische Übungen</li> <li>• Diskussion und Auswertung der Übungen / Rollenspiele in der Gruppe</li> <li>• Lernen im Austausch mit der Gruppe</li> <li>•</li> </ul>	<u>Modul 5</u> <ul style="list-style-type: none"> <li>• zu viel Theorie (8 Nennungen)</li> <li>• zu „trockener“ Lernstoff</li> <li>• zu wenig praktische Übungen</li> <li>• Verhältnis Theorie / Praxis unausgewogen</li> <li>• Zusammenhang zwischen den Theorie-Inhalten unklar</li> </ul>
<u>Modul 7</u> <ul style="list-style-type: none"> <li>• gutes Handwerkszeug</li> <li>• Verknüpfung mit dem eigenen Projekt</li> <li>• Reflektion der Phasen des Projektmanagements</li> <li>• Inhalte gut in die eigene Arbeit übertragbar</li> <li>• Arbeit im Tandem</li> <li>• kompaktes Handout / Arbeitsblätter</li> </ul>	<u>Modul 9</u> <ul style="list-style-type: none"> <li>• „echte“ Beispiele aus der Praxis fehlten</li> <li>• Theoretische Inhalte hätten vertieft werden müssen, waren in der Form nicht greifbar (6 Nennungen)</li> <li>• zu wenig neue Impulse</li> <li>• praktische Übungen fehlten</li> <li>• Diskussion in der Gruppe kam zu kurz</li> <li>• Einbettung der Themen hat gefehlt</li> <li>• zu wenig Interaktion mit der Gruppe, keine gemeinsame Erarbeitung der Theorieinhalte</li> </ul>
<u>Modul 8</u> <ul style="list-style-type: none"> <li>• Rollenspiele, gemeinsame Reflexion</li> <li>• hilfreiche Praxisbeispiele &amp; Modelle</li> <li>• Theorie und Praxis ausgewogen</li> <li>• Themen strukturiert und greifbar aufbereitet</li> <li>• gelungene Verknüpfung zur Praxis</li> </ul>	

Tabelle 2: Ergebnisse der qualitativen Analyse der offenen Kommentare des Zufriedenheitsfragebogens

### Zusammenfassung und Interpretation der Ergebnisse

Die Ergebnisse zeigen, dass die Zufriedenheit der Teilnehmer mit der Ausbildung insgesamt sehr hoch ausfällt und somit auch von einer hohen Akzeptanz der Ausbildung ausgegangen werden kann. Allerdings zeigen sich Unterschiede in der Beurteilung der Ausbildung zwischen den Modulen. So wurden die Module 3 (*Innovationsprozesse und ihre Analyse*), 5 (*Interventionen zur Förderung von Innovationen*) und 9 (*Evaluation und Verankerung*) im Vergleich signifikant schlechter bewertet, wengleich die Gesamtzufriedenheit hier immer noch hoch ist. Modul 6 (*Gesprächsführung und Moderation*) wurde hingegen besonders positiv bewertet.

Die Ergebnisse der quantitativen Analysen werden dabei gestützt von qualitativen Ergebnissen. Es fällt auf, dass für die besonders positiv bewerteten Module insbesondere das ausgewogene Verhältnis von Theorieinhalten und praktischen Übungen hervorgehoben wurde. So lobten die Teilnehmer einerseits die anschauliche Aufbereitung der Themen. Andererseits wurden praktische Übungen wie Rollenspiele und Projektarbeit als besonders hilfreich für den Transfer der Inhalte in die Praxis erlebt. Überdies wurde der Austausch in der Gruppe positiv hervorgehoben. Betrachtet man demgegenüber die negativen Aspekte für die schlechter bewerteten Module, sind es überwiegend genau diese Aspekte, die in den Modulen zu kurz kamen (Tab. 2).

Als weiteres interessantes Ergebnis auf der Zufriedenheitsebene kann berichtet werden, dass die Variabilität der Gesamtzufriedenheit zu 70% auf Unterschiede zwischen den drei Bewertungsdimensionen zurückgeführt werden kann. Im Rahmen der Einzelvergleiche konnte gezeigt werden, dass die Bewertung der Trainer über alle Module hinweg signifikant positiver ausfällt als die Bewertung der Zufriedenheit mit dem Inhalt und der Didaktik der Ausbildung. Auch in den sonst schlechter bewerteten Modulen erreichen die Trainer sehr gute Bewertungen. Der Eindruck der sehr hohen Gesamtzufriedenheit mit der Ausbildung, sollte vor diesem Hintergrund noch einmal differenziert betrachtet werden.

### **3.2.2 Überprüfung von Indikatoren für die Eignung des Ansatzes**

#### Wissenschaftliche Anforderungen

Um Hinweise über die Güte des Erhebungsverfahrens zu erhalten, wurde auf der Ebene Reaktion die Reliabilitätsschätzung für die drei Skalen vorgenommen. Tabelle 3 zeigt die Ergebnisse der Analysen.

Modul / Zeitpunkt der Messung	Skala „Trainer“		Skala „Didaktik“		Skala „Inhalt“	
	Cronbachs Alpha	Itemanzahl	Cronbachs Alpha	Itemanzahl	Cronbachs Alpha	Itemanzahl
<b>Modul 1 - t1</b>	.825	8	.726	5	.659	4
<b>Modul 2 - t2</b>	.865	8	.813	5	.386	4
<b>Modul 3 - t3</b>	.958	8	.877	5	.784	4
<b>Modul 4 - t4</b>	.812	8	.534	5	.739	4
<b>Modul 5 - t5</b>	.881	8	.888	5	.763	4
<b>Modul 6 - t6</b>	.571	8	.365	5	.548	4
<b>Modul 7 - t7</b>	.837	8	.657	5	.825	4
<b>Modul 8 - t8</b>	.905	8	.795	5	.667	4
<b>Modul 9 - t9</b>	.889	8	.466	5	.259	4
<b>Modul 10 - t10</b>	.945	8	.942	5	.776	4

Tabelle 3 Reliabilitäten der Skalen des Zufriedenheitsfragebogens zu den 10 Messzeitpunkten

Es zeigt sich, dass die Reliabilität für die Skala „Trainer“ zu allen Messzeitpunkten nach der Klassifikation von Bühner (2006) überwiegend im mittleren Bereich liegt, teilweise auch hohe Werte erreicht. Lediglich zu t6 liegt der Wert im niedrigen und damit inakzeptablen Bereich. Für die Skala „Didaktik“ unterscheiden sich die Werte für die interne Konsistenz zwischen den Messzeitpunkten zum Teil stark. Lediglich zu sechs Messzeitpunkten liegen die Werte mindestens im akzeptablen Bereich. Für die Skala „Inhalt“ zeichnet sich ein ähnliches Bild. Immerhin zu fünf Messzeitpunkten liegen die Werte mindestens im akzeptablen Bereich. Es zeigen sich demgegenüber fünf Messungen mit zum Teil sehr niedrigen und damit inakzeptablen Werten.

### Praktikabilität

Um Anhaltspunkte dafür zu bekommen, inwieweit das Vorgehen im Rahmen der Erhebung auf der Ebene als praktikabel gelten kann, wurde die Höhe des Rücklaufs der Daten betrachtet. Tabelle 4 zeigt den ursprünglichen Rücklauf der Daten.

Module	Rücklauf absolut	Rücklauf relativ (gerundet)
Modul 1	12 / 15	80 %
Modul 2	14 / 15	93 %
Modul 3	13 / 15	87 %
Modul 4	14 / 15	93 %
Modul 5	15 / 15	100 %
Modul 6	14 / 15	93 %
Modul 7	12 / 15	80%
Modul 8	13 / 15	87%
Modul 9	14 / 15	93 %
Modul 10	15 / 15	100 %

Tabelle 4 Ursprünglicher Rücklauf der Daten auf der Ebene Reaktion



Der Rücklauf des Fragebogens auf der Reaktionsebene kann insgesamt als hoch bewertet werden (mind. 80%). Allerdings zeigen sich Schwankungen zwischen den Modulen, die als zufällig interpretiert werden. In der Regel handelt es sich bei den fehlenden Fällen um die Personen, die an dem Modul nicht anwesend waren und daher auch keine Einschätzung vornehmen konnten.

### Zusammenfassung und Interpretation der Ergebnisse

Die Ergebnisse der Reliabilitätsschätzungen zeigen, dass sich die Werte für die drei Skalen stark unterscheiden. Während die Skala Trainer überwiegend Werte im mittleren bis hohen Bereich erreicht, fallen die Schätzungen für die Skalen Didaktik und Inhalt deutlich schlechter aus. Bei beiden Skalen werden zu der Hälfte der Messzeitpunkte lediglich niedrige Werte für die interne Konsistenz erreicht. Damit muss von einer eingeschränkten Güte des Verfahrens ausgegangen werden. Insbesondere für die Skalen Inhalt und Didaktik besteht Bearbeitungsbedarf.

Hinsichtlich der Praktikabilität des Verfahrens deutet die durchweg hohe Teilnahme an der Befragung auf ein praktikables Vorgehen für die Teilnehmer hin. So fehlen überwiegend die Daten von den Personen, die an dem Modul nicht anwesend waren und so sinnvollerweise auch keine Einschätzung vornehmen konnten.

## **3.3 Ebene Lernen**

### **3.3.1 Überprüfung der Wirksamkeit**

Auf der Lernebene soll überprüft werden, inwieweit die Ausbildung zum Innovationspromotor einen Zuwachs an deklarativem und prozeduralem Wissen bewirkt.

Die Überprüfung folgender Hypothesen steht zunächst im Fokus der Analysen:

*Hypothese 1: Zu mindestens zwei Messzeitpunkten unterscheiden sich die Mittelwerte für das deklarative Wissen signifikant.*

*Hypothese 2: Zu mindestens zwei Messzeitpunkten unterscheiden sich die Mittelwerte für das prozedurale Wissen signifikant.*

Bei der deskriptiven Betrachtung der Daten lassen sich zu den drei Messzeitpunkten sowohl für das deklarative als auch für das prozedurale Wissen Unterschiede zu mindestens zwei Zeitpunkten beobachten: Für das **deklarative Wissen** kann zu t1 ein recht hohes Ausgangsniveau (3,7) konstatiert werden. Zu t2 nimmt der Wert leicht ab (3,5). Zu t3 zeigt sich dann ein deutlich höherer Wert als zu t1 und zu t2 (4,2). Es ergeben sich damit folgende Differenzwerte:  $d_{t1;t2} = -0,2$ ;  $d_{t2;t3} = 0,7$ ;  $d_{t1;t3} = 0,5$ .

Der Wert für das **prozedurale Wissen** ist zu t1 mit einem Mittelwert von 3,2 etwas geringer ausgeprägt als der Wert für das deklarative Wissen und liegt im mittleren Bereich der Skala (1-5). Zu t2 zeigt sich dann ein deutlich höherer Wert (4,00). Zu t3 nimmt der Mittelwert hingegen wieder leicht ab, liegt mit einem Wert von 3,6 aber über dem Ausgangsniveau ( $d_{t1;t2} = 0,8$ ;  $d_{t2;t3} = -0,4$ ;  $d_{t1;t3} = 0,4$ ).<sup>24</sup>

Zur Überprüfung der aufgestellten Hypothesen gilt es nachfolgend die gefundenen Mittelwertsunterschiede auf Signifikanz zu prüfen<sup>25</sup>: Für das **deklarative Wissen** gibt die einfaktorielle Varianzanalyse mit Messwiederholung einen signifikanten Haupteffekt aus ( $F_{(2;24)} = 11.186$ ;  $p = .000$ ;  $\eta_p^2 = .482$ ). Hypothese 1 kann damit bestätigt werden, es gilt: *Zu mindestens zwei Messzeitpunkten unterscheidet sich der Mittelwert für das deklarative Wissen signifikant*. Die Effektgröße gibt aus, dass die Unterschiede in den Ergebnissen auf der Lernebene zu 50% auf das Training zurückgeführt werden können. Mit Hilfe der Einzelvergleiche mit Bonferroni Korrektur sollen weitere Erkenntnisse darüber gewonnen werden, zwischen welchen Zeitpunkten sich die Mittelwerte signifikant unterscheiden. Es zeigt sich, dass der Wert zu t3 signifikant höher ausfällt als die Werte zu t1 und t2 ( $p = .036$ ;  $p = .000$ ). Zwischen t2 und t1 wird der Unterschied in den Mittelwerten hingegen nicht signifikant ( $p = .960$ ).

Für das **prozedurale Wissen** wird ebenfalls ein signifikanter Haupteffekt gefunden ( $F_{(2;24)} = 11.739$ ;  $p = .000$ ;  $\eta_p^2 = .494$ ). Hypothese 2 wird angenommen: *Zu mindestens zwei Messzeitpunkten unterscheiden sich die Mittelwerte für das prozedurale Wissen signifikant*. Auch für das prozedurale Wissen lassen sich Unterschiede in den Ergebnissen auf der Lernebene zu 50% auf das Training zurückführen. Die Einzelvergleiche zeigen, dass der Mittelwert zu t2 signifikant höher ist als die Werte zu t1 und zu t3 ( $p = .005$ ;  $p = .044$ ). Zwischen t3 und t1 wird der deskriptiv gefundene Unterschied hingegen knapp nicht signifikant ( $p = .063$ ).

### Zusammenfassung und Interpretation der Ergebnisse

Die Frage, ob die Ausbildung zum Innovationspromotor einen Zuwachs an deklarativen und prozeduralem Wissen bewirkt, muss auf Grundlage der Ergebnisse differenziert beantwortet werden. Für das **deklarative Wissen** konnte ein signifikanter Zuwachs zu t3 im Vergleich zu t1 und t2 empirisch belegt werden. Dieses Ergebnis

---

<sup>24</sup> Im Anhang (Appendix C3, S. 143) ist eine differenzierte Übersicht über die Mittelwerte je Modul zu allen drei Zeitpunkten, untergliedert nach deklarativem und prozeduralem Wissen einsehbar. Da im Ergebnisteil die Überprüfung der aufgestellten Hypothesen im Vordergrund steht, wird auf eine ausführlichere Darstellung der deskriptiven Analysen an dieser Stelle verzichtet.

<sup>25</sup> Die vollständigen Analysen der Signifikanzprüfung finden sich im Anhang (Appendix C3, S. 144 ff.).

kann als Lernerfolg interpretiert werden. Allerdings ging der Wert für das deklarative Wissen von t1 zu t2 leicht zurück. Für diesen Befund sind verschiedene Erklärungen denkbar. So war das Ausgangsniveau für das deklarative Wissen schon sehr hoch (3,7). Die Teilnehmer gaben vor der Ausbildung an, bereits viel über die Inhalte der Ausbildung zu wissen. Im Rahmen der ersten fünf Module lag der Schwerpunkt dann insbesondere auf der Erarbeitung theoretischen Wissens zu dem Themenfeld der Innovation. In der vertieften Auseinandersetzung mit den Inhalten wurden die Teilnehmer für die Komplexität der Themenfelder sensibilisiert. Der zurückgehende Wert zu t2 könnte ein Indikator dafür sein, dass die Teilnehmer nach dieser vertieften Auseinandersetzung mit den Themen den Eindruck gewonnen haben, in einzelnen Wissensbereichen doch noch nicht alles zu *kennen* (*Erkenntnis des Nicht-Wissens*).

Für das **prozedurale Wissen** konnte ein signifikanter Zuwachs von t1 zu t2 empirisch belegt werden. Zu t3 ging der Wert aber wieder signifikant zurück und lag zudem auch nicht mehr signifikant über dem Ausgangsniveau. Für das prozedurale Wissen, was auch als Selbsteinschätzung auf der Verhaltensebene betrachtet werden kann, scheint sich damit das *Transferproblem* abzubilden (Kap 2.4.1.3). Der sich zunächst abzeichnende positive Lerntransfer nach den ersten fünf Modulen, wird nicht aufrechterhalten. Erst wenn die Aufrechterhaltung der Transferleistung gelingt, wird aber per Definition von Lerntransfer ausgegangen. Dennoch zeigt sich auch von t1 zu t3 ein Zuwachs an prozeduralem Wissen, sodass nicht von einem Nulltransfer ausgegangen werden kann. Der Differenzwert liegt immerhin bei 0,4 und wird nur knapp nicht signifikant. So kann abschließend festgehalten werden, dass die Ausbildung auch für das prozedurale Wissen einen Zuwachs bewirkt hat. Lediglich die Nachhaltigkeit dieses Zuwachses steht in Frage.

### 3.3.2 Überprüfung von Indikatoren für die Eignung des Ansatzes

#### Wissenschaftlicher Anspruch

Um Anhaltspunkte über die Güte des Verfahrens zu erhalten, wird nachfolgend geprüft, inwieweit der Umstand trainiert/nicht trainiert die Ergebnisse beeinflusst. Zunächst steht die Überprüfung der Hypothesen für das **deklarative Wissen** im Fokus:

*Hypothese 1: Für das deklarative Wissen zeigt sich von t1 zu t2 lediglich für die Items der bis dato trainierten Module ein signifikanter Zuwachs.*

- *H1a: Der Mittelwert des deklarativen Wissens ist für Cluster 1 zu t2 höher als zu t1.*
- *H1b: Der Mittelwert des deklarativen Wissens unterscheidet sich für Cluster 2 nicht zu den Messzeitpunkten t1 und t2.*

Vergleicht man die Mittelwerte der zu t2 bereits trainierten Module (Cluster 1), zeigt sich auch hier der Trend der bereits berichteten Ergebnisse: Für das deklarative Wissen ist der Mittelwert zu t2 leicht kleiner als zu t1 ( $d = 0,2$ ). Das Gleiche gilt für die Items aus Cluster 2 ( $d = 0,1$ ). Die Unterschiede werden für beide Cluster nicht signifikant ( $t_{(12)} = 1.180$ ;  $p = .261$ ;  $t_{(12)} = 1.015$ ;  $p = .330$ ).<sup>26</sup> Während damit H1b bestätigt werden kann - *Der Mittelwert des deklarativen Wissens unterscheidet sich für Cluster 2 nicht zu den Messzeitpunkten t1 und t2* - findet sich für H1a kein empirischer Beleg. *Der Mittelwert des deklarativen Wissens ist für Cluster 1 zu t2 also **nicht höher** als zu t1*. Auf Grundlage dieser Ergebnisse kann Hypothese 1 nicht gelten.

Auch für das **prozedurale Wissen** soll überprüft werden, inwieweit der Umstand trainiert / nicht trainiert die Ergebnisse des Lernfragebogens beeinflusst:

*Hypothese 2: Für das prozedurale Wissen zeigt sich von t1 zu t2 lediglich für die Items der bis dato trainierten Module ein signifikanter Zuwachs.*

- *H2a: Der Mittelwert des prozeduralen Wissens ist für Cluster 1 zu t2 höher als zu t1.*
- *H2b: Der Mittelwert des prozeduralen Wissens unterscheidet sich für Cluster 2 nicht zu den Messzeitpunkten t1 und t2.*

Es zeigt sich, dass auch hier die Ergebnisse die bisher diskutierten Aussagen des Gesamtvergleiches aller Items zu den drei Zeitpunkten stützen: Das prozedurale Wissen steigt zu t2 deutlich an. Während der Mittelwertsunterschied für Cluster 2 jedoch nicht signifikant wird ( $d = 0,6$ ;  $t_{(12)} = -1.578$ ;  $p = .140$ ), stellt sich der Unterschied für die zu t2 bereits trainierten Module als statistisch bedeutsam heraus ( $d = 0,8$ ;  $t_{(12)} = -5.573$ ;  $p = .000$ ;  $r = .85$ ). Der Effekt kann als groß klassifiziert werden. Die Hypothesen 2a und 2b können also bestätigt werden, es gilt: *Der Mittelwert des prozeduralen Wissens ist für Cluster 1 zu t2 signifikant höher als zu t1. Der Mittelwert des prozeduralen Wissens unterscheidet sich für Cluster 2 nicht signifikant zu den Messzeitpunkten t1 und t2*. Betrachtet man noch einmal die Differenzmaße, muss einschränkend darauf hingewiesen werden, dass sich dennoch auch für Cluster 2 ein deutlicher Zuwachs zu t2 beobachten lässt, der lediglich um den Wert 0,2 kleiner ist als der Zuwachs von t1 zu t2 für Cluster 1. Hypothese 2 wird dennoch angenommen, es gilt: *Für das prozedurale Wissen zeigt sich von t1 zu t2 lediglich für die Items der bis dato trainierten Module ein signifikanter Zuwachs.*

---

<sup>26</sup> Die vollständigen Analysen der Signifikanzprüfung finden sich im Anhang (Appendix C3, S. 146 ff.).

### Praktikabilität

Um Anhaltspunkte dafür zu bekommen, inwieweit das Vorgehen im Rahmen der Erhebung auf der Ebene Lernen praktikabel gelten kann, wird die Höhe des Rücklaufs der Daten betrachtet: Der Rücklauf auf der Lernebene kann insgesamt als hoch beurteilt werden. Zu t1 nahm die eine Person nicht an der Befragung teil, die im ersten Modul nicht anwesend war. Da damit für diese Person der Ausgangswert fehlte, wurden ihre Daten auch zu den anderen beiden Zeitpunkten nicht berücksichtigt. Bei einer weiteren Person fehlten aufgrund technischer Probleme die Daten zu t3. Laut eigener Angaben nahm die Person zwar an der Befragung teil, ihre Werte wurden aber nicht in dem Onlineportal gespeichert.

### Zusammenfassung und Interpretation der Ergebnisse

Auf Grundlage der Ergebnisse lassen sich nur sehr eingeschränkt Rückschlüsse auf die Güte des Verfahrens ziehen. Auf Grundlage der Befunde für das **deklarative Wissen** kann schlichtweg nicht beurteilt werden, wie gut die Skalen differenzieren: Der Wert für das deklarative Wissen geht von t1 zu t2 zurück, unabhängig davon, ob die Lerninhalte bis dato trainiert worden sind oder nicht. Für das **prozedurale Wissen** wird der Wissenszuwachs zwar nur für die bis zu t2 trainierten Module signifikant, was darauf hindeutet, dass die einzelnen Skalen die Lerninhalte der Module differenziert abbilden. Die Betrachtung der Differenzmaße zeigt aber, dass sich dennoch auch für die zu t2 noch nicht trainierten Inhalte ein deutlicher Zuwachs zu t2 ergibt ( $d = 0,6$ ). Dieser Wert ist lediglich um den Wert 0,2 kleiner als der Zuwachs von t1 zu t2 für die bis dato trainierten Module ( $d = 0,8$ ). Wie für das deklarative Wissen bestätigt sich auch für das prozedurale Wissen das Ergebnis des Gesamtvergleiches: Das prozedurale Wissen nimmt zu t2 deutlich zu. Dabei spielt es jedoch keine große Rolle, ob die Inhalte bis dato trainiert worden sind oder nicht. Dieses Ergebnis könnte darauf hindeuten, dass die Items des Lernfragebogens nicht systematisch zwischen den einzelnen Modulen differenzieren.

Die hohen Rücklaufquoten weisen auch auf der Lernebene auf ein praktikables Vorgehen hin. Lediglich zu zwei Messzeitpunkten fehlen jeweils die Werte einer Person.

### 3.4 Ebene Verhalten

#### 3.4.1 Überprüfung der Wirksamkeit

Auf der Ebene Verhalten soll geprüft werden, ob die Teilnahme an der Ausbildung eine Verhaltensänderung im Arbeitsverhalten der Teilnehmer bewirkt. Die Überprüfung folgender Hypothese steht im Mittelpunkt der statistischen Analysen:

*Hypothese 1: Der Mittelwert für das gegenwärtige Verhalten ist signifikant höher als der Mittelwert für das Verhalten vor einem Jahr.*

Betrachtet man die Verhaltenseinschätzung durch die Kollegen zunächst nur bezogen auf das gegenwärtige Verhalten lassen sich über die Verhaltensbereiche aller Module hinweg hohe Werte konstatieren. Der Mittelwert liegt bei 3,8. Betrachtet man nun die Mittelwerte im Vergleich fällt auf, dass der Wert für das gegenwärtige Verhalten damit deutlich höher ausfällt als der Wert für das Verhalten vor einem Jahr ( $d = 0,6$ ). Der Wert vor einem Jahr lag auf einem mittleren Niveau (3,2), während der Wert für das gegenwärtige Verhalten im oberen Bereich der Skala (1-5) liegt (3,8).<sup>27</sup>

Laut t-test für abhängige Hypothesen wird dieser Mittelwertsunterschied auch signifikant ( $t_{(6)} = 4.241$ ;  $p = .005$ ).<sup>28</sup> Dieser Effekt kann auch als praktisch bedeutsam eingestuft werden ( $r = .87$ ). Auf Grundlage dieses Ergebnisses wird Hypothese 1 angenommen, es gilt: *Der Mittelwert für das gegenwärtige Verhalten ist signifikant höher als der Mittelwert für das Verhalten vor einem Jahr.*

#### Zusammenfassung und Interpretation der Ergebnisse

Die Ergebnisse auf der Verhaltensebene deuten darauf hin, dass die Teilnahme an der Ausbildung eine Veränderung im Arbeitsverhalten der Teilnehmer bewirkt hat. So bemerken die Kollegen einen signifikanten Zuwachs in der Umsetzung der Ausbildungsinhalte im Vergleich zu dem Verhalten vor einem Jahr.

#### 3.4.2 Überprüfung von Indikatoren für die Eignung des Ansatzes

##### Wissenschaftliche Anforderungen

Um die gefundenen Effekte auf der Verhaltensebene zu validieren, werden die Ergebnisse der Fremdeinschätzung mit denen der Selbsteinschätzung bezüglich des Verhaltens verglichen:

---

<sup>27</sup> Eine ausführliche Darstellung der deskriptiven Analysen findet sich im Anhang (Appendix C4, S. 148).

<sup>28</sup> Die vollständigen Analysen der Signifikanzprüfung finden sich im Anhang (Appendix C4, S. 149).

*Hypothese 1: Die Fremd- und Selbsteinschätzung des Verhaltens stimmen überein.*

- *H1a: Der Mittelwert für das gegenwärtige Verhalten unterscheidet sich nicht signifikant von der Einschätzung des prozeduralen Wissens durch die Teilnehmer zu t3.*

Die Betrachtung der Mittelwerte zeigt, dass die Kollegen die Umsetzung der Lerninhalte der Ausbildung mit einem Wert von 3,9 leicht höher bewerten als die Teilnehmer selbst (3,6). Es ergibt sich ein Differenzwert von  $d = 0,3$ . Laut t-test für abhängige Stichproben wird dieser Unterschied nicht signifikant ( $t_{(6)} = 1.468$ ;  $p = .193$ ). Hypothese 1 kann bestätigt werden, es gilt: *Die Fremd- und Selbsteinschätzung des Verhaltens stimmen überein.*

### Praktikabilität

Der Rücklauf auf der Verhaltensebene liegt bei weniger als 50 % der Fälle. Lediglich für 7 unterschiedliche Teilnehmer liegen Fremdurteile vor. Allerdings haben für einige Teilnehmer zwei oder mehr Kollegen an der Befragung teilgenommen, sodass insgesamt 12 Fremdurteile vorlagen. Um die Vergleichbarkeit der Daten zu gewährleisten und Verzerrungen in den Ergebnissen zu vermeiden wurde aus den verschiedenen Bewertungen zu einer Person das Mittel je Item gebildet. Die Grundlage der Berechnungen waren damit lediglich 7 Fälle.

### Zusammenfassung und Interpretation der Ergebnisse

Der Abgleich zwischen Fremd- und Selbsteinschätzung ergab, dass die Kollegen und die Teilnehmer in ihren Einschätzungen bezüglich der Umsetzung der Lerninhalte der Ausbildung in der Arbeit übereinstimmen. Dieser Befund spricht für die Validität der Ergebnisse.

Die Praktikabilität des Vorgehens der Erhebung auf der Verhaltensebene ist auf Grundlage der Daten als eingeschränkt zu beurteilen. So wurden lediglich für 7 unterschiedliche Teilnehmer Einschätzungen vorgenommen.

## **3.5 Ebene Resultate**

Auf der Ebene Resultate soll die Frage beantwortet werden, welche Kennzahlen sich zur Erfassung des Ausbildungserfolges auf der Unternehmensebene eignen. Um Anhaltspunkte dafür zu bekommen, was sinnvolle Kennzahlen in den Unternehmen sein könnten, wurde vorab die Frage gestellt, welche Maßnahmen die Teilnehmer seit der Teilnahme an der Ausbildung in ihrem Unternehmen umsetzen konnten.

Die Ergebnisse wurden zunächst im Rahmen einer qualitativen Auswertung inhaltlichen Kategorien zugeordnet, die dann wiederum den in der Literatur vorgeschlagenen *weichen* und *harten* Faktoren als Indikatoren organisationaler Leistung zugeordnet wurden (Kap. 2.4.1.4). Betrachtet man zunächst die Ergebnisse zu der Frage nach umgesetzten Maßnahmen zur Innovationspromotion fällt auf, dass die Mehrheit der genannten Maßnahmen als Indikatoren des **Arbeits- und Organisationsklima** gelten können (15 Nennungen), gefolgt von Maßnahmen, die sich primär auf die **Qualität der Leistung auswirken** (14 Nennungen) (Tab. 5). Maßnahmen, wie die Einführung eines Ideenmanagement-Systems, für die sich ggf. Indikatoren zur **Steigerung der Produktivität** ableiten lassen, wurden deutlich seltener genannt (5 Nennungen).

Umgesetzte Maßnahmen zur Innovationspromotion	Häufigkeit
<b>Arbeits- und Organisationsklima</b>	<b>15</b>
4-Augengespräche mit Mitarbeitern	2
Erhöhung der Sichtbarkeit des Themas Innovation (Information der MA)	2
Austausch zwischen Abteilungen fördern (z.B. in Innovations-Zirkeln)	3
Mehr Transparenz im Ideenmanagementprozess (Information der MA)	1
Motivation des Teams	2
Präsenz zeigen als Innovationspromotor, Ansprechbarkeit signalisieren	1
Projektabschlüsse feiern	1
Stärkere Partizipation der Mitarbeiter ermöglichen	1
Wertschätzung der Mitarbeiter	1
<b>Qualität der Leistung</b>	<b>14</b>
Unterstützung und Feedback in Innovationsprojekten	3
Einführung eines Innovationsraumes: "Freiräume zum Querdenken"	3
Prozessoptimierung im Ideenmanagement	4
Einführung einer Softwareunterstützung	2
Projektsteuerung: Kick-Off-Veranstaltung	1
Systematisches Projektmanagement (längerfristige Planungen)	1
<b>Steigerung der Produktivität</b>	<b>5</b>
Einführung eines Ideenmanagement-Systems	4
Projektumsetzung in Zielvereinbarungen der Führungskräfte aufgenommen	1
<b>Keine Nennung möglich</b>	<b>3</b>
Keine direkten Maßnahmen umgesetzt	3

Tabelle 5 Ergebnisse der qualitativen Auswertung von Interviewfrage 1 „Maßnahmen der Innovationspromotion“

Betrachtet man im Vergleich dazu die Antworten zu Frage 2 (Kennzahlen der Innovationspromotion) zeigt sich ein gegensätzliches Bild. Hier überwiegen die *harten Faktoren*, wie die Anzahl neuer Ideen oder die Anzahl umgesetzter Projekte (Tab. 6). Die umgesetzten Maßnahmen im Bereich des **Arbeits- und Organisationsklima** wurden lediglich in einem Fall in Kennzahlen übertragen. Am häufigsten werden von den Teilnehmern Kennzahlen zur Reflektion ihrer Arbeit genannt, die im Rahmen des Innovations-Management-Systems in ihrem Unternehmen erhoben werden.



Mögliche Kennzahlen als Indikatoren organisationaler Leistung	Häufigkeit
Steigerung der Produktivität	10
Ideenmanagement: Anzahl und Umsetzung von Ideen	8
Anzahl umgesetzter Projekte in einem Kalenderjahr	2
Qualität der Leistung	8
Ideenmanagement: Qualität neuer Ideen	8
Arbeits- und Organisationsklima	1
Weniger Fehltage der Mitarbeiter	1
Keine Kennzahlen vorhanden	4
Keine Kennzahlen für Innovationspromotion	4

Tabelle 6 Ergebnisse der qualitativen Auswertung zu Interviewfrage 2 „Kennzahlen der Innovationspromotion“

### Zusammenfassung und Interpretation der Ergebnisse

Die Darstellung der Ergebnisse zeigt, wie schwierig es ist, die Arbeit eines Innovationspromotors in konkreten Kennzahlen auszudrücken und dementsprechend Erfolgsindikatoren zu generieren, die die Auswirkungen der Teilnahme an der Ausbildung auf Unternehmensebene belegen können. Zunächst fällt auf, dass die umgesetzten Maßnahmen zur Innovationspromotion nur zu einem geringen Teil auch in Kennzahlen ausdrücken lassen. Während die Mehrheit der genannten Maßnahmen sich am ehesten in sogenannte *weiche Faktoren* als Kennzahlen umsetzen ließe (*Arbeits- und Organisationsklima, Qualität der Leistung*), werden als Kennzahlen in den Unternehmen *harte Kriterien* bevorzugt (*Steigerung der Produktivität*). Als geläufigste Kennzahlen wurden dabei die Anzahl, Umsetzung und Qualität der Ideen im Rahmen eines Ideenmanagement-Systems genannt. Diese Kriterien eignen sich aber aufgrund des beschriebenen Problems der Kriterienkontamination (Kap. 2.4.1.4) nur bedingt dazu, die Arbeit der Innovationspromotoren zu reflektieren. So sind zahlreiche andere denkbare Entwicklungen und Einflussgrößen in der Organisation ausschlaggebend dafür, ob mehr und qualitativ hochwertigere Ideen eingereicht werden.

## 3.6 Transferbedingungen

### 3.6.1 Überprüfung der Wirksamkeit

#### Deskriptive Analysen

Zunächst interessiert die Frage, wie sich die Ausprägungen der Transfervariablen darstellen. Es werden dabei vorrangig die Ergebnisse je Skala berücksichtigt (Tab. 4), in Einzelfällen auch Ergebnisse einzelner Items diskutiert. Zunächst kann festgehalten werden, dass die Transferbedingungen insgesamt im Durchschnitt eher positiv eingeschätzt werden ( $3,5 = 0,5$  Einheiten über dem Skalenmittel).

Skala (Anzahl Items)	Mittelwert	Standardabweichung	Minimum	Maximum	Range
Motivation zum Lerntransfer (4)	3,6	,78	2,8	5,0	2,3
Erwartungsklarheit (4)	2,6	,81	1,3	4,0	2,7
Sanktionen durch Vorgesetzte (5)	4,8	,45	3,4	5,0	1,6
Unterstützung durch Vorgesetzte (5)	2,7	1,1	1,4	4,6	3,2
Unterstützung durch Kollegen (4)	3,1	,84	1,5	4,5	3,0
Möglichkeiten der Wissensanwendung (5)	4,0	,66	2,4	5,0	2,6
Persönliche Transferkapazität (4)	3,5	,91	2,0	5,0	3,0
Transferdesign (4)	3,8	,68	2,8	4,8	2,0
Trainings-Job-Übereinstimmung (5)	3,3	,52	2,6	4,2	1,6
<b>LTSI Gesamt</b>	<b>3,5</b>	<b>,53</b>	<b>2,4</b>	<b>4,1</b>	<b>1,7</b>

Tabelle 7 Ergebnisse der deskriptiven Analysen für die Skalen des GLTSI

Betrachtet man die Ergebnisse für die Variablen­gruppe *Motivation (Trainee-Charakteristika)* zeigt sich ein heterogenes Bild. Eher positiv wird die Skala *Motivation zum Lerntransfer* (3,6) eingeschätzt.<sup>29</sup> So stimmen 64% der Teilnehmer den Variablen der Skala *überwiegend zu* bis *völlig zu* und weitere 36% der Probanden stimmen immerhin noch *teilweise zu*. Die Betrachtung der einzelnen Items der Skala zeigt, dass die Teilnehmer insbesondere ihre Motivation, das Gelernte anzuwenden, als hoch einstufen (4,0). Deutlich schlechter wird hingegen die Skala *Erwartungsklarheit* bewertet (2,6). So gibt die Mehrheit der Teilnehmer (79%) an, lediglich *ein wenig* oder *teilweise* gewusst zu haben, was sie von der Ausbildung zu erwarten haben. Auch die Auswirkungen der Ausbildung auf ihre Arbeitsleistung war laut eigener Angaben nur rund 30% der Teilnehmer klar.

Für die zweite Gruppe der **Umfeldbedingungen** gehen zum einen die Bewertungen für die unterschiedlichen Skalen auseinander. Zum anderen divergieren die Einschätzungen der Teilnehmer untereinander zum Teil stark. Für die Skala *Sanktionen durch Vorgesetzte* zeigt sich aber zunächst ein recht homogenes Bild. So stimmen 93% der Teilnehmer den Variablen dieser Skala *überwiegend* bis *völlig zu* (*Die Skala wurde umgepolt, hohe Werte bedeuten nun geringe Sanktionen*). Die Teilnehmer erwarten weder Kritik an der Ausbildung (4,9), noch äußern sie Bedenken, ihr Vorgesetzter könne die Effektivität der Ausbildung anzweifeln (4,9). Auf der anderen Seite fühlen sich aber lediglich 28% der Teilnehmer von ihren Vorgesetzten dabei unterstützt, das Gelernte in der Arbeit anzuwenden. Demgegenüber geben 57% der Teilnehmer an, diese Unterstützung lediglich *ein wenig* oder *überhaupt nicht zu* erleben. Etwas positiver wird die Unterstützung durch die Kollegen eingeschätzt. Zwar stimmen wieder

<sup>29</sup> Eine Übersicht der deskriptiven Daten für die einzelnen Items findet sich im Anhang (Appendix C5, S. 150).

lediglich 36% der Teilnehmer den Variablen der Skala *überwiegend bis völlig zu*. Weitere 43% der Befragten fühlen sich aber immerhin noch *teilweise* unterstützt.

Zwei weitere Skalen, die zum Teil auch das Arbeitsumfeld betreffen, in der Struktur des GLTSI aber unter dem Terminus **Möglichkeiten** gefasst werden, erhalten deutlich positivere Bewertungen. So geben 78% der Teilnehmer an, genug Budget, Materialien und Personal zur Verfügung zu haben, um das Gelernte in ihrem Unternehmen anzuwenden (*Skala Möglichkeiten der Wissensanwendung*). Auch die *persönliche Transferkapazität* beurteilen die Teilnehmer im Durchschnitt als gut (3,5). Die Mehrheit der Teilnehmer (64%) gibt an, ihr Tagesablauf erlaube es, Dinge zu ändern und neue Kenntnisse einzubringen. Auch die Arbeitsbelastung wird von immerhin 57% der Befragten eher nicht als Hinderungsgrund für die Umsetzung der Erlernten in der Arbeit erlebt. Betrachtet man weiter die Transferbedingungen, die das Training betreffen, zeigt sich für die Skala *Transferdesign*, dass 64% der Teilnehmer die in der Ausbildung verwendeten Übungen, Beispiele und die Art der Vermittlung der Inhalte als hilfreich erleben. Weitere 36% der Teilnehmer stimmen dieser Aussage *teilweise zu*. Bezüglich der *Arbeits-Trainings-Übereinstimmung* stimmt die Mehrheit der Befragten immerhin *teilweise zu* (64%). Etwa 36% der Teilnehmer erleben die Variablen der Skala als *überwiegend zutreffend*. Für die **Möglichkeiten** zum Lerntransfer kann also insgesamt positives Bild bestätigt werden.

### Überprüfung von Zusammenhängen

In diesem Abschnitt soll nun überprüft werden, inwieweit die Transfervariablen systematisch mit den Ergebnissen auf der Lern- und Verhaltensebene variieren.

Für das **deklarative und prozedurale Wissen** und die Skalen des LTSI zeigen sich nur wenige statistisch bedeutsame Zusammenhänge.<sup>30</sup> Von den beiden auf die **Motivation** bezogenen GLTSI-Skalen korreliert die Skala *Motivation zum Lerntransfer* positiv mit dem **prozeduralen Wissen zu t2** ( $r = .614, p = .25$ ). Zu t3 wird die Korrelation für das prozedurale Wissen und die Skala knapp nicht signifikant ( $r = .539; p = .57$ ). Für die Skala *Erwartungsklarheit* können hingegen keine signifikanten Korrelationen berichtet werden.

Im Bereich der **Umfeldbedingungen** für Lerntransfer ergibt sich eine systematische Variation der Messwerte für die Skala *Sanktionen durch Vorgesetzte* und das **dekla-**

---

<sup>30</sup> Die vollständige Korrelationsmatrix ist im Anhang einsehbar (Appendix C5, S. 151). Nachfolgend sollen lediglich die signifikanten Ergebnisse berichtet werden. Daneben wird auf Korrelationen verwiesen, für die das Signifikanzniveau von  $\alpha = .05$  nur knapp überschritten wird.

**rative Wissen zu t3** ( $r = .596$ ;  $p = .32$ ). Für eine weitere Skala aus dem Bereich - die *Unterstützung durch Vorgesetzte* - wird die Korrelation zu dem prozeduralen Wissen zu t2 knapp nicht signifikant ( $r = .542$ ;  $p = .56$ ).

Für die Skalen im Bereich *Möglichkeiten* werden zwei positive Korrelationen mit den Lernerfolgsmaßen signifikant: Die Skala *Transferkapazität* korreliert signifikant positiv mit dem **deklarativen** und **prozeduralen Wissen** jeweils zu t2 ( $r = .636$ ;  $p = .019$ ;  $r = .691$ ;  $p = .009$ ). Ebenfalls eine hohe Korrelationen, die nur knapp nicht signifikant wird, zeigt sich zwischen der Skala *Trainings-Arbeits-Übereinstimmung* und dem deklarativen Wissen zu t1 ( $r = .549$ ,  $r = .52$ ).

Der Konvention nach Cohen (1988) folgend, können die gefundenen signifikanten Effekte als groß und damit auch praktisch bedeutsam eingestuft werden.

Für die Ergebnisse aus der *Fremdevaluation des Verhaltens* durch die Kollegen ergeben sich **gar keine** signifikanten Korrelationen zu den gemessenen Transfervariablen. So ist davon auszugehen, dass die Variablen kaum systematisch variieren.

#### Zusammenfassung und Interpretation der Ergebnisse

Die nachfolgende Zusammenfassung und Interpretation der Ergebnisse bezieht sich auf die beiden formulierten Fragestellungen für diesen Untersuchungskomplex. So werden die Ausprägungen der Transfervariablen immer auch schon im Hinblick auf mögliche Zusammenhänge zu den Erfolgsmaßen auf den Output-Ebenen diskutiert.

Betrachtet man zunächst die auf die *Motivation* bezogenen GLTSI-Skalen, unterscheiden sich die Ausprägungen der beiden Skalen stark voneinander. Während die Motivation zum Lerntransfer hoch eingeschätzt wird, war den Teilnehmern überwiegend unklar, was sie von dem Training zu erwarten hatten und wie es sich auf die Arbeitsleistung auswirken könnte. Da die Überprüfung der Zusammenhänge zwischen der Skala und den Lernerfolgsmaßen keine signifikanten Ergebnisse zeigt, ist für die untersuchte Maßnahme dennoch nicht davon auszugehen, dass sich die Unklarheiten in den Erwartungen an das Training hinderlich auf Transferprozesse ausgewirkt hat. Die Skala *Motivation zum Lerntransfer* korreliert hingegen signifikant positiv mit dem prozeduralen Wissen zu t2. Die hohe Motivation zur Umsetzung des Gelernten geht also mit hohen Werten auf der Lernebene einher. Da die Transferbedingungen lediglich zu einem Messzeitpunkt nach Abschluss überprüft wurden und die Lernerfolgsmaße zu drei Zeitpunkten, ist es nicht möglich differenziert zu prüfen, inwieweit die Transferbedingungen wirklich Einfluss auf die Erfolgsindikatoren auf

der Lernebene zu den jeweiligen Zeitpunkten genommen haben. Damit begrenzt das Untersuchungsdesign die Möglichkeiten, das Zusammenwirken von ergebnis- und prozessbezogenen Variablen zu beurteilen stark. Dieser Kritikpunkt muss in der Diskussion des Evaluationsansatzes noch näher beleuchtet werden.

Bei der Betrachtung der auf das *Arbeitsumfeld* bezogenen GLTSI-Skalen fällt für die Transfervariablen für sich genommen zunächst auf, dass diese von den Teilnehmern zum Teil sehr heterogen eingeschätzt werden. Auch für die Skalen untereinander unterscheiden sich die Ergebnisse zum Teil stark. Ein recht homogenes Bild zeigt sich noch für die Skala *Sanktionen durch Vorgesetzte*, von denen die deutliche Mehrheit der Teilnehmer (93%) laut eigener Angaben nicht betroffen war. Die Skala korreliert positiv mit den Lernerfolgsmaßen zu allen drei Messzeitpunkten, wobei der Zusammenhang lediglich für das deklarative Wissen zu t3 signifikant wird. Hohe Werte auf der Skala gehen also zu allen drei Zeitpunkten mit hohen Ausprägungen auf der Lernebene einher. So wenig sich die Teilnehmer also durch ihre Vorgesetzten an der Umsetzung der Inhalte der Ausbildung gehindert fühlen, so wenig fühlen sie sich auf der anderen Seite aber auch aktiv dabei unterstützt. So stimmt die Mehrheit der Teilnehmer (57%) den Variablen der Skala lediglich ein wenig oder überhaupt nicht zu. Betrachtet man die Ergebnisse für die Korrelationen für die Skala mit den Erfolgsmaßen auf der Lernebene, scheint die mangelnde Unterstützung der Vorgesetzten dem Transfer des Gelernten aber nicht systematisch im Weg zu stehen. So lassen sich kaum signifikanten Zusammenhänge ausmachen. Lediglich zu dem prozeduralen Wissen zu t2 wird ein positiver Zusammenhang berichtet, der aber knapp nicht signifikant wird. So fühlen sich ja auch immerhin knapp 30% der Teilnehmer von ihren Vorgesetzten bei der Anwendung des Gelernten unterstützt. Von den Kollegen fühlt sich die Mehrheit der Teilnehmer mindestens teilweise bei der Umsetzung des Gelernten unterstützt. Die Variablen der Skala korrelieren in keine Richtung signifikant mit den Ausprägungen für das deklarative und prozedurale Wissen.

Für die GLTSI-Skalen, die sich auf die *Möglichkeiten* beziehen, zeigt sich im Vergleich zu den bisher berichteten Ergebnissen ein positiveres Bild. Sowohl die Möglichkeiten der Wissensanwendung als auch die persönliche Transferkapazität werden von der Mehrheit der Teilnehmer positiv eingeschätzt. Die positive Einschätzung der Transferkapazität variiert darüber hinaus systematisch mit hohen Ausprägungen des deklarativen und prozeduralen Wissens jeweils zu Messzeitpunkt 2. Für die Möglichkeiten der Wissensanwendung und Ergebnisse auf der Lernebene lassen sich hin-

gegen keine signifikanten Korrelationen berichten. Auch für die Skalen Trainingsdesign und Trainings-Arbeits-Übereinstimmung liegen die Ausprägungen der Variablen überwiegend im positiven Bereich. Es zeigen sich keine signifikanten Korrelationen zu den Ergebnissen auf der Lernebene.

Insgesamt kann festhalten werden, dass die Bedingungen für den Lerntransfer in der Teilnehmergruppe unterschiedlich gut ausgeprägt waren. Die Ergebnisse der Zusammenhangsüberprüfung geben aber keine Anhaltspunkte dafür her, dass sich schlechtere Bedingungen wirklich als Barrieren für Lernen und Transfer herausstellen. So gibt es zwischen den Lernerfolgsmäßen und den Transfervariablen nur einzelne Zusammenhänge, die sich aufgrund des Untersuchungsdesigns auch nur eingeschränkt interpretieren lassen. Zwischen der Fremdbeurteilung des Verhaltens und den Transfervariablen lassen sich sogar überhaupt keine Zusammenhänge ausmachen. Dieser Befund ist erst einmal ungewöhnlich und weist darauf hin, dass der Transfer der Lerninhalte der Ausbildung in die Arbeit gelungen ist unabhängig davon, wie gut oder schlecht die Transferbedingungen für die Teilnehmer ausgeprägt waren.

### **3.6.2 Überprüfung von Indikatoren für die Eignung des Ansatzes**

#### Wissenschaftlicher Anspruch

Im Rahmen der vorliegenden Untersuchung wurde erneut die Reliabilität der GLTSSI-Skalen bestimmt. Es zeigt sich, dass die Reliabilitätswerte für die Skalen überwiegend im mittleren Bereich liegen ( $\alpha < .80$ ) Lediglich die Skala *Trainings-Arbeits-Übereinstimmung* erreicht niedrige Werte ( $\alpha > .70$ ).

#### Praktikabilität

Im Rahmen der Abschlusserhebung wurden auch die Transferbedingungen mit abgefragt. Es nahmen alle 15 Teilnehmer teil. Wie beim Lernfragebogen fehlten bei einer Person die Daten aufgrund technischer Probleme.

#### Zusammenfassung und Interpretation der Ergebnisse

Die Reliabilitätsschätzungen für das GLTSSI liegen überwiegend im mittleren Bereich und stimmen weitestgehend mit den berichteten Ergebnissen dazu in der Literatur überein (Kauffeld, 2008).

Die Rücklaufquote kann als hoch gelten. Wieder fehlten aufgrund technischer Probleme die Daten einer Person.

## 4. Diskussion

Im Vordergrund des Kapitels steht die Frage, wie die Eignung des entwickelten Evaluationsansatzes im Lichte der Befunde einzuschätzen ist. Einerseits werden dazu anhand der exemplarischen Anwendung des Ansatzes im Rahmen der Wirksamkeitsanalyse der Ausbildung zum Innovationspromotor Schlussfolgerungen über die Eignung des Ansatzes gezogen. Andererseits dienen die Befunde der systematischen Untersuchung von Indikatoren für den wissenschaftlichen Anspruch und die Praktikabilität des Ansatzes als wichtige Quelle zur Diskussion der Forschungsfrage. Ein Rückbezug von den beobachteten Sachverhalten auf den theoretisch formulierten Ansatz kann jedoch nur mit äußerster Vorsicht erfolgen, da eine vollständige empirische Überprüfung des Ansatzes im Rahmen der Untersuchung nicht möglich war. Daran anknüpfend werden Implikationen für die Forschung und die Praxis diskutiert und der Beitrag der Arbeit sowie weiterführender Forschungsbedarf resümiert.

### 4.1 Die Eignung des Ansatzes im Lichte der Befunde

#### 4.1.1 Ebene Reaktion

Auf der Ebene Reaktion konnte ein differenziertes Bild über die Zufriedenheit der Teilnehmer mit der Ausbildung gewonnen werden. Insbesondere das Längsschnittdesign mit den 10 Messzeitpunkten ermöglichte die Generierung einer soliden und differenzierten Datengrundlage. Es wurden statistisch bedeutsame Unterschiede in der Bewertung zwischen den einzelnen Modulen deutlich, die Aufschluss darüber gaben, welche Module einer Überarbeitung bedürfen. In der Kombination mit den Ergebnissen aus der qualitativen Analyse der offenen Kommentare konnten somit wichtige Anhaltspunkte und Stellschrauben zur Weiterentwicklung der Ausbildung gewonnen werden.

Im Rahmen der Wirksamkeitsanalyse ergeben sich bereits Indikatoren dafür, inwieweit es gelingt, wissenschaftlichen Anforderungen bei der Evaluation zu genügen. So spricht die Variation der Bewertungen zwischen den Modulen dafür, dass die Teilnehmer die Befragung gewissenhaft ausgefüllt und nicht einfach automatisch hohe Werte für jedes Modul eingetragen haben - bspw. aus Desinteresse an der Befragung oder aus dem Motiv der sozialen Erwünschtheit heraus. Allerdings muss einschränkend festgehalten werden, dass die Bewertungen der Zufriedenheit insgesamt sehr hoch ausfallen und die Variabilität der Messwerte pro Messzeitpunkt zum Teil ge-

ring ist. Um diese zu erhöhen und insbesondere eine höhere Differenzierung der Werte im positiven Bereich zu erreichen, könnte der Einsatz einer 7er Skala anstelle der bisherigen 5er Skala ratsam sein (Russ-Eft, 2001). Für das Erhebungsinstrument besteht noch in anderer Hinsicht Weiterentwicklungsbedarf. Hierfür kann zunächst wieder ein Befund der Wirksamkeitsanalyse herangezogen werden: So zeigt sich, dass signifikante Unterschiede in den Einschätzungen der Zufriedenheit zwischen den drei Dimensionen Trainer, Inhalt und Didaktik bestehen. Die Bewertung der Trainer fällt über alle Module hinweg signifikant positiver aus. Auch in den sonst schlechter bewerteten Modulen erreichen die Trainer sehr gute Bewertungen, was die Werte der Gesamtzufriedenheit in die Höhe treibt und von möglichem Weiterentwicklungsbedarf in den Bereichen Inhalt und Didaktik ablenkt. Bei der Erhebung der Zufriedenheit mit der Ausbildung sollte den beiden Bereichen aber ein stärkeres Gewicht verliehen werden, da sie für die Qualität eines Trainings zentral sind. Zudem ist die Dimension Trainer anfällig für Störeinflüsse. So werden Einschätzungen von Personen stärker durch affektive Größen, wie Sympathie/Antipathie, und weniger durch objektive Kriterien geleitet. Zudem kann das Motiv der sozialen Erwünschtheit in besonderer Weise zum Tragen kommen, da die Teilnehmer im Unterschied zur Bewertung des Inhaltes und der Didaktik bei der Bewertung der Trainer das Gefühl bekommen, diese persönlich zu treffen (*insbesondere im Fall schlechter Werte*). Es kann daher ratsam sein, Modifikationen an dem Erhebungsinstrument vorzunehmen und die Anzahl der Items für die Skalen Inhalt und Didaktik mindestens auf das Niveau der Anzahl der Trainerskala zu erhöhen.

Die Analysen der *Indikatoren zur Prüfung der Eignung des Verfahrens* stützen dieses Ergebnis. So zeigt sich, dass die Reliabilitäten, insbesondere für die Skalen Didaktik und Inhalt, stark schwanken und sich zu einzelnen Modulen inakzeptable Reliabilitätswerte zeigen. Für die Skala Trainer werden hingegen für alle Messungen mittlere bis gute Werte für die interne Konsistenz berichtet. Auch dieses Ergebnis kann unter anderem auf die geringere Anzahl an Items für die Skalen Inhalt und Didaktik zurückgeführt werden. Im Umkehrschluss lässt sich die Reliabilität der Skalen erhöhen, wenn für beide Skalen Items hinzugenommen werden. In der eingesetzten Form genügt das Erhebungsinstrument nicht hinreichend den wissenschaftlichen Anforderungen.

Für die *Praktikabilität* des Vorgehens auf der Reaktionsebene finden sich hingegen positive Befunde. So zeigt sich für alle Erhebungszeitpunkte ein hoher Rücklauf.



#### 4.1.2 Ebene Lernen

Auch auf der Ebene Lernen trug das Untersuchungsdesign mit den drei Messzeitpunkten maßgeblich dazu bei, ein differenziertes Bild der Wirkungen aufzeigen zu können. Ein reiner Prä-Post-Test-Vergleich hätte zu weniger aussagekräftigen Ergebnissen geführt. So machte insbesondere die Einbeziehung der Ergebnisse der Zwischenerhebung die Beurteilung des Verlaufes der Wissensentwicklung möglich.

Für das *prozedurale Wissen* lassen sich die Messergebnisse zu den drei Zeitpunkten vor allem im Hinblick auf die Stabilität der Effekte interpretieren. Der zurückgehende Wert zu t3 deutet darauf hin, dass die Umsetzung des Gelernten nicht nachhaltig gelingt. Dieses Ergebnis stützt die konzeptionellen Überlegungen zur Wahl adäquater Zeitpunkte für die Messungen (Teil III, Kap. 3.3.3). Für die Ausbildung zeigen sich, anders als bei einem punktuellen Training, nicht erst nach Abschluss der Durchführung Veränderungen auf der Lern- und Verhaltensebene. Die Umsetzung weiterer Nachher-Messungen wäre dennoch grundsätzlich wünschenswert. Neben einem hohen Aufwand schränken aber Gründe, wie die mangelnde Erreichbarkeit von Probanden über einen längeren Zeitraum nach Abschluss einer Maßnahme und ein damit einhergehender großer Drop-Out für eine bereits kleine Stichprobe, die Praktikabilität weiterer Postmessungen ein. Unter Abwägung der Diskussionspunkte kann positiv hervorgehoben werden, dass das Design auf der Lernebene bereits Aufschluss über die zeitliche Stabilität der Effekte zulässt und weitere Nachher-Messungen zwar wünschenswert, aber nicht zwingend notwendig sind.

Die Ergebnisse für das *deklarative Wissen* deuten auf einen Lerneffekt durch die Ausbildung hin (*signifikanter Zuwachs von t1 zu t3*). Unter Einbezug der Werte zu t2 wird dieser Befund noch gestützt. Der Rückgang des Wertes zu t2 lässt sich inhaltlich plausibel argumentieren und wird als Indikator dafür gewertet, dass die Teilnehmer in der vertieften Auseinandersetzung mit den Inhalten der Ausbildung zunächst erkennen, wieviel sie in den Themenfeldern doch noch nicht wissen. Folgt man dieser Argumentationslinie stützen die Befunde zu t2 die *Güte des gefundenen Effektes*. So würden Störeinflüsse wie der Hawthorne Effekt dazu führen, dass sich die Werte von t1 zu t2 erhöhen, da die Teilnehmer allein aufgrund der Tatsache, dass mit ihnen eine Intervention durchgeführt wurde, höhere Werte angeben (Teil II, Kap. 2.3.1). Ebenso unwahrscheinlich sind Verzerrungen der Ergebnisse durch das Phänomen der sozialen Erwünschtheit. Ein sozial erwünschtes Antwortverhalten würde

sich in der vorliegenden Studie mit großer Wahrscheinlichkeit in einer deutlichen Erhöhung der Werte zu t2 äußern. Die Teilnehmer möchten den angestrebten Wissenszuwachs erfüllen und nicht etwa als dumm oder faul gelten (Teil II, Kap. 2.3.1).

Die Untersuchung der *Güte des Erhebungsverfahrens* als Indikator dafür, inwieweit die Erhebung auf der Ebene wissenschaftlichen Anforderungen genügt, war im Rahmen der Analyse nur sehr eingeschränkt möglich. So wurden die Skalen des Instrumentes aus zum Teil völlig unabhängigen Lernbereichen gebildet (bspw. *Kompetenzentwicklung und innovationsförderliche Arbeitsplatzgestaltung für Modul 2*), um die globalen Lernziele der Module in ihrer Breite zu erfassen. Eine Analyse der Reliabilitäten der Skalen war damit nicht sinnvoll möglich. Hingegen sollte die Betrachtung des Wissenszuwachses von t1 zu t2 im Vergleich der bis zu t2 trainierten und der bis dato noch nicht trainierten Module Aufschluss darüber geben, inwieweit die Skalen systematisch zwischen den Lerninhalten der Module differenzieren. Weder für das deklarative noch für das prozedurale Wissen fiel der Wissenszuwachs für die bereits trainierten Module zu t2 deutlich höher aus als für die bis dato noch nicht trainierten Module. Die Ergebnisse weisen darauf hin, dass die Skalen des Lernfragebogens nicht systematisch zwischen den einzelnen Modulen differenzieren. Aus wissenschaftlicher Perspektive ist es ratsam Modifikationen an dem Erhebungsinstrument auf der Lernebene vorzunehmen. Ein Weg der Optimierung könnte darin bestehen, die Lernziele je Modul jeweils über weitere Items zu operationalisieren. Damit ließen sich homogenere Skalen bilden, für die wiederum etablierte Verfahren zur Bestimmung der Güte herangezogen werden könnten. Damit das Vorgehen weiterhin gleichzeitig praktikabel bleibt, ist eine Erhöhung der Item-Anzahl allerdings nur in Grenzen möglich. In der bisherigen Version kann die hohe Rücklaufquote als Indikator für die Praktikabilität des Vorgehens gewertet werden.

Positiv kann für das Instrument die Unterteilung in das deklarative und das prozedurale Wissens hervorgehoben werden. Damit wird eine differenzierte Einschätzung der Wissensentwicklung möglich. So zeigen sich für beide Bereiche ganz unterschiedliche Lernverläufe. Die Betrachtung eines Gesamtmittels zu allen Zeitpunkten, hätte aber dazu geführt, dass fälschlicherweise der Eindruck eines linearen Lernverlaufs entsteht ( $Mt1 = 3,5$ ;  $Mt2 = 3,7$ ;  $Mt3 = 3,9$ ). Hohe Korrelationen der Bereiche zu t2 und t3 ( $r_{12} = 0.7$ ;  $r_{13} = 0.8$ ) deuten zwar auf einen linearen Zusammenhang der beiden Wissensbereiche hin, sind infolge der vorangehenden Argumentation aber nicht so zu interpretieren, dass beide Bereiche im Prinzip das Gleiche erfassen.

### 4.1.3 Verhalten

Die Ergebnisse auf der Verhaltensebene deuten auf einen Transfererfolg der Ausbildung hin. Im Abgleich mit den Ergebnissen für das prozedurale Wissen zu t3 konnte der Befund validiert werden. So zeigte sich, dass die Selbst- und Fremdeinschätzungen bezüglich der Verhaltensänderung übereinstimmen.

Allerdings muss bei der Interpretation der Ergebnisse das Messdesign kritisch diskutiert werden. So ist davon auszugehen, dass ein retrospektives Vorgehen dazu verleitet, das gegenwärtige Verhalten automatisch *besser* zu bewerten. Insbesondere der Umstand, dass es sich bei den zu bewertenden Personen um Kollegen, teilweise um Vorgesetzte handelt, kann zu sozial erwünschtem Antwortverhalten führen. Die Tatsache, dass die E-Mail zum Befragungsportal über die jeweiligen Kollegen versandt wurde, kann diese Tendenz eher noch begünstigen. Trotz explizitem Hinweis auf die Gewährleistung der Anonymität, kann sich ein Misstrauen diesbezüglich eingestellt haben. Unabhängig von den aufgeführten Problemen ist es schwierig, das Verhalten einer Person im Nachhinein valide einzuschätzen. Ein Jahr ist dabei ein relativ langer Zeitraum. Auch in der Literatur wird betont, dass retrospektive Vorher-Nachher-Designs sehr anfällig für Störeinflüsse sind (Merzenich-Hieker, 1996).

Es wäre also ratsam die Anzahl der Messzeitpunkte auf der Verhaltensebene zu erhöhen und mindestens eine zweite Messung vorzunehmen. Optimal wäre die Umsetzung einer Prämessung, um eine Vergleichsgrundlage für mögliche Verhaltensänderungen durch die Ausbildung zu generieren. Diese lässt sich aus organisatorischen Gründen aber vermutlich nur schwer umsetzen. Ein handhabbarer Weg zur Erhebung von Ausgangswerten könnte darin bestehen, direkt zu Beginn der Ausbildung eine erste Messung umzusetzen und die Erhebung - wie auf der Lernebene geschehen - als Prämessung zu behandeln. Auch die Umsetzung einer Zwischenerhebung wäre ein denkbare Vorgehen, um einen weiteren Vergleichswert zu generieren.

Ein weiteres Problem der Erhebung auf der Ebene, welches die *Güte der Ergebnisse* einschränkt, ist die sehr kleine Stichprobe von  $N = 7$ . Die Ergebnisse der parametrischen Verfahren können daher nur unter Vorbehalt interpretiert werden.

Die niedrige Rücklaufquote deutet überdies darauf hin, dass die *Praktikabilität* des Vorgehens auf der Ebene eingeschränkt war. Betrachtet man den Rücklauf der Daten noch einmal genauer zeigt sich ein sehr heterogenes Bild. Während für etwa die Hälfte der Teilnehmer gar keine Fremdeinschätzungen vorgenommen wurden, nah-

men für einige Teilnehmer sogar mehrere Kollegen an der Befragung teil. Auf Kollegenseite reichen die möglichen Gründe für eine Nicht-Teilnahme an der Befragung prinzipiell von „kein Interesse“ über „keine Zeit“ bis hinzu „Unwohlsein bei der Bewertung von Kollegen“. Erschwerend kam hinzu, dass der Kontakt zu den Probanden nur indirekt verlief und die Teilnehmer als Vermittler fungieren mussten. Damit waren die Möglichkeiten der Einflussnahme durch den Evaluator geringer als auf den anderen Ebenen. Es ist überdies denkbar, dass die Teilnehmer selbst kein Interesse daran hatten durch ihre Kollegen bewertet zu werden und diese erst gar nicht um eine Bewertung gebeten haben. Um eine höhere Akzeptanz der Erhebung zu erreichen, sollten Zielsetzung und Vorgehen der Erhebung noch klarer herausgestellt werden.

Für die Erhebung von Verhaltensänderungen werden - wie im Zuge der konzeptionellen Überlegungen erläutert - auch häufig alternative Methoden zur klassischen Befragung gewählt. Methoden wie Rollenspiele, Simulationsübungen oder Videoanalysen wurden im Rahmen der Konzeption des *ganzheitlichen* Evaluationsansatzes aber für nicht praktikabel befunden (Teil III, Kap. 3.3.2.3). Zudem wurde das Verfahren der Lernebene als Grundlage zur Entwicklung des Instrumentes gewählt, um einen Bezug zwischen den Ergebnissen der Ebenen zu erleichtern. Dieses Vorgehen hat sich als gewinnbringend herausgestellt. So konnte die Validität der Ergebnisse im Selbstbild-/Fremdbildabgleich unterstrichen werden. Dennoch sollten alternative Wege für die Erhebung der Verhaltensebene zumindest überdacht werden, um das Verhalten eines Innovationspromotors im Unternehmen noch differenzierter erfassen zu können. Eine mögliche Methode, die von einem herkömmlichen Fragebogenformat abweicht und dennoch als praktikabel gilt, ist ein sogenannter Szenario-Fragebogen - ähnlich dem Führungsstilfragebogen von Schaper & Lieberei (2004). Die Entwicklung und Validierung eines solchen Instrumentes ist Gegenstand einer weiteren Diplomarbeit, die im Rahmen des GI:VE Projektes entsteht.<sup>31</sup>

#### **4.1.4 Transferbedingungen**

Im Rahmen der Analyse der Transferbedingungen wurden nur vereinzelt signifikante Zusammenhänge zu den Erfolgskriterien auf der Lern- und Verhaltensebene festgestellt. Es ließen sich also keine bedeutsamen Barrieren für den Lern- und Transfererfolg identifizieren. Dennoch wurden die Effekte auf den Output-Ebenen abgesichert und der Einfluss von Drittvariablen wenigstens zum Teil kontrolliert.

---

<sup>31</sup> Frau Martina Bierbichler beschäftigt sich im Rahmen ihrer Diplomarbeit mit der Entwicklung eines Verfahrens zur differenzierten Erfassung von Verhaltensweisen der Innovationspromotoren.

Der Erkenntnisgewinn durch den Einsatz des Verfahrens war allerdings geringer als erwartet, was zum Teil auch an dem Design der Erhebung lag. Da die Transferbedingungen lediglich zu einem Messzeitpunkt nach Abschluss der Ausbildung überprüft wurden, während die Erhebung der Lernerfolgsmaße zu drei Zeitpunkten erfolgte, war es nicht möglich differenziert zu prüfen, inwieweit die Transferbedingungen wirklich Einfluss auf die Erfolgsindikatoren auf der Lernebene zu den jeweiligen Zeitpunkten genommen haben. Es ist daher ratsam in der Weiterentwicklung des Evaluationsansatzes die Anzahl der Erhebungen der Transferbedingungen analog zu der Anzahl der Erhebungen auf der Lern- und der Verhaltensebene zu gestalten. Trotz eines Mehraufwandes, der sich damit ergibt, ist das Vorgehen als hinreichend *praktikabel* einzuschätzen. Die sehr hohen Rücklaufquoten deuten darauf hin, dass sich die Erhebung der Transferbedingungen in der ausgeführten Form als sehr praktikabel erwies. Auch aus *wissenschaftlicher Perspektive* sollte der Einsatz eines validierten Verfahrens im Rahmen einer ganzheitlichen Trainingsevaluation unbedingt beibehalten werden. In der vorliegenden Untersuchung wurde die gute Reliabilität der Skalen des GLTSI erneut bestätigt. So ist anzunehmen, dass das Verfahren gerade bei bestehenden Transferproblemen - *wie sie für die Ausbildung auf Grundlage der Ergebnisse nicht anzunehmen sind* - gute Diagnosemöglichkeiten bietet.

#### **4.1.5 Gesamtmodell**

Es kann zusammenfassend festgehalten werden, dass die Ergebnisse der empirischen Analysen für die Eignung des Ansatzes sprechen. So konnte auf den erhobenen Ebenen ein differenziertes Bild über die Wirkungen der Ausbildung gezeichnet werden. Unter Einbeziehung von Einflussvariablen auf den Trainings- und Transferprozess konnten die Ergebnisse auf der Output-Ebene gestützt werden.

Die *wissenschaftlichen Anforderungen* an einen ganzheitlichen Evaluationsansatz wurden allerdings nur zum Teil erfüllt. Auf allen Ebenen wurde ein Weiterentwicklungsbedarf für die Erhebungsinstrumente festgestellt. Allerdings konnte auf Grundlage der Ergebnisse für jedes Verfahren bereits konkrete und vor allem handhabbare Möglichkeiten zur Weiterentwicklung der Instrumente angeführt werden.

Aus methodischer Sicht ist darüber hinaus die geringe Stichprobengröße der Studie problematisch. Die Ergebnisse der angewandten statistischen Verfahren wie der Varianzanalyse und des t-tests, können allesamt nur unter Vorbehalt interpretiert werden.

Allerdings bietet das *messwiederholende Untersuchungsdesign* hierzu einen Ausgleich, da sich die Datengrundlage für die Berechnungen unter Einbezug der Werte zu den verschiedenen Messzeitpunkten erhöht. Die Abhängigkeit der Daten, die sich damit ergibt, fordert allerdings an vielen Stellen den Einsatz besonderer Messverfahren. Überdies sind bestimmte Analysen, wie bspw. die Bestimmung und Klassifizierung von Effektgrößen, nur eingeschränkt möglich.

Im Rahmen des Untersuchungsdesigns der Studie - mit dem Fokus auf der Umsetzung mehrerer Messzeitpunkte - konnten dennoch aufschlussreiche und differenzierte Erkenntnisse über die Wirkungen der Ausbildung gewonnen werden. Ein eindeutiger Ausschluss von Alternativerklärungen für die Ergebnisse war ohne den Einbezug einer Kontrollgruppe aber - wie erwartet - nicht möglich. Insbesondere das Auftreten von Störeinflüssen, die die interne Validität der Ergebnisse in Frage stellen wie *Reifungseffekte*, *Zeiteinflüsse* oder die *Testübung* (Teil II, Kap. 2.3.1), konnten nicht kontrolliert werden. Wie es allerdings gelingen kann für umfassende Trainingsprogramme - wie die untersuchte Maßnahme - eine adäquate Kontrollgruppe zu bilden, bleibt wohl eine der größten Herausforderungen von Trainingsevaluationen. Auf Grundlage von messwiederholenden Designs ohne Kontrollgruppe lassen sich aber - wie die vorliegende Studie zeigt - zumindest Anhaltspunkte für die Einflussnahme von Störeinflüssen gewinnen. So deuten die Ergebnisse an verschiedener Stelle darauf hin, dass ein Einfluss von Störvariablen wie dem *Hawthorne-Effekt* oder der *sozialen Erwünschtheit* eher nicht anzunehmen ist.

Neben der internen Validität der Ergebnisse interessiert die Frage, inwieweit sich die Befunde der Studie über die spezifischen Untersuchungsbedingungen hinaus verallgemeinern lassen (Teil II, Kap. 2.3). Die hohe Natürlichkeit der Untersuchungsbedingungen der vorliegenden Studie, die komplett im Feld durchgeführt wurde, erhöht die externe Validität der Befunde. Da die Stichprobe der Untersuchung aber nicht als repräsentativ gelten kann (*freiwillige Teilnahme an der Ausbildung*, *Selbstselektion*), ist dennoch von einer eingeschränkten externen Validität der Ergebnisse auszugehen (Bortz & Döring, 2006). Um die Repräsentativität der Ergebnisse zu steigern, wäre es notwendig, den entwickelten Ansatz anhand weiterer Trainingsmaßnahmen zu prüfen. Die Befunde aus Trainingsevaluationen lassen sich dennoch immer nur bedingt Verallgemeinern, da für jede Trainingsmaßnahme wieder spezifische Anpassungen im Evaluationskonzept vorgenommen werden müssen.

## 4.2 Implikationen für die Forschung

Die Ergebnisse der vorliegenden Untersuchung bieten neue Ansatzpunkte zur Diskussion der Chancen und Grenzen des Vier-Ebenen-Modells nach Kirkpatrick. Während sich das Modell aufgrund der einfachen Systematik gut als Strukturierungs- und Orientierungsgrundlage für die Untersuchung eignete, bot es wenig Anhaltspunkte für die präzise Erfassung von Erfolgskriterien auf den vier Ebenen.

Um die Forschung in Richtung möglicher Erfolgskriterien auf der Ebene Resultate für verhaltensorientierte Trainings voranzutreiben, wurden die Teilnehmer der Ausbildung um ihre Einschätzung möglicher Kriterien gebeten. Auch die Ergebnisse der vorliegenden Studie spiegeln die Schwierigkeit, speziell die *weichen Kriterien* des Erfolges für verhaltensorientierte Trainingsmaßnahmen in präzise Kriterien auf der vierten Ebene zu übersetzen (Solga, 2011a). Als mögliche Kennzahlen wurden vor allem *harte Faktoren*, wie die *Anzahl und Umsetzung neuer Ideen* im Rahmen eines Ideenmanagement-Systems genannt. Diese Kriterien eignen sich aber - wie bereits herausgestellt - aufgrund des Phänomens der Kriterienkontamination nur bedingt dazu, die Arbeit eines Innovationspromotors zu reflektieren. Zudem scheitert die Erhebung oftmals an datenschutzrechtlichen Hürden der Unternehmen oder aus Gründen des Aufwands.

Aufschlussreiche, neue Perspektiven auf mögliche Erfolgskriterien auf der vierten Ebene, eröffneten sich hingegen im Rahmen der Analyse der umgesetzten Maßnahmen zur Innovationspromotion. Für einen Großteil der hier genannten Maßnahmen - *Projektabschlüsse feiern, Mitarbeitergespräche führen, Austausch zwischen Abteilungen fördern, Einführung von Ideenräumen und neuer Software, Maßnahmen der Projektsteuerung, wie Kick-Off-Veranstaltungen u.a* - wäre die Überführung in präzise Kennzahlen denkbar, die als Erfolgskriterium auf der Ebene Resultate fungieren könnten. Dieses Vorgehen könnte sowohl wissenschaftlichen als auch praktikablen Anforderungen genügen. Es lässt sich schlussfolgern, dass sich die Suche nach praktikablen und validen Kennzahlen für die Ebene also stärker auf den Bereich der sogenannten *weichen Faktoren (Indikatoren des Arbeits- und Organisationsklimas und der Qualität der Leistung)* konzentrieren sollte, statt weiterhin hartnäckig aber erfolglos den vermeintlichen *Königsweg* der Erhebung *harter Faktoren* oder gar die Durchführung einer *Effizienzanalyse* anzusteuern. Damit ließe sich die Erhebungsquote für die Ebene - wie einhellig gefordert - deutlich steigern.

Auch im Rahmen der vorliegenden Untersuchungen sollen die Befunde überdies dazu genutzt werden, die Forschung zu den impliziten Annahmen des Modells voranzutreiben. Das *hierarchische Abhängigkeitsverhältnis* wird in der Diskussion des Modells zwar auch vereinzelt in Frage gestellt (z.B. Alliger & Janak, 1989), ein Blick auf den Diskurs zeigt aber, dass die durchaus berechtigte Kritik an der Ebene Reaktion dazu führt, dass ihr Nutzen aus dem Blickfeld gerät.

Im Rahmen der vorliegenden Untersuchung zeigte sich hingegen deutlich, dass die Ergebnisse auf der Ebene die entscheidenden Anhaltspunkte für die Weiterentwicklung der Maßnahme gaben. Gerade im Hinblick auf die Optimierungsfunktion einer Evaluation kommt der Ebene damit eine wichtige Funktion im Gesamtmodell zu. Aussagen zur Wirksamkeit einer Maßnahme lassen sich hingegen - wie wiederholt postuliert - nicht allein auf der Grundlage der Ergebnisse auf der Ebene treffen.

Die wohl umstrittenste Annahme des Modells ist - wie mehrfach herausgestellt - die kausale und lineare Abhängigkeit der vier Ebenen. Auch im Rahmen der vorliegenden Untersuchung wurde diese erneut überprüft. Es konnten keine empirischen Belege für die von Kirkpatrick postulierte Abhängigkeit der Ebenen Zufriedenheit, Lernen und Verhalten gefunden werden.<sup>32</sup> Dieses Ergebnis reiht sich in die Befunde der bereits diskutierten Forschungsarbeiten ein (Alliger & Janak, 1989; Alliger u.a., 1997; Cloquitt u.a., 2000). Die Unabhängigkeit der vier Ebenen sollte daher in der Weiterentwicklung von Evaluationsmodellen stärker beachtet werden.

### **4.3 Implikationen für die Trainingspraxis**

Eine systematische Evaluation der Wirksamkeit eines Trainings, die wissenschaftlichen Anforderungen genügt, bedeutet immer einen hohen Aufwand. Dennoch konnten im Rahmen der vorliegenden Untersuchung Lösungen aufgezeigt werden, die gleichzeitig praktikablen Ansprüchen der Trainingspraxis genügen. Nachfolgend werden erste konkrete Empfehlungen für die Gestaltung von Evaluationsvorhaben in der Trainingspraxis abgeleitet.

---

<sup>32</sup> Lediglich zwischen dem prozeduralen Wissen zu t1 und den Einschätzungen des gegenwärtigen Verhaltens durch die Kollegen zeigt sich ein signifikanter positiver Zusammenhang ( $r = .763$ ;  $p = .46$ ). Dieses Ergebnis lässt sich aber nicht sinnvoll als Zusammenhang zwischen den Ebenen Lernen und Verhalten interpretieren, da die Einschätzung des prozeduralen Wissens durch die Teilnehmer gleichzeitig die Selbsteinschätzung des Verhaltens repräsentiert. Im Rahmen der vorliegenden Untersuchung war es nicht möglich Zusammenhänge zu der Ebene Resultate zu überprüfen. Die ausführliche Darstellung der Überprüfung der Zusammenhänge findet sich im Anhang (Appendix C6, S. 154 ff.).



### 4.3.1 Evaluationsmodell

Um die Wirksamkeit einer Trainingsmaßnahme differenziert zu erfassen, empfiehlt sich die Entwicklung von Evaluationskriterien entlang der Wirkungskette des Evaluationsmodells nach Kirkpatrick und Kirkpatrick (2006). So bietet das Modell vor allem eine gute Möglichkeit zur Systematisierung und Strukturierung eines Evaluationsvorhabens. Ergänzend zu den ergebnisbezogenen Variablen sollte erhoben werden, wie die Bedingungen des Transfers für die Teilnehmer ausgeprägt sind. Wie gezeigt werden konnte lassen sich Transferbedingungen mit Hilfe des Lerntransfer-System-Inventars (GLTSI, Kauffeld, 2008) praktikabel und valide erheben. Vor allem im Fall bestehender Transferprobleme bietet das Verfahren eine gute Möglichkeit zur Identifikation von Barrieren und Stellschrauben des Transfers. Allerdings sollte darauf geachtet werden, die Anzahl der Erhebungen der Transferbedingungen analog zu der Anzahl der Erhebungen auf der Lern- und der Verhaltensebene zu gestalten. Nur so lassen sich die Ergebnisse in den Analysen auch systematisch in Verbindung bringen.

### 4.3.2 Erhebungsmethoden

#### Reaktion

Auf der Ebene Reaktion wird der Einsatz eines standardisierten Fragebogens empfohlen, der eine Differenzierung der Zufriedenheit in weitere Dimensionen (*Trainer, Inhalt, Didaktik*) vorsieht. Dabei sollte - je nach Erkenntnisinteresse - auf ein gleichgewichtiges Verhältnis der drei Dimensionen geachtet werden. Das Verfahren sollte zudem die Möglichkeit bieten, freie Kommentare zu positiven und negativen Aspekten abzugeben. Nur so wird die Feedbackfunktion des Formates optimal genutzt und es können gezielt Anhaltspunkte zur Modifikation der Trainingsmaßnahme gewonnen werden. Es gilt zu beachten, dass die Ergebnisse auf der Ebene für sich genommen keine Schlussfolgerungen über die Trainingswirksamkeit zulassen und daher immer im Kontext des Gesamtmodells betrachtet werden sollten.

#### Lernen

Auf der Lernebene empfiehlt sich ebenfalls der Einsatz eines standardisierten Fragebogens. Ein praktikabler Weg zur Entwicklung des Verfahrens besteht darin, die Fragebogenitems aus den Lernzielen der Ausbildung abzuleiten. Gerade für verhaltensorientierte Trainingsformate kann es so gelingen, die Breite der Lern- und Entwicklungsfelder adäquat mit einem einzelnen Instrument zu erfassen. Um wissen-

schaftlichen Anforderungen zu genügen, sollte darauf geachtet werden, homogene Skalen zu bilden. Stehen also verschiedene Lernziele je Modul im Fokus, sollten diese jeweils durch mehrere Items repräsentiert werden. Nur so lässt sich die Güte des Verfahrens sinnvoll bestimmen. Es bietet sich weiter eine Unterteilung in deklaratives und prozedurales Wissen an, um den Erwerb von Fachwissen und Handlungswissen differenziert abzubilden. Durch die Erfassung des Handlungswissens werden weitergehende Analysen mit den Ergebnissen der Verhaltensebene möglich.

### Verhalten

Um valide Ergebnisse auf der Verhaltensebene zu generieren, sollte bei der Erhebung von Erfolgskriterien auf Fremdurteile gesetzt werden. Als geeignete Personengruppe aus dem Funktionsfeld können die Kollegen der Teilnehmer benannt werden. So ist davon auszugehen, dass diese Veränderungen im Arbeitsverhalten der Teilnehmer in der täglichen Arbeit erfahren und implizit beobachten. Gleichzeitig ist bei dieser Personengruppe von einer vergleichsweise guten Erreichbarkeit und hohen Bereitschaft zur Teilnahme an der Befragung auszugehen. Um den Aufwand der Erhebung im Rahmen zu halten bietet sich auch hierfür der Einsatz eines standardisierten Fragebogens an. Ein praktikables Vorgehen zur Entwicklung des Verfahrens ist die Orientierung an den entwickelten Items des Lernfragebogens. Auf dieser Grundlage lassen sich konkrete Verhaltensaussagen formulieren, die sich standardisiert erfassen lassen. Damit wird es überdies leichter möglich, die Ergebnisse der Lern- und Verhaltensebene in den Analysen der Daten aufeinander zu beziehen.

### Resultate

Zur Erhebung von Erfolgskriterien auf der Ebene Resultate bietet es sich an, sogenannte *weiche Kriterien* des Erfolges, wie Indikatoren des Arbeits- und Organisationsklimas oder der Qualität der Leistung, zu quantifizieren. Die konkreten Kennzahlen müssen unter Berücksichtigung der Zielstellung des Trainings und der vorherrschenden Bedingungen in den Unternehmen abgeleitet werden.

### **4.3.3 Evaluationsdesign**

Da es im Rahmen von Trainingsevaluationen selten sinnvoll möglich ist eine Kontrollgruppe zur Trainingsgruppe zu bilden, sollte der Fokus bei der Konstruktion des Evaluationsdesigns auf die Umsetzung mehrerer Messzeitpunkte gelegt werden. Bei Trainings, die mehrere Module umfassen sollte die Einschätzung der Zufriedenheit

separat nach jedem Modul erfolgen, um zu differenzierten Ergebnissen zu gelangen. Auf den Ebenen Lernen und Verhalten sollte mindestens eine Prä- und Postmessung erfolgen. Die Umsetzung eines dritten Messzeitpunktes (bspw. eine Zwischenerhebung) ist sehr zu empfehlen, da sich nur so differenzierte Ergebnisse über Wirkungsverläufe und die Stabilität der Effekte generieren lassen.

#### **4.4 Methodische Einschränkungen der Studie**

Wenngleich eine umfassende Modellprüfung aufgrund der Rahmenbedingungen der Untersuchung nicht vorgenommen werden konnte, hätten durchaus weitere Indikatoren als Anhaltspunkte für die Eignung des Ansatzes hinsichtlich der Wissenschaftlichkeit und Praktikabilität herangezogen werden können. Die Rücklaufquote als alleiniges Kriterium zur Beurteilung der Praktikabilität des Vorgehens war nur begrenzt aussagekräftig. Es ist daher ratsam künftig weitere Indikatoren zur Erfassung dieses wichtigen Kriteriums einzubeziehen. Denkbar wäre hier bspw. die ergänzende qualitative Befragung der Teilnehmer oder die Generierung weiterer objektiver Kriterien, wie Aufwand, personelle und zeitliche Ressourcen u.a. Eine weitere interessante Quelle, die als Indikatoren für beide Anspruchsgruppen herangezogen werden könnte, sind die *fehlenden Werte*. So kann die systematische Analyse fehlender Werte prinzipiell sowohl Aufschluss über die Güte der Verfahren als auch über die Praktikabilität geben (Seidel, 2012). So können wenige und eine eher zufällige Verteilung der Werte über die Items hinweg als Indiz für eine hohe Reliabilität der Instrumente gewertet werden (Schendera, 2007). Zudem sprechen wenige fehlende Werte auch für die Praktikabilität und Akzeptanz einer Erhebung.

#### **4.5 Fazit und Ausblick**

Das Ziel der vorliegenden Arbeit bestand darin, einen Evaluationsansatz für die Trainingspraxis aufzuzeigen, der eine ausgewogene Balance zwischen wissenschaftlichem Anspruch und praktikablen Anforderungen im Feld herstellt. Die Bearbeitung der Forschungsfrage erfolgte auf zwei Ebenen: In einem ersten Schritt wurde auf konzeptioneller Ebene entlang der drei zentralen Bausteine einer Evaluation - *Evaluationsmodell, Evaluationsdesign und Erhebungsmethoden* - ein Ansatz entwickelt, der beiden Anspruchsgruppen gerecht werden sollte. In einem zweiten Schritt wurde die Eignung des Ansatzes empirisch geprüft. Dazu wurden die konzeptionellen Überlegungen in Bezug auf die Ausbildung zum Innovationspromotor als Gegenstand der exemplarischen Trainingsevaluation konkretisiert. Es konnten wichtige Erkenntnisse

über die Chancen und Grenzen des entwickelten Ansatzes gewonnen und Ansatzpunkte zur Weiterentwicklung identifiziert werden.

So ist es auf allen Erhebungsebenen gelungen, differenzierte Erkenntnisse über die Wirkungen der Ausbildung abzuleiten. Das Anliegen, dabei sowohl wissenschaftlichen als auch praktikablen Anforderungen zu genügen, wurde teilweise erfüllt. So waren der kausalen Interpretation im Rahmen des messwiederholenden Untersuchungsdesigns ohne Kontrollgruppe Grenzen gesetzt. Ebenso musste für die Erhebungsinstrumente auf allen Ebenen Weiterentwicklungsbedarf konstatiert werden. Die Praktikabilität des Vorgehens wurde überwiegend für gut befunden. Lediglich für die Verhaltensebene sind künftig weitere Schritte erforderlich, um die Praktikabilität des Vorgehens zu erhöhen und den Rücklauf der Daten zu steigern.

Dennoch ist es im Rahmen der vorliegenden Arbeit gelungen, aufzuzeigen, wie ein ganzheitlicher Evaluationsansatz aussehen kann, der eine ausgewogene Balance zwischen wissenschaftlichem Anspruch und praktikablen Anforderungen im Feld herstellt. Auf der Grundlage der Befunde konnten erste Empfehlungen für die Trainingspraxis abgeleitet werden, die ein praktikables und wissenschaftlich anspruchsvolles Vorgehen im Rahmen von Evaluationen verbinden. Ein allgemeingültiger Ansatz lässt sich dennoch nur schwer formulieren, da Trainingsevaluationen immer an die spezifischen Bedingungen der jeweiligen Maßnahme angepasst werden müssen. Damit sind auch der Validierung des Ansatzes in weiteren Studien Grenzen gesetzt. Dennoch liegt genau darin der weiterführende Forschungsbedarf. Es gilt den Nutzen und Erfolg des Ansatzes anhand der Evaluation verschiedener Trainingsmaßnahmen systematisch zu prüfen. In Anbetracht der hohen Relevanz des Themas für die Unternehmens- und Trainingspraxis wäre es wünschenswert, wenn auch die Ergebnisse der Trainings- und Transferforschung stärker dazu genutzt würden, valide Evaluationskonzepte für die Trainingspraxis zu entwickeln und Wege aufzuzeigen, wie sich diese praktikabel umsetzen lassen.

## Literatur

- Alliger, G. M. & Janak E. A. (1989). Kirkpatrick`s levels of training kriteria: thirty years later. *Personnel Psychology*, 42, 331-342.
- Alliger, G. M., Tannenbaum, S. I., Bennet, W., Traver, H. & Shotland, A. (1997). A meta-analysis oft the relations among training criteria. *Personnel Psychology*, 50, 341-358.
- Alvarez, K., Salas, E. & Garofano, C. M. (2004). An Integrated Model of Training Evaluation and Effectiveness. *Human Resource Development Review*, 3(4), pp. 385-416.
- Arthur, W., Bennett, W., 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, 234-245.
- artop-Webside (2012). Verfügbar unter [http://artop.de/3000\\_Ausbildungen/3000\\_Ausbildungen.html](http://artop.de/3000_Ausbildungen/3000_Ausbildungen.html) [zuletzt abgerufen am 24.07.12].
- Bachmann, T., Runkel, R. & Scholl, W. (2010). Ausbildung von Beraterrollen. In U. P. Kanning, L v. Rosenstiel, & H. Schuler (Hrsg.), *Jenseits des Elfenbeinturms: Psychologie als nützliche Wissenschaft*. Göttingen: Vandenhoeck & Ruprecht.
- Baldwin, T. T. & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41(1), 63–105.
- Balzer, L. (2005). *Wie werden Evaluationsprojekte erfolgreich? – Ein integrierender theoretischer Ansatz und eine empirische Studie zum Evaluationsprozess*. Landau: Empirische Pädagogik.
- Bandura, A. (1979). *Sozial-kognitive Lerntheorie*. Stuttgart: Klett-Cotta.
- Bates, R. A., Kauffeld, S. & Holton, E. F. (2007). Examining the factor structure and predictive ability of the german-version of the learning transfer systems inventory (LTSI). *Journal of European Industrial Training*, 31, 195-211.

- Bergmann, B. (2007). Selbstkonzept beruflicher Kompetenz. In J. Erpenbeck & L. v. Rosenstiel (Hrsg.), *Handbuch Kompetenzmessung* (S. 194-223). Stuttgart: Schäffer-Poeschel.
- Bergmann, B. (1999). *Training für den Arbeitsprozess: Entwicklung und Evaluation aufgaben- und zielgruppenspezifischer Trainingsprogramme*. Zürich: Hochschulverlag.
- Bergmann, B. & Sonntag, K. (2006). Transfer: Die Umsetzung und Generalisierung erworbener Kompetenzen in den Arbeitsalltag. In K. Sonntag (Hrsg.), *Personalentwicklung in Organisationen* (3. überarb. und erw. Aufl.) (S. 355-388). Göttingen: Hogrefe.
- Bethscheider, M., Höhns., G. & Münchhausen, G. (2010): *Kompetenzorientierung in der beruflichen Bildung*. Bielefeld: Bertelsmann.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. überarb. Aufl.). Heidelberg: Springer.
- Broad, M. L. & Newstrom, J. W. (1992). *Transfer of training. Action-packed strategies to ensure high payoff from training investments*. Reading, MA: Addison-Wesley.
- Buren, M. v. & Erskine, W. (2002). *ASTD State of the Industry Report*. Washington DC: ASTD.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. aktual. und erw. Aufl.). München: Pearson Studium.
- Campbell, D. T. (1971). Personal training and development. *Annual Review of Psychology*, 22, 565-602.
- 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, 141-164.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Colquitt, J. A., LePine, J. A. & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678-707.

- DeGEval – Gesellschaft für Evaluation e.V. (2008). *Standards für Evaluation* (4. unveränd. Aufl.). Mainz: DeGEval.
- Demmerle, C., Schmidt, J. M., Hess, M., Solga, M. & Ryschka, J. (2011). In J. Ryschka, M. Solga & A. Mattenklott (Hrsg.), *Praxishandbuch Personalentwicklung: Instrumente, Konzepte, Beispiele* (3. Aufl.) (S. 273-325). Wiesbaden: Gabler.
- DIE WELT (2004). Berufliche Weiterbildung in Deutschland krankt an der Umsetzung. Verfügbar unter <http://www.welt.de/print-welt/article342472/BeruflicheWeiterbildunginDeutschlandkranktanderUmsetzung.html> [zuletzt abgerufen am 20.07.12] .
- Edelmann, D. & Tippelt, R. (2004). Kompetenz – Kompetenzmessung: ein (kritischer) Überblick. *Durchblick*, 3, 7-10.
- Erpenbeck, J. & Rosenstiel, L. v. (2007). *Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (2. überarb. u. erw. Aufl.). Stuttgart: Schäffer-Poeschel.
- Field, A. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)* (3rd ed.). London: SAGE.
- Frankfurter Allgemeine Zeitung (2009). Wirkung von Weiterbildung: Erlebnis ohne Ergebnis. Verfügbar unter <http://www.faz.net/aktuell/wirtschaft/wirkung-von-weiterbildung-erlebnis-ohne-ergebnis-1923079.html> [zuletzt abgerufen am 20.07.12].
- Frankfurter Allgemeine Zeitung (2008). *Der große Trainingsschwindel*, 198, 37.
- Gemünden, H. G., Salomo, S. & Hölzle, K. (2007): Role Models for Radical Innovations in Time of Open Innovation. *Creativity and innovation management*, 16(4), 408–421.
- GI:VE-Website (2010). Verfügbar unter <http://www.vertrauenkultur-innovation.de/> [zuletzt abgerufen am 20.07.12].
- Goldstein, I. L. & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth.
- Gollwitzer, M. & Jäger, R. S. (2009): *Evaluation kompakt*. Weinheim: Beltz.

- Graf, J. (2010). *Weiterbildungsszene Deutschland 2010: Studie über den deutschen Weiterbildungsmarkt*. Bonn: managerSeminare.
- Gris, R. (2008). *Die Weiterbildungslüge: Warum Seminare und Trainings Kapital vernichten und Karrieren knicken*. Frankfurt: Campus.
- 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. 19-40). Bern: Huber.
- Hanke, U. (2005). *Evaluieren von Weiterbildung: Theoretische Grundlagen und praktische Vorschläge*. Hamburg: Kovač.
- Hauschildt, J. & Chakrabarti, A. K. (1988): Arbeitsteilung im Innovationsmanagement: Forschungsergebnisse, Kriterien und Modelle. *Zeitschrift Führung + Organisation*, 57, 378–388.
- Hauschildt, J. & Gemünden, H. G. (1999). *Promotoren - Champions der Innovation* (2. erw. Aufl.). Wiesbaden: Gabler.
- Häder, M. (2006). *Empirische Sozialforschung: Eine Einführung*. Wiesbaden: VS Verlag für Sozialforschung.
- Häring, K. (2003). *Evaluation der Weiterbildung von Führungskräften*. Wiesbaden: Deutscher Universitäts Verlag.
- Heyse, V. & Erpenbeck, J. (2004). *Kompetenztraining*. Stuttgart: Schäfer-Poeschel.
- Hochholdinger, S., Rowold, J. & Schaper, N. (2008a). Praxis- und Forschungsrelevanz von Trainings. In J. Rowold, S. Hochholdinger & N. Schaper (Hrsg.), *Evaluation und Transfersicherung betrieblicher Trainings: Modelle, Methoden und Befunde* (S. 13-27). Göttingen: Hogrefe.
- Hochholdinger, S., Rowold, J. & Schaper, N. (2008b). Ansätze zur Trainings- und Transferevaluation. In J. Rowold, S. Hochholdinger & N. Schaper (Hrsg.), *Evaluation und Transfersicherung betrieblicher Trainings. Modelle, Methoden und Befunde* (S. 30-49). Göttingen: Hogrefe.
- Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7, 5-21.



- Holton, E. F., Bates, R. A. & Ruona, W. E. A. (2000). Development of a generalized learning transfer system inventory. *Human Resources Development Quarterly*, 11, 333-360.
- Höft, S. (2006). Erfolgsüberprüfung personalpsychologischer Arbeit. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 761-796). Göttingen: Hogrefe.
- Hüttner, J. & Scholl, W. (2011). Interne Innovationspromotoren als Garanten nachhaltiger Innovationsfähigkeit. In Jeschke, S. (Hrsg.), *Innovation im Dienste der Gesellschaft – Beiträge des 3. Zukunftsforums Innovationsfähigkeit des BMBF* (S. 73-85). Frankfurt: Campus.
- Karg, U. (2006). *Betriebliche Weiterbildung und Lerntransfer: Einflussfaktoren auf den Lerntransfer im organisationalen Kontext*. Bielefeld: Bertelsmann.
- Kauffeld, S. (2010). *Nachhaltige Weiterbildung: Betriebliche Seminare und Trainings entwickeln, Erfolge messen, Transfer sichern*. Heidelberg: Springer.
- Kauffeld, S. (2006). *Kompetenzen messen, bewerten, entwickeln: ein prozessanalytischer Ansatz für Gruppen*. Stuttgart: Schäffer-Poeschel.
- Kauffeld, S., Bates, R., Holton, E. F. & Müller, C. (2008). Das deutsche Lerntransfer-System-Inventar (GLTSI): Psychometrische Überprüfung der deutschsprachigen Version. *Zeitschrift für Personalpsychologie*, 7(2), 50-69.
- Kirkpatrick, D. L. (1967). Evaluation of training. In R. L. Craig (Eds.), *Training and development handbook: A guide to human resources development* (pp. 18.1–18.27). New York, NY: McGraw-Hill.
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13, 3-26.
- Kirkpatrick, D. L. & Kirkpatrick, J. D. (2006). *Evaluating Training Programs: The Four Levels* (3rd ed.). San Francisco: Berrett-Koehler Publishers.
- Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Eds.), *Creating, implementing, and managing effective training and development. State-of-the-art lessons for practice* (pp. 331-375). San Francisco, CA: Jossey-Bass.
- Kuper, H. (in Druck). *Ansätze des Lerntransfers und Methoden der Evaluation*. Deutsche Universität für Weiterbildung (DUW).

- Lang, A. & Otto, B. (2010). Erreichen vorbereitete Verhandlungsführer mehr? Überprüfung eines Trainings zur Planung von Verhandlungen. *Zeitschrift für Arbeits- u. Organisationspsychologie*, 54(2), 92-102.
- Lenske, W. & Werner, D. (2009). Umfang, Kosten und Trends der betrieblichen Weiterbildung: Ergebnisse der IW-Weiterbildungserhebung 2008 [Elektronische Version]. *IW-Trends – Vierteljahresschrift zur empirischen Wirtschaftsforschung aus dem Institut der deutschen Wirtschaft Köln*, 36(1).
- Malmendier, M. (2003). *Kommunikations- und Verhaltenstrainings in Organisationen: Zur Interventionspraxis in der Personalentwicklung*. Frankfurt: Campus Verlag.
- Mansfeld, M. (2011). *Individuen im Innovationsmanagement*. Gabler Verlag.
- Meier, N. (2007). *Verhaltenstrainings in der Personalentwicklung: Erfolgskriterien und Transfersicherung: Eine Praxisstudie*. München: Hampp.
- Merzenich-Hieker, C. (1996). *Evaluation von Kommunikations- und Verhaltenstrainings in Organisationen: Eine empirische Studie*. Aachen: Shaker.
- Nork, M. E. (1991). *Management Training. Evaluation – Probleme – Lösungsansätze*. München: Hampp.
- Patry, J.-L. & Perrez, M. (2000). Theorie-Praxis-Probleme und die Evaluation von Interventionsmaßnahmen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien* (S. 19-40). Bern: Huber.
- Phillips, J. J. & Schirmer, F. C. (2005). *Return on Investment in der Personalentwicklung: Der 5-Stufen-Evaluationsprozess* (2. aktual. und erw. Aufl.). Heidelberg: Springer.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2006a). *SPSS-Ergänzungen. Quantitative Methoden: Einführung in die Statistik* (Bd. 2). Heidelberg: Springer.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2006b). *Quantitative Methoden: Einführung in die Statistik* (Bd. 2). Heidelberg: Springer.

- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2006a). *Quantitative Methoden: Einführung in die Statistik* (Bd. 1). Heidelberg: Springer.
- Rossi, P. H., Freeman, H. E. & Hofmann, G. (1988). *Programm-Evaluation: Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Rowold, J. (2008). *Evaluation eines unternehmensweiten Trainings zur Kundenorientierung in Call Centern auf den vier Ebenen nach Kirkpatrick*. In J. Rowold, S. Hochholdinger & N. Schaper (Hrsg.), *Evaluation und Transfersicherung betrieblicher Trainings. Modelle, Methoden und Befunde* (S. 90-99). Göttingen: Hogrefe.
- Rudinger, G. (2012). *Can one purposefully encourage innovation? An empirical evaluation of a pilot training as innovation promotor*. Nicht veröffentlichte Studienabschlussarbeit, Humboldt-Universität zu Berlin, Berlin.
- Russ-Eft, D., Preskill, H. (2001). *Evaluation in Organizations: A systematic approach to enhancing learning, performance and change*. Cambridge, Massachusetts: Perseus Publishing.
- Saborowski, Y. & Muellerbuchhof, R. (2010). Selbstmanagement-Training als Methode der Kompetenzentwicklung bei Berufseinsteigern - am Beispiel von Auszubildenden technischer Fachrichtungen. *Zeitschrift für Arbeits- u. Organisationspsychologie*, 54(2), 83-91.
- Salas, E., Cannon-Bowers, J. A. & Kozlowski, S. W. J. (1997). The science and practise 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. 357-368). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schaper, N. & Lieberei, W. (2004). *Führungsstilfragebogen FSF*. Göttingen: Hogrefe
- Scharper, N. & Sonntag, K. (2007). Weiterbildungsverhalten. In D. Frey & L. v. Rosenstiel (Hrsg.), *Markt- und Wirtschaftspsychologie: Enzyklopädie der Psychologie* (Themenbereich D, Serie 3, Bd. 6) (S. 573-648). Göttingen: Hogrefe.

- Scharpf, Robert (1999). *Training und Transfer: Lernen, Anwenden und die Bedeutung fähigkeitsbezogener Kognitionen*. München: Hampp.
- Schendera (2007). *Datenqualität mit SPSS*. München: Oldenbourg.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA: Sage.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Scholl, W. (2010). Innovationen – Wie Unternehmen neues Wissen produzieren und etablieren. In H. Hof & U. Wengenroth (Hrsg.), *Innovationsforschung – Ansätze, Methoden, Grenzen und Perspektiven* (2. Aufl.) (S. 271-300). Münster: LIT.
- Scholl, W. (2004). *Innovation und Information: Wie in Unternehmen neues Wissen produziert wird*. Göttingen: Hogrefe.
- Schuler, H. (1998). *Lehrbuch Organisationspsychologie*. Bern: Hans Huber.
- Seidel, J. (2007): *Transferkompetenz und Transfer: Theoretische und empirische Untersuchung zu den Wirksamkeitsbedingungen betrieblicher Weiterbildung* (Bildung, Arbeit, Beruf und Beratung, Bd. 3). Landau: Verlag für Empirische Pädagogik.
- Solga, M. (2011a). Evaluation von Personalentwicklung. In J. Ryschka, M. Solga & A. Mattenklott (Hrsg.), *Praxishandbuch Personalentwicklung: Instrumente, Konzepte, Beispiele* (3. Aufl.) (S. 369-394). Wiesbaden: Gabler.
- Solga, M. (2011b). Förderung von Lerntransfer. In J. Ryschka, M. Solga & A. Mattenklott (Hrsg.), *Praxishandbuch Personalentwicklung: Instrumente, Konzepte, Beispiele* (3. überarb. und erw. Aufl.) (S. 339-363). Wiesbaden: Gabler.
- Stockmann, Reinhard (2007). *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Münster: Waxmann.
- Suchmann, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York: Russell Sage Foundation.

- Taylor, P. J., Russ-Eft, D. F. & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology*, 90, 692-709.
- Taylor, P. J., Russ-Eft, D. F. & Taylor, H. (2009). Transfer of management training from alternative perspectives. *Journal of Applied Psychology*, 94, 104-121.
- Thierau-Brunner, H., Wottowa, H. & Stangel-Meseke, M. (2006). Evaluation von Personalentwicklungsmaßnahmen. In K. Sonntag (Hrsg.), *Personalentwicklung in Organisationen* (3. überarb. und erw. Auflage) (S. 329-351). Göttingen: Hogrefe.
- Tränkle, U. (1983). Fragebogenkonstruktion. In H. Feger & J. Bredenkamp (Hrsg.), *Datenerhebung. Enzyklopädie der Psychologie* (Bd. B,I,2) (S.222-301). Göttingen: Hogrefe.
- Twitchell, S., Holton, E. F. & Trott, J. R. (2001). Technical training evaluation practices in the United States. *Performance Improvement Quarterly*, 13 (3), 84-109.
- Weisweiler, S (2008). *Der Einfluss von Individuum, Trainingsmaßnahme und Umfeld - eine Untersuchung zum Transferprozess in der Weiterbildung*. Inaugural-Dissertation zur Erlangung der Doktorwürde der Philosophischen Fakultät II (Psychologie, Pädagogik und Sportwissenschaft) der Universität Regensburg.
- Wild, K.-P. & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15, 185–200.
- Willett, J. B. (1980). Some Results on Reliability for the Longitudinal Measurement of Change: Implications for the Design of Studies of Individual Growth. *Educational and Psychological Measurement*, 49, 587-602.
- Witte, Eberhard (1973): *Organisation für Innovationsentscheidungen: Das Promotoren-Modell*. Göttingen: Schwartz & Co.
- Wottowa, H. & Thierau, H. (2003). *Lehrbuch Evaluation* (3. Aufl.). Bern: Huber.

## **Eidesstattliche Versicherung**

Hiermit versichere ich, dass ich die vorliegende Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, alle Ausführungen, die anderen Schriften wörtlich oder sinngemäß entnommen wurden, kenntlich gemacht sind und die Arbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung war.

Berlin, den 26. Juli 2012

Sandra Tirre

## Anhang

<b>Appendix A: Zusammensetzung der Stichprobe</b>	112
<b>Appendix B: Erhebungsinstrumente</b>	113
Appendix B1: Erhebungsinstrument der Ebene Reaktion, Printversion	113
Appendix B2: Erhebungsinstrument der Ebene Lernen, Abbildung der Online-Version	115
Appendix B3: Erhebungsinstrument der Ebene Verhalten	122
Appendix B3.1 Abbildung der Online-Version	122
Appendix B3.2 Übersicht über die Skalen und Items des Verfahrens und die Ausgangsitems des FB Lernen	126
Appendix B4: Verwendete Version des deutschen Lerntransfer-System-Inventars (GLTSI), Online-Version	128
<b>Appendix C: Ergebnisse der Berechnungen</b>	131
Appendix C1: Ergebnisse der Überprüfung der Normalverteilungsannahme	131
Appendix C2: Analysen auf der Ebene Reaktion	137
Appendix C3: Analysen auf der Ebene Lernen	143
Appendix C4: Analysen auf der Ebene Verhalten	148
Appendix C5: Analysen der Transferbedingungen	150
Appendix C6: Gesamtmodell	154

## Appendix A: Zusammensetzung der Stichprobe

**Tabelle 1: Deskriptive Analyse des Teilnehmerkreises (in Anlehnung an Rudinger, 2012)**

		Absolut	in Prozent
<b>Höchster Bildungsabschluss</b>			
	Postdoc	1	6.7%
	Hochschulabschluss	10	66.7%
	Ausbildung	4	26.7%
<b>Gegenwärtige Position / Funktion</b>			
	Geschäftsführer	3	20%
	Leitende Positionen (z.B. Projektleiter, Teamleiter, Abteilungsleiter)	7	46.7%
	keine leitende Funktion	5	33.3%
<b>Unternehmensgröße</b>			
	Klein (< 10 MA)	2	13.3%
	Mittelständisch (< 250 MA)	2	33.3%
	Groß (< 500 MA MA)	8	53.3%
<b>Branche</b>			
	Gastronomie / Catering	2	
	Baugewerbe	1	
	Wirtschaftsdienst	2	
	Energieversorgung /-wirtschaft	1	
	Gesundheitspflege	1	
	Wohnbau	1	
	IT-Technik	1	
	Fertigungsindustrie	1	
	Mess- und Verfahrenstechnik	1	
	Reststoffverwertung	2	
	Telekommunikation	1	
	Handel	1	



## Appendix B: Erhebungsinstrumente

### Appendix B1: Erhebungsinstrument der Ebene Reaktion, Printversion

Beim ersten Teil dieses Fragebogens geht es darum, wie <u>Sie</u> die Modul Inhalte einschätzen.	Stimmt ganz und gar nicht				Stimmt voll und ganz
	1	2	3	4	5
1. Meine Erwartungen an die vermittelten Inhalte wurden erfüllt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Das Modul hat einen guten Einblick in das Themengebiet ermöglicht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Der Stoff war so interessant, dass ich eine weitere Vertiefung für notwendig halte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Ich habe viele Anregungen für eine spätere Tätigkeit als Innovationspromotor/in bekommen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Bei den nun folgenden Fragen geht es um Ihre Einschätzung des/ der Seminarleiters/in.	1: Stimmt ganz und gar nicht					5: Stimmt voll und ganz				
	Seminarleiter/in 1 Name: _____					Seminarleiter/in 2 Name: _____				
	1	2	3	4	5	1	2	3	4	5
1. Meine Erwartungen an den/die Seminarleiter/in wurden erfüllt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Bei diesem/r Seminarleiter/in würde ich gerne nochmals an einem Modul teilnehmen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Der/die Seminarleiter/in hat den Stoff anschaulich und verständlich mit uns erarbeitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Der/die Seminarleiter/in ging auf die Bedürfnisse der Teilnehmer ein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Fragen und/oder kritische Anmerkungen wurden von dem/der Seminarleiter/in stets aufgenommen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Der/die Seminarleiter/in hat die Diskussion gefördert und zu guten Ergebnissen geführt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Der/die Seminarleiter/in war bei allen Themen gut vorbereitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Der/die Seminarleiter/in hat mit verschiedenen Methoden die Inhalte sehr abwechslungsreich gestaltet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie erleben <u>Sie</u> die Darstellung des vermittelten Stoffes?	Stimmt ganz und gar nicht				Stimmt voll und ganz
	1	2	3	4	5
1. Die theoretischen Ausführungen waren informativ und verständlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Theorie und praktische Beispiele standen in einem guten Verhältnis.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Es wurde gezielt darauf hingearbeitet, das Gelernte auch in die Praxis zu übertragen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Man konnte dem Stoff des Moduls leicht folgen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Man konnte durch die Übungen richtig in die Thematik eindringen und sich mit ihr persönlich auseinandersetzen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hier bitten wir Sie um konkrete Rückmeldungen!

▫ Positive Aspekte (Besonders gefallen hat mir..., Besonders wichtig war für mich..., Besonders informativ war ....)

▫ Negative Aspekte (In diesem Modul fehlten mir ..., Zu kurz kam mir..., Besonders geärgert hat mich ....)

▫ Was haben Sie persönlich für sich gelernt – theoretisch und/oder praktisch?

Modul: .....

## Appendix B2: Erhebungsinstrument der Ebene Lernen, Abbildung der Online-Version

*Die nachfolgende Original-Instruktion der Abschlusserhebung (t3) steht exemplarisch für die Instruktionen zu den Erhebungen auf der Lernebene (t1 und t2).*

*In der Erhebung wurden neben dem in der Arbeit ausgewerteten deklarativen und prozeduralen Wissen die Erwartungen der Teilnehmer an die Ausbildung sowie ihr Bild eines idealen Innovationspromotors für weitere Fragestellungen im Projekt erfragt. Daher bezieht sich die nachfolgende Instruktion auf vier Teile, von denen aber nur die Items des deklarativen und prozeduralen Wissens aufgeführt werden (Teil I).*

*Da die Skalen des GLTSI ebenfalls im Rahmen der Abschlussbefragung erhoben wurden, findet sich nachfolgend auch schon die Instruktion für das Verfahren. Die verwendeten Skalen des GLTSI werden aber erst unter Appendix B4 berichtet.*

### **Einleitung**

Liebe Teilnehmerin, Lieber Teilnehmer,

wie angekündigt, findet nach dem 10. Modul der Ausbildung die Abschlusserhebung statt.

Der folgende Fragebogen besteht aus vier Teilen, von denen dir drei Teile bereits aus der Datenerhebung zu Beginn der Ausbildung und aus der Zwischenerhebung bekannt sind.

Im ersten Abschnitt geht es darum zu erfassen, zu welchen Themen du deine Kenntnisse erweiterst und zusätzlich praktische Erfahrungen im Berufsalltag gesammelt hast. Im zweiten Teil geht es um den idealen Innovationspromotor. Auch hier interessiert, wie deine Beurteilung nach Abschluss der Ausbildung ausfällt. Der dritte Teil greift die Erwartungen auf, die du bereits vor dem Ausbildungsbeginn zum ersten Mal eingeschätzt hast.

Im vierten Teil der Befragung soll es zusätzlich darum gehen zu erfahren, welche Rahmenbedingungen die Anwendung des Gelernten in deiner Arbeit und deinem Arbeitsumfeld beeinflussen. Wirst du bspw. durch Kollegen und Vorgesetzte in der Anwendung des Gelernten unterstützt? Verfügst du über die notwendigen Ressourcen zur Anwendung der Ausbildungsinhalte?

Um deine Anonymität zu wahren, bitte ich dich, wie gewohnt deinen persönlichen Code zu notieren. Gleichzeitig stellt die Befragungssoftware von Unipark sicher, dass ein Rückschluss auf deine IP-Adresse ausgeschlossen ist.

Die Beantwortung des Fragebogen wird ca. 25 Minuten in Anspruch nehmen. Bitte antworte spontan und verlasse dich auf dein Gefühl, wenn du dir einmal nicht so sicher bist. Es geht lediglich um deine ganz persönliche Einschätzung.

Schon einmal vorab vielen Dank für Ihre Mühe und viel Spaß bei der Teilnahme.

**Dein persönlicher Code setzt sich zusammen aus:**

- 1. den ersten zwei Buchstaben des Vornamens deiner Mutter**
- 2. dem Tag (zweistellig) deines Geburtsdatums**
- 3. dem zweiten & dritten Buchstaben deines Vornamens**

**Beispiel: Maria; 25.05.; Anna = MA25NN**

Bitte trage deinen persönlichen Code in dieses Kästchen ein:

Im ersten Teil geht es darum zu erfahren, inwiefern du Kenntnisse über bestimmte Themen besitzt und inwiefern du damit bereits Erfahrungen in deinem Praxisalltag gesammelt hast. Dazu wirst du Begriffe sehen, die sich alle aus den Inhalten der Ausbildungsmodule ergeben.

Diese Begriffe sind auf einer Skala von 1 – 5 einzuschätzen.

- 1 = trifft überhaupt nicht zu
- 2 = trifft ein wenig zu
- 3 = trifft teilweise zu
- 4 = trifft überwiegend zu
- 5 = trifft absolut zu

Zudem gilt es diese Einschätzung bezüglich zwei Aussagen abzugeben.

„Kenne ich.“ - Das ist die erste Aussage, die erscheint. Hier geht es um deine persönliche Einschätzung, inwiefern du mit diesem Thema vertraut bist, darüber etwas gelesen oder erfahren hast. Insbesondere ist hier der Bezug zum Thema Innovation wichtig. Inwieweit ist dir die Bedeutung des Themas für Innovationen bewußt.

„Setze ich um.“ - Das ist die zweite Aussage, die es einzuschätzen gilt. Hierbei geht es um die tatsächliche Anwendung im Berufsalltag. Dabei kann es sich um gesammelte Erfahrungen mit einer Methode, die aktive Umsetzung einer Technik oder den Transfer eines theoretischen Ansatzes in die Praxis handeln.

	Trifft überhaupt nicht zu	Trifft ein wenig zu	Trifft teilweise zu	Trifft überwiegend zu	Trifft absolut zu
<b>Lernen aus Fehlern.</b>					
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Nachhaltige Konfliktbewältigung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Strategische Personalentwicklung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Horizontale &amp; vertikale Kommunikation</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Aktive Unterstützung von Innovationsprozessen</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Innovationsförderliche Arbeitsplatzgestaltung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Trifft überhaupt nicht zu	Trifft ein wenig zu	Trifft teilweise zu	Trifft überwiegend zu	Trifft absolut zu
<b>Kompetenzentwicklung</b>					
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Vertrauensförderliche Maßnahmen</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Motivieren</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Nachhaltigkeit</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Mitarbeiterorientierung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Projektmanagement</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Trifft überhaupt nicht zu	Trifft ein wenig zu	Trifft teilweise zu	Trifft überwiegend zu	Trifft absolut zu
<b>Selbstreflexion</b>					
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Individueller Wissenserwerb</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Gruppensteuerung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Systematisches Monitoring</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Angemessener Umgang mit Widerständen</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Gesprächsführung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Trifft überhaupt nicht zu	Trifft ein wenig zu	Trifft teilweise zu	Trifft überwiegend zu	Trifft absolut zu
<b>Projektevaluation</b>					
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Konfliktphasen bei der Lösung beachten</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Lernen in und mit der Organisation</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Visualisierung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Moderation</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Interorganisationale Kooperation</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Trifft überhaupt nicht zu	Trifft ein wenig zu	Trifft teilweise zu	Trifft überwiegend zu	Trifft absolut zu
<b>Unternehmerisches Handeln</b>					
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Veränderungen initiieren</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Mitarbeiterbeteiligung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Konstruktiver Umgang mit Rückschlägen</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Selbstmanagement</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Zweckvolle Konfliktbehandlung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



	Trifft überhaupt nicht zu	Trifft ein wenig zu	Trifft teilweise zu	Trifft überwiegend zu	Trifft absolut zu
<b>Strategische Projektplanung</b>					
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Lernen im Team</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Systematisches Bewerten von Innovationen</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Analyse von Innovationsprozessen</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Ideenaustausch</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Aktive Innovationsförderung</b>					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Kenne ich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setze ich um.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Appendix B3: Erhebungsinstrument der Ebene Verhalten

### B3.1 Abbildung der Online-Version

#### Einleitung

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

wie angekündigt möchten wir Sie bitten, in dem nachfolgenden Fragebogen einzuschätzen, wie Sie die Umsetzung bestimmter Verhaltensweisen durch Ihre Kollegin/Ihren Kollegen im beruflichen Kontext erleben.

Es handelt sich dabei um die Themenbereiche und Lerninhalte der Ausbildung, die einen direkten oder indirekten Bezug zum Themenfeld Innovation haben. *Eine Innovation meint dabei eine relevante Änderung in Ihrem Unternehmen. Diese kann gestützt sein auf neue Erkenntnisse, soziale Verhaltensweisen, Materialien und Maschinen und hat eine direkte und/oder indirekte Verbesserung zum Ziel.*

Es geht um Ihre Beobachtungen. Erinnern Sie sich an gemeinsame Projekte und Arbeitsfelder. Wie haben Sie Ihre Kollegin/Ihren Kollegen in den verschiedenen Bereichen erlebt?

**Die Umsetzung der Verhaltensweisen ist auf einer Skala von 1-5 einzuschätzen.**

**1 = trifft überhaupt nicht zu**

**2 = trifft ein wenig zu**

**3 = trifft teilweise zu**

**4 = trifft überwiegend zu**

**5 = trifft absolut zu**

Je Bewertungsaspekt fragen wir Sie nach Ihrer Einschätzung des Verhaltens Ihrer Kollegin/Ihres Kollegen **vor einem Jahr** und zum **gegenwärtigen Zeitpunkt**.

Bitte antworten Sie spontan und verlassen Sie sich auf Ihr Gefühl, wenn Sie sich einmal nicht so sicher sind. Es geht lediglich um Ihre ganz persönliche Einschätzung.

Ihre Antworten bleiben wie bereits erwähnt anonym. Dazu möchten wir Sie wie angekündigt bitten, Ihren persönlichen Code aus der E-Mail in das Feld auf der nächsten Seite einzutragen.

Schon einmal vorab vielen Dank für Ihre Mühe und viel Spaß bei der Teilnahme.

Bitte tragen Sie den Code aus der E-Mail in dieses Kästchen ein:

Ihr Kollege/Ihre Kollegin:

**Ermuntert andere aktiv, ihre Ideen zu äußern.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Unterstützt die Auswahl von Ideen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Setzt sich für eine erfolgreiche Umsetzung neuer Ideen ein.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Organisiert die eigene Arbeitsweise systematisch (bspw. mit Hilfe von Zielsetzungs-, Durchführungs-, Entscheidungs-, Zeitplantechniken).**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Bemüht sich stets, die eigenen Standpunkte und Handlungen kritisch zu hinterfragen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Unterstützt andere aktiv dabei, sich neues Wissen anzueignen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Unterstützt Teammitglieder dabei, gemeinsam zu Lernen (bspw. durch den Austausch von Bewertungen und Sichtweisen, durch Abstimmungsprozesse).**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Bemüht sich darum, eine förderliche Lernkultur in der Organisation voranzutreiben.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Unterstützt aktiv den Wissensaustausch und die Kooperation mit anderen Unternehmen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Regt die Mitarbeiter stets zu einem offenen Austausch ihrer Ideen an.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Erfasst systematisch innovationsförderliche und- hemmende Gegebenheiten im Unternehmen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

**Regt fortwährend an, Aktivitäten und Prozesse der Organisation auch an den Bedarfen der Mitarbeiter zu orientieren.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr gegenwärtiges Verhalten	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Bemüht sich um eine umfassende Einbeziehung von Mitarbeitern in allen relevanten Unternehmensbereichen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Achtet auf eine anregende und energiereiche Arbeitsumgebung.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Regt andere dazu an, Möglichkeiten zur Kompetenz- bzw. Personalentwicklung zu nutzen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Bemüht sich stets, Veränderung anzuregen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Bemüht sich stets, Widerständen bei Veränderungen konstruktiv zu begegnen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Bemüht sich aktiv darum, ein Klima des Vertrauens im Unternehmen zu fördern (bspw. durch Offenheit, Transparenz, die Weitergabe von Informationen, Verzicht auf Kontrolle etc.).**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Ermutigt andere kontinuierlich nach Chancen für Verbesserungen zu suchen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Fördert einen konstruktiven Umgang mit Fehlern in der Organisation.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Bemüht sich darum, die Kollegen zu motivieren.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Bemüht sich um eine Kommunikation im Unternehmen über Hierarchiegrenzen hinweg.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Steuert den Prozess in Meetings und sonstigen Arbeitstreffen gezielt zum Wohle der Beteiligten und der Ergebnisse.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Unterstützt Arbeitsprozesse gezielt durch Techniken der Visualisierung (Arbeit am Flipchart, Nutzung von Metaplankarten etc.).**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Bemüht sich stets um eine professionelle Form der Gesprächsführung (bspw. durch Wertschätzung, Aktives Zuhören).**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Unterstützt Projektmitglieder und -leiter dabei, Projektmanagement systematisch zu betreiben (bspw. mit Hilfe von Methoden zur Planung, Organisation, Kalkulation, Dokumentation von Projekten).**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Unterstützt Projektgruppen dabei, Strategien, Ansätze und Methoden zur systematischen Evaluation ihrer Projekte anzuwenden.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Reagiert stets angemessen bei auftretenden Konflikten.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Bemüht sich stets um einen konstruktiven Umgang mit Rückschlägen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Bemüht sich stets um eine nachhaltige (langfristig wirksame) Form der Konfliktbewältigung im Unternehmen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Bemüht sich um eine kontinuierliche und systematische Überwachung von Innovationsprojekten im Unternehmen.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Hinterfragt frühzeitig, wie Innovationsprojekte nachhaltig umgesetzt werden können.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ihr Kollege/Ihre Kollegin:

**Berücksichtigt unternehmerische Gesichtspunkte im Innovationsprozess (z.B. Marktfaktoren, Kennzahlen, Unternehmensziele)**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Bewertet Innovationsvorhaben abschließend systematisch, um ihren Erfolg bestimmen zu können.**

	trifft überhaupt nicht zu	trifft ein wenig zu	trifft teilweise zu	trifft überwiegend zu	trifft absolut zu
Verhalten vor einem Jahr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegenwärtiges Verhalten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### B3.2 Übersicht über die Skalen und Items des Verfahrens und die Ausgangsitems des FB Lernen

Skala	Ausgangs-Item (FB Lernen)	Item FB Verhalten
Rollenklärung	Aktive Unterstützung von Innovationsprozessen	Ermuntert andere aktiv, ihre Ideen zu äußern.
Rollenklärung	Aktive Unterstützung von Innovationsprozessen	Unterstützt die Auswahl von Ideen.
Rollenklärung	Aktive Unterstützung von Innovationsprozessen	Setzt sich für eine erfolgreiche Umsetzung neuer Ideen ein.
Rollenklärung	Selbstmanagement	Organisiert die eigene Arbeitsweise systematisch (bspw. mit Hilfe von Zielsetzungs-, Durchführungs-, Entscheidungs-, Zeitplantechniken).
Rollenklärung	Selbstreflektion	Bemüht sich stets die eigenen Standpunkte und Handlungen kritisch zu hinterfragen.
Ideenmanagement	Individueller Wissenserwerb	Unterstützt andere aktiv dabei, sich neues Wissen anzueignen
Ideenmanagement	Lernen im Team	Unterstützt Teammitglieder dabei, gemeinsam zu Lernen (bspw. durch den Austausch von Bewertungen und Sichtweisen, durch Abstimmungsprozesse).
Ideenmanagement	Lernen in und mit der Organisation	Bemüht sich darum, eine förderliche Lernkultur in der Organisation voranzutreiben.
Ideenmanagement	Interorganisationale Koordination	Unterstützt aktiv den Wissensaustausch und die Kooperation mit anderen Unternehmen.
Ideenmanagement	Ideenaustausch	Regt die Mitarbeiter stets zu einem offenen Austausch ihrer Ideen an.
Innovationsprozesse und ihre Analyse	Analyse von Innovationsprozessen	Erfasst systematisch innovationsförderliche und -hemmende Gegebenheiten im Unternehmen.
Innovationsprozesse und ihre Analyse	Mitarbeiterorientierung	Regt fortwährend an, Aktivitäten und Prozesse der Organisation auch an den Bedarfen der Mitarbeiter zu orientieren.
Innovationsprozesse und ihre Analyse	Mitarbeiterbeteiligung	Bemüht sich um eine umfassende Einbeziehung von Mitarbeitern in allen relevanten Unternehmensbereichen.
Innovationsprozesse und ihre Analyse	Innovationsförderliche Arbeitsplatzgestaltung	Achtet auf eine anregende und energiereiche Arbeitsumgebung.
Innovationsprozesse und ihre Analyse	Kompetenzentwicklung	Regt andere dazu an, Möglichkeiten zur Kompetenz- bzw. Personalentwicklung zu nutzen.
Rahmenbedingungen für Innovation und Veränderung	Veränderung initiieren	Bemüht sich stets, Veränderung anzuregen.
Rahmenbedingungen für Innovation und Veränderung	Angemessener Umgang mit Widerständen	Bemüht sich stets, Widerständen bei Veränderungen konstruktiv zu begegnen.
Rahmenbedingungen für Innovation und Veränderung	Vertrauensförderliche Maßnahmen	Bemüht sich aktiv darum, ein Klima des Vertrauens im Unternehmen zu fördern (bspw. durch Offenheit, Transparenz, die Weitergabe von Informationen, Verzicht auf Kontrolle etc.).
Intervention zur Förderung von Innovation	Aktive Innovationsförderung	Ermutigt andere kontinuierlich nach Chancen für Verbesserungen zu suchen.
Intervention zur Förderung von Innovation	Lernen aus Fehlern	Fördert einen konstruktiven Umgang mit Fehlern in der Organisation.
Intervention zur Förderung von Innovation	Motivieren	Bemüht sich darum, die Kollegen zu motivieren.
Intervention zur Förderung von Innovation	Horizontale und vertikale Kommunikation	Bemüht sich um eine Kommunikation im Unternehmen über Hierarchiegrenzen hinweg.

Moderation und Gesprächsführung	Moderation	Steuert den Prozess in Meetings und sonstigen Arbeitstreffen gezielt zum Wohle der Beteiligten und der Ergebnisse.
Moderation und Gesprächsführung	Visualisieren	Unterstützt Arbeitsprozesse gezielt durch Techniken der Visualisierung (Arbeit am Flipchart, Nutzung von Metaplankarten etc.).
Moderation und Gesprächsführung	Gesprächsführung	Bemüht sich stets um eine professionelle Form der Gesprächsführung (bspw. durch Wertschätzung, Aktives Zuhören).
Projektmanagement	Anwendung von Projektmanagementmethoden	Unterstützt Projektmitglieder und -leiter dabei, Projektmanagement systematisch zu betreiben (bspw. mit Hilfe von Methoden zur Planung, Organisation, Kalkulation, Dokumentation von Projekten).
Projektevaluation	Projektevaluation	Unterstützt Projektgruppen dabei, Strategien, Ansätze und Methoden zur systematischen Evaluation ihrer Projekte anzuwenden.
Konflikte in Innovationsprozessen	Zweckvolle Konflikt-handhabung + Konfliktphasen	Reagiert stets angemessen bei auftretenden Konflikten.
Konflikte in Innovationsprozessen	Umgang mit Rückschlägen	Bemüht sich stets um einen konstruktiven Umgang mit Rückschlägen.
Konflikte in Innovationsprozessen	Nachhaltige Konfliktbewältigung	Bemüht sich stets um eine nachhaltige ( <i>langfristig wirksame</i> ) Form der Konfliktbewältigung im Unternehmen.
Verankerung im Unternehmen	Systematisches Monitoring	Bemüht sich um eine kontinuierliche und systematische Überwachung von Innovationsprojekten im Unternehmen.
Verankerung im Unternehmen	Nachhaltigkeit	hinterfragt frühzeitig, wie Innovationsprojekte nachhaltig umgesetzt werden können.
Verankerung im Unternehmen	Unternehmerisches Handeln	Berücksichtigt unternehmerische Gesichtspunkte im Innovationsprozess (z.B. Marktfaktoren, Kennzahlen, Unternehmensziele).
Verankerung im Unternehmen	Systematisches Bewerten von Innovationsvorhaben	Bewertet Innovationsvorhaben abschließend systematisch, um ihren Erfolg bestimmen zu können.

## Appendix B4: Verwendete Version des deutschen Lerntransfer-System-Inventars (GLTSI), Online-Version

*(Nachfolgend werden die Skalen zur besseren Übersicht mit Überschriften versehen; in der Befragung waren die Skalen hingegen nicht sichtbar beschriftet.)*

### Motivation zum Lerntransfer

**Die Ausbildung wird meine Produktivität erhöhen.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Ich kann es immer kaum erwarten, nach der Ausbildung wieder zur Arbeit zu gehen und das Erlernete auszuprobieren.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Ich glaube die Ausbildung wird mir helfen, meine jetzige Arbeit besser zu bewältigen.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Ich bin motiviert, wenn ich daran denke, das was ich neu gelernt habe bei der Arbeit auszuprobieren**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

### Transferdesign

**Die Aktivitäten und Übungen, die die Trainer benutzten, helfen mir, das was ich gelernt habe, bei der Arbeit anzuwenden.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Diejenigen, die die Ausbildung durchführen, verstehen, wie ich das Erlernete anwenden werde.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Die Trainer haben viele Beispiele verwendet, mit denen sie mir zeigten, wie ich das Gelernte bei der Arbeit anwenden könnte.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Die Art und Weise, wie die Trainer die Ausbildungsinhalte vermittelten, gab mir Vertrauen, das Gelernte auch anzuwenden.**

- |                       |        |                       |            |                       |        |                       |        |                       |        |
|-----------------------|--------|-----------------------|------------|-----------------------|--------|-----------------------|--------|-----------------------|--------|
| <input type="radio"/> | trifft | <input type="radio"/> | trifft ein | <input type="radio"/> | trifft | <input type="radio"/> | trifft | <input type="radio"/> | trifft |
|-----------------------|--------|-----------------------|------------|-----------------------|--------|-----------------------|--------|-----------------------|--------|

### Jobübereinstimmung

**Die in der Ausbildung verwendeten Werkzeuge, Materialien und die Ausstattung sind den Dingen, die ich bei der Arbeit nutze, sehr ähnlich.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Die in der Ausbildung angewendeten Methoden sind denen bei der Arbeit sehr ähnlich.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Es gefällt mir, dass die Ausbildungsinhalte und meine Arbeit so ähnlich sind.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Was in der Ausbildung vermittelt wurde, entspricht meinen Arbeitsanforderungen.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|

**Die in der Ausbildung verwendeten Beispielsituationen sind den Situationen sehr ähnlich, denen ich bei der Arbeit begegne.**

- |                       |                                 |                       |                           |                       |                           |                       |                             |                       |                         |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|
| <input type="radio"/> | trifft<br>überhaupt<br>nicht zu | <input type="radio"/> | trifft ein<br>wenig<br>zu | <input type="radio"/> | trifft<br>teilweise<br>zu | <input type="radio"/> | trifft<br>überwiegend<br>zu | <input type="radio"/> | trifft<br>absolut<br>zu |
|-----------------------|---------------------------------|-----------------------|---------------------------|-----------------------|---------------------------|-----------------------|-----------------------------|-----------------------|-------------------------|



Erwartungsklarheit

**Vor der Ausbildung wusste ich, welche Auswirkungen das Programm auf meine Arbeitsleistung haben sollte.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Vor Antritt der Ausbildung hatte ich einen guten Überblick über ihre Nützlichkeit im Hinblick auf meine Weiterentwicklung im Job.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Was ich von der Ausbildung zu erwarten habe, wusste ich vorher.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Die erwarteten Ergebnisse der Ausbildung waren von Anfang an klar.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

Transferkapazität

**Mein Tagesablauf erlaubt es mir Dinge zu ändern um meine neuen Kenntnisse einbringen zu können.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Meine Prioritäten müssen geändert werden, bevor ich in der Lage sein werde, das Neugelernnte anzuwenden.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Meine Arbeitsbelastung erlaubt es mir, neue Dinge, die ich gelernt habe, auszuprobieren.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Momentan passiert einfach zu viel bei der Arbeit, so dass es mir schwer fällt, die Ausbildungsinhalte auszuprobieren.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

Wissensanwendung\_Möglichkeiten

**Bei der Arbeit hindert mich ein begrenztes Budget, die in der Ausbildung erworbenen Fertigkeiten zu nutzen.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Unser gegenwärtiger Personalstand ist für mich ausreichend, um die Ausbildungsinhalte anzuwenden.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Es ist schwierig, die benötigten Materialien und das Zubehör zu bekommen, die ich brauche, um das Wissen und die Fertigkeiten, die ich in der Ausbildung erworben habe, anzuwenden.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Es sind genügend Ressourcen vorhanden, um meine in der Ausbildung erworbenen Fertigkeiten einbringen zu können.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Ich habe Möglichkeiten, die Ausbildung in meiner Arbeit zu nutzen.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

Unterstützung durch Kollegen

**Meine Kollegen schätzen es, wenn ich die in der Ausbildung neu erworbenen Fertigkeiten anwende.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Meine Kollegen ermutigen mich, die in der Ausbildung erworbenen Fertigkeiten anzuwenden.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Bei der Arbeit erwarten meine Kollegen von mir, dass ich das in der Ausbildung Erlernete anwende.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Meine Kollegen sind geduldig mit mir, wenn ich neue Fertigkeiten oder Techniken bei der Arbeit ausprobiere.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Mein Vorgesetzter trifft sich regelmäßig mit mir, um Probleme zu bearbeiten, die ich haben könnte, wenn ich das neu Erlernete ausprobiere.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Mein Vorgesetzter trifft sich mit mir, um zu diskutieren, wie ich die Fortbildungsinhalte bei der Arbeit anwenden kann.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Mein Vorgesetzter zeigt Interesse an dem, was ich in der Fortbildung lerne.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Mein Vorgesetzter lobt meine Arbeit, wenn ich die in der Ausbildung erworbenen Kenntnisse einbringe.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Mein Vorgesetzter hilft mir, basierend auf der Ausbildung realistische Ziele für meine Arbeitsleistung zu setzen.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

Einstellung\_Sanktionen\_FK

**Mein Vorgesetzter wird etwas dagegen haben, wenn ich versuche, die Ausbildungsinhalte bei der Arbeit anzuwenden.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Mein Vorgesetzter wird etwas dagegen haben, wenn ich die Techniken anwende, die ich in der Ausbildung erlernt habe.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

**Mein Vorgesetzter denkt, ich arbeite weniger effektiv, wenn ich die in der Ausbildung vermittelten Techniken nutze.**

- trifft überhaupt nicht zu  trifft ein wenig zu  trifft teilweise zu  trifft überwiegend zu  trifft absolut zu

## Appendix C: Ergebnisse der Berechnungen

### Appendix C1: Ergebnisse der Überprüfung der Normalverteilungsannahme

*Gesamtzufriedenheit Module 1-10*

#### Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von t1_G_Zufriedenheit ist normal mit Mittelwert 4.43 und Standardabweichung 0.37.	Kolmogorov-Smirnov-Test einer Stichprobe	,434	Nullhypothese behalten.
2	Die Verteilung von t2_G_Zufriedenheit ist normal mit Mittelwert 4.35 und Standardabweichung 0.48.	Kolmogorov-Smirnov-Test einer Stichprobe	,743	Nullhypothese behalten.
3	Die Verteilung von t3_G_Zufriedenheit ist normal mit Mittelwert 3.95 und Standardabweichung 0.56.	Kolmogorov-Smirnov-Test einer Stichprobe	,784	Nullhypothese behalten.
4	Die Verteilung von t4_G_Zufriedenheit ist normal mit Mittelwert 4.62 und Standardabweichung 0.28.	Kolmogorov-Smirnov-Test einer Stichprobe	,785	Nullhypothese behalten.
5	Die Verteilung von t5_G_Zufriedenheit ist normal mit Mittelwert 3.93 und Standardabweichung 0.52.	Kolmogorov-Smirnov-Test einer Stichprobe	,564	Nullhypothese behalten.
6	Die Verteilung von t6_G_Zufriedenheit ist normal mit Mittelwert 4.84 und Standardabweichung 0.18.	Kolmogorov-Smirnov-Test einer Stichprobe	,514	Nullhypothese behalten.
7	Die Verteilung von t7_G_Zufriedenheit ist normal mit Mittelwert 4.59 und Standardabweichung 0.33.	Kolmogorov-Smirnov-Test einer Stichprobe	,607	Nullhypothese behalten.
8	Die Verteilung von t8_G_Zufriedenheit ist normal mit Mittelwert 4.58 und Standardabweichung 0.39.	Kolmogorov-Smirnov-Test einer Stichprobe	,321	Nullhypothese behalten.
9	Die Verteilung von t9_G_Zufriedenheit ist normal mit Mittelwert 3.96 und Standardabweichung 0.38.	Kolmogorov-Smirnov-Test einer Stichprobe	1,000	Nullhypothese behalten.
10	Die Verteilung von t10_G_Zufriedenheit ist normal mit Mittelwert 4.52 und Standardabweichung 0.46.	Kolmogorov-Smirnov-Test einer Stichprobe	,701	Nullhypothese behalten.
11	Die Verteilung von Bewertung_Trainer_Gesamt ist normal mit Mittelwert 4.63 und Standardabweichung 0.30.	Kolmogorov-Smirnov-Test einer Stichprobe	,363	Nullhypothese behalten.
12	Die Verteilung von Bewertung_Inhalt_Gesamt ist normal mit Mittelwert 4.24 und Standardabweichung 0.30.	Kolmogorov-Smirnov-Test einer Stichprobe	,959	Nullhypothese behalten.
13	Die Verteilung von Bewertung_Didaktik_Gesamt ist normal mit Mittelwert 4.27 und Standardabweichung 0.32.	Kolmogorov-Smirnov-Test einer Stichprobe	,826	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

**Übersicht über Hypothesentest**

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von t1_Trainer_MW ist normal mit Mittelwert 4.75 und Standardabweichung 0.31.	Kolmogorov-Smirnov-Test einer Stichprobe	,257	Nullhypothese behalten.
2	Die Verteilung von t2_Trainer_MW ist normal mit Mittelwert 4.58 und Standardabweichung 0.40.	Kolmogorov-Smirnov-Test einer Stichprobe	,603	Nullhypothese behalten.
3	Die Verteilung von t3_Trainer_MW ist normal mit Mittelwert 4.35 und Standardabweichung 0.62.	Kolmogorov-Smirnov-Test einer Stichprobe	,635	Nullhypothese behalten.
4	Die Verteilung von t4_Trainer_MW ist normal mit Mittelwert 4.67 und Standardabweichung 0.27.	Kolmogorov-Smirnov-Test einer Stichprobe	,693	Nullhypothese behalten.
5	Die Verteilung von t5_Trainer_MW ist normal mit Mittelwert 4.27 und Standardabweichung 0.55.	Kolmogorov-Smirnov-Test einer Stichprobe	,877	Nullhypothese behalten.
6	Die Verteilung von t6_Trainer_MW ist normal mit Mittelwert 4.95 und Standardabweichung 0.11.	Kolmogorov-Smirnov-Test einer Stichprobe	,012	Nullhypothese ablehnen.
7	Die Verteilung von t7_Trainer_MW ist normal mit Mittelwert 4.77 und Standardabweichung 0.27.	Kolmogorov-Smirnov-Test einer Stichprobe	,568	Nullhypothese behalten.
8	Die Verteilung von t8_Trainer_MW ist normal mit Mittelwert 4.73 und Standardabweichung 0.35.	Kolmogorov-Smirnov-Test einer Stichprobe	,265	Nullhypothese behalten.
9	Die Verteilung von t9_Trainer_MW ist normal mit Mittelwert 4.45 und Standardabweichung 0.46.	Kolmogorov-Smirnov-Test einer Stichprobe	,920	Nullhypothese behalten.
10	Die Verteilung von t10_Trainer_MW ist normal mit Mittelwert 4.76 und Standardabweichung 0.48.	Kolmogorov-Smirnov-Test einer Stichprobe	,042	Nullhypothese ablehnen.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

**Übersicht über Hypothesentest**

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von t1_inhalt_MW ist normal mit Mittelwert 4.27 und Standardabweichung 0.50.	Kolmogorov-Smirnov-Test einer Stichprobe	,513	Nullhypothese behalten.
2	Die Verteilung von t2_inhalt_MW ist normal mit Mittelwert 4.15 und Standardabweichung 0.42.	Kolmogorov-Smirnov-Test einer Stichprobe	,603	Nullhypothese behalten.
3	Die Verteilung von t3_inhalt_MW ist normal mit Mittelwert 3.68 und Standardabweichung 0.60.	Kolmogorov-Smirnov-Test einer Stichprobe	,873	Nullhypothese behalten.
4	Die Verteilung von t4_inhalt_MW ist normal mit Mittelwert 4.55 und Standardabweichung 0.46.	Kolmogorov-Smirnov-Test einer Stichprobe	,566	Nullhypothese behalten.
5	Die Verteilung von t5_inhalt_MW ist normal mit Mittelwert 4.08 und Standardabweichung 0.64.	Kolmogorov-Smirnov-Test einer Stichprobe	,567	Nullhypothese behalten.
6	Die Verteilung von t6_Inhalt_MW ist normal mit Mittelwert 4.77 und Standardabweichung 0.35.	Kolmogorov-Smirnov-Test einer Stichprobe	,181	Nullhypothese behalten.
7	Die Verteilung von t7_Inhalt_MW ist normal mit Mittelwert 4.40 und Standardabweichung 0.52.	Kolmogorov-Smirnov-Test einer Stichprobe	,496	Nullhypothese behalten.
8	Die Verteilung von t8_Inhalt_MW ist normal mit Mittelwert 4.45 und Standardabweichung 0.51.	Kolmogorov-Smirnov-Test einer Stichprobe	,662	Nullhypothese behalten.
9	Die Verteilung von t9_Inhalt_MW ist normal mit Mittelwert 3.58 und Standardabweichung 0.46.	Kolmogorov-Smirnov-Test einer Stichprobe	,767	Nullhypothese behalten.
10	Die Verteilung von t10_Inhalt_MW ist normal mit Mittelwert 4.45 und Standardabweichung 0.54.	Kolmogorov-Smirnov-Test einer Stichprobe	,565	Nullhypothese behalten.

Asymptotische Signifikanzniveaus werden angezeigt. Das Signifikanzniveau ist .05.

**Übersicht über Hypothesentest**

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von t1_didaktik_MW ist normal mit Mittelwert 4.28 und Standardabweichung 0.47.	Kolmogorov-Smirnov-Test einer Stichprobe	,582	Nullhypothese behalten.
2	Die Verteilung von t2_didaktik_MW ist normal mit Mittelwert 4.33 und Standardabweichung 0.80.	Kolmogorov-Smirnov-Test einer Stichprobe	,073	Nullhypothese behalten.
3	Die Verteilung von t3_didaktik_MW ist normal mit Mittelwert 3.83 und Standardabweichung 0.65.	Kolmogorov-Smirnov-Test einer Stichprobe	,887	Nullhypothese behalten.
4	Die Verteilung von t4_didaktik_MW ist normal mit Mittelwert 4.65 und Standardabweichung 0.32.	Kolmogorov-Smirnov-Test einer Stichprobe	,390	Nullhypothese behalten.
5	Die Verteilung von t5_didaktik_MW ist normal mit Mittelwert 3.44 und Standardabweichung 0.65.	Kolmogorov-Smirnov-Test einer Stichprobe	,922	Nullhypothese behalten.
6	Die Verteilung von t6_Didaktik_MW ist normal mit Mittelwert 4.80 und Standardabweichung 0.21.	Kolmogorov-Smirnov-Test einer Stichprobe	,388	Nullhypothese behalten.
7	Die Verteilung von t7_Didaktik_MW ist normal mit Mittelwert 4.61 und Standardabweichung 0.36.	Kolmogorov-Smirnov-Test einer Stichprobe	,394	Nullhypothese behalten.
8	Die Verteilung von t8_Didaktik_MW ist normal mit Mittelwert 4.57 und Standardabweichung 0.40.	Kolmogorov-Smirnov-Test einer Stichprobe	,645	Nullhypothese behalten.
9	Die Verteilung von t9_Didaktik_MW ist normal mit Mittelwert 3.85 und Standardabweichung 0.44.	Kolmogorov-Smirnov-Test einer Stichprobe	,955	Nullhypothese behalten.
10	Die Verteilung von t10_Didaktik_MW ist normal mit Mittelwert 4.36 und Standardabweichung 0.69.	Kolmogorov-Smirnov-Test einer Stichprobe	,747	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

**Übersicht über Hypothesentest**

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von MW_t1_Erfahrung ist normal mit Mittelwert 3.22 und Standardabweichung 0.60.	Kolmogorov-Smirnov-Test einer Stichprobe	,995	Nullhypothese behalten.
2	Die Verteilung von MW_t2_Erfahrung ist normal mit Mittelwert 4.17 und Standardabweichung 0.58.	Kolmogorov-Smirnov-Test einer Stichprobe	,607	Nullhypothese behalten.
3	Die Verteilung von MW_t3_Erfahrung ist normal mit Mittelwert 3.54 und Standardabweichung 0.44.	Kolmogorov-Smirnov-Test einer Stichprobe	,942	Nullhypothese behalten.
4	Die Verteilung von MW_t1_Theorie ist normal mit Mittelwert 3.66 und Standardabweichung 0.53.	Kolmogorov-Smirnov-Test einer Stichprobe	,980	Nullhypothese behalten.
5	Die Verteilung von MW_t2_Theorie ist normal mit Mittelwert 3.54 und Standardabweichung 0.47.	Kolmogorov-Smirnov-Test einer Stichprobe	,981	Nullhypothese behalten.
6	Die Verteilung von MW_t3_Theorie ist normal mit Mittelwert 4.21 und Standardabweichung 0.62.	Kolmogorov-Smirnov-Test einer Stichprobe	,999	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

*Ebene Verhalten*

**Übersicht über Hypothesentest**

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von MW_Verhalten_vor_1Jahr ist normal mit Mittelwert 3.25 und Standardabweichung 0.65.	Kolmogorov-Smirnov-Test einer Stichprobe	,867	Nullhypothese behalten.
2	Die Verteilung von MW_Gegenwärtiges_Verhalten ist normal mit Mittelwert 3.84 und Standardabweichung 0.55.	Kolmogorov-Smirnov-Test einer Stichprobe	,993	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

Transferbedingungen

Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von Skala_Motivation_Lerntransfer_MW ist normal mit Mittelwert 4.04 und Standardabweichung 0.51.	Kolmogorov-Smirnov-Test einer Stichprobe	,624	Nullhypothese behalten.
2	Die Verteilung von Skala_Transferdesign_MW ist normal mit Mittelwert 4.00 und Standardabweichung 0.65.	Kolmogorov-Smirnov-Test einer Stichprobe	,880	Nullhypothese behalten.
3	Die Verteilung von Skala_Jobfit_MW ist normal mit Mittelwert 3.60 und Standardabweichung 0.50.	Kolmogorov-Smirnov-Test einer Stichprobe	,902	Nullhypothese behalten.
4	Die Verteilung von Skala_Erwartungsklarheit_MW ist normal mit Mittelwert 3.07 und Standardabweichung 0.62.	Kolmogorov-Smirnov-Test einer Stichprobe	,988	Nullhypothese behalten.
5	Die Verteilung von Skala_Transferkapazitaet_MW ist normal mit Mittelwert 3.93 und Standardabweichung 0.85.	Kolmogorov-Smirnov-Test einer Stichprobe	,975	Nullhypothese behalten.
6	Die Verteilung von Skala_Wissensanwendung_MW ist normal mit Mittelwert 4.31 und Standardabweichung 0.41.	Kolmogorov-Smirnov-Test einer Stichprobe	,366	Nullhypothese behalten.
7	Die Verteilung von Skala_UnterstuetzungKollegen_MW ist normal mit Mittelwert 3.57 und Standardabweichung 0.69.	Kolmogorov-Smirnov-Test einer Stichprobe	,869	Nullhypothese behalten.
8	Die Verteilung von Skala_UnterstuetzungVorgesetzte_MW ist normal mit Mittelwert 2.77 und Standardabweichung 1.06.	Kolmogorov-Smirnov-Test einer Stichprobe	,923	Nullhypothese behalten.
9	Die Verteilung von Skala_Einstellungen_Sanktionen_FK_MW ist normal mit Mittelwert 4.70 und Standardabweichung 0.64.	Kolmogorov-Smirnov-Test einer Stichprobe	,306	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.



## Appendix C2: Analysen auf der Ebene Reaktion

### Ergebnisse der deskriptiven Analysen

#### Gesamtzufriedenheit

	N	Minimum	Maximum	Mittelwert	Standardabweichung
t1_Zufriedenheit_Gesamt	15	3,83	5,00	4,4309	,37415
t2_Zufriedenheit_Gesamt	15	3,40	4,83	4,3542	,47717
t3_Zufriedenheit_Gesamt	15	2,64	4,87	3,9547	,55995
t4_Zufriedenheit_Gesamt	15	4,02	5,00	4,6233	,27777
t5_Zufriedenheit_Gesamt	15	3,03	4,85	3,9314	,51581
t6_Zufriedenheit_Gesamt	15	4,34	5,00	4,8389	,18066
t7_Zufriedenheit_Gesamt	15	3,93	5,00	4,5947	,32797
t8_Zufriedenheit_Gesamt	15	3,79	5,00	4,5842	,39066
t9_Zufriedenheit_Gesamt	15	3,15	4,53	3,9622	,38020
t10_Zufriedenheit_Gesamt	15	3,63	5,00	4,5228	,45706
Zufriedenheit GESAMT alle Module	15	3,74	4,74	4,3797	,28741

### Zufriedenheit nach Dimensionen

	N	Minimum	Maximum	Mittelwert	Standardabweichung
t1 inhalt MW	15	3,00	5,00	4,2667	,50415
t1 Trainer MW	15	4,13	5,00	4,7462	,30952
t1 didaktik MW	15	3,20	5,00	4,2800	,47087
t2 inhalt MW	15	3,25	4,75	4,1500	,42046
t2 Trainer MW	15	3,88	5,00	4,5792	,40330
t2 didaktik MW	15	3,00	5,00	4,3333	,79522
t3 inhalt MW	15	2,50	5,00	3,6833	,60109
t3 Trainer MW	15	2,81	5,00	4,3542	,62351
t3 didaktik MW	15	2,60	4,60	3,8267	,65407
t4 inhalt MW	15	3,75	5,00	4,5500	,45513
t4 Trainer MW	15	4,31	5,00	4,6667	,27003
t4 didaktik MW	15	4,00	5,00	4,6533	,31593
t5 inhalt MW	15	2,75	5,00	4,0833	,63854
t5 Trainer MW	15	3,38	5,00	4,2708	,54979
t5 didaktik MW	15	2,40	5,00	3,4400	,65115
t6 Inhalt MW	15	3,75	5,00	4,7667	,34675
t6 Trainer MW	15	4,63	5,00	4,9500	,10617
t6 Didaktik MW	15	4,40	5,00	4,8000	,21381
t7 Inhalt MW	15	3,25	5,00	4,4000	,52440
t7 Trainer MW	15	4,13	5,00	4,7708	,27209
t7 Didaktik MW	15	3,80	5,00	4,6133	,35830
t8 Inhalt MW	15	3,25	5,00	4,4500	,51060
t8 Trainer MW	15	3,88	5,00	4,7292	,35487
t8 Didaktik MW	15	4,00	5,00	4,5733	,39905
t9 Inhalt MW	15	2,75	4,25	3,5833	,45968
t9 Trainer MW	15	3,50	5,00	4,4500	,46483
t9 Didaktik MW	15	3,20	4,60	3,8533	,43731
t10 Inhalt MW	15	3,25	5,00	4,4500	,54445
t10 Trainer MW	15	3,38	5,00	4,7583	,48058
t10 Didaktik MW	15	3,00	5,00	4,3600	,68536
Gültige Werte (Listenweise)	15				

### Mittelwerte der Gesamtzufriedenheit nach Dimensionen

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Bewertung Trainer Gesamt	15	3,84	4,98	4,6275	,29878
Bewertung Inhalt Gesamt	15	3,70	4,65	4,2383	,30309
Bewertung_Didaktik_Gesamt	15	3,58	4,66	4,2733	,32308
Gültige Werte (Listenweise)	15				

## Ergebnisse der Signifikanzprüfung

### Einfaktorielle Varianzanalyse mit Messwiederholung

*Unterschiede in der Gesamtzufriedenheit Module 1-10*

#### Mauchly-Test auf Sphärität<sup>b</sup>

Maß: MASS\_1

Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Untergrenze
Messung	,004	59,026	44	,095	,495	,756	,111

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

b. Design: Konstanter Term

Innersubjektdesign: Messung

#### Tests der Innersubjekteffekte

Maß: MASS\_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat	Nichtzentralitäts-Parameter	Beobachtete Schärfe <sup>a</sup>
Messung	Sphärität angenommen	14,068	9	1,563	16,655	,000	,543	149,894	1,000
	Greenhouse-Geisser	14,068	4,459	3,155	16,655	,000	,543	74,271	1,000
	Huynh-Feldt	14,068	6,802	2,068	16,655	,000	,543	113,280	1,000
	Untergrenze	14,068	1,000	14,068	16,655	,001	,543	16,655	,966
Fehler(Messung)	Sphärität angenommen	11,826	126	,094					
	Greenhouse-Geisser	11,826	62,432	,189					
	Huynh-Feldt	11,826	95,222	,124					
	Untergrenze	11,826	14,000	,845					

a. Unter Verwendung von Alpha =

**Paarweise Vergleiche**

Maß: MASS\_1

(I)Messung	(J)Messung	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>a</sup>	95% Konfidenzintervall für die Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
1	2	.077	.074	1,000	-,224	.378
	3	.476*	.111	.035	.021	.932
	4	-,192	.073	.859	-,489	.104
	5	.500*	.121	.045	.006	.993
	6	-,408*	.090	.021	-,776	-,040
	7	-,164	.122	1,000	-,662	.334
	8	-,153	.097	1,000	-,550	.244
	9	.469*	.081	.002	.139	.799
	10	-,092	.099	1,000	-,498	.314
	2	1	-,077	.074	1,000	-,378
3		.399	.107	.098	-,037	.835
4		-,269	.108	1,000	-,709	.171
5		.423*	.094	.022	.039	.807
6		-,485*	.112	.032	-,944	-,025
7		-,241	.142	1,000	-,822	.341
8		-,230	.105	1,000	-,659	.199
9		.392*	.085	.018	.044	.740
10		-,169	.126	1,000	-,684	.346
3		1	-,476*	.111	.035	-,932
	2	-,399	.107	.098	-,835	.037
	4	-,669*	.113	.002	-,1,130	-,207
	5	.023	.162	1,000	-,640	.686
	6	-,884*	.136	.001	-,1,438	-,330
	7	-,640	.158	.053	-,1,285	.005
	8	-,629*	.116	.004	-,1,104	-,155
	9	-,007	.107	1,000	-,443	.428
	10	-,568*	.135	.040	-,1,121	-,015
	4	1	.192	.073	.859	-,104
2		.269	.108	1,000	-,171	.709
3		.669*	.113	.002	.207	1,130
5		.692*	.140	.010	.119	1,265
6		-,216*	.049	.027	-,416	-,015
7		.029	.081	1,000	-,301	.358
8		.039	.105	1,000	-,388	.466
9		.661*	.071	.000	.371	.952
10		.101	.085	1,000	-,248	.449
5		1	-,500*	.121	.045	-,993
	2	-,423*	.094	.022	-,807	-,039
	3	-,023	.162	1,000	-,686	.640
	4	-,692*	.140	.010	-,1,265	-,119
	6	-,908*	.128	.000	-,1,429	-,386
	7	-,663*	.154	.032	-,1,291	-,035
	8	-,653*	.129	.008	-,1,182	-,124
	9	-,031	.126	1,000	-,544	.482
	10	-,591	.153	.077	-,1,216	.034

6	1	408*	,090	,021	040	776
	2	485*	,112	,032	025	944
	3	884*	,136	,001	330	1,438
	4	216*	,049	,027	015	416
	5	908*	,128	,000	386	1,429
	7	244	,063	,079	-,015	503
	8	255	,104	1,000	-,169	679
	9	877*	,086	,000	527	1,227
	10	316	,088	,137	-,045	678
	7	1	164	,122	1,000	-,334
2		241	,142	1,000	-,341	822
3		640	,158	,053	-,005	1,285
4		-,029	,081	1,000	-,358	301
5		663*	,154	,032	035	1,291
6		-,244	,063	,079	-,503	015
8		011	,120	1,000	-,478	499
9		633*	,107	,002	197	1,068
10		072	,101	1,000	-,339	483
8		1	153	,097	1,000	-,244
	2	230	,105	1,000	-,199	659
	3	629*	,116	,004	155	1,104
	4	-,039	,105	1,000	-,466	388
	5	653*	,129	,008	124	1,182
	6	-,255	,104	1,000	-,679	169
	7	-,011	,120	1,000	-,499	478
	9	622*	,107	,002	186	1,058
	10	061	,140	1,000	-,511	634
	9	1	-,469*	,081	,002	-,799
2		-,392*	,085	,018	-,740	-,044
3		007	,107	1,000	-,428	443
4		-,661*	,071	,000	-,952	-,371
5		031	,126	1,000	-,482	544
6		-,877*	,086	,000	-,1,227	-,527
7		-,633*	,107	,002	-,1,068	-,197
8		-,622*	,107	,002	-,1,058	-,186
10		-,561*	,089	,001	-,924	-,198
10		1	092	,099	1,000	-,314
	2	169	,126	1,000	-,346	684
	3	568*	,135	,040	015	1,121
	4	-,101	,085	1,000	-,449	248
	5	591	,153	,077	-,034	1,216
	6	-,316	,088	,137	-,678	045
	7	-,072	,101	1,000	-,483	339
	8	-,061	,140	1,000	-,634	511
	9	561*	,089	,001	198	924

Basiert auf den geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Bonferroni.

\*. Die mittlere Differenz ist auf dem

## Unterschiede zwischen den Zufriedenheitsdimensionen

### Mauchly-Test auf Sphärizität<sup>b</sup>

Maß: MASS\_1

	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse- Geisser	Huynh-Feldt	Untergrenze
Innersubjekteffekt Dimension	,749	3,761	2	,153	,799	,886	,500

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

b. Design: Konstanter Term

Innersubjektdesign: Dimension

### Tests der Innersubjekteffekte

Maß: MASS\_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta- Quadrat	Nichtzentralitäts- Parameter	Beobachtete Schärfe <sup>a</sup>
Dimension	Sphärizität angenommen	1,391	2	,695	36,882	,000	,725	73,764	1,000
	Greenhouse- Geisser	1,391	1,598	,870	36,882	,000	,725	58,954	1,000
	Huynh-Feldt	1,391	1,772	,785	36,882	,000	,725	65,358	1,000
	Untergrenze	1,391	1,000	1,391	36,882	,000	,725	36,882	1,000
Fehler(Dimension)	Sphärizität angenommen	,528	28	,019					
	Greenhouse- Geisser	,528	22,378	,024					
	Huynh-Feldt	,528	24,809	,021					
	Untergrenze	,528	14,000	,038					

a. Unter Verwendung von Alpha =

### Paarweise Vergleiche

Maß: MASS\_1

(I)Dimension	(J)Dimension	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>a</sup>	95% Konfidenzintervall für die Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
1	2	,389 <sup>*</sup>	,061	,000	,222	,556
	3	,354 <sup>*</sup>	,042	,000	,239	,470
2	1	-,389 <sup>*</sup>	,061	,000	-,556	-,222
	3	-,035	,044	1,000	-,156	,086
3	1	-,354 <sup>*</sup>	,042	,000	-,470	-,239
	2	,035	,044	1,000	-,086	,156

Basiert auf den geschätzten Randmitteln

\*. Die mittlere Differenz ist auf dem

a. Anpassung für Mehrfachvergleiche: Bonferroni.

## Appendix C3: Analysen auf der Ebene Lernen

### Ergebnisse der deskriptiven Analysen

**Mittelwerte und Standardabweichung, Wissen gesamt und unterteilt in deklaratives und prozedurales Wissen zu den drei Messzeitpunkten**

	N	Mittelwert	Standardabweichung
MW_t1_prozedurales Wissen	13	3,2421	,45594
MW_t2_prozedurales Wissen	13	3,9814	,65599
MW_t3_prozedurales Wissen	13	3,5769	,43214
MW_t1_deklaratives Wissen	13	3,6921	,39731
MW_t2_deklaratives Wissen	13	3,5166	,49868
MW_t3_deklaratives Wissen	13	4,2457	,52837
MW_t1_GESAMT	13	3,4671	,32317
MW_t2_GESAMT	13	3,7490	,53562
MW_t3_GESAMT	13	3,9113	,45082
Gültige Werte (Listenweise)	13		

**Mittelwerte pro Modul, unterteilt in deklaratives und prozedurales Wissen**

Modul	Ergebnisse zu t1		Ergebnisse zu t2		Ergebnisse zu t3	
	Deklarativ	Prozedural	Deklarativ	Prozedural	Deklarativ	Prozedural
1	4,1	3,8	3,8	4,5	4,7	4,0
2	3,6	3,4	3,4	4,0	4,1	3,6
3	3,5	3,1	3,6	4,2	4,2	3,5
4	3,7	3,0	3,3	4,2	4,2	3,4
5	4,1	3,6	3,7	4,3	4,3	4,0
6	3,8	3,2	3,4	4,0	4,3	3,7
7	3,6	2,9	3,4	3,9	4,4	3,4
8	3,6	3,1	3,8	3,6	4,2	3,5
9	3,6	3,0	3,3	4,0	4,1	3,2
<b>Gesamt</b>	<b>3,7</b>	<b>3,2</b>	<b>3,5</b>	<b>4,0</b>	<b>4,3</b>	<b>3,6</b>

## Ergebnisse der Signifikanzprüfung

### Einfaktorielle Varianzanalyse mit Messwiederholung

*Deklaratives Wissen t1-t3*

#### Mauchly-Test auf Sphärität<sup>b</sup>

Maß: MASS\_1

Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Untergrenze
Messung	,769	2,887	2	,236	,812	,922	,500

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

b. Design: Konstanter Term

Innersubjektdesign: Messung

#### Tests der Innersubjekteffekte

Maß: MASS\_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat	Nichtzentralitäts-Parameter	Beobachtete Schärfe <sup>a</sup>
Messung	Sphärität angenommen	3,765	2	1,883	11,186	,000	,482	22,372	984
	Greenhouse-Geisser	3,765	1,625	2,317	11,186	,001	,482	18,176	963
	Huynh-Feldt	3,765	1,843	2,043	11,186	,001	,482	20,619	977
	Untergrenze	3,765	1,000	3,765	11,186	,006	,482	11,186	866
Fehler(Messung)	Sphärität angenommen	4,039	24	,168					
	Greenhouse-Geisser	4,039	19,499	,207					
	Huynh-Feldt	4,039	22,119	,183					
	Untergrenze	4,039	12,000	,337					

a. Unter Verwendung von Alpha =

#### Paarweise Vergleiche

Maß: MASS\_1

(I) Messung	(J) Messung	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>a</sup>	95% Konfidenzintervall für die Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
1	2	,175	,169	,960	-,295	,646
	3	-,554*	,187	,036	-1,074	-,033
2	1	-,175	,169	,960	-,646	,295
	3	-,729*	,118	,000	-1,058	-,400
3	1	,554*	,187	,036	,033	1,074
	2	,729*	,118	,000	,400	1,058

Basiert auf den geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Bonferroni.

\*. Die mittlere Differenz ist auf dem



Prozedurales Wissen t1-t3

**Mauchly-Test auf Sphärizität<sup>b</sup>**

Maß: MASS\_1

Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Untergrenze
Messung	,785	2,656	2	,265	,823	,937	,500

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

b. Design: Konstanter Term

Innersubjektdesign: Messung

**Tests der Innersubjekteffekte**

Maß: MASS\_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Nichtzentralitäts-Parameter	Beobachtete Schärfe <sup>a</sup>
Messung	Sphärizität angenommen	3,564	2	1,782	11,739	,000	23,477	988
	Greenhouse-Geisser	3,564	1,647	2,164	11,739	,001	19,330	972
	Huynh-Feldt	3,564	1,875	1,901	11,739	,000	22,004	983
	Untergrenze	3,564	1,000	3,564	11,739	,005	11,739	881
Fehler(Messung)	Sphärizität angenommen	3,643	24	152				
	Greenhouse-Geisser	3,643	19,761	184				
	Huynh-Feldt	3,643	22,494	162				
	Untergrenze	3,643	12,000	304				

a. Unter Verwendung von Alpha =

**Paarweise Vergleiche**

Maß: MASS\_1

(I) Messung	(J) Messung	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>a</sup>	95% Konfidenzintervall für die Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
1	2	-,739*	,184	,005	-1,251	-,228
	3	-,335	,126	,063	-,686	,016
2	1	,739*	,184	,005	,228	1,251
	3	,405*	,142	,044	,009	,800
3	1	,335	,126	,063	-,016	,686
	2	-,405*	,142	,044	-,800	-,009

Basiert auf den geschätzten Randmitteln

\*. Die mittlere Differenz ist auf dem

a. Anpassung für Mehrfachvergleiche: Bonferroni.

### t-test für abhängige Stichproben

Wissenszuwachs zu t2 - Cluster 1 vs. Cluster 2

Deklaratives Wissen: Cluster 1 (zu t2 trainierte Module)

#### Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Cluster1_t1_Deklarativ	3,7823	13	,48938	,13573
	Cluster1_t2_Deklarativ	3,5546	13	,43373	,12030

#### Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren 1	Cluster1_t1_Deklarativ & Cluster1_t2_Deklarativ	13	-,132	,666

#### Test bei gepaarten Stichproben

	Gepaarte Differenzen					T	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
				Untere	Obere			
Paaren 1 Cluster1_t1_Deklarativ - Cluster1_t2_Deklarativ	,22769	,69557	,19292	-,19264	,64802	1,180	12	,261

Deklaratives Wissen: Cluster 2 (zu t2 noch nicht trainierte Module)

#### Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Cluster2_t1_Deklarativ	3,6298	13	,32029	,08883
	Cluster2_t2_Deklarativ	3,4808	13	,59241	,16430

#### Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren 1	Cluster2_t1_Deklarativ & Cluster2_t2_Deklarativ	13	,456	,117

#### Test bei gepaarten Stichproben

	Gepaarte Differenzen					T	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
				Untere	Obere			
Paaren 1 Cluster2_t1_Deklarativ - Cluster2_t2_Deklarativ	,14904	,52950	,14686	-,17094	,46901	1,015	12	,330

*Prozedurales Wissen: Cluster 1 (zu t2 trainierte Module)*

**Statistik bei gepaarten Stichproben**

	Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1 Cluster1_t1_Prozedural	3,3938	13	50323	13957
Cluster1_t2_Prozedural	4,2185	13	48492	13449

**Korrelationen bei gepaarten Stichproben**

	N	Korrelation	Signifikanz
Paaren 1 Cluster1_t1_Prozedural & Cluster1_t2_Prozedural	13	417	156

**Test bei gepaarten Stichproben**

	Gepaarte Differenzen					T	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
				Untere	Obere			
Paaren Cluster1 t1 Prozedural - Cluster1_t2_Prozedural	-,82462	,53353	,14798	-1,14703	-,50220	-5,573	12	,000

*Prozedurales Wissen: Cluster 2 (zu t2 noch nicht trainierte Module)*

**Statistik bei gepaarten Stichproben**

	Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1 Cluster2_t1_Prozedural	3,0593	13	44234	12268
Cluster2_t2_Prozedural	3,6202	13	1,24772	34605

**Korrelationen bei gepaarten Stichproben**

	N	Korrelation	Signifikanz
Paaren 1 Cluster2 t1 Prozedural & Cluster2_t2_Prozedural	13	100	744

**Test bei gepaarten Stichproben**

	Gepaarte Differenzen					T	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
				Untere	Obere			
Paaren Cluster2 t1 Prozedural - Cluster2_t2_Prozedural	-,56090	1,28125	,35535	-1,33515	21335	-1,578	12	,140

## Appendix C4: Analysen auf der Ebene Verhalten

### Ergebnisse der deskriptiven Analysen

Verhalten gegenwärtig / vor einem Jahr pro Modul

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Gegenwärtiges Verhalten_Modul1	7	2,33	4,67	3,7302	,78530
Verhalten vor 1 Jahr_Modul1	7	2,44	4,17	3,2698	,57056
Gegenwärtiges Verhalten_Modul2	7	2,20	4,70	3,7286	,79313
Verhalten vor 1 Jahr_Modul2	7	2,00	4,30	3,1429	,82028
Gegenwärtiges Verhalten_Modul3	7	2,80	4,50	3,7000	,55076
Verhalten vor 1 Jahr_Modul3	7	2,00	4,20	3,0571	,86189
Gegenwärtiges Verhalten_Modul4	7	3,33	4,33	3,8571	,37796
Verhalten vor 1 Jahr_Modul4	7	2,67	4,17	3,4524	,53328
Gegenwärtiges Verhalten_Modul5	7	3,50	4,88	4,1250	,43899
Verhalten vor 1 Jahr_Modul5	7	2,75	4,50	3,6607	,64837
Gegenwärtiges Verhalten_Modul6	7	3,00	5,00	3,9524	,77408
Verhalten vor 1 Jahr_Modul6	7	2,33	4,33	3,2381	,80999
Gegenwärtiges Verhalten_Modul7	7	1,50	4,50	3,1429	,98802
Verhalten vor 1 Jahr_Modul7	7	1,00	4,00	2,7143	,99403
Gegenwärtiges Verhalten_Modul8	7	3,33	5,00	4,1190	,58305
Verhalten vor 1 Jahr_Modul8	7	2,83	4,17	3,4762	,49468
Gegenwärtiges Verhalten_Modul9	7	3,00	4,75	3,8571	,62678
Verhalten vor 1 Jahr_Modul9	7	2,25	4,38	3,0714	,82871
Gültige Werte (Listenweise)	7				

## Ergebnisse der Signifikanzprüfung

### t-test für abhängige Stichproben

*Vergleich gegenwärtiges Verhalten / Verhalten vor einem Jahr*

#### Statistik bei gepaarten Stichproben

	Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1 MW_Gegenwärtiges_Verhalten	3,8430	7	,54751	,20694
MW_Verhalten_vor_1Jahr	3,2459	7	,65203	,24644

#### Korrelationen bei gepaarten Stichproben

	N	Korrelation	Signifikanz
Paaren 1 MW_Gegenwärtiges_Verhalten & MW_Verhalten_vor_1Jahr	7	,821	,024

#### Test bei gepaarten Stichproben

	Gepaarte Differenzen					T	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
				Untere	Obere			
Paaren 1 MW_Gegenwärtiges_Verhalten - MW_Verhalten_vor_1Jahr	,59709	,37247	,14078	,25261	,94156	4,241	6	,005

*Selbstbild-Fremdbild-Abgleich*

#### Statistik bei gepaarten Stichproben

	Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1 MW_Gegenwärtiges_Verhalten	3,8430	7	,54751	,20694
MW_prozedurales_Wissen_t3	3,5460	7	,43877	,16584

#### Korrelationen bei gepaarten Stichproben

	N	Korrelation	Signifikanz
Paaren 1 MW_Gegenwärtiges_Verhalten & MW_prozedurales_Wissen_t3	7	,428	,338

#### Test bei gepaarten Stichproben

	Gepaarte Differenzen					T	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
				Untere	Obere			
Paaren 1 MW_Gegenwärtiges_Verhalten - MW_prozedurales_Wissen_t3	,29697	,53540	,20236	-,19818	,79213	1,468	6	,193

## Appendix C5: Analysen der Transferbedingungen

### Deskriptive Analysen für die einzelnen Items

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Produktivitätssteigerung	14	1	5	3,36	1,151
Vorfreude, Erlerntes auszuprobieren	14	1	5	3,21	1,051
Bessere Arbeitsbewältigung	14	3	5	3,71	,726
Motivation zur Anwendung	14	3	5	4,07	,829
Übungen	14	3	5	3,93	,730
Kompetente Trainer	14	2	5	3,36	,929
Beispiele	14	2	5	3,79	,893
Vermittlung	14	2	5	4,00	,877
Materialien	14	2	4	3,29	,726
Methoden	14	2	4	3,29	,726
Gefallen	14	2	5	3,14	,949
Anforderungen	14	2	4	3,29	,611
Beispielsituationen	14	2	5	3,57	,852
Auswirkungen auf die Arbeit	14	1	4	2,36	1,082
Berufliche Entwicklung	14	1	5	2,71	1,204
Training	14	1	4	2,57	,852
Ergebnisse	14	1	4	2,71	,994
Tagesablauf	14	2	5	3,71	,994
Prioritäten	14	1	5	3,36	1,216
Arbeitsbelastung	14	2	5	3,57	,938
Arbeitsgeschehen	14	2	5	3,36	1,082
Budget	14	3	5	4,07	,917
Personal	14	2	5	3,79	,893
Materialien	14	3	5	4,50	,650
Ressourcen	14	2	5	3,79	,893
Möglichkeiten	14	2	5	3,93	,917
Schätzen bei Anwendung	14	1	5	3,36	1,008
Ermutigung	14	1	5	3,14	1,027
Erwartung der Anwendung	14	1	4	2,86	,864
Geduld	14	1	5	3,07	1,141
Problembearbeitung	14	1	5	2,21	1,528
Anwenden Inhalte	14	1	5	2,21	1,477
Interesse	14	2	5	3,57	,938
Lob	14	1	5	3,00	1,109
realistische Zielsetzung	14	1	5	2,43	1,399
Ablehnung	14	3	5	4,79	,579
Techniken	14	4	5	4,79	,426
Effektivität	14	3	5	4,86	,535
Hilfreich	14	3	5	4,57	,756
Kritik	14	4	5	4,93	,267
Gültige Werte (Listenweise)	14				

## Überprüfung von Zusammenhängen

### *Transferbedingungen und Ebene Lernen*

		Skala_Transferdesign	Skala_Jobfit	Skala_Wissensanwendung	Skala_Transferkapazität	Skala_Motivation_Lerntransfer	Skala_Erwartungsklarheit	Skala_UnterstuetzungKollegen	Skala_UnterstuetzungVorgesetzte	Skala_Einstellungen_Saktionen_FK
MW_t1_Erfahrung	Korrelation nach Pearson	-,283	-,200	,114	,502	,186	-,254	-,328	-,193	,499
	Signifikanz (2-seitig)	,348	,512	,712	,080	,543	,403	,273	,527	,083
	N	13	13	13	13	13	13	13	13	13
MW_t2_Erfahrung	Korrelation nach Pearson	,116	,111	,167	,691**	,614*	,447	-,004	,542	,452
	Signifikanz (2-seitig)	,705	,717	,585	,009	,025	,126	,991	,056	,121
	N	13	13	13	13	13	13	13	13	13
MW_t3_Erfahrung	Korrelation nach Pearson	-,127	-,030	-,004	,396	,539	-,181	-,099	,122	,512
	Signifikanz (2-seitig)	,679	,922	,989	,180	,057	,554	,747	,691	,074
	N	13	13	13	13	13	13	13	13	13
MW_t1_Theorie	Korrelation nach Pearson	-,166	,549	-,069	-,004	,383	-,029	-,118	-,438	,308
	Signifikanz (2-seitig)	,587	,052	,822	,989	,196	,924	,700	,135	,307
	N	13	13	13	13	13	13	13	13	13
MW_t2_Theorie	Korrelation nach Pearson	-,242	,006	-,029	,636*	,452	,015	-,012	,327	,476
	Signifikanz (2-seitig)	,426	,985	,925	,019	,121	,960	,968	,275	,100
	N	13	13	13	13	13	13	13	13	13
MW_t3_Theorie	Korrelation nach Pearson	-,026	-,199	-,210	,266	,350	-,024	-,413	,300	,596
	Signifikanz (2-seitig)	,933	,515	,490	,379	,241	,938	,161	,320	,032
	N	13	13	13	13	13	13	13	13	13

\*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant. \*\*. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Variablengruppen nach GLTSI  
 Möglichkeiten  
 Motivation  
 Umfeld

Korrelation wird auf einem Signifikanzniveau von .05 signifikant  
 Korrelation wird auf einem Signifikanzniveau von .05 knapp überschritten  
 positive Korrelation ( $r > .1$ )  
 negative Korrelation ( $r > .1$ )

*Transferbedingungen und Ebene Verhalten*

		Skala_Motivation_ Lerntransfer	Skala_Transfer- design	Skala_Jobfit	Skala_Erwartungs- klarheit	Skala_Transfer- kapazitaet	Skala_Wissens- anwendung	Skala_Unterstuet- zung_Kollegen	Skala_Unterstuet- zung_Vorgesetzte	Skala_Sanktionen_ FK	LTSI_Gesamt
MW_Gesamt_Verhalten gegenwärtig	Korrelation nach Pearson	,666	-,532	,427	-,277	,413	,445	-,266	-,148	,253	,191
	Signifikanz (2-seitig)	,102	,219	,339	,547	,357	,318	,564	,779	,628	,681
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul1	Korrelation nach Pearson	,376	-,356	,225	-,162	,563	,646	-,190	,184	-,033	,278
	Signifikanz (2-seitig)	,406	,433	,628	,729	,189	,117	,684	,727	,950	,546
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul2	Korrelation nach Pearson	,647	-,391	,134	-,560	,300	,537	-,470	-,373	,077	-,032
	Signifikanz (2-seitig)	,116	,386	,775	,191	,513	,214	,287	,467	,885	,946
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul3	Korrelation nach Pearson	,413	-,539	,505	-,557	,071	,322	-,297	-,670	,121	-,069
	Signifikanz (2-seitig)	,191	,212	,248	,194	,879	,482	,518	,146	,819	,883
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul4	Korrelation nach Pearson	,536	-,057	,000	-,185	,654	,690	-,275	,122	,181	,352
	Signifikanz (2-seitig)	,214	,904	1,000	,691	,111	,086	,551	,818	,731	,438
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul5	Korrelation nach Pearson	,606	-,441	,339	,000	,642	,413	-,086	,089	,788	,483
	Signifikanz (2-seitig)	,072	,322	,456	1,000	,120	,358	,854	,866	,063	,272
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul6	Korrelation nach Pearson	,552	-,542	,713	-,035	,268	,124	-,110	-,053	,333	,237
	Signifikanz (2-seitig)	,199	,209	,072	,941	,561	,791	,815	,921	,519	,609
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul7	Korrelation nach Pearson	,444	-,588	,536	-,188	,237	,157	-,324	-,094	,249	,073
	Signifikanz (2-seitig)	,318	,165	,215	,686	,608	,737	,478	,859	,634	,877
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul8	Korrelation nach Pearson	,194	-,461	,606	-,008	,244	,348	,114	,142	-,223	,209
	Signifikanz (2-seitig)	,677	,297	,149	,986	,598	,444	,808	,788	,671	,653
	N	7	7	7	7	7	7	7	6	6	7
Verhalten_gegenwärtig_Modul9	Korrelation nach Pearson	,705	-,560	,370	-,103	,408	,202	-,142	-,091	,627	,233
	Signifikanz (2-seitig)	,077	,072	,414	,827	,364	,664	,762	,864	,183	,616
	N	7	7	7	7	7	7	7	6	6	7



### **Ergebnisse der Reliabilitätsschätzung**

Skala	Cronbachs Alpha	Itemanzahl
Motivation zum Lerntransfer	.829	4
Erwartungsklarheit	.786	4
Sanktionen durch Vorgesetzte	.892	5
Unterstützung durch Vorgesetzte	.899	5
Unterstützung durch Kollegen	.849	4
Möglichkeiten der Wissensanwendung	.824	5
Persönliche Transferkapazität	.883	4
Transferdesign	.804	4
Trainings-Job-Übereinstimmung	.682	5



*Ebene Reaktion und Ebene Verhalten*

		Zufriedenheit_GESAMT_alle_Module
Gegenwärtiges_Verhalten	Korrelation nach Pearson	,008
	Signifikanz (2-seitig)	,987
	N	7

*Ebene Lernen und Ebene Verhalten*

Deklaratives Wissen

		Deklaratives Wissen_t1	Deklaratives Wissen_t2	Deklaratives Wissen_t2
Gegenwärtiges_Verhalten	Korrelation nach Pearson	,598	,562	,205
	Signifikanz (2-seitig)	,156	,189	,660
	N	7	7	7

Prozedurales Wissen

		Prozedurales Wissen_t1	Prozedurales Wissen_t2	Prozedurales Wissen_t3
Gegenwärtiges_Verhalten	Korrelation nach Pearson	,763	,220	,429
	Signifikanz (2-seitig)	,046	,636	,336
	N	7	7	7

\*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.