

Masterarbeit im Fach Psychologie

Evaluation der Führungsausbildung am Institut der Feuerwehr NRW

Entwicklung eines Fragebogens zur Bewertung
der Gruppen- und Zugführerlehrgänge

Vorgelegt von

Stephanie Babel

im Mai 2016

Betreuung

**PD Dr. Meinald Thielsch
Prof. Dr. Guido Hertel**

Inhaltsverzeichnis

Abbildungs- und Tabellenverzeichnis	II
Vorbemerkung zum Sprachgebrauch	III
Zusammenfassung	1
1 Einleitung	2
2 Theoretischer und Empirischer Hintergrund	3
2.1 Führungsausbildung am IdF NRW	3
2.2 Evaluation	5
2.3 Was macht einen guten Lehrgang aus?	7
2.4 Kompetenzerwerb	10
2.5 Zusammenfassende Intention dieser Arbeit	13
3 Studie I	14
3.1 Methode	14
3.2 Ergebnisse	17
3.3 Diskussion	21
4 Itemkonstruktion	24
5 Studie II	25
5.1 Methode	26
5.2 Ergebnisse	27
5.3 Diskussion und Itemauswahl	28
6 Studie III	29
6.4 Methode	29
6.5 Ergebnisse	34
6.6 Diskussion	44
7 Allgemeine Diskussion	50
7.4 Faktoren guter Lehre am IdF NRW	50
7.5 Testgüte	52
7.6 Limitationen und weitere Forschungsimplikationen	54
7.7 Fazit	55
Referenzen	57
Anhang	64
Plagiatserklärung	88

Abbildungs- und Tabellenverzeichnis

Abbildung 1	10
Tabelle 1	19
Tabelle 2	20
Tabelle 3	21
Tabelle 4	30
Tabelle 5	31
Tabelle 6	37
Tabelle 7	40
Tabelle 8	42
Tabelle 9	43

Vorbemerkung zum Sprachgebrauch

Aus Gründen der besseren Lesbarkeit wird in dieser Arbeit ausschließlich die männliche Sprachform (z. B. Teilnehmer) verwendet. Ich möchte an dieser Stelle jedoch explizit darauf hinweisen, dass diese Form geschlechtsunabhängig zu verstehen ist. Alle Inhalte beziehen sich in gleichem Maße sowohl auf Frauen als auch auf Männer. Denn auch wenn es sich bei der Feuerwehr weiterhin um eine „Männerdomäne“ handelt (Horwath, 2013), jeweils 92–96% der Lehrgangsteilnehmer am IdF NRW männlich sind (Fuchs, Zündorf, & Vöge, persönliches Gespräch, Dezember 2015) und infolgedessen auch in dieser Arbeit überwiegend Männer befragt wurden, so sind zunehmend auch Frauen aktiv und erfolgreich bei der Feuerwehr tätig (Horwath, 2013).

Zusammenfassung

Am Institut der Feuerwehr Nordrhein-Westfalen (IdF NRW) werden jährlich ca. 1500 Führungskräfte ausgebildet. In mehrwöchigen Lehrgängen werden ihnen Fach- und Führungskompetenzen vermittelt, die sie befähigen sollen, eine Einsatzmannschaft in stressigen und mitunter hochrisikoreichen Situationen sicher zu führen. Doch wie gut sind die Lehrgänge wirklich? Vor dem Hintergrund dieser Frage wird in der vorliegenden Arbeit ein Fragebogen entwickelt, mit dem sich die Güte der Lehrgänge zunächst aus Teilnehmersicht beurteilen lässt. Dies erfolgt anhand dreier aufeinanderfolgender empirischer Studien sowie unter Einbezug von Forschungsbefunden und Instrumente zur Lehrevaluation im Hochschulkontext (z. B. Rindermann, 2001). In Studie I wird herausgearbeitet, was eine gute Führungsausbildung am IdF NRW ausmacht. Hierzu werden Lehrgangsteilnehmer und Dozenten ($n = 64$) zu den aus ihrer Sicht erfolgskritischen Merkmalen eines guten Lehrgangs interviewt. Die Antworten werden inhaltsanalytisch ausgewertet und Dimensionen guter Lehre im feuerwehrtechnischen Ausbildungskontext herausgestellt. Um diese abzubilden werden erprobte Items aus bestehenden Lehrevaluationsinstrumenten sowie auch kontextspezifisch, neukonstruierte Items zusammengestellt und in Studie II von Lehrgangsteilnehmern und Dozenten ($n = 33$) hinsichtlich ihrer Verständlichkeit, Relevanz und Vollständigkeit beurteilt. Basierend auf den Beurteilungen wird eine begründete Auswahl an Items für eine Erstversion des Fragebogens zusammengestellt und in Studie III umfassend pilotiert ($n = 243$). Infolge von Item- und Verteilungsanalysen wird die Itemmenge weiter reduziert. Die verbleibenden Items werden einer exploratorischen Faktorenanalyse unterzogen, um schließlich die empirisch zugrundeliegende Faktorenstruktur des Fragebogens zu ermitteln. Dabei zeigen sich deutliche Übereinstimmungen mit den Erfolgsfaktoren guter Hochschullehre (z. B. Souvignier & Gold, 2002). Zugleich zeigen sich jedoch auch kontextspezifische Faktoren. Für den Fragebogen werden mit Blick auf die ermittelte Faktorenstruktur schließlich zehn Skalen mit jeweils drei bis fünf Items gebildet, die als wesentlich für die Bewertung von Lehrveranstaltungen am IdF NRW angesehen werden. Die Skalen werden ersten psychometrischen Untersuchungen unterzogen. Limitationen der drei Studien und Implikationen für weitere Forschung werden diskutiert.

Schlüsselwörter: Evaluation, Lehrevaluation, Feuerwehr, Gruppenführer, Zugführer, Ausbildung, Veranstaltungsbeurteilung, Qualität, Erfolgsfaktoren, Dimensionen guter Lehre

1 Einleitung

Die vorliegende Arbeit ist im Rahmen eines Kooperationsprojektes zwischen dem Institut der Feuerwehr Nordrhein-Westfalen (IdF NRW) und der Organisations- und Wirtschaftspsychologie an der Universität Münster (OWMs) entstanden. Der Impuls kam dabei vom IdF NRW, der zentralen Aus- und Fortbildungsstätte für Feuerwehrangehörige und Mitglieder von Krisenstäben des Landes NRW, mit Sitz in Münster (Homepage des IdF NRW, 2016a). Hier werden jedes Jahr mehrere hundert Lehrveranstaltungen durchgeführt, die sich sowohl an berufliche als auch an ehrenamtliche Angehörige von öffentlichen und privaten Feuerwehren richten (ebd.). Das Angebot ist vielfältig und reicht von ein- oder mehrtägigen Seminaren zu verschiedenen Spezialthemen (z. B. *Ölschadenbeseitigung auf Straßenflächen*) bis hin zu mehrwöchigen Qualifizierungslehrgängen für Ausbilder und Führungskräfte (Homepage des IdF NRW, 2016b). Letztere umfassen die Ausbildung von Führungskräften ab der dritten Führungsebene, sprich vom Gruppenführer an aufwärts.

Nachdem die Führungsausbildung am IdF NRW zuletzt im Jahr 2006¹ umfassend evaluiert und überarbeitet wurde (Löchteken & Fuchs, 2015), steht das seither etablierte Ausbildungskonzept der Gruppen- und Zugführerlehrgänge² aktuell erneut zur Diskussion. Als Reaktion auf die im vergangenen Jahr laut gewordenen Forderungen nach einer Verkürzung der Lehrgangsdauer, einer Vereinheitlichung der Ausbildung von beruflichen und ehrenamtlichen Führungskräften sowie einer methodisch flexibleren und den Praxistransfer stärker fördernden Lehrgangsgestaltung (ebd.) haben Verantwortliche am IdF NRW in den vergangenen Monaten bereits neue Unterrichtskonzepte erarbeitet. Diese werden in dem Zeitraum von Februar bis Juli 2016 in Form von Pilotlehrgängen erstmalig praktisch erprobt.

Die umfassende Begleitung dieser Pilotdurchführungen, in Form einer externen und wissenschaftlich fundierten Evaluation stellt dabei den zentralen Gegenstand des vom IdF NRW initiierten Kooperationsprojektes dar. Ziel dabei ist es, die Veränderungen zwischen dem alten und dem neuen Lehrgang zu bewerten, die Güte der neuen Lehrkonzepte zu beurteilen sowie eventuelle Schwachstellen und damit einhergehende, weiterführende Verbesserungspotenziale aufzuzeigen (Löchteken & Fuchs, 2015). Hierzu bedarf es eines fundierten Evaluationsinstruments, mit dem sich die qualitätsentscheidenden und erfolgsrelevanten Aspekte einer erfolgreichen Führungsausbildung am IdF NRW zuverlässig erfassen lassen.

Da es in der Literatur jedoch keine Hinweise auf bereits bestehende Instrumente zur Lehrgangsevaluation im feuerwehrtechnischen Ausbildungsbetrieb oder vergleichbaren Kontexten gibt, ist es Ziel und Inhalt dieser Arbeit, zunächst ein solches Instrument zu entwickeln.

¹ Ergebnisse dieser Evaluation liegen den Projektverantwortlichen des IdF aktuell nicht mehr vor und bleiben in dieser Arbeit entsprechend unberücksichtigt.

² Ein Gruppenführer ist befähigt im Einsatz eine maximal acht Mann starke Mannschaft zu leiten, während ein Zugführer berechtigt ist eine bis zu 22-köpfige Einsatzmannschaft zu führen (FwDV 3, 2008).

2 Theoretischer und Empirischer Hintergrund

Effektiv führen, wenn es brennt! – Mit diesem Ziel werden am IdF NRW jährlich etwa 1500 haupt- und ehrenamtliche Führungskräfte auf Gruppen- und Zugführerebene ausgebildet (Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalen, 2014; Sievers, 2015). In mehrwöchigen Lehrgängen erwerben die angehenden Führungskräfte Kenntnisse und Fertigkeiten, die sie befähigen Einsätze erfolgreich zu leiten (AG-BIII, 2007; AG-FIII, 2005; AG-F IV, 2007).

Dabei meint die Leitung von Einsätzen gemäß der Feuerwehrdienstvorschrift (FwDV) 100 „das gesamtverantwortliche Handeln für eine Einsatzstelle und für die dort eingesetzten Einsatzkräfte“ (FwDV 100, 2013, S. 6). Den Führungskräften der Feuerwehr kommt also insofern eine entscheidende Rolle zu, als dass sie am Einsatzort sämtliche Verantwortung tragen, sowohl für die Koordination und Steuerung aller Prozessabläufe, als auch für die Sicherheit der Gesellschaft und ihrer Einsatzmannschaft (ebd.). In zumeist kritischen, komplexen und mitunter hochrisikoreichen Situationen, welche nach Mistele (2007) durch hohe Dynamiken, Zeitdruck, Intransparenz, suboptimale Informationsversorgung sowie durch empfundenen Stress und hohe Eigengefährdung gekennzeichnet sein können, müssen sie Entscheidungen treffen, Befehle geben und hierdurch Handlungen initiieren, die die weitere Entwicklung der Einsatzsituation maßgeblich bestimmen – im positiven, aber auch im negativen Sinne (Bräuer, 2014). So können fehlerhafte Entscheidungen katastrophale Folgen für die Gesundheit und das Leben von Menschen haben oder zu einer überdurchschnittlichen Gefährdung von Tier und Umwelt führen (Dietrich & Childress, 2004).

Es liegt demnach im Interesse der Allgemeinheit, dass Führungskräfte der Feuerwehr gut ausgebildet sind, um auch in kritischen Situationen bestmögliche Entscheidungen treffen und durch ihr Handeln eine größtmögliche Sicherheit für die Gesellschaft gewährleisten zu können. Es wird erwartet, dass sie den Herausforderungen unbestimmter und dynamischer Einsatzsituationen gewachsen sind, Einsatzlagen erfassen und diese dann möglichst fehlerfrei bewältigen können (Mistele, 2007). Aufgrund dessen kommt der Qualität der Führungsausbildung am IdF NRW eine besonders hohe Bedeutung zu.

2.1 Führungsausbildung am IdF NRW

Die Fähigkeit, kritische Einsatzsituationen erfolgreich zu bewältigen, wird nach Mistele (2007) durch ein Zusammenspiel von *formellen* sowie vor allem auch *informellen Lernprozessen* ausgebildet bzw. gefördert. Formelles Lernen findet durch die systematische und organisierte Vermittlung von festgelegten Lerninhalten, zumeist in Form von explizitem Theoriewissen, statt (Dehnbostel, 2003). Bei dem informellen Lernen stellt sich hingegen ein Lernergebnis ein, „ohne dass es von vorneherein bewusst angestrebt wird“ (ebd., S. 5). Dies kann durch eine bewusste Reflexion von Erfahrungen und Handlungen geschehen oder aber durch implizite Lernprozesse, die völlig unbewusst und unreflektiert ablaufen. Während Ers-

teres in einem Erkenntnisgewinn resultiert, äußert sich Letzteres durch implizites Handlungswissen (ebd.).

Im Einklang damit fußt die Ausbildung der Gruppen- und Zugführer am IdF NRW sowohl auf formellem als auch auf informellem Lernen (vgl. Mistele, 2007). So wird den angehenden Führungskräften im ersten Teil ihrer zwei- bis dreiwöchigen Ausbildung zunächst explizites theoretisches Wissen zu verschiedenen Einsatz- und Führungsthemen (z. B. „Mitarbeiterführung“, „einsatzbezogene Führungsgrundsätze und Führungsvorgang“, „Einsatznachbesprechung“) vermittelt (AG-BIII, 2007; AG-FIII, 2005; AG-F IV, 2007). Die theoretische Wissensvermittlung erfolgt hierbei überwiegend durch Frontalunterricht³ (ebd.), visuell unterstützt durch PowerPoint-Präsentationen sowie teilweise auch durch kurze Videosequenzen. Stellenweise wird das vermittelte Wissen auch durch Simulationen an einer Planspielplatte veranschaulicht. Dabei handelt es sich um ein Hilfsmittel, mit dem sich Szenarien anhand einer Modelllandschaft darstellen lassen⁴. Ferner sollen die Lehrgangsteilnehmer ihr explizites Wissen bei Gruppenarbeiten mit anschließender Ergebnispräsentation oder -diskussionen, die während des Unterrichts regelmäßig zu verschiedenen Schwerpunktthemen angeleitet werden, gezielt erweitern bzw. vertiefen können. Um die Unterrichtsinhalte nachzubereiten und ihr explizites Theoriewissen weiter zu vertiefen, erhalten die Lehrgangsteilnehmer außerdem ein umfangreiches Nachschlagewerk (inkl. Lernunterlagen und Hintergrundliteratur), sowie regelmäßig auch Übungsaufgaben zu den im Unterricht behandelten oder weiterführenden Themen, die sie außerhalb des Unterrichtes eigenverantwortlich bearbeiten können. Hierbei handelt es sich um die sogenannten EVA-Aufgaben (wobei EVA als Abkürzung für eigenverantwortliches Arbeiten steht; ebd.).

Während der erste Teil des Lehrgangs überwiegend theoriebasiert verläuft⁵, werden im zweiten Teil realitätsnahe Einsatzsimulationen im Außengelände des IdF NRW durchgeführt (IdF NRW Ausbildungsplan –Musterplan BIII, 2012; IdF NRW Ausbildungsplan –Musterplan FIII, 2013; IdF NRW Ausbildungsplan –Musterplan FIV, 2016). Im Zuge dieser Einsatzübungen können die angehenden Gruppen- und Zugführer ihr neuerworbenes theoretisches Einsatz- und Führungswissen praktisch anwenden bzw. erproben und dabei um eigene Erfahrungen sowie (implizites) Handlungswissen erweitern. Während durch klar definierte Lernin-

³ Jeder Lehrgang wird dabei in der Regel von einem zweiköpfigen Dozententeam unterrichtet.

⁴ Die Planspielplatte kommt unter anderem auch für das *Planspiel* zum Einsatz. Dabei handelt es sich um eine Unterrichtsmethode, bei der die Lehrgangsteilnehmer unter Vorgabe einer auf der Platte dargestellten Lage verschiedene Entscheidungs- und Handlungsalternativen durchspielen und diese hinsichtlich ihrer möglichen Konsequenzen reflektieren und bewerten sollen (vgl. Internetseite des IdF NRW, 2016). Da das Planspiel nur in den alten Zugführerlehrgängen durchgeführt wird, soll es in dieser Arbeit nicht gesondert betrachtet werden und stattdessen unter dem Begriff *Planübung* subsummiert werden. Planübung meint damit allgemein die Veranschaulichung und/oder Nachbereitung von Einsatzlagen und verschiedenen Lösungsansätzen anhand der Planspielplatte, so wie es auch in den Gruppenführerlehrgängen und den neuen Zugführerlehrgängen anzutreffen ist (vgl. Fuchs et al., persönliches Gespräch, Dezember 2015).

⁵ In den neuen Lehrgängen wird es eine größere Abwechslung zwischen der theoretischen Wissensvermittlung und den praktischen Übungen geben, mit dem Ziel das theoretische Wissen unmittelbar umzusetzen zu bringen, um so den Praxistransfer zu fördern (Löchteken, 2015a).

halte und -ziele sowie dem hohen didaktischen Strukturierungsgrad weiterhin formelle Lernprozesse gefördert werden, werden durch die Einsatzübungen vordergründig auch informelle Lernprozesse angeregt (d. h. Lernen durch Erfahrung; vgl. Dehnbostel, 2003). Ähnlich den realen Einsätzen findet im Anschluss an jede Übung eine Einsatznachbesprechung statt, in der die Einsatzsituation und -handlungen gemeinsam reflektiert und bewertet werden. Die Reflexions- und Feedbackprozesse bieten dabei eine weitere Basis für erfahrungsbasiertes Lernen (vgl. Mistele, 2007).

Sowohl die alten als auch die neuen Lehrgänge werden schließlich durch eine theoretische Prüfung abgeschlossen (AG-BIII, 2007; AG-FIII, 2005; AG-F IV, 2007; Lehrgangsbeschreibung-BIII-Pilot, 2015; Lehrgangsbeschreibung-FIII-Pilot, 2015; Lehrgangsbeschreibung-FIV, 2015). In den alten Lehrgängen wird diese noch um eine praktische Prüfung ergänzt (ebd.). Die Prüfungen werden im Rahmen dieser Arbeit jedoch nicht weiter thematisiert. Vielmehr sei hierfür auf die Arbeit von Busjan (2016) verwiesen.

Doch wie gut sind die Lehrgänge –im Sinne einer erfolgreichen Ausbildung von Führungskräften am IdF NRW– tatsächlich? Um diese Frage speziell auch vor dem Hintergrund der einleitend dargestellten Neuausrichtung beantworten zu können, bedarf es einer wissenschaftlich fundierten Evaluation der alten und neuen Lehrgänge.

2.2 Evaluation

2.2.1 Definition

Was aber bedeutet Evaluation? – In der Literatur stößt man auf eine Vielzahl unterschiedlicher Definitionen des Begriffs Evaluation (vgl. Stockmann, 2002). So steht Evaluation im alltäglichen Sprachgebrauch zunächst einmal für eine „sach- und fachgerechte Bewertung“ (www.duden.de). Im wissenschaftlichen Kontext verstehen viele Autoren darunter konkreter die „systematische Untersuchung des Nutzens oder Wertes eines Gegenstandes“, wobei mit Gegenstand unter anderem auch Programme und Maßnahmen (wie Trainings, Aus- und Weiterbildungen) gemeint sind (Deutsche Gesellschaft für Evaluation e.V., 2002, S. 13). Fasst man zum Beispiel die Definition von Rindermann (2001) und die Definition von Döring und Bortz (2016) zusammen, meint wissenschaftliche Evaluation die empirische Untersuchung (d. h. Datensammlung) und systematische Analyse (d. h. Datenauswertung) einer Maßnahme, mit der Absicht, diese in ihrer Durchführung und Wirksamkeit umfassend zu bewerten sowie gegebenenfalls zu modifizieren. Die Evaluation einer Maßnahme kann dabei zu unterschiedlichen Zeitpunkten, hinsichtlich verschiedener Kriterien und auf Basis unterschiedlicher Datenquellen erfolgen (ebd.).

2.2.2 Das Vier-Ebenen-Modell von Kirkpatrick

Nach dem bekannten Modell von Kirkpatrick (1959) sollte die Evaluation eines Trainings (bzw. einer Aus- oder Weiterbildungsmaßnahme) auf vier aufeinanderfolgenden Ebenen erfolgen, wobei jede dieser Ebenen mit einem bestimmten Evaluationskriterium korrespondiert.

So gilt es auf der ersten Ebene zunächst die *Reaktion* der Teilnehmer auf das Training zu erfassen. Reaktion meint nach Kirkpatrick (1959) die subjektive, gefühlsmäßige Bewertung – unmittelbar nach Abschluss der Maßnahme. Zugrundeliegende Dimensionen stellen hierbei die empfundene Akzeptanz, Zufriedenheit oder Nützlichkeit des Trainings dar (Wirtz, 2014). Als Erhebungsinstrument für die sogenannte *Reaktionsebene* empfehlen Kirkpatrick und Kirkpatrick (2006) den Einsatz eines standardisierten Fragebogens, mit dem sich die Reaktionen der Teilnehmer auf den unterschiedlichen Dimensionen quantifizieren lassen.

Auf der zweiten Ebene (–der sogenannten *Lernebene*–) soll beurteilt werden, ob die Teilnehmer die Lerninhalte und -ziele der Maßnahme erfolgreich bewältigt haben (Kirkpatrick, 1959). Hierzu muss erfasst werden, inwiefern durch die Teilnahme an dem Training Wissen erlernt, Fähigkeiten entwickelt und Einstellungen verändert wurden (Kirkpatrick & Kirkpatrick, 2006). Eine objektive Messung des Lernzuwachses ist dabei durch einen Prä-Post-Vergleich von Wissen, Fähigkeiten und Einstellungen der Teilnehmer möglich. Dies setzt voraus, dass sowohl *vor* als auch unmittelbar *nach* dem Training umfassende und fallspezifische Leistungs- und Einstellungstests durchgeführt werden (Wirtz, 2014). Alternativ kann der subjektive Lernerfolg auf Basis von Selbsteinschätzungen der Teilnehmer oder Fremdeinschätzungen durch den Trainer oder einen Beobachter beurteilt werden (Holling, 1999; Kirkpatrick & Kirkpatrick, 2006). Hierzu werden die Teilnehmer (bzw. Trainer oder Beobachter) durch den Einsatz eines standardisierten Fragebogens am Ende des Trainings gebeten, ihren eigenen Lernfortschritt (bzw. den Fortschritt der Teilnehmer) hinsichtlich Kompetenzerwerb, Wissenszuwachs und Einstellungsänderung einzuschätzen (ebd.).

Auf der dritten Ebene (–der sogenannten *Verhaltensebene*–) soll geprüft werden, ob infolge des Trainings eine Verhaltensänderung im beruflichen Alltag stattgefunden hat (Kirkpatrick, 1959). Dahinter steht die Frage, inwieweit ein Transfer der Lerninhalte in die Praxis stattgefunden hat und das Gelernte eine entsprechende Umsetzung findet (Kirkpatrick & Kirkpatrick, 2006). Als Messverfahren eignen sich hierfür Beobachtungen am Arbeitsplatz sowie Befragungen von Vorgesetzten oder Kollegen (Holling, 1999; Kauffeld, 2006).

Auf der vierten und letzten Ebene (–der sogenannten *Resultatebene*–) gilt es schließlich zu erheben, welche Auswirkungen die Trainingsmaßnahme für die Organisation als Ganzes hat (Kirkpatrick & Kirkpatrick, 2006). Das heißt, es soll festgestellt werden, ob durch die Maßnahme übergeordnete, organisationale Ziele besser erreicht werden (z. B. im Sinne einer Kosten- oder Fehlerreduktion oder im Sinne einer Qualitäts- oder Produktivitätssteigerung; Holling, 1999).

Vor dem Hintergrund dieser vier aufeinander aufbauenden Ebenen einer umfassenden Evaluation nach Kirkpatrick (1959) soll in dieser Arbeit ein Fragebogen entwickelt werden, mit dem sich die Führungslehrgänge am IdF NRW zunächst hinsichtlich der ersten beiden Ebenen (d. h. *Reaktions-* und *Lernebene*) beurteilen lassen. Dies stellt insofern einen ersten

wichtigen Schritt im Evaluationsprojekt dar, als dass sich in der deutsch- und englischsprachigen Literatur bisher keine Hinweise auf eine dokumentierte Evaluation von Lehrgängen im feuerwehrtechnischen (oder einem vergleichbaren) Aus- und Weiterbildungskontext finden lassen und folglich weder auf ein bestehendes Instrument noch auf spezifische Erfolgskriterien zurückgegriffen werden kann. Mit dem neukonzipierten Fragebogen sollen einerseits die gefühlsmäßige Beurteilung des Lehrgangs durch die Teilnehmer (vgl. Holling, 1999) sowie andererseits auch der selbsteingeschätzte Lernerfolg standardisiert erfasst werden können.

2.3 Was macht einen guten Lehrgang aus?

Die individuelle Zufriedenheit mit der Ausbildung, die empfundene Nützlichkeit und den subjektiven Lernerfolg der Teilnehmer zu erfragen, reicht für eine umfassende Evaluation der Lehrgänge allein nicht aus (vgl. Dunkin & Barnes, 1986). Es handelt sich hierbei um Zielkriterien, mit denen sich die Güte eines Lehrgangs insgesamt bewerten, nicht aber differenziert beschreiben lässt (vgl. Rindermann, 2001). Um spezifische Stärken und Schwächen der Gruppen- und Zugführerlehrgänge zu identifizieren, gilt es vielmehr auch diejenigen Faktoren zu erfassen, die die Güte dieser Lehrgänge –und damit auch die oben genannten Zielkriterien– maßgeblich bedingen. Doch welche sind das? – An dieser Stelle stellt sich die Frage danach, was überhaupt eine gute Führungsausbildung am IdF NRW ausmacht? Oder formuliert als eine erste Forschungsfrage, die es in dieser Arbeit zu beantworten gilt:

(1a) Was sind qualitäts- und erfolgsentscheidende Faktoren eines guten Führungslehrgangs am IdF NRW?

Mit Blick auf die in Abschnitt 2.1 geschilderte methodische Ausgestaltung der Lehrgänge am IdF NRW lassen sich hier einige Ähnlichkeiten zu verschiedenen Formen universitärer Lehre erkennen. Dies bezieht sich primär auf die Theorieeinheiten (v.a. durch Frontalunterricht, Gruppenarbeiten und -diskussionen) im ersten Teil der Lehrgänge. Diese weisen zum Teil Vorlesungs- und ansonsten überwiegend Seminarcharakter auf. Die erkennbaren Ähnlichkeiten sind insofern von Interesse, als dass es eine Vielzahl fundierter Forschungsarbeiten zur Evaluation von Hochschullehre gibt (vgl. Braun & Gusy, 2006; Souvignier & Gold, 2002). So soll in dieser Arbeit auf im universitären Kontext identifizierten Bedingungsvariablen guter Lehre sowie damit einhergehend auch auf etablierte Instrumente zur Lehrveranstaltungsevaluation zurückgegriffen werden.

2.3.1 Lehrveranstaltungsevaluation im universitären Kontext

Zur Evaluation von universitären Lehrveranstaltungen existiert mittlerweile eine Vielzahl unterschiedlicher Fragebogeninventare. Eine Zusammenstellung von verschiedenen Inventare im deutschsprachigen Raum –einschließlich ihrer Entwicklungen und psychometrischen Qualitäten– findet sich in den Übersichtsarbeiten von Braun und Gusy (2006) sowie Souvignier und Gold (2002), auf die an dieser Stelle zunächst verwiesen sei.

Mit Blick auf die Dimensionen, die den etablierten Fragebogeninventaren jeweils zugrunde liegen und anhand derer die Qualität der Lehrveranstaltungen beurteilt werden soll, lassen sich gewisse Unterschiede zwischen den einzelnen Fragebögen erkennen. So unterscheiden sich einige Inventare zunächst darin, wie viele Dimensionen sie in welchem Umfang erfassen (vgl. Zumbach, Spinath, Schahn, Friedrich, & Kögel, 2007). Während beispielsweise dem *Heidelberger Inventar zur Lehrveranstaltungsevaluation II* (HILVE II; Rindermann, 2001) 20 Dimensionen zugrunde liegen, die durch insgesamt 51 Items abgebildet werden, umfassen andere Inventare, wie das *Kommunikationsinstrument für die Evaluation von Lehrveranstaltungen* (KIEL; Gediga et al., 2000), der *Fragebogen zur Evaluation von Vorlesungen* (FEVOR; Staufenbiel 2000) sowie auch der *Münsteraner Fragebogen zur Evaluation von Seminaren* (MFE-Sr; Thielsch & Hirschfeld, 2012^a) hingegen nur jeweils vier Dimensionen, die insgesamt durch 14 bis 22 Items dargestellt werden. Darüber hinaus setzt sich der *Münsteraner Fragebogen zur Evaluation von Vorlesungen* (MFE-Vr; Thielsch & Hirschfeld, 2012^b) im Kern nur aus drei Dimensionen und 19 Items zusammen, die jedoch bedarfsgemäß um verschiedene Zusatzskalen (z. B. zu Gruppenarbeiten, Hausaufgaben und/ oder Moderationen) ergänzt werden können.

Die Fragebögen unterscheiden sich dabei nicht nur hinsichtlich ihres Umfangs, sondern auch in der Berücksichtigung und Gewichtung verschiedener Aspekte von Lehrveranstaltungsqualität (vgl. Souvignier & Gold, 2002) sowie in der Benennung einzelner Skalen. Um die Qualität von Lehrveranstaltungen zu beurteilen, erfasst der FEVOR beispielsweise die Aspekte „Planung und Darstellung“, „Interessantheit und Relevanz verdeutlichen“, „Umgang mit Studierenden“ sowie „Schwierigkeit und Umfang“ (Staufenbiel, 2000). Beim KIEL liegt der Schwerpunkt hingegen auf den Aspekten „Verständlichkeit“, „Gliederung der Veranstaltung“, „Bewertung der Materialien“ sowie „Interesse der Studierenden“ (Gediga et al., 2000), während der MFE-Sr nach den Aspekten „Dozent und Didaktik“, „Überforderung“, „Teilnehmer“ und „Materialien“ fragt (Thielsch & Hirschfeld, 2012a). Für eine vollständige Übersicht über die für diese Arbeit relevanten Inventare sowie deren Skalen sei auf den Anhang A verwiesen.

Die Psychologen Souvignier und Gold (2002) kommen in ihrer Übersichtsarbeit nach einer Sichtung deutschsprachiger Lehrevaluationsinstrumenten zu dem Schluss, dass trotz unterschiedlicher Skalen insgesamt jedoch

„ein breiter Konsens darüber besteht, welche Faktoren als wesentlich zur Bewertung der Qualität universitärer Lehre angesehen werden: den Kern bildet das *Dozentenhandeln* – differenziert nach didaktischen Maßnahmen und sozial-interaktiven Verhaltensweisen; es wird um Angaben zu den *Inhalten* und zum *Anforderungsniveau* von Lehrveranstaltungen, zu den *organisatorischen Rahmenbedingungen* und zum *studentischen Lernverhalten* ergänzt“ (S. 265).

Hierzu ordnen sie die unterschiedlichen Skalen der etablierten Inventare jeweils einem dieser Faktoren zu (siehe hierzu Tabelle 2 in Souvignier & Gold, 2002).

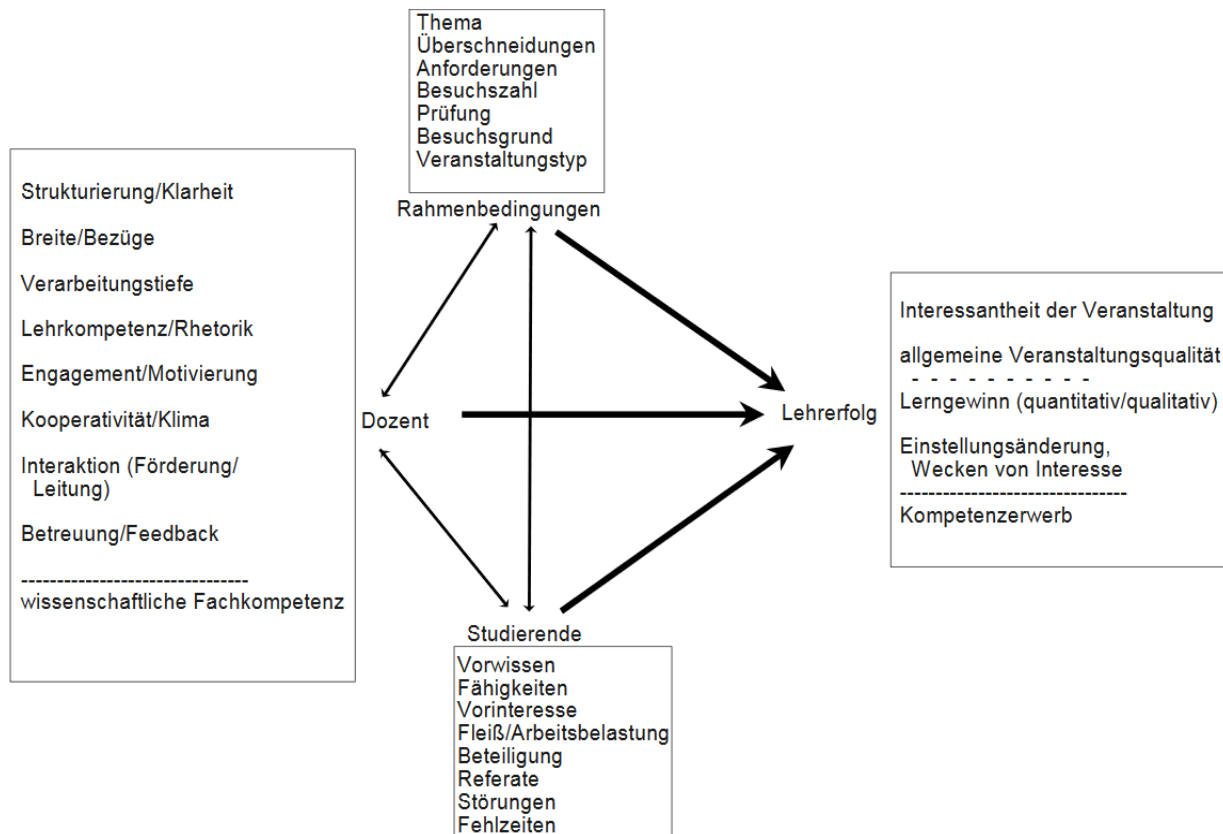


Abbildung 1. Multidimensionales Bedingungsmodell des Lehrerfolgs (aus Rindermann 2001).

Die von Souvignier und Gold (2002) herausgestellten Faktoren spiegeln dabei auch die Hauptkomponenten des viel zitierten *Multidimensionalen Bedingungsmodells des Lehrerfolgs* von Rindermann (1998, 2001, 2003) wider. Das Modell von Rindermann stellt insofern eine wichtige Grundlage dieser Arbeit dar, als dass es eine gute Übersicht über verschiedene Bedingungsvariablen guter Lehrveranstaltungen bietet (siehe Abbildung 1), auf die im Zuge der Fragebogenkonstruktion gegebenenfalls zurückgegriffen werden soll. Das Modell gründet in der Annahme, dass gute Lehrveranstaltungen, die einen hohen Lehrerfolg haben, aus einem komplexen Zusammenspiel von Determinanten seitens des Lehrenden sowie der Studierenden resultieren und von äußeren Rahmenbedingungen beeinflusst werden (ebd.). Lehrerfolg wird dabei als ein mehrdimensionales Konstrukt verstanden, das sowohl die Qualität und Interessantheit der Veranstaltung als auch den Wissenszuwachs, die Einstellungsänderung und den Kompetenzerwerb auf Seiten der Studierenden umfasst (Rindermann, 2003).

Zu den erfolgsentscheidenden Determinanten guter Lehre zählen auf Seiten des Lehrenden in erster Linie Struktur und Klarheit. Hinzu kommen „verständliche Erklärungen und Darstellungen, rhetorische Kompetenz, Motivierung, Verarbeitungstiefe, Zeit- und Schwierigkeitsmanagement, Engagement, Kooperativität/Kommunikation und in Seminaren das Leiten von Diskussionen“ (Rindermann, 2003, S. 236). Des Weiteren begünstigen „vorgebildete[n], interessierte[n] und engagierte[n] Studenten sowie günstige Rahmenbedingungen in Form von ausreichend Zeit, Platz, adäquater Anforderungshöhe und inhaltsangemessenen Veran-

staltungsformen (den) Lernerfolg“ (Rindermann, 2001, 64). Dabei interagieren die einzelnen Determinanten stets miteinander und wirken folglich nicht nur direkt sondern auch indirekt auf den Erfolg der Lehrveranstaltung. So kann zum Beispiel die wahrgenommene Anforderung sowohl durch die Erklärungs- und Darstellungsweise des Dozenten als auch durch das Vorwissen der Studenten positiv oder negativ beeinflusst werden (ebd.). Eine vollständige Auflistung aller Determinanten zeigt Abbildung 1.

Da die Führungslehrgänge am IdF NRW in ihrer Ausgestaltung und Durchführung große Ähnlichkeiten zu den Seminaren und Vorlesungen im universitären Kontext aufweisen (vgl. Abschnitte 2.1 und 2.3), stellt sich an dieser Stelle die Frage, ob und inwiefern die von Rindermann (1998, 2001, 2003) zusammengetragenen Bedingungsvariablen des Lehrerfolgs auch für den feuerwehrtechnischen Ausbildungsbetrieb am IdF NRW zutreffend sind. Denn sollte es hierbei Übereinstimmungen geben, kann bei der Fragebogenkonstruktion gezielt auf validierte Skalen oder entsprechende Items etablierter Lehrevaluationsinventare zurückgegriffen werden (siehe Anhang A). Demnach wird die erste Forschungsfrage dieser Arbeit um eine weitere Frage ergänzt, die es ebenfalls im ersten Schritt der Fragebogenkonstruktion, sprich noch vor der Zusammenstellung der Items, zu beantworten gilt:

(1b) Gibt es Gemeinsamkeiten zwischen den im universitären Kontext identifizierten Bedingungsvariablen erfolgreicher Lehre und den erfolgskritische Faktoren eines guten Führungslehrgangs am IdF NRW? Und wenn ja, welche sind das bzw. wo liegen gegebenenfalls auch Unterschiede?

2.4 Kompetenzerwerb

Neben den erfolgskritischen Faktoren eines guten Lehrgangs, soll mit dem neukonzipierten Fragebogen auch der selbsteingeschätzte Lernerfolg der Teilnehmer erfasst werden (vgl. Lernebene nach Kirkpatrick, 1959; Abschnitt 2.2.2). Dies ist bei bestehenden Evaluationsinstrumenten im Allgemeinen nicht der Fall. So werden „für die Messung der Qualität einer Lehrveranstaltung typischerweise Prozessdaten, kaum Strukturdaten und so gut wie keine Ergebnisdaten erhoben“, wie Braun, Gusy, Leidner und Hannover (2008, S. 4) in ihrer Arbeit bemängeln. Während Prozessdaten den gestalterischen Verlauf einer Veranstaltung und damit primär auch das Dozentenhandeln abbilden, beziehen sich Strukturdaten in erster Linie auf die materielle Ausstattung sowie auf weitere Kontextvariablen der Lehrveranstaltung (siehe Braun et al., 2008 für Beispielitems). Ergebnisdaten beschreiben hingegen „den Outcome, d. h. den Erfolg einer Lehrveranstaltung durch (den) selbsteingeschätzten Kompetenzerwerb der Studierenden“ (ebd., S. 4).

Der selbsteingeschätzte Lernerfolg soll in dieser Arbeit als eine Facette des Lehrerfolgs (vgl. Abschnitt 2.3.1.; Rindermann, 2003) unter anderem auch zu Validierungszwecken ausführlich erfasst werden. So sollen die Teilnehmer neben einer allgemeinen Einschätzung

ihres Lernerfolges (in Form eines Wissenszuwachses) auch die wahrgenommene Weiterentwicklung spezifischer Kompetenzen beurteilen. Für den universitären Kontext entwickelten Braun und Kollegen (2008) hierfür das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen (BEvaKomp). Mit Blick auf die hier erfassten Kompetenzen lässt sich das BEvaKomp jedoch nicht ohne weiteres auf den feuerwehrtechnischen Ausbildungsbetrieb übertragen. Vielmehr müssen hier andere Kompetenzen berücksichtigt werden – *aber welche?*

Eine Antwort auf diese Fragen soll aus der Arbeit von Mistele (2007) abgeleitet werden. In seiner Dissertation identifiziert er sogenannte verlässlichkeitsbeeinflussende Faktoren, die es Organisationen, die in ihrer Arbeit zumeist mit „dynamischen, unsicheren Umwelten, komplexen Aufgaben und unerwarteten, unvorhersehbaren Aufgaben und Situationsveränderungen konfrontiert sind“, ermöglichen, zuverlässig zu entscheiden und zu handeln (ebd., S. 33). Hierzu zählt unter anderem auch die Feuerwehr – und dabei speziell die Führungskräfte, die in kritischen Einsatzsituationen weitreichende Entscheidungen verantwortungsvoll treffen und anleiten müssen. Bei den verlässlichkeitsbeeinflussenden Faktoren nach Mistele (2008) handelt es sich größtenteils um solche Fähigkeiten, die es im Zuge der Führungsausbildung am IdF NRW zu fördern gilt (vgl. AG-BIII, 2007; AG-FIII, 2005; AG-F IV, 2007). Vor diesem Hintergrund eignet sich ein Großteil der von Mistele identifizierten Faktoren für die Erfassung des subjektiven Lernerfolges im Sinne eines Kompetenzerwerbs. Im Folgenden sollen nun diejenigen Faktoren dargestellt werden, die bei der Konzeption des Fragebogens berücksichtigt werden sollen.

2.4.1 Verlässlichkeitsbeeinflussende Faktoren

Einen ersten verlässlichkeitsbeeinflussenden Faktor stellt die frühzeitige und ganzheitliche *Wahrnehmungsfähigkeit* von Umweltsituationen dar (Mistele, 2007). So müssen speziell die Führungskräfte der Feuerwehr in der Lage sein, Gefahrenlagen sowie „Situationsveränderungen möglichst ganzheitlich und schnell zu erfassen [und] zu verarbeiten“, um fehlerfreie Handlungsentscheidungen zu treffen (ebd., S. 166). Damit einhergehend stellt die erkenntnisgesteuerte *Entscheidungsfähigkeit* einen weiteren verlässlichkeitsbeeinflussenden Faktor dar (ebd.). Da wichtige Handlungsentscheidungen zumeist unter einem extremen Zeitdruck getroffen werden müssen, ist es in der Regel nicht möglich, zunächst „systematisch verschiedene Handlungsalternativen aufzustellen, zu bewerten und sich für eine dieser Alternativen zu entscheiden“ (ebd., S. 80). Vielmehr werden Entscheidungen erkenntnisgesteuert getroffen, das heißt, dass neue oder unbekannte Situationen an Hand bestimmter Merkmale als bekannt identifiziert und entsprechend eingeordnet werden (vgl. Mustererkennung nach Klein, 2003). Damit erhalten Situationen einen prototypischen Charakter, aus dem sich erfolgsversprechende Handlungsentscheidungen ableiten lassen (ebd.).

Des Weiteren stellen eine gute Selbsteinschätzungsfähigkeit einerseits sowie eine hohe Motivation und organisationales Commitment andererseits weitere verlässlichkeitsbeeinflussende Faktoren dar (Mistele, 2007). *Selbsteinschätzungsfähigkeit* meint dabei, die eigenen Grenzen zu kennen (z. B. hinsichtlich Ermüdung, Stress oder Belastung) und sich dieser auch in fordernden Einsatzsituationen bzw. bei schwierigen Handlungsentscheidungen stets bewusst zu sein (ebd.). Der Faktor *Motivation und organisationales Commitment* meint hingegen eine tiefe innere Überzeugung und ein großes Interesse an den Aufgaben und Zielen der Organisation, sprich der Feuerwehr (ebd.). Dies äußert sich wiederum in einer hohen Identifikation mit der Organisation und impliziert eine hohe Leistungsbereitschaft (vgl. Van Dick, 2004). Dieser Faktor ist vor allem auch deshalb interessant, da es sich bei einem Großteil der am IdF NRW jährlich ausgebildeten Führungskräfte um ehrenamtlich tätige Feuerwehrmänner handelt (Lehrgangskatalog IdF NRW, 2016), die am Einsatzort die gleiche Verantwortung tragen wie ihre hauptamtlichen Kollegen, dafür jedoch nicht gleichermaßen monetär entlohnt werden.

Um die Gruppe oder den Zug auch in kritischen Situationen effektiv zu führen, bedarf es außerdem einer ausgeprägten *Kooperationsfähigkeit*. Mistele (2007) subsummiert unter diesem verlässlichkeitsbeeinflussenden Faktor unter anderem *Kommunikations-* sowie *Teamfähigkeit*. Die Führungskraft muss ein gutes Verständnis davon haben, wie die Einsatzmannschaft bestmöglich zusammenarbeitet und welche Rolle ihr selbst dabei zukommt. Sie muss verstehen, welche Informationen die Einsatzmannschaft von ihr benötigt und muss in der Lage sein, klare und verständliche Anweisungen erteilen zu können (i.S. der Befehlsgebung; vgl. hierzu weiterführend Sievers, 2015).

Damit einhergehend benennt Mistele (2007) die *Führungs- und Managementfähigkeit* als einen weiteren verlässlichkeitsbeeinflussenden Faktor. Führen im Einsatz bedeutet gemäß der FwDV 100 „die Einflussnahme auf die Entscheidungen und das Verhalten anderer Menschen mit dem Zweck, mittels steuerndem und richtungsweisendem Einwirken vorgegebene und aufgabenbezogene Ziele zu verwirklichen. Das bedeutet, andere zu veranlassen, das zu tun, was zur Erreichung des gesetzten Zieles erforderlich ist“ (S. 6). Im übertragenen Sinne bedeutet Führung im Einsatz demnach steuern und koordinieren. Im Einklang damit subsummiert Mistele (2007) unter diesem Faktor einerseits *Planungs- und Koordinationsfähigkeit* sowie andererseits *Assertiveness*, das heißt „die Fähigkeit seine Meinung und Situationseinschätzung gegenüber anderen [...] kundzutun und zu vertreten“ (S. 92).

Einen weiteren verlässlichkeitsbeeinflussenden Faktor stellt das *arbeitsbezogene erfahrungsbasierte Lernen* dar (Mistele, 2007). Mit dem Ziel, die eigene Leistungsfähigkeit stetig zu verbessern, fußt diese Form des Lernens auf bewussten Reflexions- und Feedbackprozessen. So werden im Rahmen von Einsatznachbesprechungen erlebte Situationen, Handlungen, Probleme sowie eventuelle Fehler mit der Mannschaft reflektiert, analysiert und evaluiert, um so gemeinsam aus diesen zu lernen (ebd.). Führungskräfte sollten den Nutzen ausführlicher Reflexions- und Feedbackprozesse im Zuge der Einsatznachbesprechungen

kennen und diese nutzen, um sich aktiv Rückmeldung von ihrer Mannschaft einzuholen und erfahrungsbasierte Lernprozesse zu initiieren (ebd.; vgl. Ott, 2013).

2.5 Zusammenfassende Intention dieser Arbeit

Nachdem also zunächst die erfolgskritischen Faktoren einer guten Führungsausbildung am IdF NRW (vgl. Forschungsfrage 1a; Abschnitt 2.2.2) ermittelt und mit den Bedingungsvariablen erfolgreicher Hochschullehre abgeglichen werden (vgl. Forschungsfrage 1b; Abschnitt 2.3.1), soll in dieser Arbeit sodann ein Fragebogen entwickelt werden, mit dem sich die Güte der Gruppen- und Zugführerlehrgänge zuverlässig beurteilen lässt. Dazu soll neben Informationen seitens des IdF NRW (vgl. Abschnitt 2.1) auf die verlässlichkeitsbeeinflussende Forschung von Mistele (vgl. Abschnitt 2.4.1) und –soweit möglich– auf bestehende Lehrevaluationsinventare (vgl. Abschnitt 2.3.1; Anhang A) zurückgegriffen werden. Der konzipierte Fragebogen soll in dieser Arbeit zudem erprobt und hinsichtlich der klassischen Haupt- und Nebengütekriterien beurteilt werden (z. B. Moosbrugger & Kelava, 2008). Diesbezüglich ergeben sich abschließend die beiden Forschungsfragen:

(2a) Erfüllt der entwickelte Fragebogen die zentralen Gütekriterien Reliabilität und Validität?

(2b) Erfüllt der entwickelte Fragebogen zudem auch wesentliche Nebengütekriterien wie Verständlichkeit, Durchführungsökonomie und Zumutbarkeit?

2.5.1 Multimethodales Studiendesign

Zu diesem Zweck wurden in der vorliegenden Arbeit drei aufeinanderfolgende Studien durchgeführt. In der ersten Studie (**Studie I**; vgl. Abschnitt 3) wurden Dozenten und Lehrgangsteilnehmern am IdF NRW zu den aus ihrer Sicht erfolgskritischen Merkmalen einer guten Führungsausbildung am IdF qualitativ befragt. Basierend auf den Ergebnissen wurde – unter Rückgriff auf bereits vorhandene sowie selbstgenerierte Items – eine erste Version des Fragebogens erstellt. In der zweiten Studie (**Studie II**; vgl. Abschnitt 4) wurden die zusammengetragenen Items im Hinblick auf Relevanz, Vollständigkeit und Verständlichkeit überprüft. Hierzu wurde der Fragebogen den Dozenten sowie einigen Lehrgangsteilnehmern am IdF NRW vorgelegt, um diesen hinsichtlich der genannten Kriterien zu beurteilen. Auf Basis der Rückmeldungen wurde die Erstversion entsprechend modifiziert. Die dritte Studie (**Studie III**; vgl. Abschnitt 5) stellt schließlich die Pilotierung des Fragebogens in mehreren alten sowie neuen Gruppen- und Zugführerlehrgängen am IdF NRW dar. Basierend auf den Ergebnissen wurde zunächst eine Itemselektion durchgeführt. Anschließend wurde die empirisch zugrundeliegende Faktorenstruktur mittels einer exploratorischen Faktorenanalyse ermittelt, entsprechende Skalen gebildet und erste Validierungen durchgeführt.

3 Studie I

Ziel dieser Studie war es, die zentralen Bedingungsvariablen guter Führungsausbildung am IdF NRW zu erfassen (→ Forschungsfrage 1a) sowie basierend darauf, entsprechende Gemeinsamkeiten und Unterschiede zu den von Rindermann (2001) postulierten Erfolgsfaktoren guter Hochschullehre herauszustellen (→ Forschungsfrage 1b). Einhergehend mit der Aussage von Rindermann, dass „der direkteste Weg, Kennzeichen guter Lehre zu finden, [...] darin [bestehe], Dozenten und Studenten, Personen also, die tagtäglich gute oder schlechte Lehre, Vorträge und Veranstaltungen erleben, zu befragen“ (S. 64), wurden hierfür zunächst qualitative Befragungen von Dozenten und Lehrgangsteilnehmern am IdF NRW durchgeführt.

3.1 Methode

3.1.1 Stichprobe

Im Zuge der Befragung wurden zunächst drei Teilnehmer eines Zugführerlehrgangs (100% männlich) sowie fünf Dozenten und Prüfer (100% männlich) am IdF NRW interviewt. Die interviewten Lehrgangsteilnehmer waren zwischen 40 und 50 Jahre alt ($MW = 43,7$; $SD = 5,5$). Das Alter der interviewten Dozenten und Prüfer variierte zwischen 33 und 56 Jahren ($MW = 42,2$; $SD = 8,8$).

Zusätzlich nahmen 56 Lehrgangsteilnehmer (100% männlich) an der Befragung teil, indem sie ihre Meinung durch das Ausfüllen eines standardisierten Fragebogens kundtaten (vgl. Abschnitt 3.1.2). Von ihnen absolvierten 55% eine Gruppen- und 45% einen Zugführerlehrgang. Das Alter der Teilnehmer lag dabei zwischen 23 und 50 Jahren ($MW = 33,9$; $SD = 7,4$).

Insgesamt wurden in dieser Studie demnach $n = 64$ Personen aus verschiedenen Zielgruppen und mittels unterschiedlicher Verfahren zu den erfolgskritischen Merkmalen guter Lehre am IdF NRW befragt. Alle Personen nahmen freiwillig teil.

3.1.2 Material und Durchführung

In dieser Studie kamen sowohl teilstrukturierte Interviews als auch ein standardisierter Fragebogen zum Einsatz. Während in beiden Fällen im Kern dieselben Fragen gestellt wurden, sollte die Interviewmethode eine flexiblere und stellenweise ins Detail gehende Erfassung erfolgsrelevanter Determinanten ermöglichen. Zudem sollten sprachliche oder inhaltliche Verständnisprobleme erkannt und geklärt werden können. Da aus zeitlichen Gründen jedoch nur eine begrenzte Anzahl an Interviews mit Lehrgangsteilnehmern geführt werden konnte (vgl. Abschnitt 3.1.1) wurde ergänzend auf die Fragebogenmethode zurückgegriffen. Durch den Einsatz des standardisierten Fragebogens sollte eine größere Anzahl an Teilnehmern, aus zum Teil unterschiedlichen Lehrgängen sowie mit durchaus verschiedenen Erfahrungen und Ansichten auf eine zeitlich flexiblere und ökonomische Weise befragt wer-

den (vgl. Schmidt-Atzert & Amelang, 2012). Ziel dabei war es, die individuellen Erfahrungen und Konzepte guter Lehrgänge möglichst teilnehmerübergreifend und breit zu erfassen.

Während die Interviews zu festen, vorab vereinbarten Terminen stattfanden (d. h. bei den Lehrgangsteilnehmern jeweils am letzten Tag ihrer zweiwöchigen Ausbildung und bei den Dozenten je nach ihrer zeitlichen Verfügbarkeit) und 15 bis 20 Minuten andauerten⁶, wurden die Fragebögen am Ende der ersten Ausbildungswoche in den verschiedenen Lehrgänge verteilt. Die Teilnehmer wurden gebeten, den Bogen innerhalb der nächsten Tage auszufüllen und am letzten Tag der Ausbildung in gesammelter Form an den Lehrgangsleiter zurückzugeben. Die Befragungszeitpunkte wurden bewusst so gewählt, dass die Teilnehmer in beiden Fällen gegebenenfalls auch auf Erfahrungen aus dem von ihnen besuchten Führungslehrgang zurückgreifen konnten.

In Anbetracht dessen, dass durch die Interviews und den Fragebogen im Wesentlichen dieselbe Befragung realisiert wurde, wird auf eine getrennte inhaltliche Darstellung von Interviewleitfaden und Fragebogen verzichtet. Für eine differenzierte Darstellung der Materialien sei auf Anhang B (Leitfaden) und Anhang C (Fragebogen) verwiesen. Stattdessen sollen im Folgenden Aufbau und Inhalte der Befragung methodenübergreifend dargestellt werden.

Inhalt der Befragung. Die Befragung bestand im Wesentlichen aus zwei Teilen. Im ersten Teil wurden die Teilnehmer und Dozenten gebeten, sich an eine besonders gute oder eine besonders schlechte Lehrveranstaltung zurückzuerinnern, an der sie irgendwann zuvor einmal teilgenommen oder mitgewirkt haben. Sie wurden instruiert, sich hierfür gezielt einige Minuten Zeit zu nehmen, um sich die Veranstaltung umfassend ins Gedächtnis zu rufen und darüber nachzudenken, warum diese in ihren Augen besonders gut oder besonders schlecht war. Ihre Überlegungen konnten die Befragten mit dem Interviewer teilen bzw. in einem dafür vorgesehen Feld verschriftlichen⁷. Basierend auf ihren Erfahrungen sollten die Teilnehmer und Dozenten anschließend die sieben wichtigsten Aspekte benennen, die aus ihrer Sicht eine erfolgreiche Lehrveranstaltung bedingen (vgl. Grabbe, 2003). In diesem ersten Teil der Befragung wurde also zunächst offen und erfahrungsbasiert nach den erfolgskritischen Aspekten eines guten Lehrgangs gefragt. Hierdurch sollte sichergestellt werden, dass der Blickwinkel der befragten Teilnehmer und Dozenten nicht vorzeitig eingeschränkt bzw. ver-

⁶ An dieser Stelle sei angemerkt, dass sich sowohl das Interview als auch der Fragebogen insgesamt aus zwei Blöcken zusammensetzte. Im ersten Block wurden die Dozenten und Teilnehmer zu den Merkmalen guter Lehre befragt, was den Gegenstand dieser Studie darstellt und im Folgenden aufgezeigt wird. Im zweiten und unmittelbar daran anschließenden Block wurden die Dozenten und Teilnehmer zu den Merkmalen einer guten *Prüfung* befragt. Diese Befragung wurde von Jessica Busjan (2016) durchgeführt und ist Gegenstand ihrer Masterarbeit, auf die an dieser Stelle verwiesen sei. Die beiden Befragungen wurden aus ökonomischen Gründen in Kombination durchgeführt. Da der zweite Block ebenfalls etwa 15 Minuten andauerte, belief sich die Gesamtdauer des Interviews somit auf etwa 30-35 Minuten.

⁷ Diese Frage sollte in erster Linie dem Einstieg in die Thematik dienen, sodass die Antworten im Rahmen der Auswertung nur sekundär betrachtet wurden. Das heißt, die Schilderungen wurden nicht explizit ausgewertet, sondern lediglich im Hinblick auf relevante Aspekte gescannt, die durch die sonstigen Ergebnisse noch nicht abgebildet wurden.

zerrt wurde, sondern vielmehr in alle Richtungen gedacht wurde und infolgedessen auch alle relevanten Aspekte zur Sprache kamen.

Im zweiten Teil der Befragung wurde der Fokus dann gezielt auf die von Rindermann (2001) postulierten Dimensionen erfolgreicher Lehre gerichtet (siehe Abbildung 1). Vor dem Hintergrund dieses Modells wurden die Teilnehmer und Dozenten gefragt, wodurch sich einerseits ein guter Dozent auszeichnet. Ergänzend wurden in den Interviews konkret auch Aspekte eines guten Lehrstils erfragt (vgl. Grabbe, 2003; siehe Anhang B). Des Weiteren wurde danach gefragt, wie andererseits auch die Teilnehmer den Erfolg einer Lehrveranstaltung beeinflussen und welche Rahmenbedingungen erfolgsbegünstigend sind. In den Interviews sollte auch erfasst werden, was die zentralen Lernziele sind, die es im Rahmen der Gruppen- bzw. Zugführerausbildung zu erreichen gilt. Abschließend erhielten alle Befragten die Möglichkeit weitere Aspekte zu ergänzen, die bis dahin noch nicht benannt wurden, welche im Zuge einer Lehrgangsevaluation aber durchaus relevant sein könnten – entweder weil sie den Lehrerfolg direkt oder indirekt beeinflussen oder weil sie das Urteil eines Lehrgangsteilnehmers über die Qualität der Lehre denkbar verzerren können (= Biasvariablen; vgl. Grabbe, 2003).

3.1.3 Datenaufbereitung und -analyse

Auf eine Audioaufzeichnung der Interviews wurde aufgrund verschiedener Überlegungen (z. B. Anonymität) verzichtet. Stattdessen wurden die von den Dozenten und Teilnehmern schlagwortartig genannten Merkmale während des Interviews wörtlich mitgeschrieben. Längere Aussagen wurden in Anlehnung an die Technik der *Zusammenfassung und induktiven Kategorienbildung* im Zuge einer qualitativen Inhaltsanalyse nach Mayring (2015, S. 69 ff.) zunächst paraphrasiert, generalisiert und auf ihre Kernaussagen reduziert. In einem weiteren Schritt wurde dann für jede Frage (bzw. Dimension) ein an den Ähnlichkeiten der Antworten orientiertes Kategoriensystem erstellt (vgl. Abschnitt 3.2). Die Kategoriensysteme wurden anhand der aus den Fragebögen gewonnenen Antworten hinsichtlich Vollständigkeit und Generalisierbarkeit überprüft.

Schließlich wurden alle Antworten, sowohl aus den Interviews als auch aus den Fragebögen, den jeweiligen Kategorien zugeordnet und entsprechende Häufigkeitsverteilungen berechnet. Hierfür wurden die Daten vorab in Microsoft Excel tabellarisch aufbereitet (siehe digitaler Anhang). Die Zuordnung wurde dabei einerseits von der Verfasserin der vorliegenden Arbeit vorgenommen. Um adäquate Interrater-Reliabilitäten (d. h. Beurteiler-Übereinstimmungen) für die inhaltsanalytischen Auswertungen sicherzustellen, wurden die Aussagen andererseits auch durch eine unabhängige, zweite Person kategorisiert. Die zweite Person war ebenfalls Psychologin (B.Sc.) und mit der Methode der qualitativen Inhaltsanalyse nach Mayring (2015) bereits gut vertraut. Inhaltlich wurde sie von der Verfasserin auf ihre Aufgabe vorbereitet, indem ihr die Verfasserin die einzelnen Kategorien anhand von Beispielen erläu-

terte. Zudem wurden ihr Hintergrundinformationen zur Führungsausbildung am IdF NRW gegeben, um die Angaben besser verstehen und einordnen zu können. Offene Fragen wurden im Vorfeld geklärt.

Nachdem die beiden Beurteilerinnen die Antworten jeweils unabhängig voneinander den Kategorien zugeordnet hatten, wurden ihre Zuordnungen miteinander verglichen. Um die Übereinstimmung der beiden Beurteilerinnen zu erfassen, wurde –da es sich um nominalskalierte Daten handelt– für alle Dimensionen der Koeffizient *Cohens Kappa* κ berechnet (Döring & Bortz, 2016; Cohen, 1960). Alle Berechnungen wurden mit dem Statistikprogramm IBM SPSS Statistics Version 23 durchgeführt.

3.2 Ergebnisse

Die hier berichteten Ergebnisse beziehen sich, sofern nicht anders geschrieben, stets auf die gesamte Stichprobe (vgl. Abschnitt 3.1.1). Ergänzend finden sich im digitalen Anhang für jede Dimension jeweils auch dozenten-, gruppen- sowie zugführerspezifische Häufigkeiten und Relevanzwerte.

3.2.1 Erfolgskritische Aspekte guter Lehre

Auf die Frage nach den sieben wichtigsten Aspekten guter Lehre (→ Forschungsfrage 1a) benannten die befragten Dozenten im Mittel jeweils 5,0 ($SD = 1,9$) und die Lehrgangsteilnehmer im Mittel jeweils 5,2 erfolgskritische Aspekte ($SD = 1,8$). Insgesamt ergaben sich somit 330 Aspekte, die für die induktive Kategorienbildung herangezogen werden konnten (siehe digitaler Anhang). Angelehnt an das von Mayring (2015) vorgeschlagene induktive Vorgehen wurden schließlich fünf Kategorien gebildet, denen die Aspekte jeweils zugeordnet werden sollten. Die von den beiden Beurteilern unabhängig voneinander vorgenommenen Zuordnungen wiesen eine zufallsbereinigte Übereinstimmung von $\kappa = 0,85$ auf. Dieser Wert spricht nach einer Klassifikation von Greve und Wentura (1997) für eine gute bis ausgezeichnete Interrater-Reliabilität (da $> 0,75$; siehe S. 111).

Knapp die Hälfte (d. h. 48%) der zusammengetragenen Argumente wurde der ersten Kategorie –*Dozent*– zugeordnet. Diese Kategorie umfasst verschiedene Verhaltensweisen und Merkmale der Dozenten, die aus Sicht der Befragten den Erfolg der Lehre beeinflussen. Darunter fallen in erster Linie didaktische Maßnahmen. Die Dozenten müssen die Inhalte klar, verständlich sowie möglichst praxisnah vermitteln können. Sie müssen abwechslungsreich und flexibel in der Wahl der Methode sein und über ein gutes Zeit- und Schwierigkeitsmanagement verfügen. Sie sollten engagiert und motiviert sein und auf ein nötiges Fach- und Erfahrungswissen zurückgreifen können. Zudem fallen hierunter auch sozial-interaktive Verhaltensweisen. So sollten die Dozenten für eine angenehme Arbeitsatmosphäre sorgen, den Teilnehmern fair und kooperativ begegnen und auf die Bedürfnisse des Lehrgangs eingehen, um schließlich einen maximalen Lehrerfolg zu ermöglichen.

Die zweite Kategorie –*Rahmenbedingungen*–, der insgesamt 26% der zugrundeliegenden Argumente zugeordnet wurden, lässt erkennen, dass aus Sicht der Befragten (v.a. aus Sicht der Lehrgangsteilnehmer; siehe digitaler Anhang) auch die Ausstattung sowie Organisation des IdF NRW einen entscheidenden Einfluss auf die Güte einer Lehrveranstaltung hat. Hierunter fallen die Qualität und Verfügbarkeit von Übungsmaterialien bzw. -objekten, funktionierende Technik sowie angemessene und gut ausgestattete Räumlichkeiten. Ergänzt wird dies um strukturelle Aspekte wie Lehrgangszeiten und Gruppengrößen. Außerdem fällt hierunter die Güte von Unterkunft und Verpflegung während der zweiwöchigen Lehrgangsdauer.

Die dritte und die vierte Kategorie umfassen jeweils erfolgskritische Determinanten seitens der Lehrgangsteilnehmer. Die dritte Kategorie –*Gruppe*–, der 11% der genannten Aspekte zugeordnet wurden, beschreibt dabei das kameradschaftliche Miteinander der Teilnehmer. Für einen erfolgreichen Lehrgang ist es wichtig, dass die Gruppe zusammenhält. Die Teilnehmer müssen sich gegenseitig respektieren und unterstützen. Es ist wichtig, dass sich jeder auf die anderen verlassen kann und eine gute Zusammenarbeit möglich ist.

Der vierten Kategorie –*Teilnehmer*– wurden insgesamt 6% der betrachteten Aspekte zugeordnet. Dabei handelt es sich um verschiedene Merkmale und Verhaltensweisen des einzelnen Teilnehmers. Aus Sicht der Befragten begünstigen einerseits die individuelle Vorbereitung, das Vorwissen und -interesse des Teilnehmers den Lehrerfolg. Andererseits trägt jeder Einzelne durch seine aktive Mitarbeit sowie durch sein diszipliniertes (Lern-)verhalten zum Erfolg des Lehrgangs bei.

Der fünften Kategorie –*Lehrerfolg*– wurden insgesamt 7% der Aspekte zugeordnet. Dabei handelt es sich um solche Aspekte, die das Outcome eines guten Lehrgangs beschreiben. Der Lehrerfolg äußert sich aus Sicht der Befragten unter anderem durch einen Wissenszuwachs, den Erwerb relevanter Kompetenzen und einer persönlichen Weiterentwicklung. Hinzu kommen das Bestehen der Prüfung sowie ein Erreichen persönlicher Ziele als Erfolgskriterien guter Lehre.

Lediglich 1% der zusammengetragenen Aspekte konnte keiner der gebildeten Kategorien zugeordnet werden. Hierbei handelte es sich primär um Merkmale einer guten Prüfung.

3.2.2 Guter Dozent

Auf die Frage danach, wodurch sich ein guter Dozent auszeichnet (vgl. Abschnitt 3.1.2) zählte jeder befragte Dozent im Mittel 4,8 Merkmale ($SD = 0,8$) und jeder Lehrgangsteilnehmer im Mittel 3,1 Merkmale ($SD = 1,8$) auf. Insgesamt ergaben sich somit 209 Aspekte, auf Basis derer nach dem induktiven Vorgehen von Mayring (2015) sieben Kategorien gebildet wurden (siehe digitaler Anhang). Diese werden in Tabelle 1 vorgestellt. Die zufallsbereinigte Übereinstimmung der von den beiden Beurteilern unabhängig voneinander vorgenommenen Zuordnungen lag hierbei bei $\kappa = 0,90$. Gemäß der Klassifikation von Greve und Wentura (1997, S. 111) spricht dies für eine gute bis ausgezeichnete Interrater-Reliabilität ($da > 0,75$).

Tabelle 1

Kategoriensystem „Guter Dozent“.

Kategorie	Beispiele	Anteil
Persönlichkeit & Sozialkompetenz	Freundlich , kooperativ, respektvoll; Klima	(27%)
Fachkompetenz & Erfahrung	Theoriewissen; praktische Einsatzerfahrung	(23%)
Lehrkompetenz & -methoden	Abwechslungsreiche, anschauliche Stoffvermittlung	(17%)
Rhetorik & Verständlichkeit	Angemessene, verständliche Sprache	(10%)
Engagement & Motivierung	Vorbereitung; Freude am Job; Spaß machen	(9%)
Interaktion & Feedback	Auf den Lehrgang eingehen; Offen für Anregungen	(9%)
Struktur & Klarheit	Zielklarheit; Roter Faden; Transparenz	(4%)
Dozententeam	Aufeinander abgestimmtes, einheitliches Vorgehen	(1%)

Anmerkung. In Klammern findet sich derjenige Anteil aller Antworten, die der jeweiligen Kategorie zugeordnet werden konnten. Jede Antwort konnte einer dieser Kategorien zugeordnet werden.

Guter Lehrstil. Auf die ergänzende Frage nach einem guten Lehrstil, die lediglich in den Interviews gestellt wurde (vgl. Abschnitt 3.1.2), benannten die Dozenten sowie auch die Lehrgangsteilnehmer jeweils 1-2 Charakteristiken ($MW = 1,4$; $SD = 0,5$). Insgesamt konnten 11 Aspekte aus dem Datenmaterial extrahiert und analysiert werden. Eine Kategorisierung dieser Aspekte ergab, dass primär eine methodisch *abwechslungsreiche* sowie *flexibel* an die Bedürfnisse des Lehrgangs angepasste Inhaltsvermittlung (46% der genannten Aspekte) einen guten Lehrstil ausmacht. Des Weiteren sollte der Dozent die Inhalte auf eine *anschauliche* (27%), *authentische* (18%) sowie *interaktive* (9%) Weise vermitteln (siehe digitaler Anhang). Die Interrater-Reliabilität für die Zuordnung der charakterisierenden Aspekte zu diesen vier Kategorien lag mit einem Wert von $\kappa = 0,88$ ebenfalls in einem guten bis ausgezeichneten Bereich (da $> 0,75$; vgl. Greve & Wentura, 1997, S. 111).

3.2.3 Guter Teilnehmer

Auf die Frage danach, wie die Teilnehmer den Erfolg einer Lehrveranstaltung beeinflussen (vgl. Abschnitt 3.1.2), berichteten die Dozenten im Mittel jeweils 4,5 ($SD = 1,5$) und die Lehrgangsteilnehmer im Mittel jeweils 1,8 ($SD = 1,3$) mögliche Einflussfaktoren. Infolgedessen ergaben sich hier 132 Aspekte, die analysiert und induktiv in acht Kategorien zusammengefasst bzw. abgebildet werden konnten (nach Mayring, 2015; siehe digitaler Anhang). Das resultierende Kategoriensystem zeigt Tabelle 2. Die Interrater-Reliabilität betrug in diesem Fall $\kappa = 0,92$ und ist (da $> 0,75$) nach Greve und Wentura (1997, S. 111) als gut bis ausgezeichnet zu interpretieren.

3.2.4 Gute Rahmenbedingungen

Auf die Frage nach erfolgsentscheidenden Rahmenbedingungen einer guten Lehrveranstaltung (vgl. Abschnitt 3.1.2) benannten die Dozent im Mittel jeweils 4,2 ($SD = 1,8$) und die

Tabelle 2

Kategoriensystem „Guter Teilnehmer“.

Kategorie	Beispiele	Anteil
Beteiligung & Engagement	Sich einbringen; aktiv mitarbeiten	(28%)
(Vor-)Interesse & Lernbereitschaft	Offen für Neues; Spaß und Motivation	(20%)
Vorbereitung & Vorwissen	Praktische Erfahrung; theoretische Vorbereitung	(17%)
Gruppenzusammenhalt & -arbeit	Austausch; Zusammenhalt; Unterstützung	(16%)
Disziplin & Respekt	Regelkonformes Verhalten (z. B. Pünktlichkeit)	(11%)
Nachbereitung	EVA; Nachschlagewerk nutzen	(3%)
Transfervermögen	Gelerntes umsetzen können	(3%)

Anmerkung. In Klammern findet sich derjenige Anteil aller Antworten, die der jeweiligen Kategorie zugeordnet werden konnten. Die übrigen 2% der Antworten wurden der Kategorie „Sonstiges“ zugeordnet und beschreiben primär Persönlichkeitseigenschaften der Teilnehmer.

Lehrgangsteilnehmer im Mittel jeweils 2,3 (SD = 1,8) entsprechende Bedingungen. Insgesamt wurden 158 Aspekte zusammengetragen und analysiert (siehe digitaler Anhang). Dabei stellten 14% dieser Aspekte dozentenspezifische Merkmale und 11% der genannten Aspekte teilnehmerspezifische Merkmale dar, die allesamt bereits in Abschnitt 3.2.2 bzw. Abschnitt 3.2.3 aufgezeigt wurden.

Die übrigen 119 Aspekte wurden jeweils einer von fünf, der im Zuge des induktiven Vorgehens nach Mayring (2015), gebildeten Kategorien zugeordnet. Das zugrundeliegende Kategoriensystem wird durch Tabelle 3 wiedergegeben. Der Vergleich der von den beiden Beurteilern unabhängig voneinander vorgenommenen Zuordnungen resultierte in einer zufallsbereinigten Übereinstimmung von $\kappa = 0,96$, was nach Greve und Wentura (1997) für eine gute bis ausgezeichnete Interrater-Reliabilität spricht (da $> 0,75$; S. 111).

3.2.5 Biasfaktoren

Auf die Frage nach Aspekten, die das Urteil von Teilnehmern über die Güte einer Lehrveranstaltung darüber hinaus verzerren könnten, benannten die Dozenten im Mittel jeweils 1,5 (SD = 1,5) und die Lehrgangsteilnehmer im Mittel jeweils 0,6 (SD = 1,0) mögliche Biasfaktoren. Insgesamt wurden 46 aus Sicht der Befragten urteilsverzerrende Aspekte aus dem Datenmaterial extrahiert und analysiert (siehe digitaler Anhang). Neben dozentenspezifischen (28%; vgl. Abschnitt 3.2.2) und teilnehmerspezifischen Determinanten (22%; vgl. Abschnitt 3.2.2) sowie bereits benannten Rahmenbedingungen (24%; vgl. Abschnitt 3.2.4), stellen hier die faire Gestaltung und das Bestehen der *Prüfung* (11%) sowie *externe Störeinflüsse* (z. B. Wetter und Lärm; 15%) mögliche Biasfaktoren dar. Die Beurteiler-Übereinstimmung hinsichtlich der Zuordnung der Aspekte zu den gebildeten Kategorien betrug in diesem Fall $\kappa = 0,86$ und ist damit nach Greve und Wentura (1997) als gut bis ausgezeichnet zu interpretieren.

Tabelle 3

Kategoriensystem „Gute Rahmenbedingungen“.

Kategorie	Beispiele	Anteil
Räumliche & materielle Ausstattung	Modern, funktionsfähig; bedarfsgemäß verfügbar	(23%)
Unterbringung	Verpflegung; Unterkunft; Freizeitangebote	(21%)
Zeitlicher Rahmen & Ablauf	Planung und Struktur; Lehrgangszeiten; Pausen	(17%)
Vorbereitung & Informationen	Bereitstellung von Vorabinformationen	(11%)
Inhalte & Anforderungsniveau	Schwierigkeit der Inhalte; Aktualität der Themen	(3%)

Anmerkung. In Klammern findet sich derjenige Anteil aller Antworten, die der jeweiligen Kategorie zugeordnet werden konnten. Die übrigen 25% der Antworten auf diese Fragen stellten dozenten- oder teilnehmerspezifische Merkmale dar.

3.2.6 Lernziele

Im Rahmen der Dozenteninterviews wurde zudem die Frage nach zentralen Lernzielen der Gruppen- und Zugführerausbildung gestellt (vgl. Abschnitt 3.1.2). Die befragten Dozenten benannten insgesamt 13 Ziele ($MW = 2,6$; $SD = 0,9$), die auf Ähnlichkeiten hin analysiert wurden. Resultierend zeigte sich, dass das Erlangen von (Entscheidungs-)Sicherheit in Einsatzsituationen (39%) sowie der Erwerb relevanter Führungskompetenzen (31%) wesentliche Ausbildungsziele darstellen. Ergänzt werden diese um den Erwerb von Fachwissen (15%) und dem Anstoß einer persönlichen Weiterentwicklung bzw. Reifung (15%). Die Interrater-Reliabilität lag in diesem Fall bei $\kappa = 0,89$ und ist damit ebenfalls als gut zu bis ausgezeichnet interpretieren (vgl. Kategorisierung nach Greve & Wentura, 1997, S. 111).

3.3 Diskussion

Ziel dieser Studie war es, die sowohl aus Dozenten- als auch aus Teilnehmersicht zentralen qualitäts- und erfolgsentscheidenden Faktoren guter Führungsausbildung am IdF NRW zu erfassen. Dazu wurde sowohl auf die Interview- als auch auf die Fragebogenmethode zurückgegriffen. Im Rahmen der Interviews sollten die Antworten jeweils (je nach Bedarf) spezifisch hinterfragt werden können, um so ein umfassendes Verständnis relevanter Determinanten guter Lehrveranstaltungen zu erlangen. Dies blieb jedoch aus. Aufgrund zeitlicher Limitationen konnten lediglich die Kern- und Ergänzungsfragen (siehe Abschnitt 3.1.2) gestellt werden.

Zudem erwies sich die Interviewmethode zur Befragung der Lehrgangsteilnehmer auch deshalb als weniger gewinnbringend als im Vorfeld angenommen, da die Teilnehmer insgesamt nur wenig konkrete Ideen hatten, was die Qualität einer guten Lehrveranstaltung ausmacht, bzw. diese nur nach längerem Überlegen benennen konnten. Auf die Dozenten traf dies nicht zu. Im Gegensatz zu den interviewten Teilnehmern hatten sie sehr klare Vorstellungen darüber, was das Gelingen einer guten Lehrveranstaltung voraussetzt und konnten

diese während der Interviews auch unmittelbar abrufen und umfassend wiedergeben. Infolgedessen erwies sich der ergänzende Einsatz von Fragebögen in der Gruppe der Lehrgangsteilnehmer als äußerst sinnvoll. So konnten hierdurch einerseits die Meinungen vieler Teilnehmer auf eine effiziente Weise erfasst und berücksichtigt werden. Andererseits zeigen die Ergebnisse, dass durch den ergänzenden Einsatz der Fragebögen schließlich auch auf Seiten der Teilnehmer eine große Anzahl an verschiedenen erfolgskritischen Aspekten zusammengetragen wurde. Dies wird unter anderem auch darauf zurückgeführt, dass die Teilnehmer beim Ausfüllen der Fragebögen die nötige Zeit hatten, um über bedingende Merkmale einer guten Lehrveranstaltung nachzudenken.

Als Antwort auf Forschungsfrage 1a (siehe Abschnitt 2.3) konnten auf Basis der Angaben aller befragten Dozenten und Teilnehmer in dieser Studie fünf Hauptkategorien bzw. Dimensionen einer guten Führungsausbildung am IdF NRW identifiziert werden. Mit Blick auf Forschungsfrage 1b (siehe Abschnitt 2.3.1) ist festzustellen, dass die identifizierten Kategorien deutlich mit den von Rindermann (2001) postulierten Dimensionen erfolgreicher Lehre im universitären Kontext überlappen und diese lediglich um eine weitere ergänzen. So resultieren Qualität und Erfolg der Lehrveranstaltung im feuerwehrtechnischen Ausbildungskontext ebenfalls aus einem Zusammenspiel verschiedener Determinanten seitens der Dozenten sowie der Teilnehmermerkmale und äußeren Rahmenbedingungen (vgl. Abbildung 1). Ergänzend kommt hier noch der Einfluss der Gruppe hinzu. Während Studenten in Lehrveranstaltungen zumeist nicht unmittelbar auf ihre Kommilitonen angewiesen sind und die Zusammenarbeit aller Veranstaltungsbesucher nur in seltenen Fällen erforderlich oder gar erfolgsentscheidend ist, gilt im Zuge der Führungsausbildung am IdF NRW Gegenzugliches. Hier sind die Lehrgangsteilnehmer in der Bewältigung der ihnen in Theorie und Praxis gestellten Aufgaben regelmäßig aufeinander angewiesen. Jeder Teilnehmer muss sich auf seine Kameraden verlassen können und die nötige Unterstützung erfahren. So stellen –anders als im universitären Kontext– der Zusammenhalt und die gute Zusammenarbeit der Lehrgangsgruppe einen weiteren, zentral erfolgsentscheidenden Faktor der Ausbildungsgüte dar.

Als Antwort auf Forschungsfrage 1a (siehe Abschnitt 2.3) konnten in einem zweiten Schritt jeweils auch spezifische Subkategorien der Dimensionen Dozenten, Teilnehmer und Rahmenbedingungen herausgearbeitet werden. Mit Blick auf Forschungsfrage 1b (siehe Abschnitt 2.3.1) ist festzustellen, dass auch die herausgearbeiteten Subkategorien in weiten Teilen mit den von Rindermann (2001) aufgezeigten Determinanten guter Hochschullehre überlappen (vgl. Abbildung 1). So bedingen die Dozenten einerseits durch ihre fachlich-didaktische Kompetenzen sowie andererseits auch durch ihre sozial-interaktive Verhaltensweisen die Güte der Lehrveranstaltung (siehe Tabelle 1; vgl. Souvignier & Gold, 2002). Dabei agieren sie stets im Team, was einen wesentlichen Unterschied zu universitären Lehr-

veranstaltungen darstellt. Die daraus resultierend relevanten Aspekte werden ergänzend durch die Subkategorie *Dozententeam* abgebildet.

Auf Seiten der Teilnehmer determinieren –wie auch im universitären Kontext– verschiedene Facetten von Lernbereitschaft, Kompetenz und Verhalten den Lehrerfolg (siehe Tabelle 2; vgl. Hattie, 2008). Ergänzend kommen hier *Gruppenzusammenhalt und -arbeit* als eine weitere Kategorie hinzu. Dabei wurde der Einfluss der Gruppe zuvor bereits als eine Hauptdimension identifiziert, so dass sich nun die Frage stellt, ob das kameradschaftliche Miteinander sowie das sich gegenseitige Unterstützen der Lehrgangsteilnehmer tatsächlich eine eigenständige Dimension oder doch eine Subkategorie der Kategorie *Teilnehmer* darstellt. In jedem Falle scheint der Zusammenhalt der Gruppe für die Güte einer Lehrveranstaltung am IdF NRW erfolgsentscheidend zu sein.

Die zur Spezifizierung erfolgskritischer Rahmenbedingungen identifizierten Subkategorien erweisen sich als sehr kontextspezifisch. So kommt hierbei in erster Linie der räumlichen sowie materiellen Ausstattung des IdF NRW eine entscheidende Bedeutung zu. Dies betrifft einerseits die Unterrichtsräume und die eingesetzten Lehrmaterialien sowie die Übungsobjekte im Außengelände. Diese müssen bedarfsgemäß in einem guten Zustand zur Verfügung stehen. Andererseits bezieht sich dies auch auf den Internatsbereich des IdF NRW, in dem die Teilnehmer für die zweiwöchige Lehrgangsdauer untergebracht sind. Darüber hinaus stellen die zeitliche Planung und Lehrgangsvorbereitungen durch das IdF NRW sowie die vermittelten Inhalte und Anforderungen erfolgskritische Rahmenbedingungen dar (siehe Tabelle 3). Letzteres findet sich dabei auch in den von Rindermann (2001) genannten äußeren Rahmenbedingungen.

Während es insgesamt nur wenige Antworten auf die Frage nach den Biasvariablen gab, benannten die Befragten bereits zuvor (d. h. auf die Frage nach einem guten Dozent, einem guten Teilnehmer oder guten Rahmenbedingungen) verschiedene Verzerrungsfaktoren. Mit Blick in die Lehrevaluationsliteratur (z. B. Rindermann, 2001; Wolf, Spiel, & Pellert, 2001) stellen bereits genannte Merkmale des Dozenten (z. B. Sympathie), der Teilnehmer (z. B. Interesse, Motivation und Erwartung vor der Veranstaltung) sowie identifizierte Rahmenbedingungen (z. B. Ausstattung, Veranstaltungsgröße, Schwierigkeit und Arbeitsaufwand) neben den hier genannten externen Störfaktoren und der Prüfung ebenfalls mögliche Biasfaktoren dar. Sie können das Urteil eines Teilnehmers über die Güte der Lehrveranstaltung verzerren, ohne diese tatsächlich zu beeinflussen (vgl. ebd.). Um zu erfassen, ob und inwiefern die verschiedenen Variablen das Urteil über die Güte der Ausbildungslehrgänge am IdF NRW letztendlich tatsächlich verzerren, müssen die identifizierten Merkmale im Zuge der Evaluation ebenfalls berücksichtigt werden.

Die von den Dozenten benannten Lernziele spiegeln darüber hinaus in Teilen die von Miste (2007) identifizierten Faktoren verlässlichen Handelns wider (vgl. Abschnitt 2.4.1). Infolgedessen soll bei der Konstruktion des Evaluationsinstruments wie geplant auf die Arbeit von Miste zurückgegriffen werden und der Erwerb der von ihm postulierten Kompetenzen (bzw.

verlässlichkeitsbeeinflussenden Faktoren) als ein mögliches Kriterium des Lehrerfolgs erfasst werden.

Mit Blick auf die Neuausrichtung der Führungsausbildung und den damit einhergehenden Veränderungen der Lehrveranstaltungen (vgl. Abschnitt 1) sei an dieser Stelle noch Folgendes anzumerken: Während *Inhalt und Anforderungsniveau* der Lehrveranstaltung den Ergebnissen dieser Befragung zur Folge nur eine untergeordnete und wenig benannte Rahmenbedingung guter Lehre darstellen, sind diese für die Evaluation der (alten vs. neuen) Führungsausbildung am IdF NRW durchaus von großem Interesse. Grund hierfür sind die zeitbedingten Veränderungen der Veranstaltungsinhalte in den neuen Lehrgängen und das damit einhergehende Interesse daran, wie diese von den Teilnehmern wahrgenommen werden (Löchteken & Vöge, persönliches Gespräch, Oktober, 2015). Konkret stellt sich zum Beispiel die Frage danach, wie die Teilnehmer das Lernpensum in den neuen (vs. alten) Lehrgängen bewerten (vgl. Löchteken, 2015b). Da die Inhalte größtenteils durch den Lehrplan des IdF NRW vorgegeben werden, zum Teil aber auch durch die Dozenten lehrgangsspezifisch und situativ angepasst werden können, soll die Kategorie *Inhalt* im Folgenden zunächst als eigenständige Kategorie mit den Facetten *Umfang, Tiefe, Tempo, Schwierigkeit* und *Relevanz* betrachtet werden. Gleiches gilt für die unterschiedlichen Unterrichtsmethoden (d. h. Gruppen-, Plan- und Einsatzübungen sowie EVA; vgl. Abschnitt 2.1), die im Zuge der Befragung nur selten konkret benannt wurden, für das IdF NRW im Zuge einer Lehrgangsevaluation aber dennoch von großem Interesse sind (z. B. Löchteken, 2015a).

4 Itemkonstruktion

Die Ergebnisse der ersten Studie zeigen, dass bei einer Evaluation der Führungsausbildung am IdF NRW einerseits ähnliche Bereiche und Merkmale zu berücksichtigen sind, wie sie auch durch Theorien und Fragebögen zur universitären Lehrevaluation aufgezeigt werden (vgl. Abschnitt 2.3.1). Andererseits gilt es jedoch auch weitere, für den feuerwehrtechnischen Ausbildungskontext spezifische Aspekte zu erfassen (vgl. Abschnitt 3.3).

Vor diesem Hintergrund konnte bei der Erstellung des Fragebogens zum einen auf bestehende Lehrevaluationsinventare mit guten Testgütekriterien zurückgegriffen werden (siehe Anhang A für eine Übersicht). So wurden unter anderem aus den Arbeiten von Gediga und Kollegen (2000), Grötemeier und Thielsch (2010a, 2010b), Hirschfeld und Thielsch (2009a und 2009b), Staufenbiel (2000), Thielsch und Hirschfeld (2012a, 2012b), Wolf und Kollegen (2001) sowie aus der Arbeit von Zumbach und Kollegen (2007) jeweils vereinzelt passende Items übernommen (siehe Anhang D für eine Übersicht). In den meisten Fällen mussten diese noch geringfügig verändert werden. Darunter fiel zunächst die sprachliche sowie inhaltliche Anpassung an den feuerwehrtechnischen Ausbildungskontext. So wurde zum Beispiel „Student/in“ durch „Lehrgangsteilnehmer“ und „Vorlesung“ sowie auch „Seminar“ durch „Lehrveranstaltung“ ersetzt. Zudem wurden die Items, sofern dies noch nicht der Fall war,

extrem formuliert (z. B. durch den Zusatz des Attributs sehr: „Die Lehrmethoden waren [*sehr*] abwechslungsreich“; vgl. Hirschfeld & Thielsch, 2009a). Hierdurch sollten Deckeneffekte – und die damit einhergehenden unzureichende Differenzierungsfähigkeit – aufgrund *zu leichter* Items von vornherein vermieden werden (vgl. Wolf et al., 2001, S. 99). Des Weiteren wurden einige Items um ein einleitendes *Ich finde...* ergänzt, „um dem subjektiven Charakter der Bewertung Rechnung zu tragen und zu vermeiden, dass generalisierte Haltungen in die Bewertung einfließen“ (ebd., S. 98). Schließlich wurden die Items, wenn nötig, umgepolt, sodass alle Items, die zur Erfassung einer Kategorie herangezogen wurden, in die gleiche Richtung (d. h. vorzugsweise positiv) gepolt waren. Dies sollte anschließende Analysen vereinfachen. Eine umfassende Darstellung aller vorgenommenen Anpassungen findet sich im digitalen Anhang. Insgesamt wurden so 68 Items zusammengetragen, mit denen sich die identifizierten Kategorien bereits in großen Teilen abbilden ließen.

Um schließlich auch die für den feuerwehrtechnischen Ausbildungskontext spezifischen Aspekte abbilden zu können, wurden unter Berücksichtigung der Regeln der Itemkonstruktion (z. B. Bühner, 2011; Moosbrugger & Kelava, 2007; Thielsch, Lenzner & Melles, 2012) zum anderen auch neue Items generiert. Neben den bereits genannten Kriterien wurde hierbei besonders auf eine kurze und verhaltensnahe Formulierung der Items geachtet. Doppelte Verneinungen wurden, soweit möglich, vermieden (ebd.). Insgesamt wurden so 84 neue Items konstruiert (siehe Anhang D für eine Übersicht), von denen 14 die von Mistele (2007) definierten Kompetenzen umschreiben (vgl. Abschnitt 2.4.1).

Die Idee dabei war, zunächst eine große Anzahl an Items zusammenzutragen, um alle bisher erfassten und angenommenen Determinanten guter Lehre umfassend zu berücksichtigen. Wie die tabellarische Übersicht in Anhang D erkennen lässt, wurden dabei für die identifizierten Kategorien zum Teil unterschiedlich viele Items zusammengestellt. Dies ist darauf zurückzuführen, dass die Kategorien unterschiedlich umfangreich sind und für eine vollständige Beschreibung jeweils mehr oder weniger Aspekte zu berücksichtigen sind. Erst in den darauffolgenden Schritten sollte die Menge der Items dann basierend auf weiteren Befragungen auf ihr Wesentliches reduziert werden (siehe Studie II und Studie III).

5 Studie II

„Die Einbindung der Betroffenen [gilt allgemein] als wesentliche Voraussetzung für das Gelingen von Evaluationsprojekten“ (Wolf et al., 2001, S.90). Vor diesem Hintergrund wurden die Items vor Erstellung der Fragebogenerstversion sowohl den Verantwortlichen als auch den Betroffenen (d. h. Dozenten sowie Lehrgangsteilnehmern) am IdF NRW zur kritischen Begutachtung vorgelegt. Dabei sollten sie jedes Item hinsichtlich Relevanz und Verständlichkeit beurteilen sowie die aus ihrer Sicht noch fehlenden Aspekte ergänzen. Infolgedessen sollte die Verständlichkeit und Passung der zusammengestellten Items sowie eine Berück-

sichtigung aller für das IdF NRW relevanter Aspekte sichergestellt werden. Zudem sollten irrelevante Items identifiziert und die Itemmenge entsprechend reduziert werden können.

5.1 Methode

5.1.1 Stichprobe

An der Befragungen nahmen sieben Dozenten (86% männlich) sowie 26 Teilnehmer eines Gruppenführerlehrgangs (96% männlich) teil, sodass insgesamt $n = 33$ Personen am IdF NRW befragt wurden. Die befragten Dozenten waren zwischen 27 und 45 Jahre alt ($MW = 38,9$; $SD = 6,0$). In der Gruppe der Teilnehmer variierte das Alter zwischen 21 und 43 Jahren ($MW = 30,5$; $SD = 6,9$). Alle Personen nahmen freiwillig teil.

5.1.2 Material

Zur Befragung der Dozenten und Teilnehmer wurde ein selbstkonzipierter, standardisierter Fragebogen eingesetzt. Dieser umfasste zunächst eine ausführliche Aufklärung über die Ziele und Inhalte dieser Befragung (siehe Anhang E). Damit einhergehend wurden die Dozenten und Teilnehmer explizit instruiert, bei der Beurteilung der ihnen dargebotenen Items durchaus kritisch zu sein und schließlich nur die aus ihrer Sicht wirklich relevanten Aspekte einer guten Lehrveranstaltung als solche herauszustellen.

Insgesamt umfasste der Fragebogen 131 Items⁸, die von allen Befragten zum einen hinsichtlich ihrer Relevanz für die Bewertung der Güte einer Lehrveranstaltung sowie zum anderen auch hinsichtlich ihrer Verständlichkeit beurteilt werden sollten. Die zu beurteilenden Aussagen wurden kontextspezifisch in Gegenwartsform präsentiert (siehe Anhang E). Als Bewertungsformat wurde eine vierstufige Likert-Skala mit den Abstufungen *unwichtig*, *eher unwichtig*, *eher wichtig* und *wichtig* gewählt. Ergänzt wurde dies um die Antwortkategorie *nicht verständlich*, die entweder zusätzlich (d. h. wenn der Inhalt eines Items zum Beispiel durchaus wichtig, die Formulierung dessen jedoch unverständlich war) oder exklusiv (d. h. wenn die Aussage gänzlich unverständlich und eine Relevanzeinstufung infolgedessen unmöglich war) angekreuzt werden konnte.

Um etwaige Reihenfolgeeffekte in den Bewertungen erkennen und ausgleichen zu können, wurden zwei Versionen des Fragebogens erstellt und eingesetzt. Die beiden Versionen beinhalteten dieselben Items, gaben diese jedoch in entgegengesetzter Reihenfolge wieder. Jeder Befragte beurteilte die Items also entweder anhand von Version A oder anhand von Version B.

Auf die Beurteilung der Items folgten in beiden Versionen schließlich noch drei offene Fragen. So wurden die Dozenten und Teilnehmer gefragt, (1) was aus ihrer Sicht weitere Aspekte sind, die es im Zuge einer Veranstaltungsbeurteilung am IdF NRW zu berücksichti-

⁸ Wie die Zahl erkennen lässt, wurden hierbei nicht alle zuvor zusammengestellten Items ($m = 147$; siehe Anhang C) berücksichtigt. Vielmehr sollten nur die Items im Aussageformat bewertet werden (d. h. keine offenen Fragen). Zudem wurden weder Kontroll- noch Zusatzitems berücksichtigt (siehe digitaler Anhang).

gen gilt, (2) welche weiteren Ziele sie durch ihren Besuch der Lehrveranstaltung erreichen und (3) welche noch nicht benannten Kompetenzen sie während der Ausbildung erwerben oder ausbauen wollen.

5.1.3 Durchführung

Um sicherzustellen, dass alle Befragten bei der Beurteilung der Items auf ein Mindestmaß an Lehrgangserfahrung am IdF NRW zurückgreifen können, wurde die Befragung der Teilnehmer erst am letzten Unterrichtstag durchgeführt. Für die Beantwortung des Fragebogens hatte jeder Teilnehmer 30 Minuten Zeit. Die Dozenten erhielten den Fragebogen am selben Tag, hatten aufgrund ihrer unterschiedlichen zeitlichen Verfügbarkeiten jedoch die Möglichkeit, diesen zu einem beliebigen Zeitpunkt innerhalb der nächsten zwei Tage auszufüllen.

5.1.4 Datenanalyse

Um mögliche Reihenfolgeeffekte erkennen und in weiteren Analysen berücksichtigen zu können, wurde eine multivariate Varianzanalyse (MANOVA) mit der Reihenfolge, in der die Items präsentiert wurden (d. h. Version A vs. Version B), als festen Faktor und den Relevanzbewertungen aller 131 Items als abhängige Variablen durchgeführt. Auf dieses Verfahren wurde zurückgegriffen, da es, sofern die Stichprobengrößen in den zu vergleichenden Gruppen in etwa gleich und ausreichend groß sind (d. h. mindestens 10 Fälle pro Gruppe), als robust gegenüber Abweichungen von der Normalverteilung und heterogene Varianzen gilt (vgl. Döring & Bortz, 2016).

Darüber hinaus wurden die Relevanz- und Verständlichkeitsbewertungen deskriptiv statistisch analysiert. Es wurden relevante Verteilungsparameter berechnet und Häufigkeitsverteilungen analysiert, um darauf basierend eine Itemselektion vorzunehmen. Alle Berechnungen wurden mit dem Statistikprogramm IBM SPSS Statistics Version 23 durchgeführt.

5.2 Ergebnisse

Die MANOVA zeigte, dass es keinen Effekt der Reihenfolge auf die Relevanzbeurteilung der Items gab ($F(1,31) = 0,50$, $p = 0,83$). Das heißt, es gab keinen signifikanten Unterschied zwischen den anhand von Version A ($n = 17$) und den anhand von Version B ($n = 16$) vorgenommenen Bewertungen der Items.

Die Analyse der Relevanz- und Verständlichkeitsbewertungen ergab, dass zunächst 22 der Items von jeweils 3% der Befragten und sieben Items von jeweils 6% der Befragten als *nicht verständlich* beurteilt wurden (siehe digitaler Anhang). Darüber hinaus wurde kein Item von mehr als 2 Personen (d. h. von mehr als 6% aller Befragten) als *nicht verständlich* beurteilt. Des Weiteren zeigte die Analyse, dass alle 131 Items im Mittel (MW) mit Werten zwischen 2,7 und 3,9 ($SD = 0,3$ bis 1,1) bewertet und damit als *eher wichtig* bis *wichtig* beurteilt wurden. Dabei hatten sechs der Items einen $MW \leq 3,0$ (aber $\geq 2,7$) und 57 Items einen $MW \geq 3,5$ (aber $\leq 3,9$). Mit Ausnahme von Item 106 (vgl. Anhang D; Schiefe = 0) waren die Items zudem durchweg rechtssteil (bzw. linksschief) verteilt. Auf eine detaillierte, itemspezifische

Ausführung der Verteilungsparameter wird an dieser Stelle verzichtet. Hierfür sei auf die Excel-Datei „StudieII Häufigkeitsanalysen“ im digitalen Anhang verwiesen, die für jedes Item eine grafisch aufbereitete Häufigkeitsverteilung der Relevanz- und Verständlichkeitsbewertungen sowie relevante Verteilungsparameter zeigt.

Die offenen Fragen blieben zu 95% unbeantwortet. Darüber hinaus wurden in den Freitextkommentaren durchweg bereits berücksichtigte Aspekte (2%), Ziele (2%) und Kompetenzen (2%) noch einmal betont, konkretisiert oder mit anderen Worten wiedergegeben.

5.3 Diskussion und Itemauswahl

Ziel dieser Studie war es, basierend auf den Relevanz- und Verständlichkeitsurteilen von Dozenten und Teilnehmern, eine angemessene und begründete Auswahl an Items für die Erstversion des Fragebogens treffen zu können. Zudem sollte sichergestellt werden, dass diese Items dem spezifischen Kontext entsprechend verständlich formuliert waren und zudem alle relevanten Aspekte berücksichtigt wurden.

Die beiden letzten Punkte konnten durch die Ergebnisse dieser Studie unmittelbar bestätigt werden. So wurde einerseits keines der 131 präsentierten Items von einem nennenswerten Anteil aller Befragten als *unverständlich* beurteilt. Andererseits benannten die Dozenten und Teilnehmer –wenn überhaupt– nur bereits berücksichtigte Erfolgsmerkmale guter Lehre am IdF NRW, sodass anzunehmen war, dass durch die zusammengestellten Items bereits alle relevanten Aspekte, Kompetenzen und Ziele abgebildet wurden. Infolgedessen mussten die Items weder inhaltlich oder sprachlich verändert noch um weitere Items ergänzt werden.

Der erste Punkt (d. h. die Reduktion der Items) folgte hingegen nicht unmittelbar aus den Beurteilungen der Befragten. Obwohl die Dozenten und Teilnehmer explizit dazu aufgefordert worden waren, die Items kritisch zu beurteilen und dabei nur die aus ihrer Sicht wirklich wichtigen Aspekte selektiv als solche zu markieren, wurde jedes der 131 Items insgesamt betrachtet als *wichtig* oder *eher wichtig* bewertet. Wie weitere Verteilungsparameter zeigten, variierten die Urteile der Befragten dabei zumeist nur gering. Folglich war es nicht möglich, die Itemmenge ausschließlich anhand einer statistischen Entscheidungsregel (z. B. einem vorgegebenen Cut-Off-Wert) angemessen zu reduzieren. Stattdessen musste die Itemselektion letztendlich auch anhand von inhaltlichen Überlegungen erfolgen. So wurden all diejenigen Items, die von den Befragten insgesamt als *eher wichtig* (und eben nicht als *wichtig*) beurteilt worden waren, gemeinsam mit einem im Projekt eingebundenen Evaluationsexperten mit langjähriger Erfahrung kritisch diskutiert. Basierend auf seiner Evaluationsexpertise und gemeinsamen inhaltlichen Abwägungen konnte die Itemmenge schließlich um 23 Items reduziert werden (siehe Anhang D). Dabei wurden zum Beispiel solche Items ausgeschlossen, die einem anderen, als *wichtiger* bewerteten Item sehr ähnlich waren und darüber hinaus vermutlich keinen wesentlichen inhaltlichen Mehrwert bringen würden.

Insgesamt konnten jedoch nur verhältnismäßig wenige Items begründet eliminiert werden, sodass die Erstversion weiterhin mehr als 100 Items umfassen sollte. Mit Blick auf die Test-

gütekriterien Durchführungsökonomie und Zumutbarkeit (vgl. Abschnitt 2.5; Bühner, 2011) wäre eine stärkere Reduktion wünschenswert gewesen. Dies war auf Basis der Datenlage zu diesem Zeitpunkt allerdings noch nicht möglich. Zudem war aus Gesprächen mit Projektverantwortlichen seitens des IdF NRW ersichtlich geworden, dass diese vor dem Hintergrund der derzeitigen Neuausrichtung der Führungslehrgänge durchaus an einer derart umfangreichen Befragung interessiert waren (z. B. Löchteken & Vöge, persönliches Gespräch, Oktober, 2015). So konnte der Fragebogen im Zuge der Projektphase zunächst in seiner ganzen Länge eingesetzt und erprobt werden. Mit dem Ziel, daraus ein alltagstaugliches Standardinstrument zur Evaluation der Gruppen- und Zugführungslehrgänge am IdF NRW zu entwickeln, sollte der Fragebogen langfristig (d. h. basierend auf den Ergebnissen weiterer Erprobungen; siehe Studie 3) deutlich verkürzt werden.

6 Studie III

Basierend auf den Ergebnissen von Studie II wurde eine erste Version des Fragebogens zur Evaluation der Führungsausbildung am IdF NRW erstellt (siehe Anhang F). Diese sollte in der vorliegenden Studie über die verschiedenen Lehrgangsformen hinweg pilotiert werden. Ziel dabei war es zunächst, die einzelnen Items hinsichtlich ihrer unterschiedlichen Verteilungscharakteristiken zu untersuchen, um basierend darauf eine weitere begründete Itemselektion vornehmen zu können. Zudem sollte die zugrundeliegende Faktorenstruktur (d. h. die tatsächlichen Beurteilungsdimensionen guter Lehre am IdF NRW; → Forschungsfrage 1a) exploratorisch ermittelt werden. Über eine faktorenanalytische Validierung hinaus sollten in dieser Studie noch weitere erste Validierungen des Fragebogens anhand unterschiedlicher Konstrukte und Kriterien erfolgen.

Basierend auf den Ergebnissen dieser Studie sollte schließlich ein standardmäßig einsetzbares Evaluationsinstrument entwickelt werden, welches den in Abschnitt 2.5 genannten Haupt- und Nebengütekriterien ausreichend genügt (→ Forschungsfragen 2a & b).

6.4 Methode

6.4.1 Stichprobe

Insgesamt 263 Teilnehmer aus elf verschiedenen Gruppen- und Zugführerlehrgängen am IdF NRW gaben den Fragebogen ausgefüllt zurück, was einer Ausschöpfungsquote von 97% entsprach. Weil sie der anonymen Nutzung ihrer Angaben nicht zugestimmt hatten ($n = 10$), sie mehr als 10% der aussagebasierten Items nicht beantwortet hatten ($n = 7$; vgl. Roth, 1994) oder ihr Antwortmuster in gewisser Weise auffällig war (z. B. monoton oder unrealistische, $n = 3$), konnten die Datensätze von 20 Teilnehmern in weiteren Analysen nicht berücksichtigt werden.

Alle nachfolgend berichteten Analysen und Ergebnisse basieren damit auf einer Stichprobe von $n = 243$ Teilnehmern aus unterschiedlichen Führungslehrgängen (siehe Tabelle 4).

Tabelle 4

Übersicht über die Anzahl berücksichtigter Teilnehmer (sowie Lehrgänge) aufgeschlüsselt nach den sechs unterschiedlichen Lehrgangsformen.

	Gruppenführer		Zugführer
	ehrenamtlich	beruflich	ehrenamtlich
Alte Lehrgangsform	141 (6)	26 (1)	18 (1)
Neue Lehrgangsform	17 (1)	18 (1)	23 (1)

Anmerkung. Die unterschiedlichen Stichprobenumfänge kommen einerseits dadurch zustande, dass die einzelnen Lehrgangsformen jeweils unterschiedlich häufig stattfinden. So findet der ehrenamtliche Gruppenführerlehrgang mitunter jede bis jede zweite Woche statt, während alle anderen Lehrgangsformen während des Erhebungszeitraums von Ende Januar bis Mitte April 2016 jeweils nur einmal stattfanden. Andererseits werden in den Lehrgängen zum Teil unterschiedlich viele Teilnehmerplätze angeboten. Während an den alten Gruppenführerlehrgängen jeweils 26 Leute teilnehmen können, bieten die neuen Gruppenführerlehrgänge nur 20 Plätze. An den Zugführerlehrgängen können jeweils 24 Personen teilnehmen.

Von ihnen waren 95,5% männlich und 3,3% weiblich. Die übrigen 1,2% gaben ihr Geschlecht nicht an. Das Alter der Teilnehmer variierte zwischen 21 und 55 Jahren ($MW = 31,8$; $SD = 6,5$). Eine Übersicht über den höchsten Schulabschluss sowie den aktuellen Ausbildungsstand der Befragten zeigt Tabelle 5. Darüber hinaus waren zum Befragungszeitpunkt 70,4% der Teilnehmer ehrenamtlich und 14% der Teilnehmer hauptamtlich bei der Feuerwehr tätig. Die übrigen 15,6% gaben an, aktuell sowohl haupt- als auch ehrenamtlich tätig zu sein. Die mittlere Einsatzerfahrung belief sich unter allen Befragten dabei auf 13,8 Jahren ($SD = 8,2$), mit durchschnittlich 9 Einsätzen pro Monat ($SD = 15,5$).

6.4.2 Material

Der neukonzipierte Fragebogen kam in Form einer 13-seitigen, selbsterklärenden Paper-Pencil Version zum Einsatz. Diese umfasste zunächst einleitende Informationen zum Hintergrund der Befragung sowie ausführliche Instruktionen und wichtige Bearbeitungshinweisen (siehe Anhang F). Es folgten sechs Fragen zu verschiedenen demografischen Variablen (d. h. Alter, Geschlecht, Bildungsstand, Einsatzerfahrung) sowie die Anleitung zur Erstellung eines persönlichen Teilnahmecodes⁹. Als weitere Kontrollvariable wurde an dieser Stelle auch die aktuelle Stimmung der Teilnehmer anhand der von Jäger (2004) konstruierten 5-stufigen Smiley-Skala erfragt.

Darauf folgten 116 Items, die basierend auf den Ergebnissen von Studie I und Studie II zusammengestellt worden waren und erfolgskritische Aspekte guter Lehre am IdF NRW abbildeten. Hierbei handelte es sich um unterschiedliche Aussagen, deren Stimmigkeit zumeist anhand einer 7-stufigen Likert-Skala (von 1 = *stimme gar nicht zu* bis 7 = *stimme vollkom-*

⁹ Da alle Teilnehmer an den jeweils darauffolgenden Tagen (über diese Arbeit hinausgehend) zudem von Busjan (2016) zu den mündlichen und schriftlichen Prüfungen befragt wurden, sollte es durch diesen Code möglich sein, die Angaben aus den unterschiedlichen Befragungen eines jeden Teilnehmers einander zuordnen zu können (ohne dabei die Gewährleistung von Anonymität einschränken zu müssen).

Tabelle 5

Übersicht über den Bildungsstand der Stichprobe ($n = 243$).

Höchster Schulabschluss		Aktueller Ausbildungsstand	
Keinen Schulabschluss	0,0%	Keine abgeschlossene Berufsausbildung	1,2%
Haupt-/Volksschulabschluss	11,5%	Noch in Ausbildung oder Student/in	2,9%
Realschulabschluss/ Mittlere Reife	39,1%	Abgeschlossene Lehre oder Meisterbrief	67,1%
Fachhochschulreife	22,6%	Berufsakademie-, Fachhochschul- oder Universitätsabschluss	22,2%
Abitur	24,3%	Promotion oder höher	0,8%
Sonstiger Schulabschluss	1,2%	Sonstiger beruflicher Ausbildungsstand	5,3%

Anmerkung. Aufgrund fehlender Angaben, summieren sich die Prozentzahlen einer Spalte nicht zu 100% auf.

men zu) zu beurteilen war. Das 7-stufige Antwortformat wurde gewählt, um möglichst viel Varianz in den Ergebnissen zu erzeugen. Alle Stufen wurden verbal benannt und schließlich noch um die Antwortkategorie 0 = *nicht sinnvoll beantwortbar* ergänzt, um somit die Reliabilität und Validität des Fragebogens von vorneherein zu erhöhen (vgl. Thielsch et al., 2012). Hiervon ausgenommen waren lediglich fünf Items, von denen vier Items (d. h. Item 60, 88, 90 & 105¹⁰, siehe Anhang D) ein dichotomes Antwortformat (Nein/Ja) aufwiesen. Ein weiteres Item (d. h. Item 58, siehe Anhang D) verlangte eine Häufigkeitseinschätzung auf einer 5-stufigen Rating-Skala (mit 1 = *nie*, 2 = *selten*, 3 = *gelegentlich*, 4 = *oft*, 5 = *immer*).

Um die Auswertung der Antwortmuster möglichst einfach zu halten (d. h. aufwendige und fehleranfällige Umpolungen einzelner Items zu vermeiden), wurde der Großteil aller Items so formuliert, dass höhere Werte jeweils eine bessere Bewertung repräsentierten. Hiervon ausgenommen waren lediglich die neun Items der Dimension *Inhalt* (siehe Anhang D). Um doppelte Verneinungen und umständliche Formulierungen zu vermeiden (vgl. Thielsch et al., 2012), empfahl es sich, diese Items einheitlich negativ zu polen.

Für eine Globalbeurteilung der Lehrveranstaltung wurden die Teilnehmer (in Anlehnung an bestehende Inventare; z. B. Staufenberg, 2000) anschließend gebeten, diese mit einer Schulnote von 1 (= *sehr gut*) bis 6 (= *ungenügend*) zu bewerten.

Ergänzend sollten sie angeben, ob sie die besuchte Veranstaltung allgemein weiterempfehlen würden (dichotomes Format: Ja/Nein). Die beschriebenen Items wurden schließlich noch um fünf offene Fragen ergänzt (vgl. Item 148–152 im Anhang D), mit denen konkret solche Aspekte erfragt wurden, die für den Vergleich des alten und neuen Lehrkonzeptes zusätzlich relevant sein können (vgl. Löchteken, 2015). Neben Fragen zur allgemeinen Ver-

¹⁰ An dieser Stelle sei darauf hingewiesen, dass die Items in den Fragebögen unterschiedliche Nummerierungen aufweisen. Dies ist primär der Tatsache geschuldet, dass die Itemmenge im Verlauf dieser Arbeit immer wieder verändert und mitunter auch in variierenden Reihenfolgen präsentiert wurde. Um Missverständnissen vorzubeugen, werden die Items in dieser Arbeit jedoch über die Studien hinweg einheitlich benannt. Grundlage hierfür stellen die tabellarischen Auflistung und die daraus hervorgehenden Nummerierung im Anhang D dar.

anstellungskritik und Verbesserungsvorschlägen wurden die Teilnehmer gefragt, was ihre Ziele waren und ob sie diese erreicht haben bzw. in welchen Bereichen sie sich noch unsicher fühlen. Diese Fragen stellen den Gegenstand weiterer Forschung dar (siehe hierfür Loberg, 2016) und werden im Zuge dieser Arbeit folglich nicht weiter berücksichtigt.

Der Fragebogen endete schließlich mit einer Einwilligungserklärung zur Datennutzung. Die Teilnehmer mussten explizit ankreuzen, ob ihre Angaben anonymisiert in die wissenschaftliche Auswertung eingehen dürfen oder nicht.

6.4.3 Durchführung

Die Teilnehmer beantworteten den Fragebogen kurz vor Abschluss des Lehrgangs, jedoch vor den Prüfungen, am Ende eines letzten Unterrichtstages. Der Befragungszeitpunkt wurde bewusst so gewählt, dass die Teilnehmer den Lehrgang umfassend beurteilen konnten, ihr Urteil jedoch nicht bereits durch etwaige Prüfungserfahrungen und -ergebnisse verzerrt war. Die mittlere Bearbeitungszeit betrug ca. 25 Minuten.

6.4.4 Datenaufbereitung und -analyse

Vor Durchführung der Analysen wurden die negativ formulierten Items der Dimension *Inhalt* (d. h. Item 45, 47, 48, 49, 51, 52, 54, 56 und 57) zunächst revers kodiert, sodass höhere Werte auch hier einheitlich für eine bessere Bewertung standen.

Alle nachfolgend berichteten Analysen wurden mit dem Statistikprogramm IBM SPSS Statistics Version 23 durchgeführt.

Itemselektion. Die Items sollten annähernd normalverteilt sein (vgl. Bühner, 2006). Um dies zu überprüfen, wurden für jedes Item deskriptive Verteilungsparameter (z. B. Mittelwert, Standardabweichung, Schiefe sowie Kurtosis) berechnet und analysiert. Ein besonderes Augenmerk lag dabei einerseits auf möglichen Boden- und Deckeneffekten (d. h. links- oder rechtssteilen Verteilungen), die zeigten, dass ein Item entweder zu schwer oder zu leicht war (ebd.). Die Schwierigkeit eines Items gibt in diesem Fall den Grad der Zustimmung an (z. B. Döring & Bortz, 2016). Zu schwer bedeutet dabei, dass die große Mehrheit der Befragungsteilnehmer diesem Item nicht zugestimmt hat, während ein sehr leichtes Item angibt, dass die Mehrheit aller Befragten diesem eindeutig zugestimmt hat.

Andererseits interessierten auch zwei- oder mehrgipflige Verteilungsmuster, die darauf hindeuten, dass ein Item von den Teilnehmern mitunter unterschiedlich aufgefasst worden war (ebd.). Hierfür wurden Häufigkeitsdiagramme hinzugezogen. Des Weiteren sollten sehr homogene Items identifiziert werden, die hoch miteinander korrelieren und sich zudem auch inhaltlich sehr ähnlich waren.

Basierend auf den Ergebnissen dieser Analysen sollte die Itemmenge anhand klar definierter Selektionskriterien schrittweise reduziert werden, sodass schließlich nur ausgewählte und hinreichend gut verteilte Items in die darauffolgenden Analysen eingingen.

Exploratorische Faktorenanalysen. Um die tatsächliche, empirisch zugrundeliegende Faktorenstruktur der Items zu ermitteln, wurden exploratorische Faktorenanalysen durchgeführt. Hierzu wurden die verbleibenden Items vorab in vier Gruppen aufgeteilt. Die erste Gruppe umfasste all diejenigen Items, die nicht auf einer 7-stufigen Likert-Skala zu beantworten waren. Diese Items konnten im Zuge der exploratorischen Faktorenanalyse nicht berücksichtigt werden. Die zweite Gruppe umfasste all diejenigen Items, mit denen potentielle Biasvariablen erfasst wurden, sprich solche Merkmale, „die grundsätzlich nicht mit der Lehrqualität zu tun haben, jedoch [Teilnehmer] systematisch bei der Bewertung beeinflussen“ (Wolf et al., 2001, S. 91). Wie in Abschnitt 3.3 dargestellt fallen hierunter sowohl Merkmale des Dozenten (z. B. Sympathie) sowie Merkmale der Teilnehmer (Motivation und Vorinteresse) und verschiedene Rahmenbedingungen (z. B. räumliche und materielle Ausstattung, Lehrgangszeiten und -größe sowie Prüfungsrelevanz der Inhalte). Auch diese Variablen fanden keinen Eingang in die exploratorische Faktorenanalyse.

Der dritten Gruppe wurden schließlich diejenigen Items zugeordnet, mit denen Facetten des Outcomes einer erfolgreichen Lehrveranstaltung auf Teilnehmerseite (d. h. der subjektive Lerngewinn und Kompetenzerwerb; vgl. hierzu Abbildung 1 in Abschnitt 2.3.1) erfasst wurden. Mit diesen Items wurde eine erste exploratorische Faktorenanalyse durchgeführt. Die Idee dabei war es, solche Erfolgsfaktoren zu identifizieren, die schließlich (–wenn aufgrund der gegebenen Stichprobenabhängigkeit auch nur bedingt–) als ein mögliches Validierungskriterium hinzugezogen werden konnten.

Die übrigen Items bildeten die vierte Gruppe. Hierbei handelte es sich um diejenigen Items, die Aussagen zu verschiedenen qualitätsbedingende Aspekte (bzw. Determinanten) guter Lehrveranstaltungen am IdF NRW umfassten und denen die Teilnehmer jeweils auf einer 7-stufigen Likert-Skala zustimmen mussten. Mit diesen Items wurde eine weitere exploratorische Faktorenanalyse durchgeführt. Basierend auf den errechneten Faktorladungen wurde zunächst eine weitere Itemselektion vorgenommen (z. B. aufgrund von Doppelladungen), um schließlich substantielle Skalen zu bilden. Zur Analyse der Reliabilität der Skalen wurde die interne Konsistenz über Cronbachs Alpha bestimmt. Zudem wurden Interkorrelationen der Skalen berechnet.

Bias-Effekte. Nachdem zunächst alle denkbaren Biasvariablen in den Fragebogen aufgenommen worden waren (siehe Anhang G), sollte auch hier nun eine Reduktion auf die tatsächlichen, das Urteil der Teilnehmer verzerrenden Variablen stattfinden. Hierzu wurden zunächst korrelative Zusammenhänge zwischen den möglichen Bias-Items und der Gesamtbeurteilung der Lehrveranstaltung durch die Teilnehmer berechnet. Die Gesamtbeurteilung der Veranstaltung erfolgte in Form einer Schulnote (von 1 = *sehr gut* bis 6 = *ungenügend*; vgl. Item 145). Um die Interpretation der Ergebnisse zu vereinfachen und um Missverständnissen vorzubeugen, wurde dieses Item zunächst umgepolt (→ Item 145*), sodass auch hier ein höherer Wert eine positivere Bewertung anzeigte. Auch wenn anhand von Korrelationen keinerlei Aussagen über Kausalitäten getroffen werden können, geben sie einen ersten Hinweis auf mögliche Einflüsse.

Erste Validierungen. Anschließend wurden erste Validierungen der gebildeten Skalen durchgeführt. Hierzu wurden einerseits die korrelativen Zusammenhänge der Skalen mit verschiedenen Kriterien berechnet (→ Kriteriumsvaliditäten). Wie bereits erwähnt, stellten die im Zuge der ersten Faktorenanalyse herausgearbeiteten Erfolgskriterien (d. h. die sogenannten Outcome-Skalen) ein erstes Validierungskriterium dar. Des Weiteren wurden auch die Schulnoten, mit denen die Teilnehmer die Lehrveranstaltung insgesamt bewertet hatten (Item 145*), sowie der subjektiv eingeschätzte Lernerfolg (Item 123) zur Ermittlung der konkurrenten Validität des Fragebogens herangezogen (vgl. Bühner, 2011).

Die mündlichen und schriftlichen Prüfungsnoten der Teilnehmer fungierten als ein weiteres Validierungskriterium. Diese wurden in der Arbeit von Busjan (2016) erfragt und zunächst über die persönlichen Teilnehmercodes in den Datensatz dieser Arbeit integriert. Anschließend wurden Korrelationen zwischen den Veranstaltungsbewertungen der Teilnehmer und ihren Prüfungsnoten berechnet, um erste Aussagen über die prognostische Validität des Fragebogens treffen zu können (vgl. Bühner, 2011).

Andererseits wurde über die Ermittlung von korrelativen Zusammenhängen zwischen der Gesamtbenotung der Lehrveranstaltung und den einzelnen gebildeten Skalen einerseits und der zu Befragungsbeginn angegebenen Stimmung der Teilnehmer andererseits (siehe Abschnitt 6.1.2) auch diskriminante Validierungen durchgeführt. Ziel dabei war es, sicherzustellen, dass die Beurteilung des Lehrgangs nicht primär nur die momentane Stimmung des Teilnehmers zum Bearbeitungszeitpunkt widerspiegelt (vgl. Bühner, 2011).

6.5 Ergebnisse

6.5.1 Itemselektion

Die Itemselektion erfolgte in vier Schritten, denen jeweils konkrete Ausschlusskriterien zugrunde lagen¹¹. Im ersten Schritt wurden zunächst diejenigen Items eliminiert, die von mehr als 5% der Teilnehmer nicht beantwortet bzw. als nicht beantwortbar beurteilt worden waren. Dies betraf drei Items (→ Item 82, 85 und 93).

Im zweiten Schritt wurden die Verteilungen der Itemantworten analysiert. Diese waren, mit Ausnahme von Item 47 und Item 66, durchweg rechtssteil verteilt (d. h. Schiefe < 0). Nach Bühner (2006) liegt jedoch erst dann eine substantielle Verletzung der Normalverteilung und eine damit einhergehende Nicht-Eignung für faktorenanalytische Betrachtungen vor, wenn Schiefe und Exzess außerhalb der von West, Finch und Curran (1995) festgelegten Grenzen liegen. Nach den Autoren West und Kollegen sollte die Schiefe $< |2|$ und der Exzess $< |7|$ sein. Infolgedessen wurden in diesem Schritt alle Items mit einer Schiefe $> |2|$ und/oder einem Exzess $> |7|$ eliminiert. Dies betraf acht Items (→ Item 1, 10, 30, 80, 91, 110, 143 und 144).

¹¹ Ergänzend sei hier auf die Excel-Datei „Studie III Itemselektion“ verwiesen. Diese umfasst alle relevanten Verteilungsparameter und bietet zudem eine gute Übersicht über die schrittweise vollzogenen Selektionen.

Hohe Item-Interkorrelationen und inhaltliche Überlegungen stellten ein drittes Ausschlusskriterium dar. Korrelierten zwei Items höher als 0,75 miteinander und waren sich diese Items zudem auch inhaltlich sehr ähnlich (–sodass eine Berücksichtigung beider Items keinen bedeutsamen Mehrwert brachte–), wurde eines der beiden Items ausgeschlossen. Die Auswahl erfolgte dabei basierend auf statistischen Verteilungskennwerten sowie inhaltlichen Überlegungen. Insgesamt wurden in diesem Schritt fünf weitere Items eliminiert (→ Item 7, 40, 41, 83 und 124).

Im vierten Schritt wurden alle noch verbliebenen Items, die sehr leicht waren ($MW \geq 6$) und zugleich nur eine geringe Streuung ($SD < 1$) aufwiesen ($m = 29$), von vier Experten begutachtet. Items, die von den Experten als unabdingbar für eine Lehrgangsevaluation am IdF NRW beurteilt wurden, wurden in der Itemmenge belassen ($m = 18$). Die übrigen 11 Items (→ Item 2, 3, 5, 9, 18, 19, 28, 71, 92, 95 und 136) wurden eliminiert. In diesem Schritt (d. h. basierend auf Verteilungsanalysen und Experteneinschätzungen) konnten zudem auch zwei dichotome Items (→ Item 88 und 105) aus der Itemmenge ausgeschlossen werden.

Insgesamt wurde die Itemmenge somit zunächst um 29 (–davon zwei dichotome–) Items reduziert. In Folge dieser Reduktion umfasste die Dimension *Einsatzübungen* (Methode 4, vgl. Anhang E) schließlich nur noch zwei Items. Da jedoch mindestens drei Items pro ange-dachtem Faktor vorliegen sollten (z. B. Bühner, 2006), wurden diese beiden Items (→ Item 79 und 81) im Folgenden ebenfalls nicht weiter berücksichtigt.

Nach Selektion und anschließender Gruppierung der Items (siehe hierfür Anhang G), gingen 17 Items in die erste und 50 Items in die zweite exploratorische Faktorenanalyse ein (vgl. hierzu auch Abschnitt 6.1.4).

6.5.2 Exploratorische Faktorenanalysen

Nach West und Kollegen (1995) liefert eine Faktorenanalyse dann interpretierbare Ergebnissen, wenn die Stichprobe $n > 200$ ist und Schiefe und Exzess aller Items jeweils innerhalb der bereits thematisierten Grenzen liegen (siehe Abschnitt 6.2.1). Während letzteres durch den zweiten Schritt der vorausgegangenen Itemselektion explizit sichergestellt wurde, ist auch ersteres mit einer zugrundeliegende Stichprobengröße von $n = 243$ gegeben.

Faktorenanalyse I. In die erste Faktorenanalyse (vgl. Abschnitt 6.1.4) gingen insgesamt 17 –den subjektiven Lerngewinn bzw. Kompetenzerwerb erfassende– Items ein. Mit einem MSA-Wert von 0,91 zeigte der KMO-Test nach Bühler (2010, S. 347) eine sehr gute Eignung des vorhandenen Datensatzes zur Durchführung einer Faktorenanalyse an. Die exploratorische Faktorenanalyse (Hauptachsenanalyse mit obliquer Promax-Rotation) legte nach dem Kaiser-Guttman-Kriterium (Eigenwert > 1 ; Guttman, 1954; Kaiser, 1960) eine Drei-Faktoren-Lösung nahe. Der MAP-Test¹² sprach hingegen sowohl in seiner Originalversion (Velicer,

¹² Kurz für Minimum-Average-Partial-Test (Velicer, 1976; Velicer, Eaton, & Fava, 2000). Zur Durchführung des MAP-Tests wurde das Skript von O'Connor (2000) für das Statistikprogramm IBM SPSS Statistics verwendet.

1976) als auch in der revidierten Fassung (Velicer et al., 2000) für zwei Faktoren¹³. In Übereinstimmung damit legte auch der Scree-Test nach Catell (1966) die Extraktion von zwei Faktoren nahe. Inhaltliche Überlegungen stützten die Zwei-Faktoren-Lösung, weshalb diese gewählt wurde. Durch diese Lösung konnten 56,8% der Varianz aufgeklärt werden.

Dabei luden zunächst zehn Items einfach auf dem ersten Faktor (Eigenwert = 8,01) und fünf Items auf dem zweiten Faktor (Eigenwert = 1,6). Die übrigen zwei Items luden sowohl auf dem ersten als auch auf dem zweiten Faktor substantiell und ließen sich folglich keinem der beiden Faktoren statistisch eindeutig zuordnen. Gestützt durch inhaltliche Überlegungen wurden diese doppelt ladenden Items (→ Item 131 und 142) schließlich aus der Itemmenge entfernt. Eine Übersicht über die Ladungen findet sich in Anhang H.

Skalenbildung. Um eindeutig interpretierbare und überschaubare Skalen zu bilden, wurden die jeweils auf einem Faktor ladenden Items anschließend zueinander in Bezug gesetzt und statistisch sowie inhaltlich analysiert. Hierdurch konnte zunächst die Menge der zehn auf den ersten Faktor ladenden Items um fünf Items (→ Item 134, 135, 137, 139 und 140) reduziert werden. Die übrigen fünf Items bilden jeweils unterschiedliche Kompetenzen ab, die es während der Führungsausbildung am IdF NRW zu erwerben bzw. zu stärken gilt (vgl. Mistele, 2007; Abschnitt 2.4.1). Gemeinsam bilden sie die als *Kompetenzerwerb* benannte erste Outcome-Skala. Zudem konnte auch die Menge der fünf auf den zweiten Faktor ladenden Items um zwei Items (→ Item 123 und 141) reduziert werden. Die verbleibenden drei Items erfassen, inwiefern sich die angehenden Führungskräfte durch ihre Teilnahme am Lehrgang auf bevorstehende Einsätze vorbereitet fühlen und ob sie das Gelernte auf ihre Heimatstelle übertragen können. Infolgedessen bilden sie die als *Vorbereitung & Transfer* betitelte zweite Outcome-Skala.

Die gebildeten Skalen korrelieren mit $r = 0,46$ nur gering. Nachfolgend werden die beiden Skalen inhaltlich und statistisch beschrieben (siehe ergänzend auch Tabelle 6).

Skala I.1 Kompetenzerwerb. Die erste Outcome-Skala erfasst den durch die Teilnehmer selbsteingeschätzten Erwerb relevanter Führungs- und Handlungskompetenzen, welche zu Beginn dieser Arbeit aus der Verlässlichkeitsforschung von Mistele (2007) abgeleitet wurden. Darunter fallen zunächst verbesserte *Wahrnehmungs-* und *Entscheidungsfähigkeiten* (vgl. hierzu Abschnitt 2.4.1), die einerseits durch Item 129 („Durch meine Teilnahme am Lehrgang gelingt es mir jetzt besser, Gefahrenlagen frühzeitig zu erkennen“) sowie andererseits durch Item 130 („Durch meine Teilnahme am Lehrgang fällt es mir jetzt leichter, Entscheidungen in kritischen Situationen zu treffen“) erfasst werden. Des Weiteren bildet die Skala durch Item 132 („Durch meine Teilnahme am Lehrgang kenne ich meine persönlichen Grenzen besser“) sowie Item 133 („Durch meine Teilnahme am Lehrgang traue ich mir besser zu, in stressigen Situationen ruhig zu bleiben“) das Ausmaß einer wahrgenommenen Verbesserung der

¹³ Aufgrund des starken ersten Faktors wurde von der Durchführung einer Parallelanalyse abgesehen (vgl. Bühner, 2011). Stattdessen empfahl sich eine Inspektion des Screeplots (d. h. Scree-Test nach Catell, 1966).

Tabelle 6

Übersicht über die Outcome-Skalen.

Skala	Cronbachs α	$MW (SD)$	Items	r_{it}
(I.1) Kompetenzerwerb	0,85	5,7 (0,8)	Item 129	0,68
			Item 130	0,70
			Item 132	0,68
			Item 133	0,69
			Item 138	0,53
(I.2) Vorbereitung & Transfer	0,79	5,6 (0,9)	Item 126	0,69
			Item 127	0,71
			Item 128	0,60

Selbsteinschätzungsfähigkeit ab (vgl. hierzu Abschnitt 2.4.1). Mit Items 138 („Durch meine Teilnahme am Lehrgang kann ich korrektere und verständlichere Anweisungen geben“) erfasst die Kompetenzerwerbs-Skala schließlich auch Verbesserungen hinsichtlich der *Kooperations- und Kommunikationsfähigkeit* der Teilnehmer (vgl. hierzu Abschnitt 2.4.1).

Die fünf Items dieser Skala zeigen Trennschärfen ($0,53 \leq r_{it} \leq 0,70$; siehe Tabelle 6), die nach Moosbrugger und Kelava (2008)¹⁴ als gut zu bezeichnen sind. Mit einem Wert von Cronbachs $\alpha = 0,85$ weist die Skala insgesamt eine ebenfalls gute interne Konsistenz auf (Bortz & Döring, 2006). Verteilungskennwerte lassen erkennen, dass die Skala linksschief verteilt ist (Schiefe = -1,6).

Skala I.2 Vorbereitung & Transfer. Die zweite Outcome-Skala erfasst, inwiefern sich die angehenden Führungskräfte durch ihre Teilnahme am Lehrgang befähigt fühlen, bevorstehende Einsätze in ihrer Heimatstelle zu leiten. Hierzu wird nach der nötigen Vorbereitung (Item 126: „Ich fühle mich auf meinen nächsten Einsatz, den ich als Gruppen-/ Zugführer leiten werde, sehr gut vorbereitet“), der erlangten Sicherheit (Item 127: „Durch die praktischen Übungen im Lehrgang habe ich die nötige Sicherheit gewonnen, einen Einsatz als Gruppen-/ Zugführer zu leiten“) sowie dem allgemeinen Transfer (Item 128: „Ich kann das Gelernte auf meine Heimatstelle übertragen“) gefragt.

Alle drei Items weisen Trennschärfen auf ($0,60 \leq r_{it} \leq 0,70$; siehe Tabelle 6), die nach Moosbrugger und Kelava (2008) als gut zu bezeichnen sind. Insgesamt weist die Skala eine annähernd gute interne Konsistenz auf (siehe Tabelle 6; Bortz & Döring, 2006). Verteilungskennwerte lassen zudem erkennen, dass die Skala linksschief verteilt ist (Schiefe = -1,1).

¹⁴ Die Autoren sprechen von guten Trennschärfen, wenn diese zwischen 0.4 und 0.7 liegen (S.84).

Faktorenanalyse II und Skalenbildung. In die zweite Faktorenanalyse (vgl. Abschnitt 6.1.4) gingen insgesamt 50 Items ein. Der KMO-Test nach Bühler (2010, S. 347) zeigte mit einem MSA-Wert von 0,88 eine gute Eignung des vorhandenen Datensatzes zur Durchführung einer Faktorenanalyse an. Die exploratorische Faktorenanalyse (Hauptachsenanalyse mit obliquen Promax-Rotation) legte nach dem Kaiser-Guttman-Kriterium (Eigenwert > 1) eine Elf-Faktoren-Lösung nahe. Der MAP-Test sprach in seiner Originalversion (Velicer, 1976) hingegen für acht und in der revidierten Fassung (Velicer et al., 2000) für zehn Faktoren. Der Scree-Test nach Catell (1966) legte die Extraktion von sechs oder acht Faktoren nahe. Gegeben der Tatsache, dass die Acht-Faktoren-Lösung in einfacher zu interpretierenden Faktorladungen resultierte, wurde schließlich diese Lösung gewählt. Durch diese Lösung konnten 61,0% der Varianz aufgeklärt werden. Lediglich ein Item (\rightarrow Item 36) lud auf keinem der acht extrahierten Faktoren und wurde aufgrund inhaltlicher Überlegungen bei der Skalenkonstruktion nicht weiter berücksichtigt.

Skala II.1 Dozentenverhalten. Insgesamt 15 Items luden primär auf dem ersten Faktor (Eigenwert = 13,8). Fünf dieser Items (\rightarrow Item 11, 13, 20, 23 und 26) luden zusätzlich auf einem der anderen sieben Faktoren (jeweils mit $\lambda > 0,3$). Gestützt durch inhaltliche Überlegungen wurden die doppelt ladenden Items, mit Ausnahme von Items 23 und Item 26, nachfolgend nicht weiter berücksichtigt. Eine Übersicht über die Ladungen zeigt Anhang H. Um eine eindeutig interpretierbare und überschaubare erste Skala zu bilden, wurden die verbleibenden elf Items zueinander in Bezug gesetzt und statistisch sowie inhaltlich gegeneinander gehalten.

Nach einem begründeten Ausschluss von weiteren sieben Items (\rightarrow Item 6, 12, 18, 22, 27, 29 und 43) resultierten schließlich fünf Items, die gemeinsam die Skala *Dozentenverhalten* bilden. Die Skala erfasst mit Item 19 („Der sprachliche Ausdruck der Dozenten war verständlich“), Item 23 („Ich finde, die Dozenten gingen auf Fragen und Anregungen der Teilnehmer angemessen ein“), Item 26 („Ich finde, die Dozenten gaben nützliches Feedback“), Item 28 („Die Dozenten motivierten mich, mich einzubringen“) sowie Item 42 („Ich finde, die Dozenten haben ihr Vorgehen gut miteinander abgestimmt“) verschiedene Aspekte des Dozentenhandelns. Dabei weisen alle vier Items Trennschärfen auf ($0,56 \leq r_{it} \leq 0,67$; siehe Tabelle 7), die nach Moosbrugger und Kelava (2008) als gut zu bezeichnen sind. Mit einem Wert von Cronbachs $\alpha = 0,80$ weist die Skala insgesamt eine gute interne Konsistenz auf (Bortz & Döring, 2006).

Skala II.2 Überforderung. Auf dem zweiten Faktor (Eigenwert = 4,4) luden primär sieben Items, von denen nur ein Item (\rightarrow Item 49) eine Doppelladung aufwies und zusätzlich auf dem siebten Faktor lud. Eine Übersicht über die Ladungen zeigt Anhang H. Basierend auf inhaltlichen und statistischen Abwägungen wurde dieses Item gemeinsam mit vier weiteren Items (\rightarrow Item 20, 54, 51 und 56) von weiteren Betrachtungen ausgeschlossen.

Es resultieren drei Items, die verschiedene Facetten der Inhaltsvermittlung abbilden und gemeinsam die Skala *Überforderung* darstellen. Mit Item 45 („Der Umfang der zu lernenden

Inhalte hat mich überfordert“), Item 48 („Das Tempo der Stoffvermittlung war zu hoch für mich“) sowie Item 52 („Die Dozenten passten die vermittelten Inhalte zu wenig an den Wissensstand der Teilnehmer an“) erfasst die Skala die Angemessenheit von Umfang, Tempo und Anforderungsniveaus der vermittelten Inhalte aus Sicht der Teilnehmer. Tabelle 7 zeigt, dass alle drei Items gute Trennschärfen aufweisen (vgl. Moosbrugger und Kelava, 2008). Die interne Konsistenz der Skala gilt nach der Arbeit von Cortina (1994) mit einem Wert von Cronbachs $\alpha = 0,77$ als befriedigend.

Skala II.3 Gruppenarbeit. Auf dem dritten Faktor (Eigenwert = 2,8) luden sieben Items, von denen nur eins (\rightarrow Item 66) eine Doppelladung aufwies und zusätzlich auch auf dem siebten Faktor lud. Eine Übersicht über die Ladungen zeigt Anhang H. Statistisch sowie inhaltlich begründet wurde dieses Item gemeinsam mit drei weiteren Items (\rightarrow Item 68, 70 und 73) eliminiert.

Die resultierenden drei Items (d. h. Item 67: „Ich habe bei den Gruppenarbeiten viel von anderen Teilnehmern gelernt“, Item 69: „Alle Teilnehmer haben sich, entsprechend ihrer Kompetenzen, an den Gruppenarbeiten beteiligt“ und Item 72: „Die Gruppenarbeiten haben den Austausch zwischen den Teilnehmern gefördert“), erfassen Aspekte der Zusammenarbeit, die durch die Unterrichtsmethode *Gruppenarbeit* gefördert werden sollen. Infolgedessen bilden diese drei Items die gleichnamige Skala. Wie Tabelle 7 erkennen lässt, liegen die Trennschärfen der drei Items in dem von Moosbrugger und Kelava (2008) als gut definierten Bereich. Mit einem Wert von Cronbachs $\alpha = 0,75$ gilt die interne Konsistenz der Skala nach Cortina (1994) ebenfalls als befriedigend.

Skala II.4 Struktur. Insgesamt vier Items luden primär auf dem vierten Faktor (Eigenwert = 2,3). Eine Übersicht über die Ladungen zeigt Anhang H. Weil das Item 32 im Wesentlichen eine Zusammenfassung der übrigen drei Items darstellte, blieb dieses bei der Skalenbildung unberücksichtigt.

Die übrigen drei Items (d. h. Item 33: „Ich konnte im Verlauf der Lehrveranstaltung die Gliederung immer nachvollziehen“, Item 34: „Ich finde, die Lehrveranstaltung gab einen guten Überblick über das Themengebiet“ und Item 35: „Zu Beginn der Lehrveranstaltung wurden die Lernziele deutlich dargestellt“) erfassen die Transparenz, Gliederung und Zielklarheit der Lehrveranstaltung und formen infolgedessen die Skala *Struktur*. Wie Tabelle 7 zu entnehmen ist, liegen die Trennschärfen der drei Items alle in dem von Moosbrugger und Kelava (2008) als gut definierten Bereich. Mit einem Wert von Cronbachs $\alpha = 0,76$ gilt die interne Konsistenz dieser Skala als befriedigend (vgl. Cortina, 1994).

Skala II.5 Planübung. Auf dem fünften Faktor (Eigenwert = 2,1) luden insgesamt vier Items einfach (vgl. Anhang H), die schließlich allesamt bei der Skalenbildung berücksichtigt wurden. Die vier Items (d. h. Item 75: „Ich habe bei der Planübung viel gelernt“, Item 76: „Die Dozenten gaben mir nützliches Feedback zu meinen Leistungen bei der Planübung“, Item 77: „In der Planübung konnte ich das neuerworbene Wissen anwenden“ und Item 78: „Der Schwierigkeitsgrad der Planübung war angemessen“) erfassen unterschiedliche Aspekte der

Tabelle 7

Übersicht über die erfolgsbeschreibenden Skalen.

Skala	Cronbachs α	$MW (SD)$	Items	r_{it}
(II.1) Dozentenverhalten	0,80	6,2 (6,7)		
			Item 19	0,59
			Item 23	0,62
			Item 26	0,67
			Item 28	0,58
			Item 42	0,56
(II.2) Überforderung	0,77	5,3 (1,1)		
			Item 45	0,69
			Item 48	0,70
(II.3) Gruppenarbeit	0,75	5,6 (0,9)		
			Item 52	0,46
(II.4) Struktur	0,76	5,8 (0,8)		
			Item 67	0,60
			Item 69	0,61
(II.5) Planübung	0,86	5,8 (0,9)		
			Item 72	0,57
(II.6) EVA	0,76	5,7 (0,8)		
			Item 33	0,58
			Item 34	0,66
(II.7) Didaktik	0,78	5,4 (1,1)		
			Item 35	0,56
(II.8) Gruppe	0,74	6,3 (0,6)		
			Item 75	0,65
			Item 76	0,70
			Item 77	0,79
			Item 78	0,69
(II.9) EVA	0,76	5,7 (0,8)		
			Item 59	0,56
			Item 62	0,56
			Item 63	0,49
			Item 64	0,51
(II.10) Gruppe	0,74	6,3 (0,6)		
			Item 65	0,56
(II.11) Didaktik	0,78	5,4 (1,1)		
			Item 15	0,65
			Item 16	0,67
(II.12) Gruppe	0,74	6,3 (0,6)		
			Item 17	0,59
(II.13) Gruppe	0,74	6,3 (0,6)		
			Item 94	0,50
			Item 96	0,69
(II.14) Gruppe	0,74	6,3 (0,6)		
			Item 97	0,53

Unterrichtsmethode *Planübung* und bilden folglich die gleichnamige Skala. Mit Ausnahme von Item 77 weisen die Items gute Trennschärfen auf (vgl. Moosbrugger & Kelava, 2008). Item 79 weist hingegen eine sehr hohe Trennschärfe auf, was bedeutet, dass dieses Item sehr ähnlich misst wie die gesamte Skala. Aus inhaltlichen Gründen (d. h. basierend auf Expertenurteilen) wurde dieses Item dennoch in die Skala aufgenommen. Mit einem Wert von Cronbachs $\alpha = 0,86$ weist die Skala eine gute Reliabilität auf (Bortz & Döring, 2006).

Skala II.6 EVA. Fünf Items luden gleichermaßen eindeutig auf dem sechsten Faktor (Eigenwert = 2,0). Eine Übersicht über die Ladungen zeigt Anhang H. Da die Items unterschiedliche, aber durchweg relevante Aspekte der Methode EVA abbilden, wurde von einer Item-Reduktion abgesehen. Stattdessen gingen alle fünf Items (d. h. Item 59: „Ich hatte ausreichend Zeit, die EVA-Aufgaben zu bearbeiten“, Item 62: „Der Umfang der EVA-Aufgaben war angemessen“, Item 63: „Die EVA-Aufgaben haben mein Verständnis der Inhalte weiter vertieft“, Item 64: „Der Schwierigkeitsgrad der EVA-Aufgaben war angemessen“ und Item 65: „Die Aufgabenstellungen der EVA-Aufgaben waren verständlich“) in die Bildung der Skala *EVA* ein. Wie der Tabelle 7 zu entnehmen ist, weisen die Items durchweg gute Trennschärfen auf (vgl. Moosbrugger & Kelava, 2008). Die interne Konsistenz der Skala gilt nach Cortina (1994) mit Cronbachs $\alpha = 0,77$ als befriedigend.

Skala II.7 Didaktik. Insgesamt vier Items luden primär auf dem siebten Faktor (Eigenwert = 1,6). Eine Übersicht über die Ladungen zeigt Anhang H. Aufgrund von Doppelladung und inhaltlichen Diskrepanzen wurde ein Item (\rightarrow Item 47) von weiteren Betrachtungen ausgeschlossen.

Die übrigen drei Items erfassen Aspekte der methodischen und didaktischen Gestaltung eines Lehrgangs (z. B. die Wahl der Methoden oder das Verhältnis von praktischen und theoretischen Anteilen) und stellen infolgedessen die als *Didaktik* betitelte Skala dar. Konkret handelt es sich hierbei um Item 15 („Ich finde, theoretische und praktische Anteile standen in einem angemessenen Verhältnis zueinander“), Item 16 („Es gab ausreichend Gelegenheiten, das Gelernte praktisch zu erproben“) und Item 17 („Ich finde, die Lehrmethoden waren gut auf unsere Teilnehmergruppe abgestimmt“). Die Trennschärfen dieser drei Items (siehe hierfür Tabelle 7) liegen ebenfalls in dem von Moosbrugger und Kelava (2008) als gut definiertem Bereich. Mit einem Wert von Cronbachs $\alpha = 0,78$ weist die Skala eine befriedigende interne Konsistenz auf (vgl. Cortina, 1994).

Skala II.8 Gruppe. Auf dem achten Faktor (Eigenwert = 2,1) luden schließlich drei Items einfach (vgl. Anhang H), die allesamt bei der Skalenbildung berücksichtigt wurden. Die drei Items thematisieren die Aktivität der anderen Teilnehmer, die gegenseitige Unterstützung sowie den Zusammenhalt der Teilnehmergruppe. Resultierend bilden sie die sogenannte Skala *Gruppe*. Konkret handelt es sich hierbei um Item 94 („Die anderen Teilnehmer brachten sich aktiv ein“), Item 96 („Die Teilnehmer unterstützen sich gegenseitig“) sowie Item 97 („Ich finde, es herrschte ein guter Zusammenhalt im Lehrgang“), die alle drei jeweils gute

Tabelle 8

Interkorrelationen der erfolgsbedingenden Skalen.

	Skala II.1	Skala II.2	Skala II.3	Skala II.4	Skala II.5	Skala II.6	Skala II.7	Skala II.8
Skala II.1	1	0,28**	0,37**	0,45**	0,51**	0,25**	0,48**	0,36**
Skala II.2		1	0,03	0,29**	0,26**	0,27**	0,49**	0,23**
Skala II.3			1	0,33**	0,33**	0,26**	0,24**	0,34**
Skala II.4				1	0,35**	0,29**	0,43**	0,33**
Skala II.5					1	0,33**	0,39**	0,32**
Skala II.6						1	0,27**	0,24**
Skala II.7							1	0,38**
Skala II.8								1

Anmerkungen: ** $p < 0,01$ (zweiseitig). Skala II.1 = Dozentenverhalten; Skala II.2 = Überforderung; Skala II.3 = Gruppenarbeit; Skala II.4 = Struktur; Skala II.5 = Planübung; Skala II.6 = EVA; Skala II.7 = Didaktik; Skala II.8 = Gruppe.

Trennschärfen aufweisen (siehe hierfür Tabelle 7; vgl. Moosbrugger & Kelava, 2008). Die interne Konsistenz dieser Skala ist mit einem Wert von Cronbachs $\alpha = 0,74$ als befriedigend zu werten (vgl. Cortina, 1994).

Verteilungscharakteristiken der Skalen. Die Verteilungsparameter zeigen darüber hinaus, dass alle acht Skalen linksschief verteilt sind (Schiefe $\leq -0,7$). Alle Skalen weisen zudem einen Mittelwert $\geq 5,3$ sowie eine Standardabweichung $\leq 1,1$ auf (siehe Tabelle 7). Tabelle 8 zeigt die Inter-Korrelationen der Skalen. Diese liegen durchweg in einem geringen bis mittleren Bereich (d. h. $r \leq 0,51$).

6.5.3 Bias-Effekte

Mit Ausnahme von vier Items (\rightarrow Item 38, 86, 99 und 100) korrelierten die benannten Bias-Items signifikant mit der Gesamtbenotung des Lehrgangs ($0,18 \leq r \leq 0,59$; $p < 0,01$; vgl. Anhang J). In den meisten Fällen korrelierten sie dabei nur gering mit dem Gesamturteil. Lediglich drei Items wiesen eine mittlere ($r \geq 0,30$; Item 84, 113 und 121) und zwei Items eine hohe Korrelation mit der Gesamtbenotung des Lehrgangs auf ($r \geq 0,50$; Item 119 und 125). Diese fünf Items erfassen die selbsteingeschätzte Vorbereitung des Teilnehmers auf den Lehrgang (Item 84) und die Prüfung (Item 125), die Angemessenheit von Teilnehmergröße (Item 119) und lehrgangsbedingtem Zeitaufwand (Item 113), sowie eventuelle Ablenkungen des Teilnehmers (z. B. durch private oder berufliche Verpflichtungen; Item 121). Basierend auf statistischen sowie inhaltlichen Überlegungen wurden diese fünf Items beibehalten, während die übrigen Items (\rightarrow Item 37, 57, 103, 104, 111, 112 und 122) als weniger bedeutsam identifiziert wurden und infolge der Itemreduktion im Fragebogen keine weitere Berücksichtigung finden sollten.

Tabelle 9

Korrelationen zwischen den erfolgsbedingenden Skalen und den vier Lehrerfolgskriterien.

	Kompetenz- erwerb (Skala I.1)	Vorbereitung & Transfer (Skala I.2)	Gesamt- benotung (Item 145*)	Subjektiver Lerngewinn (Item 123)
Dozentenverhalten	0,34**	0,34**	0,43**	0,33**
Überforderung	0,15*	0,26**	0,36**	0,24**
Gruppenarbeit	0,21**	0,26**	0,18**	0,41**
Struktur	0,33**	0,37**	0,48**	0,43**
Planübung	0,45**	0,38**	0,41**	0,32**
EVA	0,29**	0,32**	0,33**	0,38**
Didaktik	0,24**	0,50**	0,53**	0,33**
Gruppe	0,17*	0,25**	0,29**	0,29**

Anmerkung: * $p < 0,05$ (zweiseitig), ** $p < 0,01$ (zweiseitig).

6.5.4 Erste Validierungen

Kriteriumsvaliditäten. Die Korrelationen zwischen den vorab definierten prädiktiven Skalen und den ersten vier in Abschnitt 6.1.4 benannten Validierungskriterien werden in Tabelle 9 dargestellt. Dabei zeigten sich durchweg signifikante Zusammenhänge zwischen den einzelnen Skalen und den unterschiedlichen Kriterien ($r > 0,15$; $p < 0,05$ bzw. $p < 0,01$), wobei die Höhe der Korrelationen dabei durchaus variierte. Die höchsten Zusammenhänge ergaben sich zwischen der Skala Didaktik und der Gesamtbenotung einerseits ($r = 0,53$) sowie dem vorausgesagten Transfer andererseits ($r = 0,50$). Die Skalen Struktur und Dozentenverhalten korrelierten ebenfalls am stärksten mit der Gesamtbenotung ($r = 0,48$ und $r = 0,43$). Gleiches galt für die Skala Überforderung ($r = 0,36$), die jedoch insgesamt nur geringe Zusammenhänge mit den Kriterien aufwies (vgl. Tabelle 9). Dahingegen korrelierte die Skala Planübung am höchsten mit dem selbsteingeschätzten Kompetenzerwerb ($r = 0,45$), während die Skalen Gruppenarbeit und EVA am deutlichsten mit dem subjektiven Lerngewinn korrelierten ($r = 0,41$ und $r = 0,38$). Die Skala Gruppe weist hingegen durchweg nur sehr niedrige Korrelationen mit den definierten Kriterien auf (vgl. Tabelle 9).

Prognostische Validität. Es zeigten sich zunächst keine signifikanten Korrelationen zwischen den verschiedenen Globalurteilen (d. h. der Gesamtbenotung der Lehrveranstaltung sowie der Selbsteinschätzung von Lerngewinn, Kompetenzerwerb und Vorbereitung auf die Praxis bzw. Transfer) einerseits und den mündlichen sowie schriftlichen Prüfungsnoten der Teilnehmer andererseits ($r < 0,21$; siehe digitaler Anhang). Ergänzend zeigten Regressionsanalysen, dass die acht Skalen (= Prädiktoren) gemeinsam nur 29,5% der Varianz in den

schriftlichen Noten sowie 10,1% der Varianz in den mündlichen Noten aufklären konnten. Letzteres wurde zudem nicht signifikant (siehe digitaler Anhang).

An dieser Stelle sei jedoch darauf hingewiesen, dass diese Berechnungen auf stark reduzierten Stichproben basieren. So gaben nur 38,3% der Befragten ihre schriftliche und 41,7% der Befragten ihre mündliche Prüfungsnote an.

Divergente Validität. Der Zusammenhang zwischen der Gesamtbenotung der Teilnehmer und ihrer Stimmung lag mit einer Korrelation von $r = 0,27$ ($p < 0,001$) im unterem mittleren Bereich. Mit Ausnahme der Skala Dozentenverhalten ($r = 0,09$) zeigten sich zudem auch signifikante Zusammenhänge zwischen den definierten Skalen und der Stimmung der Teilnehmer (siehe digitaler Anhang). Mit $0,13 \leq r \leq 0,34$ lagen diese Zusammenhänge in einem geringen bis mittleren Bereich.

6.6 Diskussion

Ziel dieser Studie war es, basierend auf einer umfassenden empirischen Erprobung der Fragebogenerstversion, eine begründete Itemselektion vorzunehmen, um schließlich ein wissenschaftlich fundiertes Standardinstrument zur Evaluation der Führungsausbildung am IdF NRW zu entwickeln. Hierzu wurden die Items zunächst hinsichtlich ihrer Verteilungsscharakteristiken analysiert. Es wurde ersichtlich, dass dem Großteil aller Items im Mittel jeweils im positiven Sinne (eher bis vollkommen) zugestimmt wurde. Ungeachtet dessen, dass die meisten Items bereits absichtlich extrem formuliert waren (vgl. Abschnitt 4), wurden die unterschiedlichen Aspekte der Lehre am IdF NRW durchweg positiv bewertet. Zudem war das Meinungsbild der Befragten recht homogen (d. h. es gab nur wenig Varianz in den Antworten), obwohl die Pilotierung bereits mit einer breiten Menge an Teilnehmern aus ganz unterschiedlichen Lehrgängen (d. h. verschiedenste Lehrgangsformen einerseits sowie Dozententeams andererseits) durchgeführt worden war und Selektionseffekte aufgrund der hohen Ausschöpfungsquoten (vgl. Abschnitt 6.1.1) ebenfalls auszuschließen waren.

Bei 21 Items war das Bild schließlich so eindeutig, dass diese dem IdF NRW im Zuge des groß angelegten Evaluationsprojektes zwar durchaus positiv zurückgemeldet wurden (vgl. hierzu Loberg, 2016), im Zuge einer Standardevaluation aber aufgrund der stark eingeschränkten Variant keinen Mehrwert bringen würden und infolgedessen nicht weiter berücksichtigt wurden. Weitere Item-Reduktionen erfolgten vor dem Hintergrund unterschiedlicher statistischer sowie inhaltlicher Überlegungen.

6.6.1 Skalenbildung

Die verbleibenden Items ließen sich mittels zweier exploratorischer Faktorenanalysen zu insgesamt zehn Dimensionen erfolgreicher Lehrveranstaltungen am IdF NRW gruppieren. So brachte eine erste Faktorenanalyse, in die all diejenigen Items eingingen, mit denen sogenannte Outcome-Variablen (sprich Aspekte des Lehrerfolgs; vgl. Rindermann, 2001) erfasst werden, zunächst zwei Dimensionen hervor. Diese wurden nach inhaltlicher Inspektion

als „Kompetenzerwerb“ und „Vorbereitung & Transfer“. Eine zweite Faktorenanalyse, in die die Erfolgsdeterminanten (sprich den Lehrerfolg bedingende Aspekte) eingingen, brachte weitere acht Dimensionen hervor. Hierbei handelte es sich um die Dimensionen „Dozentenverhalten“, „Überforderung“, „Gruppenarbeit“, „Struktur“, „Planübung“, „EVA“ und „Didaktik“. Die Dimensionen umfassten dabei zunächst bis zu 15 Items.

Um die herausgearbeiteten Dimensionen mit dem Fragebogen angemessen abzubilden bzw. zu erfassen, wurde für jede Dimension eine Auswahl an drei bis fünf Items getroffen und zu einer gleichnamigen Skala zusammengefasst. Die Auswahl der Items erfolgte statistisch sowie inhaltlich (d. h. unter Rückgriff auf Expertenmeinungen) begründet. Dabei wurde besonders darauf geachtet, dass auch unterschiedliche Facetten einer Dimension berücksichtigt wurden, auch wenn dies zu Lasten der internen Konsistenz der jeweiligen Skala ging. Dennoch weisen alle zehn Skalen letztendlich befriedigende bis gute Reliabilitäten auf.

6.6.2 Zusatzskala „Einsatzübung“

Aufgrund der vorher definierten Ausschlusskriterien fanden die vier Items zur Bewertung der Einsatzübungen (d. h. Item 79, 80, 81 und 82) keinen Eingang in die exploratorische Faktorenanalyse. Als eine zentrale Unterrichtsmethode stellt die Einsatzübung jedoch einen wichtigen Bestandteil der Führungsausbildung am IdF NRW dar (z. B. AG-BIII, 2007; AG-FIII, 2005; AG-F IV, 2007; Lehrgangsbeschreibung-BIII-Pilot, 2015; Lehrgangsbeschreibung-FIII-Pilot, 2015; Lehrgangsbeschreibung-FIV, 2015) und sollte im Zuge einer Lehrveranstaltungsevaluation entsprechend unbedingt berücksichtigt werden. Folglich werden die vier benannten Items ebenfalls in den Standardfragebogen aufgenommen (siehe Anhang K). Da es sich hierbei inhaltlich um die gleichen Items handelt wie zur Bewertung der Planübung (vgl. Skala II.5 Planübung) und dabei lediglich das Wort „Planübung“ durch das Wort „Einsatzübung“ ausgetauscht wurde, ist zu vermuten, dass diese vier Items eine entsprechend ähnliche Skala bilden. Dies gilt es jedoch in Folgestudien zu prüfen.

6.6.3 Interpretation der Skalenwerte

Zur Beurteilung der herausgearbeiteten Dimensionen guter Lehre am IdF NRW wurden zunächst Skalenwerte (d. h. die Mittelwerte aus den jeweils drei bis fünf zugrundeliegenden Items einer Skala) gebildet. Infolge ihrer Berechnungsweise konnten die Skalenwerte ebenfalls zwischen eins und sieben variieren (vgl. 7-stufige-Likert-Skala des Fragebogens). Dabei sind Skalenwerte ≥ 5 als positive Bewertung der jeweiligen Dimension zu interpretieren, während Skalenwert ≤ 3 auf eine negative Bewertung hindeuten. Ein Skalenwert = 4 repräsentiert hingegen eine neutrale Bewertung der dahinterstehenden Dimension. Insgesamt gilt: Je höher ein Skalenwert, umso besser die Bewertung der dahinterstehenden Dimensionen – vorausgesetzt die Items 45, 48 und 52 der Skala Überforderung wurden vorher adäquat umgepolt.

Zudem wurden die Standardabweichungen und Schiefe der Skalen berechnet. Die Verteilungskennwerte der Skalen verdeutlichten nochmals, dass die Lehrveranstaltungen am IdF

NRW auf unterschiedlichen Dimensionen durchweg positiv beurteilt wurden. Das Meinungsbild war zudem sehr homogen. Diese Befunde stehen im Einklang mit den Ergebnissen der Lehrevaluationsforschung im universitären Kontext. Auch hier zeigten sich immer wieder, dass Vorlesungen und Seminare von Studierenden eher positiv als negativ bewertet werden (z. B. Rindermann, 1996; Pellert, Spiel, Wolf, Popper & Poschauko, 2000).

Abschließend sei noch angemerkt, dass es sich aufgrund der mitunter sehr starken inhaltlichen Heterogenität der Items einer jeweiligen Skala ebenfalls empfiehlt, neben der Berechnung und Rückmeldung von Skalenwerten (d. h. Mittelwerten) stets auch die Ergebnisse auf Itemebene zu betrachten.

6.6.4 Interkorrelation der Skalen

Zwischen den herausgearbeiteten Skalen zeigten sich mit einer Ausnahme (d. h. zwischen Skala II.2 „Überforderung“ und Skala II.3 „Gruppenarbeit“) durchweg signifikant positive Zusammenhänge. Die mittelhohen Interkorrelationen der Skalen lagen dabei in einem Bereich, der auch in Arbeiten zur universitären Lehrevaluation berichtet und vielfach diskutiert wurde (z. B. Rindermann, 2001; Staufenbiel, 2000). Dabei nehmen einige Autoren an, dass die Bewertungen einzelner Dimensionen keineswegs unabhängig erfolgen, sondern letztendlich allesamt Ausdruck eines globalen Zufriedenheitsurteils hinsichtlich der Lehrveranstaltungsqualität sind (vgl. Rindermann, 1996; Souvignier & Gold, 2002). Demgegenüber steht die Position, dass Studierende durchaus in der Lage sind, die unterschiedlichen Dimensionen guter Lehre differenziert zu beurteilen und „studentische Rückmeldungen [schließlich] adäquate Abbildungen der Multidimensionalität von Lehre darstellen“ (Souvignier & Gold, 2002, S. 270; vgl. auch Marsh, 1983).

Auf Basis der vorliegenden Daten kann nicht ausgeschlossen werden, dass die Beurteilungen der unterschiedlichen Dimensionen durchweg durch einen globalen (Zufriedenheits-) Faktor beeinflusst werden, der die signifikanten Interkorrelationen der Skalen erklärt. Dennoch werden die Zusammenhänge hier in erster Linie auf eine tatsächliche Abhängigkeit der unterschiedlichen Dimensionen zurückgeführt. So wird angenommen, dass die Dimensionen nicht nur direkt, sondern auch indirekt auf den Erfolg einer Lehrveranstaltung wirken, indem sie sich zum Teil durchaus wechselseitig bedingen (vgl. hierzu auch Abschnitt 2.3.1)

6.6.5 Erste Validierungen

Kriteriumsvaliditäten. Es zeigten sich durchweg signifikant positive Zusammenhänge zwischen den acht erfolgsbedingenden Skalen einerseits sowie den vier Lehrerfolgskriterien (d. h. dem selbsteingeschätzten Kompetenzerwerb, dem antizipierten Transfer, der Gesamtbewertung des Lehrgangs und dem subjektiven Lerngewinn) andererseits.

Der selbsteingeschätzte Kompetenzerwerb (Skala I.1) zeigte dabei den höchsten Zusammenhang mit der Beurteilung der Planübung (Skala II.5). Je positiver die verschiedenen Erfolgsaspekte der Unterrichtsmethode Planübung (d. h. die Möglichkeit, neuerworbenes Wissen anwenden zu können, nützliches Feedback zu bekommen und viel zu lernen sowie

auch die Angemessenheit des Schwierigkeitsgrades; vgl. Abschnitt 2.1) bewertet wurden, desto höher wurde auch der Erwerb der relevanten Fach- und Führungskompetenzen eingeschätzt. Da es sich hierbei ausschließlich um einen korrelativen Zusammenhang handelt, lässt sich an dieser Stelle nur mutmaßen, dass durch eine gute Planübung der Kompetenzerwerb der Teilnehmer gefördert wird.

Die empfundene Vorbereitung auf die Praxis sowie der antizipierte Transfer (Skala I.2) weisen einen hohen Zusammenhang mit der Beurteilung der Didaktik der Dozenten (Skala I.7) auf. Das heißt, dass ein angemessenes Verhältnis von theoretischen und praktischen Anteilen sowie ausreichend Gelegenheiten, das Gelernte zu erproben und eine gute Anpassung der Lehrmethoden an die Bedürfnisse der jeweiligen Teilnehmergruppe insgesamt mit einer besser eingeschätzten Vorbereitung auf die Praxis bzw. antizipiertem Transfer einhergeht.

Die Gesamtbenotung des Lehrgangs (Item 145*) weist neben einem hohen Zusammenhang mit der Didaktik (Skala I.7) ebenfalls mittelhohe Zusammenhänge mit den Beurteilungen von Struktur (Skala I.4) sowie dem sozial-interaktiven Verhalten der Dozenten (Skala I.1) auf. So gilt zunächst: Je teilnehmer- und praxisnaher die Gestaltung des Lehrgangs, umso besser das Gesamturteil. Des Weiteren gilt auch: Je klarer und strukturierter die Lehrveranstaltung und/oder je besser die Bewertung verschiedener Aspekte des Dozentenverhaltens, umso besser das Gesamturteil.

Der subjektive Lerngewinn (Item 123) zeigt schließlich den höchsten Zusammenhang mit den Beurteilungen von Struktur (Skala I.4) sowie der Unterrichtsmethode Gruppenarbeit (Skala I.3). Das heißt einerseits, je nachvollziehbarer die Gliederung der Veranstaltung, je deutlicher die Darstellung der Lernziele und je besser der Überblick, der über das Themengebiet gegeben wurde, bewertet wurden, umso besser wurde auch der subjektive Lerngewinn beurteilt. Andererseits gilt auch: Je besser die Gruppenarbeit hinsichtlich verschiedener Erfolgsaspekte (d. h. der Förderung des Austausches untereinander und des Lernens voneinander) bewertet wurden, desto positiver wurde der subjektive Lerngewinn eingeschätzt.

Insgesamt betrachtet lassen sich die verschiedenen Facetten des Lehrerfolgs auf Basis der vorliegenden Daten jeweils durch unterschiedliche Skalen am besten voraussagen. In Folgestudien gilt es die Befunde dieser ersten Validierung regressionsanalytisch zu prüfen.

Prognostische Validität. Dahingegen zeigten sich keine signifikanten Zusammenhänge zwischen den vier Lehrerfolgskriterien (d. h. dem selbsteingeschätzten Kompetenzerwerb, dem antizipierten Transfer, der Gesamtbenotung des Lehrgangs und dem subjektiven Lerngewinn) einerseits sowie den mündlichen und schriftlichen Prüfungsnoten der Teilnehmer andererseits. Zudem klärten die acht erfolgsbedingenden Skalen gemeinsam nur einen sehr geringen Anteil der Varianz in den Prüfungsnoten der Befragten auf. Folglich lassen sich anhand der Lehrgangsbeurteilungen keine Aussagen über den Prüfungserfolg der Teilnehmer machen. Allerdings basieren diese Befunde auf einer sehr eingeschränkten Stichprobe.

Dies ist in Teilen zunächst darauf zurückzuführen, dass in den neuen Lehrgangsformen auf die Vergabe von Noten verzichtet wird und die Befragten aus diesen Lehrgängen entsprechend keine Angaben machen konnten. Jedoch machten auch die Teilnehmer aus den alten Lehrgängen nur vereinzelt Angaben zu ihren mündlichen und/oder schriftlichen Noten. Infolgedessen sind Selektionseffekte und daraus resultierende Ergebnisverzerrungen an dieser Stelle nicht auszuschließen.

Divergente Validität. Es zeigten sich signifikant positive Zusammenhänge zwischen der Bewertung der Lehrveranstaltung (d. h. sowohl die Gesamtbenotung als auch die Beurteilung der einzelnen Dimensionen) und der zu Befragungsbeginn angegebenen Stimmung der Teilnehmer. Diese Zusammenhänge sind jedoch durchweg eher gering. Infolgedessen ist zwar nicht auszuschließen, dass die Urteile der Teilnehmer mitunter auch durch ihre aktuelle Stimmung beeinflusst werden. Dennoch wäre ein solcher Einfluss –wenn überhaupt– nur sehr gering, sodass hingegen auszuschließen ist, dass die Beurteilung einer Lehrveranstaltung lediglich die aktuelle Stimmung eines Teilnehmers widerspiegelt. Die Ergebnisse liefern damit erste Hinweise auf eine divergente Validität.

6.6.6 Biasvariablen

In dieser Studie wurde außerdem auch eine Reihe potentieller Biasvariablen untersucht. Dabei zeigten sich sowohl Übereinstimmungen mit den im universitären Kontext identifizierten Biasvariablen, als auch einige kontextspezifische, potenziell urteilsverzerrenden Merkmale, „die grundsätzlich nichts mit der Lehrqualität zu tun haben“ (Wolf et al., 2001, S. 91).

So zeigte sich in Übereinstimmung mit der Literatur zur universitären Lehrveranstaltungsevaluation, dass die als mehr oder weniger angemessen bewertete Gruppengröße eines Lehrgangs (Item 119) einen hohen positiven Zusammenhang mit der Gesamtbeurteilung der Veranstaltung aufwies (vgl. Marsh, 1987; Staufenbiel, 2000). Ein ebenfalls hoher positiver Zusammenhang zeigte sich zudem zwischen der empfundenen Vorbereitung auf die bevorstehenden Prüfungen (Item 125) und dem Gesamturteil. Darüber hinaus ergaben sich mittlere positive Zusammenhänge zwischen der Gesamtbeurteilung der Lehrveranstaltung einerseits sowie der subjektiv empfundenen Vorbereitung des Teilnehmers auf die Veranstaltung (Item 84), der mit der Ausbildung verbundene Zeitaufwand (Item 113; vgl. Rindermann, 2001) und der wahrgenommenen Möglichkeit sich voll und ganz auf den Lehrgang zu konzentrieren (Item 121) andererseits.

In Anbetracht dessen, dass es sich hierbei jedoch nur um korrelative Zusammenhänge handelt, können an dieser Stelle (noch) keine Aussagen dazu getroffen werden, inwiefern die benannten Variablen das Urteil der Lehrgangsteilnehmer tatsächlich verzerren. Dies gilt es, in Folgestudien an einer unabhängigen Stichprobe regressionanalytisch zu überprüfen. Hierfür und um eventuelle Bias-Effekte entsprechend kontrollieren zu können, wurden die vier Items in den Standardevaluationsbogen aufgenommen.

Die übrigen Biasvariablen fanden hingegen keinen Eingang in den Standardfragebogen, weil sie entweder nur sehr geringe oder keine bedeutsamen Zusammenhänge mit der Gesamtbeurteilung des Lehrgangs aufwiesen. Darunter fielen einerseits die erkennbare Prüfungsrelevanz der Inhalte (Item 57; vgl. Grabbe, 2003), die Qualität und Verfügbarkeit von Übungsmaterialien (Item 103, 104 und 105; vgl. Rindermann, 2001), die als angemessen empfundenen Unterrichts-, Lern- und Pausenzeiten (Item 37 und 112; vgl. Moosbrugger & Schweizer, 2002), sowie auch der auf den Teilnehmern lastende Druck von Außenstehenden, diesen Ausbildungslehrgang erfolgreich zu absolvieren (Item 122). Andererseits fielen darunter auch Zeitverluste durch Absprachen und Materialbeschaffungen (Item 38), die Vorbereitung von der Heimatstelle (Item 86) sowie die Zufriedenheit mit Unterbringung (Item 86) und Verpflegung (Item 100) durch das IdF NRW.

6.6.7 Einschränkungen der Ergebnisse

Die Ermittlung der Faktorenstruktur erfolgte auf Basis von $n = 243$ Datensätzen. Eine hinreichend große Stichprobe stellt eine wesentliche Voraussetzung dafür dar, dass eine Faktorenanalyse verlässliche und interpretierbare Ergebnisse liefert (z. B. Beavers, Lounsbury, Richards, Huck, Skolits & Esquivel, 2013; Hirschfeld, von Brachel & Thielsch, 2014). Allerdings existiert in der Literatur kein klarer Konsens darüber, ab wann eine Stichprobe als hinreichend groß gilt. Während die Autoren West und Kollegen (1995) sowie auch Gorsuch (1983) von mindestens 200 Fällen sprechen, fordern Lawley und Maxwell (1971) eine Stichprobengröße, die die Anzahl der Variablen um mindestens 51 Fälle übersteigt (d. h. $n \geq 17 + 51$ für die erste sowie $n \geq 50 + 51$ für die zweite Faktorenanalyse). Beide Kriterien wurden in dieser Studie erfüllt.

Andere Autoren formulieren hingegen deutlich strengere Voraussetzungen, denen die vorliegende Studie nicht gerecht wird. So spricht Norušis (2005) von mindestens 300 Fällen, während Bryant und Yarnold (1995) für jedes Item mindestens 10 Datensätze sowie ein Verhältnis zwischen Teilnehmern und der Anzahl an Items („Subjects-to-Variables ratio“; STV ratio) fordern, welches nicht kleiner als fünf ist. Dabei trifft Letzteres zwar für die erste (mit einem STV ratio von 14,29) nicht aber für die zweite Faktorenanalyse dieser Studie (mit einem STV ratio von 4,86) zu. Auch Suhr (2006) setzt neben einer weniger strengen und allgemeinen Mindestgrenze von 100 Datensätzen ebenfalls voraus, dass das Verhältnis von Teilnehmern und Items mindestens einen Wert von fünf aufweist. Comrey und Lee (1992) kategorisieren die Eignung der Stichprobe zur Durchführung einer Faktorenanalyse bei $n = 50$ als *sehr schlecht* („very poor“), bei $n = 100$ als *schlecht* („poor“), bei $n = 200$ als *mittelmäßig* („fair“), bei $n = 300$ als *gut* („good“), bei $n = 500$ als *sehr gut* („very good“) und bei $n > 1000$ schließlich als *exzellent* („excellent“).

Vor dem Hintergrund der unterschiedlichen Konventionen lässt sich festhalten, dass die zugrundeliegende Stichprobe gemäß einiger Autoren durchaus hinreichend ist, um verlässliche und interpretierbare Ergebnisse zu erzielen. Der Meinung anderer Autoren zufolge wäre

jedoch eine größere Stichprobe erforderlich gewesen, um die Stabilität der Faktoren als gesichert annehmen zu können. Dies sollte in Folgestudien berücksichtigt werden.

Des Weiteren wird die Aussagekraft der Ergebnisse durch die nicht vorhandene Unabhängigkeit der zugrundeliegenden Stichprobe eingeschränkt. So basieren sowohl die beiden Faktorenanalysen als auch die Validitätsberechnungen der gebildeten Skalen auf ein und derselben Stichprobe, die für erste explorative Zwecke künstlich getrennt wurde. Weitere Validierungen sollten auf den Daten einer unabhängigen Stichprobe erfolgen.

7 Allgemeine Diskussion

Ziel dieser Arbeit war es, ein Instrument zur Evaluation der Gruppen- und Zugführerlehrgänge am IdF NRW zu entwickeln. Hierzu wurden basierend auf den Befragungen von Experten zunächst qualitäts- und erfolgsentscheidende Faktoren guter Lehre im feuerwehrtechnischen Ausbildungskontext identifiziert (vgl. Forschungsfrage 1a).

In Studie I konnten zunächst unterschiedliche Determinanten herausgestellt werden, die für den Erfolg einer Lehrveranstaltung aus Teilnehmer- und Dozentensicht potentiell relevant sind. Diese wiesen neben einigen fachspezifischen Aspekten zudem deutliche Gemeinsamkeiten mit den von Rindermann (2001; vgl. Abbildung 1) postulierten Erfolgsfaktoren guter Hochschullehre auf (→ Forschungsfrage 1b). Infolgedessen konnte bei der Erstellung des Fragebogens in großen Teilen auf bereits erprobte und validierte Items aus bestehenden Lehrevaluationsinventaren zurückgegriffen werden (vgl. Abschnitt 4; Anhang A). Zur Erfassung der spezifischen Aspekte wurden außerdem auch einige neue Items konstruiert (vgl. Anhang D). Nachdem die zusammengetragenen Items von Experten (d. h. Lehrgangsteilnehmern und Dozenten des IdF NRW) zunächst hinsichtlich Relevanz, Verständlichkeit sowie Vollständigkeit beurteilt wurden (vgl. Studie II), wurde die resultierende Fragebogenerstversion in verschiedenen Lehrgängen umfassend pilotiert (vgl. Studie III). Auf Basis dieser Ergebnisse wurde die Itemmenge begründet reduziert und schließlich die empirisch zugrundeliegende Faktorenstruktur ermittelt (→ Forschungsfrage 1a).

7.4 Faktoren guter Lehre am IdF NRW

Es resultierten zunächst zwei Faktoren, die verschiedene Facetten von Lehrerfolg (vgl. hierzu Abbildung 1) abbilden. So stellt der erste Faktor den Erwerb spezifischer Fach- und Führungskompetenzen (Skala I.1; vgl. auch Abschnitt 2.4) und der zweite Faktor die selbst-eingeschätzte Vorbereitung der angehenden Führungskräfte auf die Praxis sowie den antizipierten Transfer (Skala I.2) dar. Während diese beiden Faktoren in der vorliegenden Arbeit primär für Validierungszwecke herausgearbeitet und hinzugezogen wurden, sollen sie auch im Rahmen der Standardevaluation der Führungsausbildung am IdF NRW weiterhin berücksichtigt werden. So sollen entgegen der von Braun und Kollegen (2008) geäußerten Kritik an bestehenden Lehrevaluationsinventaren mit dem neuentwickelten Instrument nicht nur Prozess- und Strukturdaten sondern eben auch Ergebnisdaten erfasst werden (vgl. Abschnitt

2.4). Zudem stellt die Erfassung derartiger Ergebnisdaten eine gute Möglichkeit dar, den Lernzugewinn der angehenden Führungskräfte durch ihre Teilnahme am Lehrgang subjektiv zu erfassen, allgemeine Defizite zu erkennen und die Ausbildung dahingehend zu verbessern. Dies stellt vor allem in den neuen Lehrgangsformen einen deutliche Mehrwert dar, in denen die praktische Abschlussprüfung sowie auch die darauffolgende Einsatznachbesprechung am Ende eines Lehrgangs wegfällt (vgl. hierzu Busjan, 2016), durch die die Reifung der Teilnehmern sowie auch allgemein bestehende Defizite bisher zumindest in gewissen Maßen erkannt werden konnten.

Es resultierten acht weitere Faktoren, die als wesentlich zur Bewertung der Güte von Lehrveranstaltungen am IdF NRW angesehen werden (→ Forschungsfrage 1a). Dabei handelt es sich um erfolgsbedingende Faktoren, von denen vier mit den von Souvignier und Gold (2002) herausgearbeiteten, allgemeinen Bewertungsfaktoren der Qualität universitärer Lehre übereinstimmen (→ Forschungsfrage 1b). So zeigte sich, dass auch im feuerwehrtechnischen Ausbildungskontext zunächst dem Dozentenhandeln eine wesentliche Bedeutung zukommt. Dieses lässt sich –wie auch von Souvignier und Gold für den universitären Kontext gezeigt– in sozial-interaktives Verhalten der Dozenten (Skala II.1) einerseits, sowie der Didaktik (Skala II.7) und eine klare Strukturierung der Lehrveranstaltung (Skala II.4) andererseits unterteilen. Neben den verschiedenen Facetten des Dozentenhandelns machten das Anforderungsniveau und der Inhalt einer Lehrveranstaltung am IdF NRW (–im Sinne einer Überforderung; Skala II.2) einen weiteren wichtigen Bedingungsfaktor der Veranstaltungsgüte aus. Diese lässt sich nach dem Modell von Rindermann (2001) als Rahmenbedingung bezeichnen (vgl. hierzu Abbildung 1) und stellt nach Souvignier und Gold auch im universitären Kontext einen der wesentlichen Faktoren zur Veranstaltungsbeurteilung dar.

Die übrigen vier der acht erfolgsbedingenden Faktoren guter Lehre am IdF NRW stellen hingegen, für den feuerwehrtechnischen Ausbildungskontext spezifische Aspekte dar und weisen folglich keine Übereinstimmungen mit den Determinanten erfolgreicher Hochschullehre auf (→ Forschungsfrage 1a). Hierbei handelt es sich zunächst um den Faktor Gruppe (Skala II.8). Wie bereits in Abschnitt 3.2.1, beschrieben entscheidet letztendlich auch der Zusammenhalt einer Teilnehmergruppe über die Güte der Veranstaltung. Während dieser für den Lehrfolg im universitären Kontext zumeist nur wenig bis gar nicht ausschlaggebend ist (vgl. Abschnitt 3.3), ist es für die Güte einer Lehrveranstaltung am IdF NRW durchaus erfolgsentscheidend, dass sich jeder Teilnehmer gemäß seiner Kompetenzen einbringt und sich jeder immer auf seine Kameraden verlassen kann (vgl. Expertenbefragung; Studie II).

Die anderen drei Faktoren stellen schließlich kontextspezifische Unterrichtsmethoden dar, die im Zuge der Führungsausbildung am IdF NRW angewandt werden (z. B. AG-BIII, 2007; AG-FIII, 2005; AG-F IV, 2007) und bei der Bewertung der Güte einer Lehrveranstaltung folglich ebenfalls eine wichtige Rolle spielen. Konkret handelt es sich hierbei um die Gruppenarbeit (Skala II.3), die Planübung (Skala II.5) sowie das eigenverantwortliche Arbeiten (= EVA; Skala II.6). Ergänzt werden diese noch um die Methode der Einsatzübung, die im Rahmen der vorliegenden Arbeit bisher zwar nicht faktorenanalytisch bestätigt werden konnte, jedoch

inhaltlich relevant ist und als weitere Skala in Folgestudien ebenfalls berücksichtigt und erprobt werden soll (vgl. Abschnitt 6.3).

Eine Überlegung hierbei ist es, die Methoden-Skalen langfristig als eigene Module zu gestalten, die vom IdF NRW jeweils nach Bedarf und Interesse miterfasst werden können. Hierzu gilt es zunächst in Folgestudien zu untersuchen, welcher Stellenwert diesen Skalen zukommt, sprich welchen Anteil der Varianz sie jeweils aufklären.

7.5 Testgüte

Ein weiteres Ziel dieser Arbeit war es, schließlich auch die Güte des neuentwickelten Fragebogens –anhand der klassischen Hauptgütekriterien einerseits (→ Forschungsfrage 2a) sowie ausgewählter Nebengütekriterien andererseits (→ Forschungsfrage 2b)– zu beurteilen.

7.5.1 Objektivität

Durchführungsobjektivität. Die Bearbeitung des Fragebogens wurde durch die verantwortlichen Dozenten eines Lehrgangs und damit jeweils durch verschiedene Personen angeleitet. Unterschiede in der Anleitung waren demnach nicht auszuschließen. Um dennoch ein gewisses Maß an Durchführungsobjektivität gewährleisten zu können, erfolgte die Instruktion der Teilnehmer einheitlich schriftlich. So erhielten alle Teilnehmer zu Beginn standardmäßig dieselben ausführlichen Informationen zum Hintergrund dieser Befragung sowie klare Bearbeitungshinweise (vgl. Anhang F). Dies sorgte für eine gewisse Durchführungsobjektivität und sollte auch in zukünftigen Standardevaluationen beibehalten werden.

Zudem wurden die Dozenten angewiesen, die Evaluation jeweils am Ende eines letzten Unterrichtstages durchzuführen (vgl. Abschnitt 6.1.3). Der Durchführungsort blieb hingegen unkontrolliert. So erfolgten (–wie sich leider erst zum Ende der Pilotierungsphase herausstellte–) einige Lehrgangsbeurteilungen (planmäßig) im Unterrichtsraum unter Anwesenheit des Dozenten, während andere individuell (weniger planmäßig) im eigenen Zimmer oder Gemeinschaftsraum und damit unbeaufsichtigt durchgeführt wurden. Diese Tatsache schränkt die Durchführungsobjektivität durchaus ein. Um dies in Zukunft zu vermeiden und die Durchführungsobjektivität entsprechend zu erhöhen, ist darauf zu achten, dass die Teilnehmer den Fragebogen unter den gleichen, kontrollierten Bedingungen ausfüllen.

Um ferner auszuschließen, dass die Anwesenheit der Dozenten die Antworten der Teilnehmer gegebenenfalls verzerrt, wird vorgeschlagen, dass die Evaluation jeweils durch eine unabhängige Person (z. B. einen verantwortlichen Evaluationsbeauftragten) begleitet wird.

Auswertungs- und Interpretationsobjektivität. In Anbetracht dessen, dass die Items der Skalen durchweg ein eindeutiges, geschlossenes Antwortformat aufweisen, die Teilnehmer zu Beginn explizit aufgefordert werden, jeweils nur eine Zahl anzukreuzen (d. h. keine Kreuze zwischen den Zahlen zu machen) und sich die Skalenwerte schließlich eindeutig durch das Aufsummieren und Mitteln der zugrundeliegenden und skalenweise präsentierten Items berechnen lassen, ist eine Auswertungsobjektivität gegeben (z. B. Bühner, 2015: Dö-

ring & Bortz, 2016). Es ist jedoch darauf zu achten, dass die Items der Skala Inhalt zunächst umgepolt werden, damit auch hier ein höherer Wert für eine bessere Bewertung steht.

Zur Interpretation der Skalenwerte kann bisher nur die 7-stufige Likert-Skala inklusive der verbalen Labels herangezogen werden. Es existieren hingegen noch keine Normwerte anhand derer ein Ergebnis (bzw. Skalenwert) einheitlich zum Beispiel als durchschnittlich, über- oder unterdurchschnittlich gut interpretiert werden kann. Standardisierte Interpretationen sind infolgedessen noch nicht möglich (vgl. Bühner, 2015).

7.5.2 Reliabilität und Validität

Die internen Konsistenzen der Skalen variieren in einem Bereich, der auch in Arbeiten zur universitären Lehrevaluation berichtet wird (siehe hierfür Anhang A) und sprechen somit für hinreichend reliable Skalen (z. B. Bortz & Döring, 2006). Gleiches gilt für die Trennschärfen der Items (vgl. hierzu Souvignier und Gold, 2002). In Folgestudien sollen jedoch noch weitere Untersuchungen zu den Reliabilitäten der Skalen durchgeführt werden. So gilt es einerseits die berechneten Reliabilitäten an einer unabhängigen Stichprobe zu überprüfen. Andererseits sind weitere Reliabilitätsmaße (v.a. zur Stabilität; vgl. Staufenbiel, 2002) zu berechnen und vergleichend hinzuziehen – besonders auch vor dem Hintergrund, dass die Items einer Skala zum Teil durchaus unterschiedliche Aspekte einer Dimension erfassen.

In Anbetracht dessen, dass die zugrundeliegenden Items zunächst auf Basis von qualitativen Befragungen einiger Lehrgangsteilnehmer und Dozenten am IdF NRW entwickelt wurden (Studie I) und im Anschluss außerdem von einigen anderen Lehrgangsteilnehmern und Dozenten als relevant beurteilt wurden (Studie II), ist zudem von einer hohen Inhaltsvalidität auszugehen. Darüber hinaus liefern die in Studie III berechneten Korrelationen unter den erfolgsbedingenden Skalen einerseits sowie verschiedenen Qualitäts- und Erfolgskriterien andererseits erste Hinweise auf mitunter gute Kriteriumsvaliditäten (vgl. Abschnitt 6.3.5; z. B. Bühner, 2015). Des Weiteren ergeben sich jedoch weder Hinweise auf eine divergente Validität der Skalen gegenüber der aktuellen Stimmung der Teilnehmer noch Hinweise auf eine prognostische Validität der Skalen hinsichtlich der schriftlichen sowie praktischen Prüfungsergebnisse der Teilnehmer. Diese vorläufigen Ergebnisse gilt es jedoch in Folgestudien an einer unabhängigen Stichprobe umfassend zu überprüfen.

7.5.3 Verständlichkeit, Durchführungsökonomie und Zumutbarkeit

Die sprachliche und inhaltliche Verständlichkeit der Items wurde explizit in Studie II überprüft (vgl. Abschnitt 5) und galt somit bereits bei der Pilotierung der Erstversion als gegeben.

Die Durchführungsökonomie und damit einhergehende Zumutbarkeit wurde hingegen erst nach und nach verbessert, indem die Itemmenge im Laufe dieser Arbeit immer wieder begründet reduziert wurde. So konnte die Itemmenge für den Standardfragebogen schließlich auf 51 relevante Items (davon 46 auf einer 7-stufigen Likert-Skala zu beurteilende Aussagen, ein dichotomes Item, zwei Globalurteile sowie zwei offene Fragen) begrenzt werden. Vor dem Hintergrund der schon mehrfach zitierten Arbeit von Gold und Souvignier (2002) liegt der Umfang des neuentwickelten Fragebogens damit bereits in einem durchaus vertretbaren

und zumutbaren Bereich. Nach Sichtung deutschsprachiger Inventare zur Lehrveranstaltungsevaluation unterteilen die Autoren bestehende Verfahren in zwei Gruppen: „Während auf der einen Seite mit Blick auf die Akzeptanz der Fragebögen eine Beschränkung auf 16 bis 20 Items vorgenommen wird, liegt die Itemzahl der anderen Inventare bei 37 bis 55, um eine höhere inhaltliche Breite der Fragen gewährleisten zu können“ (ebd., S. 270). Langfristig sollte der Umfang des Fragebogens zugunsten einer verbesserten Akzeptanz demnach noch weiter reduziert werden. Dies ist besonders auch deshalb empfehlenswert, als dass die Teilnehmer den Fragebogen stets am Ende des letzten Unterrichtstages ausfüllen sollen und die Bearbeitungszeit demnach von ihrer letzten Vorbereitungszeit auf die bevorstehenden Prüfungen abgeht.

7.6 Limitationen und weitere Forschungsimplicationen

Im Laufe dieser Arbeit wurden an verschiedenen Stellen bereits immer wieder Limitationen der vorliegenden Arbeit sowie damit einhergehende Implikationen für weitere Forschung aufgezeigt. Zusammengefasst gilt es dabei zunächst die herausgearbeitete Faktorenstruktur des Standardfragebogens zu überprüfen und mittels einer konfirmatorischen Faktorenanalyse zu bestätigen. Hierzu sollten der Fragebogen in seiner deutlich verkürzten Form in einer hinreichend großen Anzahl an Lehrgängen (vgl. Abschnitt 6.3.7) eingesetzt und die Ergebnisse faktorenanalytisch analysiert werden. Die Antworten auf die offenen Fragen sollten zudem dahingehend analysiert werden, ob durch die Kürzung des Fragebogens ein wichtiger Aspekt der Bewertung guter Lehre am IdF NRW unberücksichtigt bleibt.

Um darüber hinaus die psychometrische Qualität des neuentwickelten Fragebogens zu untermauern, müssten in Folgestudien außerdem weitere Untersuchungen zur Reliabilität und Validität durchgeführt werden (vgl. Abschnitt 7.2.2). Dabei wäre es zum Beispiel auch interessant, statt dem subjektiven und antizipierten Lerngewinn (d. h. dem vorzeitig eingeschätzten Kompetenzerwerb und Transfer sowie der vermeintlichen Vorbereitung auf die bevorstehende Funktion als Gruppen- bzw. Zugführers) vielmehr auch den tatsächlichen Lerngewinn (d. h. den tatsächlichen Kompetenzerwerb, das Gelingen des Transfers sowie das sichere Agieren als Gruppen- oder Zugführer) als Erfolgs- und Validierungskriterium hinzuzuziehen. Hierfür ist es allerdings erforderlich die Teilnehmer auch über die Lehrveranstaltung hinaus zu begleiten und sie nach ihren ersten Einsätzen als Gruppen- bzw. Zugführer nochmals dahingehend zu befragen, inwiefern sie durch die Ausbildung am IdF NRW tatsächlich auf die Herausforderungen der Praxis vorbereitet wurden und inwieweit es ihnen gelungen ist, das erworbene Handlungswissen in Einsatzsituationen erfolgreich anzuwenden. Schließlich zeigt sich der tatsächliche Erwerb von relevanten Fach- und Führungskompetenzen erst durch Einsatzerfahrungen und darauffolgende Reflexionen (vgl. Miste, 2007).

Die zusätzliche Erfassung von Lern-, Transfer- und Ergebniskriterien (i.S. einer erfolgreichen Einsatzbewältigung) ist vor allem aber auch deshalb empfehlenswert, als dass die auf Basis des neuentwickelten Fragebogens vorgenommene Lehrveranstaltungsbeurteilung im

Sinne der Klassifikation von Evaluationskriterien nach Kirkpatrick (1959) in erster Linie die Akzeptanz der Teilnehmer widerspiegelt (vgl. Vier-Ebenen-Modell in Abschnitt 2.2.2). Dabei stellt die Akzeptanz einer Lehrveranstaltung unter den Teilnehmern ein durchaus wichtiges, für eine vollständige Evaluation jedoch kein allein hinreichendes Kriterium dar. Durch die Selbsteinschätzung von Kompetenz- und Wissenserwerb, die der Fragebogen den angehenden Führungskräften ebenfalls abverlangt, lassen sich zudem auch erste Aussagen über den Lerngewinn in Folge einer Veranstaltungsteilnahme (und damit hinsichtlich eines weiteren Evaluationskriteriums nach Kirkpatrick, 1959) treffen. Allerdings handelt es sich hierbei zunächst nur um subjektive Einschätzungen, die in Folgestudien durch objektivere Maße, wie einer Fremdbeurteilung oder standardisierten Prä-Post-Vergleichen (vgl. hierzu Abschnitt 2.2.2) zu validieren sind. Jedoch stellt auch der Lerngewinn ein wichtiges, aber ebenfalls kein hinreichendes Kriterium für eine umfassende Lehrveranstaltungsevaluation dar. Vielmehr müssen für eine vollständige Evaluation neben den Akzeptanz- und Lernkriterien schließlich auch Transfer- und Ergebniskriterien berücksichtigt werden. Diese lassen sich jedoch erst nach ersten Einsatzerfahrungen erfassen – was dafür spricht, die Teilnehmer in Folgestudien auch über den Lehrgang hinaus zu begleiten.

Interessant wäre es zudem, die Lehrveranstaltungen in Folgestudien jeweils auch durch die verantwortlichen Dozenten beurteilen zu lassen und diese mit den Veranstaltungsbeurteilungen durch die Teilnehmer abzugleichen, um eventuelle Unterschiede erkennen und herausstellen zu können. Außerdem gilt es zu explorieren, ob sich der Fragebogen (bzw. einzelne Skalen) gegebenenfalls auch zur Evaluation von anderen Lehrgängen am IdF NRW oder allgemein im feuerwehrtechnischen oder vergleichbaren Ausbildungskontext eignet.

7.7 Fazit

Obwohl es im Interesse der Gesamtbevölkerung liegt, dass die Führungskräfte der Feuerwehr gut ausgebildet werden (vgl. Abschnitt 2), finden sich in der Literatur bisher keine empirischen oder theoretischen Forschungsarbeiten in diesem Bereich. Während es eine Vielzahl an erprobten Instrumenten zur Beurteilung von Vorlesungen und Seminaren im universitären Kontext gibt (vgl. Abschnitt 2.3.1), existieren in der Fachliteratur hingegen keine Hinweise auf ein derartiges Instrument zur Evaluation von Lehrveranstaltungen im feuerwehrtechnischen Ausbildungskontext. Vor diesem Hintergrund liefert die vorliegende Arbeit einen wichtigen Forschungsbeitrag, indem sie zunächst zentrale Erfolgsfaktoren eines Lehrgangs für angehende Führungskräfte der Feuerwehr identifiziert und auf Basis dessen schließlich einen theoretisch sowie empirisch fundierten Fragebogen zur Evaluation der Güte von Gruppen- und Zugführerlehrgänge am IdF NRW entwickelt.

Damit erweitert die vorliegende Arbeit die Forschung zu guter Lehre im universitären Kontext auf den feuerwehrtechnischen Ausbildungskontext aus. Zudem liefert die Arbeit eine gute Grundlage für eine umfassende Evaluation sowie einer gegebenenfalls damit einhergehenden Optimierung einzelner Bestandteile der Führungsausbildung am IdF NRW.

Auch wenn die Bestätigung der Faktorenstruktur sowie weitere Prüfungen hinsichtlich der psychometrischen Qualität des neuentwickelten Fragebogens noch ausstehen, kann der Fragebogen in seiner jetzigen Version bereits zur Evaluation der Lehrgänge am IdF NRW eingesetzt werden (vgl. Abschnitt 7.5). Denn auch wenn in Zukunft durchaus weitere Anpassungen des Fragebogens möglich sind, so erfüllt der Fragebogen nicht zuletzt auf Ebene einzelner Items bereits eine wichtige Rückmelde- und Informationsfunktion hinsichtlich entscheidender Qualitätsaspekte.

Referenzen

- AG-BIII – LFV NRW, AGBF NRW, WFV NRW, & IdF NRW (2007). *Lernziele für die Ausbildung zum Gruppenführer in der Berufsfeuerwehr*. Abgerufen am 13. März 2016 unter http://www.idf.nrw.de/service/downloads/pdf/biii_lzk_2008.pdf.
- AG-FIII – LFV NRW, AGBF NRW, WFV NRW, IdF NRW (2005). *Lernziele für die Ausbildung zum Gruppenführer (Freiwillige Feuerwehr)*. Abgerufen am 13. März 2016 unter http://www.idf.nrw.de/service/downloads/pdf/rderlass_im_20051221_lernziele.pdf.
- AG-FIV – LFV NRW, AGBF NRW, WFV NRW, & IdF NRW (2007). *Lernziele für die Ausbildung zum Zugführer (Freiwillige Feuerwehr)*. Abgerufen am 13. März 2016 unter http://www.idf.nrw.de/service/downloads/pdf/lernziele_zugfuehrer_20070925.pdf.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, 18, 1–13.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Bräuer, M. M. (2014). *Führung im Kontext von lebenskritischen Situationen und Hochleistung: Eine empirische Analyse anhand ausgewählter Einsatzeinheiten von Bundeswehr und technischem Hilfswerk* (1. Aufl.). Lohmar: Josef Eul Verlag.
- Braun, E., & Gusy, B. (2006). Perspektiven der Lehrevaluation. In G. Krampen & H. Zayer (Hrsg.), *Didaktik und Evaluation in der Psychologie* (S. 152–166). Göttingen: Hogrefe.
- Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, 54, 30–42.
- Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Hrsg.), *Reading and understanding multivariate statistics* (S. 99–136). Washington, DC: American Psychological Association.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). München: Pearson Studium.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Busjan, J. N. (2016). *Evaluation von Prüfungen in den Führungslehrgängen der Feuerwehr* (Unveröffentlichte Masterarbeit). Westfälische Wilhelms-Universität, Münster.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1, 245–276.

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cortina, J. M. (1993): What is coefficient alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78, 98–104.
- Dehnbostel, P. (2003). *Informelles Lernen: Arbeitserfahrungen und Kompetenzerwerb aus berufspädagogischer Sicht*. Überarbeiteter Vortrag anlässlich der 4. Fachtagung des Programms 'Schule – Wirtschaft/Arbeitsleben' am 18./19.September in Neukirchen/Pleisse, Deutschland.
- Deutsche Gesellschaft für Evaluation e.V. (2002): *Standards für Evaluation*. Köln: Zimmermann Druck.
- Dietrich, R., & Childress, T. M. (2004). *Group interaction in high risk environments*. Aldershot, UK: Gower Publishing Limited.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5.Aufl.). Heidelberg: Springer.
- Dunkin, M. J., & Barnes, J. (1986). Research on teaching in higher education. In M. C. Wittrock (Hrsg.), *Handbook of Research on Teaching* (3. Aufl., S. 754–777). New York City, NY: Macmillan Publishers.
- Feuerwehرداریsvorschrift (FwDV) 100: Führung und Leitung im Einsatz*. (2003). Stuttgart: Kohlhammer, Deutscher Gemeindeverlag.
- Gediga, G., von Kannen, K., Schnieder, F., Köhne, S., Luck, H., & Schneider, B. (2000). *KIEL: Ein Kommunikationsinstrument für die Evaluation von Lehrveranstaltungen*. Bissendorf: Methodos.
- Gorsuch, R. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, 25, 33–39.
- Grabbe, Y. (2003). *Zwischenbericht: Lehrevaluation. Konstruktion eines Fragebogens für Vorlesungen und Seminare*. Westfälische Wilhelms-Universität Münster. Abgerufen am 17. November 2015 unter http://www.psy.uni-muenster.de/evaluation/1.4/Downloads_files/zwischenbericht.pdf.
- Greve, W., & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung*. Weinheim: Beltz.

- Grötemeier, I., & Thielsch, M. T. (2010a). Münsteraner Fragebogen zur Evaluation – Zusatzmodul Gruppenarbeit (MFE-ZGr). In A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 14.0. Bonn: GESIS.
- Grötemeier, I., & Thielsch, M. T. (2010b). Münsteraner Fragebogen zur Evaluation – Zusatzmodul Hausaufgaben (MFE-ZHa). In A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 14.0. Bonn: GESIS.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149–161.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hirschfeld, G., & Thielsch, M. T. (2009a). Münsteraner Fragebogen zur Evaluation von Seminaren (MFE-S). In A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 13.0. Bonn: GESIS.
- Hirschfeld, G., & Thielsch, M. T. (2009b). Münsteraner Fragebogen zur Evaluation von Vorlesungen (MFE-V). In A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 13.0. Bonn: GESIS.
- Hirschfeld, G., von Brachel, R., & Thielsch, M. T. (2014). Selecting items for Big Five questionnaires: At what sample size do factor loadings stabilize? *Journal of Research in Personality*, 53, 54–63.
- Holling, H. (1999). Evaluation eines komplexen Fortbildungsprogramms zur Steigerung der beruflichen Kompetenz. In H. Holling & G. Gediga (Hrsg.), *Evaluationsforschung* (S. 1–33). Göttingen: Hogrefe.
- Homepage des IdF NRW (2016a). *Über Uns – Organisation*. Abgerufen am 13. März 2016 unter <http://www.idf.nrw.de/ausbildung/katalog/katalog.php>.
- Homepage des IdF NRW (2016b). *Ausbildung – Lehrgangskatalog*. Abgerufen am 13. März 2016 unter <http://www.idf.nrw.de/ausbildung/katalog/katalog.php>.
- Horwath, I. (2013). *Gleichstellung im Feuerwehrwesen: "Gut Wehr!" und die HeldInnen von heute*. Wiesbaden: Springer.
- IdF NRW Ausbildungsplan – Musterplan B III (2012). Abgerufen am 15. März 2016 unter http://www.idf.nrw.de/ausbildung/katalog/dokumente/musterausbildungsplan_biii.pdf.
- IdF NRW Ausbildungsplan – Musterplan FIII (2013). Abgerufen am 15. März 2016 unter http://www.idf.nrw.de/ausbildung/katalog/dokumente/musterausbildungsplan_fiii_26.pdf.
- IdF NRW Ausbildungsplan – Musterplan FIV (2016). Zur Verfügung gestellt durch das IdF NRW am 15. März 2016.

- IdF NRW Fragebogen zur Veranstaltungsbewertung*. Zur Verfügung gestellt durch das IdF NRW am 10. Dezember 2015.
- Jäger, R. (2004). Konstruktion einer Ratingskala mit Smilies als symbolische Marken. *Diagnostica*, 50, 31–38.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.
- Kauffeld, S. (2006). *Kompetenzen messen, bewerten, entwickeln*. Stuttgart: Schäffer-Poeschel.
- 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*. San Francisco: Berrett-Koehler Publishers.
- Klein, G. (2003). *Natürliche Entscheidungsprozesse*. Paderborn: Junfermann.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London, UK: Butterworth-Heinemann.
- Lehrgangsbeschreibung-BIII-Pilot* (2015). Unveröffentlichtes Konzeptpapier. Zur Verfügung gestellt durch das IdF NRW am 30. Oktober 2015.
- Lehrgangsbeschreibung-FIII-Pilot* (2015). Unveröffentlichtes Konzeptpapier. Zur Verfügung gestellt durch das IdF NRW am 30. Oktober 2015.
- Lehrgangsbeschreibung-FIV-Pilot* (2015). Unveröffentlichtes Konzeptpapier. Zur Verfügung gestellt durch das IdF NRW am 30. Oktober 2015.
- Loberg, L. (2016). *Evaluation der Führungskräftelehrgänge am Institut der Feuerwehr NRW: Vergleich zweier Ausbildungskonzepte* (Unveröffentlichte Masterarbeit). Westfälische Wilhelms-Universität, Münster.
- Löchteken, T. (2015a). *Modularisierung. Zusammenfassung des Arbeitsstandes* (Unveröffentlichter Projektbericht). Institut der Feuerwehr Nordrhein-Westfalen, Münster.
- Löchteken, T. (2015b). *Evaluation* (Unveröffentlichtes Arbeitspapier). Institut der Feuerwehr Nordrhein-Westfalen, Münster.
- Löchteken, T., & Döring, J. (2015). *Beitrag zur Neuausrichtung der Führungsausbildung in NRW. Einheitliche und modulare Führungsausbildung* (Unveröffentlichter Konzeptentwurf). Institut der Feuerwehr Nordrhein-Westfalen, Münster.
- Löchteken, T., & Fuchs, P. (2015). *Neuentwicklungen in der Gruppen- und Zugführerausbildung am Institut der Feuerwehr in Nordrhein-Westfalen* (Unveröffentlichter Konzeptentwurf). Institut der Feuerwehr Nordrhein-Westfalen, Münster.

- Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, 75, 150–166.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International journal of educational research*, 11, 253–388.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalen (2014). *Gefahrenabwehr in Nordrhein-Westfalen: Jahresbericht 2013*. Abgerufen am 25. Januar 2016 unter <http://www.mik.nrw.de/themen-aufgaben/ Gefahrenabwehr-feuerwehr-katastrophenschutz /jahresbericht.html>.
- Mistele, P. (2007). *Faktoren des verlässlichen Handelns: Leistungspotenziale von Organisationen in Hochrisikoumwelten* (1. Aufl.). Wiesbaden: Deutscher Universitäts-Verlag.
- Mistele, P. (2008). Teamarbeit und Führung in Hochleistungsteams. In C. Buerschaper & S. Starke (Hrsg.), *Führung und Teamarbeit in kritischen Situationen* (S. 20–40). Frankfurt am Main: Verlag für Polizeiwissenschaften.
- Moosbrugger, H., & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- Moosbrugger, H., & Schweizer, K. (2002). Evaluationsforschung in der Psychologie. *Zeitschrift für Evaluation*, 1, 19–37.
- Norušis, M. J. (2005). *SPSS 13.0 Statistical Procedures Companion*. Chicago, IL: SPSS, Inc.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, 32, 396–402.
- Ott, M. (2013). *Besser Führen durch bessere Einsatznachbesprechungen*. Abgerufen am 18. März 2016 unter <http://www.feuerwehrleben.de/besser-fuehren-durch-bessere-einsatz-nach-besprechungen>.
- Pellert, A., Spiel, C., Wolf, P. H., Popper, V., & Poschauko, U. (2000). *Ist-Stand Lehrevaluation an Österreichischen Universitäten mit Schwerpunkt der Karl-Franzens-Universität Graz – Projektbericht*. Universität Graz, Österreich.
- Rindermann, H. (1996). Zur Qualität studentischer Lehrveranstaltungsevaluationen: Eine Antwort auf Kritik an der Lehrevaluation. *Zeitschrift für Pädagogische Psychologie*, 10, 129–145.

- Rindermann, H. (1998). Das Münchner multifaktorielle Modell der Lehrveranstaltungsqualität: Entwicklung, Begründung und Überprüfung. *Beiträge zur Hochschulforschung*, 3, 189–224.
- Rindermann, H. (2001). Die studentische Beurteilung von Lehrveranstaltungen – Forschungsgegenstand und Implikationen. In C. Spiel (Hrsg.), *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (S. 61–88). Münster: Waxmann.
- Rindermann, H. (2003). Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*, 2, 233–256.
- Robertz, A. C. (2015). *Konstruktion, Erprobung und erste Validierung eines Fragebogens zur Evaluation von Personaltrainings und intergenerativ konzipierten Personalentwicklungsprogrammen* (Unveröffentlichte Masterarbeit). Westfälische Wilhelms-Universität, Münster.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Journal of Personnel Psychology*, 47, 537–560.
- Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik* (5.Aufl.). Berlin: Springer.
- Sievers, S. C. (2015). *Effektiv führen, wenn es brennt! Eine Untersuchung des feuerwehrspezifischen Führungsprozesses „Führungsvorgang“ mittels Critical Incident Technique*. (Unveröffentlichte Masterarbeit). Westfälische Wilhelms-Universität, Münster.
- Souvignier, E., & Gold, A. (2002). Fragebögen zur Lehrevaluation: Was können sie leisten. *Zeitschrift für Evaluation*, 1, 265–280.
- Staufenbiel, T. (2000). Fragebogen zur Evaluation von universitären Lehrveranstaltungen durch Studierende und Lehrende. *Diagnostica*, 46, 169–181.
- Stockmann, R. (2002). Qualitätsmanagement und Evaluation – Konkurrierende oder sich ergänzende Konzepte? *Zeitschrift für Evaluation*, 1, 20–243.
- Suhr, D. (2006). Exploratory or Confirmatory Factor Analysis. *SAS Users Group International Conference* (S. 1–17). Cary, NC: SAS Institute, Inc.
- Suhr, D. (2006). Exploratory or Confirmatory Factor Analysis. *SAS Users Group International Conference* (S. 1–17). Cary: SAS Institute, Inc.
- Thielsch, M. T., Dusend T., & Grötemeier, I. (2010). Münsteraner Fragebogen zur Evaluation von Tutorien (MFE-T). In A. Glöckner-Rist (Hrsg.). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 14.0. Bonn: GESIS.

- Thielsch, M. T., & Hirschfeld, G. (2012a). Münsteraner Fragebogen zur Evaluation von Seminaren – revidiert (MFE-Sr). In A. Glöckner-Rist (Hrsg.). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 15.0. Bonn: GESIS.
- Thielsch, M. T., & Hirschfeld, G. (2012b). Münsteraner Fragebogen zur Evaluation von Vorlesungen - revidiert (MFE-Vr). In A. Glöckner-Rist (Hrsg.). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 15.0. Bonn: GESIS.
- Thielsch, M. T., Lenzner, T. & Melles, T. (2012). Wie gestalte ich gute Items und Interviewfragen? In M. T. Thielsch & T. Brandenburg (Hrsg.), *Praxis der Wirtschaftspsychologie II: Themen und Fallbeispiele für Studium und Praxis* (S. 221–240). Münster: MV Wissenschaft.
- van Dick, R. (2003). *Commitment und Identifikation mit Organisationen*. Göttingen: Hogrefe.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321–327.
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Hrsg.), *Problems and solutions in human assessment* (S. 41–71). Boston, MA: Kluwer Academic Publishers.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Hrsg.), *Structural equation modelling, concepts, issues and applications* (S. 56–75). Thousand Oaks: Sage.
- Wirtz, M. A. (Hrsg.), (2014). *Dorsch-Lexikon der Psychologie*. (17. Aufl.). Bern: Hans Huber.
- Wolf, P., Spiel, C., & Pellert, A. (2001). Entwicklung eines Fragebogens zur globalen Lehrveranstaltungsevaluation – ein Balanceakt zwischen theoretischem Anspruch, Praktikabilität und Akzeptanz. In C. Spiel (Hrsg.), *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (S. 89–109). Münster: Waxmann.
- Zumbach, J., Spinath, B., Schahn, J., Friedrich, M., & Kögel, M. (2007). Entwicklung einer Kurzskala zur Lehrevaluation. In M. Krämer, S. Preiser & K. Brusdeylins (Hrsg.), *Psychologiedidaktik und Evaluation VI* (S. 317–325). Göttingen: V & R unipress.

Anhang

Anhang A65

Anhang B67

Anhang C68

Anhang D69

Anhang E78

Anhang F.....79

Anhang G80

Anhang H81

Anhang I.....82

Anhang J.....84

Anhang K845

Anhang A

Übersicht über die Inventare auf die in dieser Arbeit zurückgegriffen wurde

Name des Instruments	Autoren (Jahr der Veröffentlichung)	Skalen	Testgüte
Fragebogen zur Evaluation von Seminaren (FESEM)	Staufenbiel (2000)	(1) Planung & Darstellung (2) Interessantheit & Relevanz verdeutlichen (3) Umgang mit Studierenden (4) Schwierigkeit & Umfang (5) Qualität der Referate	Interne Konsistenz $\alpha = .52$ bis $.78$
Fragebogen zur Evaluation von Vorlesungen (FEVOR)	Staufenbiel (2000)	(1) Planung & Darstellung (2) Interessantheit & Relevanz verdeutlichen (3) Umgang mit Studierenden (4) Schwierigkeit & Umfang	Interne Konsistenz $\alpha = .67$ bis $.77$
Fragebogen zur Lehrveranstaltungsevaluation (FB zur LVE)	Wolf et al. (2001)	(1) Fachliche & didaktische Kompetenz (2) Lernhilfen (3) Struktur & Organisation (4) Engagement (5) Gleichbehandlung (6) Umfeld (7) Studierendenbezogene Variablen	k. A.
Ein Kommunikationsinstrument für die Evaluation von Lehrveranstaltungen (KIEL)	Gediga et al. (2000)	(1) Verständlichkeit (2) Gliederung der Veranstaltung (3) Bewertung der Materialien (4) Interesse der Studierenden	Interne Konsistenz $\alpha = .68$ bis $.83$
Kurzskala zur Lehrevaluation	Zumbach et al. (2007)	(1) Form & Struktur (2) Merkmale des/der Dozent/in (3) Umfang & Relevanz (4) Lernerfolg	Interne Konsistenz $\alpha = .58$ bis $.73$

Münsteraner Fragebogen zur Evaluation von Seminaren (MFE-S)	Hirschfeld & Thielsch (2009a)	(1) Didaktik (2) Schwierigkeit (3) Selbststudium	Interne Konsistenz $\alpha = .57$ bis $.86$
Münsteraner Fragebogen zur Evaluation von Seminaren – revidiert (MFE-Sr)	Thielsch & Hirschfeld (2012)	(1) Dozent & Didaktik (2) Überforderung (3) Teilnehmer (4) Materialien	Interne Konsistenz $\alpha = .77$ bis $.89$
Münsteraner Fragebogen zur Evaluation von Tutorien (MFE-T)	Thielsch, Dusend, & Grötemeier (2010)	(1) Unterstützung (2) Bedarf	Interne Konsistenz $\alpha = .70$ bis $.83$
Münsteraner Fragebogen zur Evaluation von Vorlesungen (MFE-V)	Hirschfeld & Thielsch (2009b)	(1) Dozentenauftreten (2) Didaktik	Innere Konsistenz $\alpha = .78$ bis $.80$
Münsteraner Fragebogen zur Evaluation – Zusatzmodul Gruppenarbeiten (MFE-ZGr)	Grötemeier & Thielsch (2010a)		k. A.
Münsteraner Fragebogen zur Evaluation – Zusatzmodul Hausaufgaben (MFE-ZHa)	Grötemeier & Thielsch (2010b)		k. A.
Münsteraner Fragebogen zur Evaluation – Zusatzmodul Moderation (MFE-ZMo)	Thielsch & Stegemöller (2012)		Innere Konsistenz $\alpha = .86$
Fragebogens zur Evaluation von Personaltrainings und intergenerativ konzipierten Personalentwicklungsprogrammen	Robertz (2015)	(1) Dozent & Didaktik (2) Wissen & Praxisbezug (3) Lernkontext (4) Transfer	Innere Konsistenz $\alpha = .80$ bis $.85$

Anhang B





Leitfaden für das teilstrukturierte Interview (Studie I; Dozentenversion)

<div style="display: flex; justify-content: space-between; align-items: center;"> </div> <p style="text-align: center;">Interviewleitfaden (TEIL A): „Guter Lehrgang“ (DOZENT)</p> <p>Wir schätzen es sehr, dass Sie sich bereit erklärt haben, uns mit Ihrer Teilnahme an dieser Befragung zu unterstützen. Herzlichen Dank. Die Befragung findet im Rahmen eines Kooperationsprojektes zwischen dem Institut der Feuerwehr (IdF) Nordrhein-Westfalen und der Westfälischen Wilhelms-Universität (WWU) Münster statt. Das Projekt wurde im Zuge der Neuausrichtung der Führungsausbildung am IdF ins Leben gerufen, mit dem Ziel die Veränderungen in den Lehrgängen und Prüfungen ab Anfang des nächsten Jahres begleitend zu evaluieren.</p> <p>Im Vorfeld dieser Evaluation interessieren wir uns zunächst dafür, was aus Ihrer Dozentensicht einen guten Lehrgang sowie eine gute Prüfung ausmacht. Unsere Befragung besteht daher aus zwei Teilen. Im Teil A wollen wir Sie zu verschiedenen Aspekten eines guten Lehrgangs befragen. Dieser erste Teil wird etwa 20 Minuten dauern. Im Teil B wollen wir Sie zu verschiedenen Aspekten guter Prüfungen befragen. Dieser zweite Teil wird etwa 15 Minuten dauern.</p> <p>Es gibt dabei keine richtigen oder falschen Antworten. Wir interessieren uns für Ihre Erfahrungen und Ihre ehrliche Meinung.</p> <p>Ihre Antworten werden anonym und über alle Befragten (Teilnehmer & Dozenten) hinweg aggregiert ausgewertet, sodass später kein Rückschluss auf Ihre Person möglich ist. Es werden nur Aussagen über die Gesamtheit aller Befragten gemacht und es werden dabei an keiner Stelle Namen genannt. Für die Auswertung Ihrer Angaben benötigen wir lediglich Ihr Alter und Ihr Geschlecht.</p> <p>Wenn Sie im Laufe der Befragung Rückfragen oder Anmerkungen haben, äußern Sie diese jederzeit gerne. Haben Sie zum jetzigen Zeitpunkt schon irgendwelche Fragen?</p> <p style="text-align: right;">Seite 1</p>	<div style="display: flex; justify-content: space-between; align-items: center;"> </div> <p>Alter: _____ Jahre</p> <p>Geschlecht: <input type="checkbox"/> weiblich <input type="checkbox"/> männlich</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>Zu Beginn möchten wir Sie bitten, sich ganz konkret an eine, aus Ihrer Sicht, gute Lehrveranstaltung zu erinnern, die Sie in der Vergangenheit als Dozent gegeben haben. Nehmen Sie sich dazu gerne kurz einige Minuten Zeit und geben Sie mir Bescheid, sobald Sie sich eine Lehrveranstaltung in Erinnerung gerufen haben.</p> <p>Warum war diese Lehrveranstaltung aus Ihrer Sicht gut?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>Was macht für Sie einen erfolgreichen Lehrgang aus?</p> <p>Nennen Sie uns bitte, die aus Ihrer Sicht 7 wichtigsten Aspekte einer erfolgreichen Lehrveranstaltung:</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%;">1.</td> <td style="width: 25%;"></td> <td style="width: 25%;">5.</td> <td style="width: 25%;"></td> </tr> <tr> <td>2.</td> <td></td> <td>6.</td> <td></td> </tr> <tr> <td>3.</td> <td></td> <td>7.</td> <td></td> </tr> <tr> <td>4.</td> <td></td> <td colspan="2"></td> </tr> </table> </div> <p style="text-align: right;">Seite 2</p>	1.		5.		2.		6.		3.		7.		4.			
1.		5.															
2.		6.															
3.		7.															
4.																	
<div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p>Was zeichnet aus Ihrer Sicht einen guten Dozenten/ eine gute Dozentin aus?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p>Ggf. Durch welche Aspekte zeichnet sich der Lehrstil einer guten Dozentin/ eines guten Dozenten aus?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <div style="border: 1px solid black; padding: 5px;"> <p>Welchen Einfluss haben die Teilnehmer/innen auf den Erfolg einer Lehrveranstaltung?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <p style="text-align: right;">Seite 3</p>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p>Was sind entscheidende Rahmenbedingungen für eine erfolgreiche Lehrveranstaltung?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p>Gibt es Ihrer Meinung nach noch weitere Aspekte, die das Urteil einer Lehrgangsteilnehmerin/ eines Lehrgangsteilnehmers über die Güte der Lehrveranstaltung beeinflussen oder gar verzerren könnte? Wenn ja, welche?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p>Was die zentralen Lernziele im Rahmen Ihrer Zugführer-/Gruppenführerausbildung?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <div style="border: 1px solid black; padding: 5px;"> <p>Gibt es weitere Aspekte, über die wir bisher nicht gesprochen haben, die wir im Rahmen einer Lehrgangsevaluation aber unbedingt berücksichtigen sollten?</p> <div style="border: 1px solid black; height: 100px; margin-top: 5px;"></div> </div> <p style="text-align: right;">Seite 4</p>																

Anmerkungen: Die Teilnehmerversion beinhaltet dieselben Fragen und unterscheidet sich lediglich in der Anrede von der dargestellten Dozentenversion. Beide Versionen finden sich zudem im digitalen Anhang. Bei dem Teil B handelt es sich um die Befragung zu einer guten Prüfung, die unmittelbar an den Teil A anschloss. Hierfür sei auf die Arbeit von Busjan (2016) verwiesen.

Anhang C

Fragebogen (Studie I)

<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="text-align: center;">  <p>WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER</p> </div> <div style="text-align: center;">  <p>IdF</p> </div> </div> <p style="text-align: center;">Fragebogen: „Guter Lehrgang – Gute Prüfung“</p> <p>Wir schätzen es sehr, dass Sie sich bereit erklärt haben, uns mit Ihrer Teilnahme an dieser Befragung zu unterstützen. Die Befragung findet im Rahmen eines Kooperationsprojektes zwischen dem Institut der Feuerwehr (IdF) Nordrhein-Westfalen und der Westfälischen Wilhelms-Universität (WWU) Münster statt. Das Projekt wurde im Zuge der Neuausrichtung der Führungsausbildung am IdF ins Leben gerufen, mit dem Ziel die Veränderungen in den Lehrgängen und Prüfungen ab Anfang des nächsten Jahres begleitend zu evaluieren.</p> <p>Im Vorfeld dieser Evaluation interessieren wir uns heute zunächst dafür, was für Sie als Lehrgangsteilnehmer einen guten Lehrgang (Teil A) sowie eine gute Prüfung (Teil B) ausmacht.</p> <p>Diese Befragung erfolgt selbstverständlich anonym. Um ihre Angaben jedoch auswerten zu können, benötigen wir von Ihnen folgende demografische Daten:</p> <p>Alter: _____ Jahre</p> <p>Geschlecht: <input type="checkbox"/> weiblich <input type="checkbox"/> männlich</p> <p style="text-align: center;">Wenn Sie Fragen oder Anregungen haben, wenden Sie sich gerne persönlich oder per Email an uns (s_babi02@uni-muenster.de). Wir danken Ihnen ganz herzlich für Ihre Teilnahme,</p> <p style="text-align: center;">Jessica Busjan und Stephanie Babel (Studentinnen der WWU Münster)</p> <p style="text-align: right;">Seite 1</p>	<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="text-align: center;">  <p>WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER</p> </div> <div style="text-align: center;">  <p>IdF</p> </div> </div> <p style="text-align: center;">Teil A (Lehrgang)</p> <p>Bitte nehmen Sie sich zunächst einige Minuten Zeit, um sich an eine Lehrveranstaltung zu erinnern, an der Sie in der Vergangenheit teilgenommen haben.</p> <p>Bitte kreuzen Sie an, ob dies eine <input type="radio"/> gute oder <input type="radio"/> schlechte Lehrveranstaltung war.</p> <p>Warum war diese Lehrveranstaltung aus Ihrer Sicht gut bzw. schlecht?</p> <div style="border: 1px solid black; height: 100px; width: 100%;"></div> <p style="text-align: right;">Seite 2</p>
<div style="background-color: #f2f2f2; padding: 5px; margin-bottom: 5px;"> Was macht für Sie einen erfolgreichen Lehrgang aus? Nennen Sie uns bitte, die aus Ihrer Sicht 7 wichtigsten Aspekte eines erfolgreichen Lehrgangs: </div> <div style="border: 1px solid black; padding: 5px;"> 1. 2. 3. 4. 5. 6. 7. </div> <div style="background-color: #f2f2f2; padding: 5px; margin-top: 5px;"> Bitte stellen Sie sich nun eine gute Dozentin/ einen guten Dozenten vor. Durch welche Aspekte zeichnet sich diese/r aus? </div> <div style="border: 1px solid black; height: 150px; width: 100%;"></div> <p style="text-align: right;">Seite 3</p>	<div style="background-color: #f2f2f2; padding: 5px; margin-bottom: 5px;"> Welchen Einfluss haben Sie als Teilnehmer/innen auf den Erfolg einer Lehrveranstaltung? </div> <div style="border: 1px solid black; height: 80px; width: 100%;"></div> <div style="background-color: #f2f2f2; padding: 5px; margin-top: 5px;"> Was sind entscheidende Rahmenbedingungen für eine erfolgreiche Lehrveranstaltung? </div> <div style="border: 1px solid black; height: 80px; width: 100%;"></div> <div style="background-color: #f2f2f2; padding: 5px; margin-top: 5px;"> Gibt es Ihrer Meinung nach noch weitere Aspekte, die das Urteil einer Lehrgangsteilnehmerin/ eines Lehrgangsteilnehmers über die Güte der Lehrveranstaltung beeinflussen oder gar verzerren könnte? Wenn ja, welche? </div> <div style="border: 1px solid black; height: 80px; width: 100%;"></div> <p style="text-align: right;">Seite 4</p>

Anmerkungen: Der Fragebogen setzt sich aus zwei Teilen zusammen. Für diese Arbeit ist jedoch nur der erste Teil (Teil A) von Relevanz. Der zweite Teil (Teil B) fragt nach Aspekten von guten Prüfungen. Hiefür sei auf die Arbeit von Busjan (2016) verwiesen. In seiner Originalversion findet sich der Fragebogen zudem auch im digitalen Anhang.

Anhang D

Itembatterie

Dimension	Items	Adaptiert nach...
[Dozent_01] Auftreten, Persönlichkeit & Sozialkompetenz	1. Die Dozenten verhielten sich den Teilnehmern gegenüber freundlich und respektvoll.	Staufenbiel, 2000 (FE-VOR)
	2. Die Dozenten hatten für die Belange der Teilnehmer ein offenes Ohr.	Hirschfeld & Thielsch, 2009 (MFE-V)
	3. Ich finde, die Dozenten waren offen für Kritik.	Zumbach et al., 2007 (Kurzskala zur LE)
	4. Die Dozenten waren mir sehr sympathisch.	*
	5. Es herrschte eine gute Arbeitsatmosphäre.	Staufenbiel, 2000 (FE-SEM)
[Dozent_02] Fachkompetenz & Erfahrung	6. Die Dozenten konnten adäquat auf die Fragen der Teilnehmer antworten.	Grabbe, 2003
	7. Ich finde, die Dozenten wirkten fachlich sehr kompetent.	Wolf et al., 2001 (FB zur LVE)
	8. Ich finde, die Dozenten wirkten selbstsicher.	Grabbe, 2003
	9. Die Dozenten kannten sich mit den Inhalten der Lehre sehr gut aus.	Grabbe, 2003
	10. Die Dozenten konnten den Unterricht mit ihren eigenen Einsatzerfahrungen gut ergänzen.	*
[Dozent_03] Lehrkompetenz, Lehr- & Lernmethoden	11. Die Dozenten haben das Thema interessant aufgearbeitet.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	12. Die Dozenten benutzten oft Beispiele, die zum Verständnis der Lehrinhalte beitrugen.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	13. Die eingesetzten Lehrmethoden (z. B. Präsentation, Gruppenarbeit) trugen zum Verständnis der Inhalte bei.	Hirschfeld & Thielsch, 2009 (MFE-S)
	14. Die Lehrmethoden waren sehr abwechslungsreich.	Hirschfeld & Thielsch, 2009 (MFE-S)
	15. Ich finde, theoretische und praktische Anteile standen in einem angemessenen Verhältnis zueinander.	Robertz, 2015
	16. Es gab ausreichend Gelegenheiten, das Gelernte praktisch zu erproben.	*
	17. Ich finde, die Lehrmethoden waren gut auf unsere Teilnehmergruppe abgestimmt.	Robertz, 2015
	18. Ich profitiere von „Aha-Erlebnissen“ durch sehr eindrückliche Übungen oder Beispiele.	Robertz, 2015

[Dozent_04] Rhetorik & Verständlichkeit	19. Der sprachliche Ausdruck der Dozenten war verständlich.	Thielsch & Stegemöller, 2012 (MFE-ZMo)
	20. Die Dozenten erläuterten Fachbegriffe ausreichend.	*
	21. Die Dozenten waren redesicher.	Thielsch & Stegemöller, 2012 (MFE-ZMo)
	22. Die Dozenten fassten schwierige Sachverhalte prägnant zusammen.	Thielsch & Stegemöller, 2012 (MFE-ZMo)
[Dozent_05] Interaktion & Feedback	23. Ich finde, die Dozenten gingen auf Fragen und Anregungen der Teilnehmer angemessen ein.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	24. Ich finde, die Dozenten ließen ausreichend Raum für Diskussionen.	Robertz, 2015
	25. Die Teilnehmer konnten eigene Erfahrungen und Ansichten angemessen einbringen.	*
	26. Ich finde, die Dozenten gaben nützliches Feedback.	Thielsch & Stegemöller, 2012 (MFE-ZMo)
[Dozent_06] Engagement & Motivierung	27. Ich finde, die Dozenten waren engagiert.	Thielsch & Stegemöller, 2012 (MFE-ZMo)
	28. Die Dozenten motivierten mich, mich einzubringen.	*
	29. Ich hatte den Eindruck, dass die Dozenten gut vorbereitet waren.	Robertz, 2015
	30. Ich finde, die Dozenten waren am Lernerfolg der Teilnehmer interessiert.	Wolf et al., 2001 (FB zur LVE)
	31. Den Dozenten war es wichtig, dass alle Teilnehmer die Prüfung bestehen.	*
[Dozent_07] Struktur & Klarheit	32. Ich finde, die Lehrveranstaltung war klar strukturiert.	Zumbach et al., 2007 (Kurzskala zur LE)
	33. Ich konnte im Verlauf der Lehrveranstaltung die Gliederung immer nachvollziehen.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	34. Ich finde, die Lehrveranstaltung gab einen guten Überblick über das Themengebiet.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	35. Zu Beginn der Lehrveranstaltung wurden die Lernziele deutlich dargestellt.	Gediga et al., 2000 (KIEL)
	36. Ich finde, die Dozenten teilten die zur Verfügung stehende Zeit gut ein.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	37. Ich finde, Lern- und Pausenzeiten waren sinnvoll eingeteilt.	Robertz, 2015
	38. Es ging wenig Zeit für organisatorische Dinge (z. B. Absprachen, Materialbeschaffung) verloren.	*
	39. Es ging wenig Zeit für örtliche/räumliche Wechsel verloren.	*

[Dozent_08] Dozententeam	40. Ich finde, die Dozenten waren gut aufeinander abgestimmt.	*
	41. Die Dozenten verfolgten ein einheitliches Lehrkonzept.	*
	42. Ich finde, die Dozenten haben ihr Vorgehen gut miteinander abgestimmt.	*
	43. Ich habe die Dozenten als gleich kompetent erlebt. [Kontrollitem: Wurde entsprechend in der Vorbefragung nicht abgefragt]	*
[Inhalt_01] Umfang	44. (-) Der Umfang der vermittelten Themen hat mich überfordert.	*
	45. (-) Der Umfang der zu lernenden Inhalte hat mich überfordert.	*
[Inhalt_02] Tiefe	46. (-) Ich finde, manche Themen wurden zu oberflächlich behandelt.	*
	47. (-) Ich hätte mich mit einigen Themen gerne intensiver beschäftigt.	*
[Inhalt_03] Tempo	48. (-) Das Tempo der Stoffvermittlung war zu hoch für mich.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	49. (-) Zentrale Inhalte wurden nicht wiederholt.	*
	50. (-) Ich hatte Probleme dem Unterricht zu folgen.	*
[Inhalt_04] Schwierigkeit	51. (-) Die Inhalte der Lehrveranstaltung waren zu schwierig für mich.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	52. (-) Die Dozenten passten die vermittelten Inhalte zu wenig an den Wissensstand der Teilnehmer an.	*
	53. (-) Die Anforderungen dieser Lehrveranstaltung waren zu hoch.	Wolf et al., 2001 (FB zur LVE)
	54. (-) Ich hatte zu wenige Vorkenntnisse, um der Lehrveranstaltung folgen zu können.	Gediga et al., 2000 (KIEL)
[Inhalt_05] Relevanz	55. (-) Die Inhalte sind für meine berufliche Praxis nicht wichtig.	IdF NRW Fragebogen zur Veranstaltungsbewertung
	56. (-) Die Dozenten verdeutlichten die praktische Relevanz des Stoffes zu wenig.	Staufenbiel, 2000 (FESEM)
	57. (-) Mir persönlich war unklar, welche Inhalte prüfungsrelevant sind.	*

[Methode_01] Eigenverantwortliches Arbeiten [EVA]	58. Ich habe die EVA-Aufgaben ... gemacht. [Antwortvorgaben: nie/selten/gelegentlich/oft/immer]	*
	59. Ich hatte ausreichend Zeit die EVA-Aufgaben zu bearbeiten.	*
	60. Ich würde anderen Teilnehmern empfehlen, die EVA-Aufgaben zu bearbeiten. [Antwortvorgaben: Ja, Nein]	*
	61. Die EVA-Aufgaben haben mir als Vorbereitung auf die Prüfung sehr geholfen.	*
	62. Der Umfang der EVA-Aufgaben war angemessen.	Grötemeier & Thielsch, 2010 (MFE-ZHa)
	63. Die EVA-Aufgaben haben mein Verständnis der Inhalte weiter vertieft.	Grötemeier & Thielsch, 2010 (MFE-ZHa)
	64. Die Aufgabenstellungen der EVA-Aufgaben waren verständlich.	Grötemeier & Thielsch, 2010 (MFE-ZHa)
	65. Der Schwierigkeitsgrad der EVA-Aufgaben war angemessen.	Grötemeier & Thielsch, 2010 (MFE-ZHa)
[Methode_02] Gruppenarbeiten	66. Die EVA-Aufgaben wurden ausreichend nachbesprochen.	Grötemeier & Thielsch, 2010 (MFE-ZHa)
	67. Ich habe bei den Gruppenarbeiten viel von anderen Teilnehmern gelernt.	Grötemeier & Thielsch, 2010 (MFE-ZGr)
	68. Die Unterstützung der Dozenten bei den Gruppenarbeiten war ausreichend.	Grötemeier & Thielsch, 2010 (MFE-ZGr)
	69. Alle Teilnehmer haben sich, entsprechend ihrer Kompetenzen, an den Gruppenarbeiten beteiligt.	*
	70. Die Gruppenarbeiten wurden ausreichend nachbesprochen.	*
	71. Mir war an einem guten Ergebnis der Gruppenarbeiten gelegen.	Grötemeier & Thielsch, 2010 (MFE-ZGr)
	72. Die Gruppenarbeiten haben den Austausch zwischen den Teilnehmern gefördert.	*
	73. Die Leitfragen/Aufgabenstellungen der Gruppenarbeiten waren verständlich.	*
	74. Die für die Gruppenarbeiten vorgegebene Zeit war ausreichend.	*

[Methode_03] Planübung	75. Ich habe bei der Planübung viel gelernt.	Hirschfeld & Thielsch, 2009 (MFE-S)
	76. Die Dozenten gaben mir nützliches Feedback zu meinen Leistungen bei der Planübung.	*
	77. In der Planübung konnte ich das neuerworbene Wissen anwenden.	*
	78. Der Schwierigkeitsgrad der Planübung war angemessen.	*
[Methode_04] Einsatzübung	79. Ich habe bei den Einsatzübungen viel gelernt.	Hirschfeld & Thielsch, 2009 (MFE-S)
	80. Die Dozenten gaben mir nützliches Feedback zu meinen Leistungen in der Einsatzübung.	*
	81. In den Einsatzübungen konnte ich das neuerworbene Wissen anwenden.	*
	82. Der Schwierigkeitsgrad der Einsatzübung war angemessen.	*
[Teilnehmer_01] Vorbereitung, Vorwissen	83. Ich hatte ausreichend Zeit mich auf diesen Lehrgang vorzubereiten.	*
	84. Ich fühlte mich auf diesen Lehrgang sehr gut vorbereitet.	*
	85. Es wurden im Vorfeld ausreichend Materialien zur Vorbereitung auf diesen Lehrgang angeboten.	*
	86. Ich wurde von meiner Heimatstelle sehr gut vorbereitet.	*
	87. Ich habe ein Seminar zur Vorbereitung auf diesen Lehrgang besucht. [Antwortvorgaben: Ja, Nein]	*
	88. Ich habe das Angebot vom IdF zur Vorbereitung auf diesen Lehrgang genutzt. [Antwortvorgaben: Ja, Nein]	*
[Teilnehmer_02] (Vor-)Interesse, Lernbereitschaft	89. Die Inhalte dieser Lehrveranstaltung haben mich bereits vor Besuch dieser Veranstaltung sehr interessiert.	Wolf et al., 2001 (FB zur LVE)
	90. Ich nehme an dieser Lehrveranstaltung freiwillig teil. [Antwortvorgaben: Ja, Nein]	*
	91. Ich halte den Besuch dieser Lehrveranstaltung für mein weiteres dienstliches Fortkommen für sehr sinnvoll.	Wolf et al., 2001 (FB zur LVE)
	92. Ich war sehr motiviert meine Fähigkeiten zu erweitern.	*

[Gruppe_01] Mitarbeit	93. Die anderen Teilnehmer waren gut auf diesen Lehrgang vorbereitet.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	94. Die anderen Teilnehmer brachten sich aktiv ein.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	95. Die anderen Teilnehmer verfolgten den Lehrgang aufmerksam und mit Interesse.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
[Gruppe_02] Zusammenhalt	96. Die Teilnehmer unterstützen sich gegenseitig.	*
	97. Ich finde, es herrschte ein guter Zusammenhalt im Lehrgang.	*
	98. Ich konnte mich auf die anderen Teilnehmer immer verlassen.	*
[Rahmenbedingungen_01] Unterbringung	99. Mit der Unterkunft im Internatsbereich des IdF war ich voll zufrieden.	IdF NRW Fragebogen zur Veranstaltungsbewertung
	100. Mit der Verpflegung durch das IdF war ich voll zufrieden.	IdF NRW Fragebogen zur Veranstaltungsbewertung
	101. Mit dem Freizeitangebot des IdF war ich voll zufrieden.	IdF NRW Fragebogen zur Veranstaltungsbewertung
[Rahmenbedingungen_02] Räumliche & materielle Ausstattung	102. Ich finde, die bereitgestellten Unterlagen waren verständlich.	*
	103. Die Qualität der in der Lehrveranstaltung eingesetzten Medien (<i>Folien, Filme, Skizzen, etc.</i>) war gut.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	104. Die in der Lehrveranstaltung verwendeten Medien (<i>Folien, Filme, Skizzen, etc.</i>) trugen zum Verständnis der Inhalte bei.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	105. Die benötigten Materialien (<i>Fahrzeuge, Ausstattung, etc.</i>) waren immer in ausreichender Anzahl verfügbar. [Antwortvorgaben: Ja, Nein]	*
	106. Ich finde, die Materialien (<i>Fahrzeuge, Ausstattung, etc.</i>) waren auf dem neusten Stand. [Antwortvorgaben: Ja, Nein]	*
	107. Ich finde die Menge des Materials, das zur Vertiefung der Inhalte bereitgestellt wurde, war ... [Antwortvorgaben: zu gering, angemessen, zu groß]	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	108. Ich finde die Menge des Materials, das zur Prüfungsvorbereitung bereitgestellt wurde, war ... [Antwortvorgaben: zu gering, angemessen, zu groß]	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	109. Ich finde, die räumliche Ausstattung der Veranstaltungsräume war angemessen.	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	110. Die Anzahl der verfügbaren Übungsobjekte war angemessen.	*
	111. Die Übungshalle/-plätze standen dem Lehrgang bedarfsgerecht zur Verfügung.	*

[Rahmenbedingungen_03] (Zeitlicher) Ablauf	112. Die täglichen Unterrichtszeiten waren angemessen.	*
	113. Der mit dem Lehrgang verbundene Zeitaufwand war passend für mich.	Robertz, 2015
	114. [Zusatzitem bei zweigeteilter Lehrgangsform:] Aus organisatorischen Gründen war es leichter für mich, den Lehrgang in zweigeteilter Form zu besuchen.	*
	115. [Zusatzitem bei zweigeteilter Lehrgangsform:] (-) Durch die zeitliche Unterbrechung des Lehrgangs habe ich viele Inhalte vergessen.	*
	116. [Zusatzitem bei zweigeteilter Lehrgangsform:] Während der zeitlichen Unterbrechung des Lehrgangs habe ich Inhalte nachgearbeitet.	*
	117. [Zusatzitem bei zweigeteilter Lehrgangsform:] Während der zeitlichen Unterbrechung des Lehrgangs habe ich mich auf die Prüfung vorbereitet.	*
	118. [Zusatzitem bei zweigeteilter Lehrgangsform:] (-) Ich hätte den Lehrgang lieber an einem Stück besucht.	*
[Rahmenbedingungen_04] Teilnehmer	119. Die Größe der Teilnehmergruppe war angemessen.	*
	120. Ich finde, die Teilnehmergruppe war gut zusammengestellt.	Robertz, 2015
[Rahmenbedingungen_05] Lehrgang	121. Ich konnte mich voll und ganz auf den Lehrgang konzentrieren.	*
	122. Ich hatte keinen Druck von zuhause (z. B. durch Familie, Heimatfeuerwehr und/oder Arbeitgeber).	*
[Lehrerfolg_01] Lernerfolg	123. Ich habe in der Lehrveranstaltung sehr viel gelernt.	Gediga et al., 2000 (KIEL)
	124. Ich konnte mein Wissen deutlich erweitern.	Robertz, 2015
	125. Ich fühle mich auf die Prüfung sehr gut vorbereitet.	*
	126. Ich fühle mich auf meinen nächsten Einsatz, den ich als Gruppen-/Zugführer leiten werde, sehr gut vorbereitet.	*
	127. Durch die praktischen Übungen im Lehrgang habe ich die nötige Sicherheit gewonnen, einen Einsatz als Gruppen-/Zugführer zu leiten.	*
	128. Ich kann das Gelernte auf meine Heimatstelle übertragen.	*

[Lehrerfolg_02]
 Kompetenzerwerb


129. Durch meine Teilnahme am Lehrgang gelingt es mir jetzt besser, Gefahrenlagen frühzeitig zu erkennen. [>Wahrnehmungsfähigkeit]	*	(in Anlehnung an Misteale, 2007)
130. Durch meine Teilnahme am Lehrgang fällt es mir jetzt leichter, Entscheidungen in kritischen Situationen zu treffen. [>Entscheidungsfähigkeit]	*	(in Anlehnung an Misteale, 2007)
131. Durch meine Teilnahme am Lehrgang kann ich meine eigene Fähigkeit als Gruppen-/Zugführer besser einschätzen. [>Selbsteinschätzungsfähigkeit]	*	(in Anlehnung an Misteale, 2007)
132. Durch meine Teilnahme am Lehrgang kenne ich meine persönlichen Grenzen besser. [>Selbsteinschätzungsfähigkeit: eigene Grenzen kennen]	*	(in Anlehnung an Misteale, 2007)
133. Durch meine Teilnahme am Lehrgang traue ich mir besser zu, in stressigen Situationen ruhig zu bleiben. [>Selbsteinschätzungsfähigkeit: Stressresistenz]	*	(in Anlehnung an Misteale, 2007)
134. Durch meine Teilnahme am Lehrgang bin ich jetzt noch motivierter, die Rolle des GF/ZF zu übernehmen. [>Motivation & organisationales Commitment]	*	(in Anlehnung an Misteale, 2007)
135. Durch meine Teilnahme am Lehrgang bin ich darin bestärkt, Führungsverantwortung zu übernehmen. [>Motivation & organisationales Commitment]	*	(in Anlehnung an Misteale, 2007)
136. Durch meine Teilnahme am Lehrgang habe ich ein besseres Verständnis davon, welche Informationen mein Trupp von mir benötigt. [>Teamhandeln]	*	(in Anlehnung an Misteale, 2007)
137. Durch meine Teilnahme am Lehrgang ist mir klarer geworden, was meine Gruppe/mein Zug von mir erwartet. [>Kooperationsfähigkeit]	*	(in Anlehnung an Misteale, 2007)
138. Durch meine Teilnahme am Lehrgang kann ich korrektere und verständlichere Anweisungen geben. [>Kooperationsfähigkeit: Kommunikation]	*	(in Anlehnung an Misteale, 2007)
139. Durch meine Teilnahme am Lehrgang, sehe ich stärker die Sinnhaftigkeit von Einsatznachbesprechungen. [>Arbeitsbezogenes, erfahrungsbasiertes Arbeiten]	*	(in Anlehnung an Misteale, 2007)
140. Durch meine Teilnahme am Lehrgang werde ich Einsatznachbesprechungen erfolgreich leiten können. [>Arbeitsbezogenes, erfahrungsbasiertes Arbeiten]	*	(in Anlehnung an Misteale, 2007)
141. Ich kann meine Arbeit jetzt besser planen und koordinieren. [>Führung & Managementfähigkeit: Planung & Koordination]	*	(in Anlehnung an Misteale, 2007)
142. Ich kann meine Entscheidungen jetzt besser durchsetzen. [>Führung & Managementfähigkeit: Assertiveness]	*	(in Anlehnung an Misteale, 2007)
143. Ich beherrsche den Führungsvorgang jetzt besser. [Führung]	*	
144. Ich beherrsche wichtige Führungstechniken jetzt besser. [Führung]	*	

[Leherfolg_03] Globalurteil	145. Ich würde dieser Lehrveranstaltung insgesamt folgende Schulnote geben: [Antwortvorgaben: 1 („sehr gut“); 2 („gut“); 3 („befriedigend“); 4 („ausreichend“); 5 („mangelhaft“); 6 („ungenügend“)]	Staufenbiel, 2000 (FEVOR/FESEM)
	146. Ich würde diesen Lehrgang weiterempfehlen. [Antwortvorgaben: Ja, Nein]	Thielsch & Hirschfeld, 2012 (MFE-Sr)
	147. Haben Sie Ihre Ziele erreicht? [Antwortvorgaben: Ja, Nein]	*
Offene Fragen	148. Warum wollen Sie Gruppen-/Zugführer werden?	*
	149. Warum haben Sie diesen Lehrgang besucht? Was waren Ihre Ziele?	*
	150. Was war gut an der Veranstaltung?	Gediga et al., 2000 (KIEL)
	151. Was hat Ihnen an der Lehrveranstaltung nicht gefallen? Und welche Verbesserungen schlagen Sie vor?	Gediga et al., 2000 (KIEL)
	152. In welchen Teilgebieten des Lehrstoffes fühlen Sie sich noch unsicher?	Gediga et al., 2000 (KIEL)


Anmerkungen. Die rot gekennzeichneten Items wurden im Anschluss an Studie II ausgeschlossen und fanden demzufolge keine Berücksichtigung in Studie III. Die grau markierten Items wurden für den von Loberg (2016) vorgenommenen Vergleich verschiedener Lehrgangsformen entwickelt. Da diese jedoch gezielt nur in einem Lehrgang zum Einsatz kamen, wurden diese in der vorliegenden Arbeit nicht weiter thematisiert. Bei den mit einem Sternchen (*) markierten Items handelt es sich um neue und gänzlich selbstkonstruierte Items. Die mit einem Minus (-) gekennzeichneten Items sind negativ gepolte Items und müssen für die Auswertung und Interpretation entsprechend zunächst umgepolzt werden. Sofern keine anderen Antwortoptionen angegeben sind, sind die einer Items auf Likert-Skala (von 1 = *stimme gar nicht zu* bis 7 = *stimme vollkommen zu* sowie 0 = *nicht sinnvoll beantwortbar*) zu beurteilen.

Anhang E

Fragebogen (Studie II; Teilnehmerversion A) – Auszug



Westfälische
Wilhelms-Universität
Münster



IdF

Fragebogen „Guter Lehrgang“

Liebe/r Lehrgangsteilnehmer/in,

Wir schätzen es sehr, dass Sie sich bereit erklärt haben, uns mit Ihrer Teilnahme an dieser Befragung zu unterstützen. Diese findet im Rahmen eines Kooperationsprojektes zwischen dem Institut der Feuerwehr (IdF) Nordrhein-Westfalen und der Westfälischen Wilhelms-Universität (WWU) Münster statt. Das Projekt wurde im Zuge der Neuausrichtung der Führungsausbildung am IdF ins Leben gerufen, mit dem Ziel die Güte der alten und neuen Lehrgänge am IdF zu beurteilen.

Im Vorfeld interessieren wir uns zunächst dafür, **was für Sie als Lehrgangsteilnehmer/ eine gute Lehrveranstaltung ausmacht und anhand welcher Kriterien Sie dies bewerten.**

Hierfür werden wir Ihnen nachfolgend einige Aussagen präsentieren, die für den Erfolg einer Lehrveranstaltung mehr oder weniger relevant sein können. Bitte lesen Sie sich die Aussagen aufmerksam durch und geben Sie auf der dahinterstehenden Skala jeweils an, **für wie wichtig Sie diese Aussage halten, wenn es darum geht, die Güte einer Lehrveranstaltung zu bewerten.**

Welche Aspekte sind aus Ihrer Sicht für den Erfolg einer Lehrveranstaltung wirklich wichtig und welche Aspekte sind weniger bzw. gar nicht wichtig? – Machen Sie Ihre Kreuze bitte nur in den dafür vorgesehenen Kästchen und nicht zwischen den Kästchen. Wenn Sie eine Aussage sprachlich oder inhaltlich nicht verstehen, lassen Sie uns dies bitte unbedingt wissen, indem Sie das Kästchen „nicht verständlich“ ankreuzen.

Bitte beachten Sie:


- Es gibt bei dieser Befragung keine richtigen oder falschen Antworten. Vielmehr interessieren wir uns für **Ihre ganz persönliche Meinung.**
- Unser Ziel ist es, mit Hilfe Ihrer Angaben, **eine gute Auswahl** an geeigneten Fragen zusammenzustellen, mit denen die Führungslehrgänge am IdF zukünftig beurteilt werden können. Auch wenn Ihnen auf den ersten Blick alle Aussagen wichtig erscheinen sollten, **möchten wir Sie bitten kritisch zu sein. Helfen Sie uns, die Vielzahl der hier präsentierten Aussagen auf die wirklich zentralen Aspekte guter Lehre zu reduzieren.**
- Aus Gründen der besseren Lesbarkeit wird nachfolgend jeweils nur die männliche Form (Dozenten/ Teilnehmer) verwendet, die weibliche Form ist dabei jeweils mit eingeschlossen.
- **Diese Befragung erfolgt selbstverständlich anonym.** Um Ihre Angaben jedoch auswerten zu können, benötigen wir von Ihnen folgende demografische Daten:

Alter: _____ Jahre **Geschlecht:** ☐ weiblich ☐ männlich


Wenn Sie Fragen oder Anregungen haben, wenden Sie sich gerne per Email an uns (s_babi02@uni-muenster.de). **Im Namen aller Beteiligten danke ich Ihnen ganz herzlich für Ihre Teilnahme,**

Stephanie Babel, Studentin der WWU Münster

Seite 1



Westfälische
Wilhelms-Universität
Münster



IdF

Für die Güte einer Lehrveranstaltung ist der Aspekt ...

	unwichtig	eher unwichtig	eher wichtig	wichtig	nicht verständlich
1. Die Lehrveranstaltung ist klar strukturiert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Die Teilnehmer können die Gliederung im Verlauf der Lehrveranstaltung immer nachvollziehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Die Lehrveranstaltung gibt einen guten Überblick über das Themengebiet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Zu Beginn der Lehrveranstaltung werden die Lernziele deutlich dargestellt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Die Dozenten teilen die zur Verfügung stehende Zeit gut ein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Lern- und Pausenzeiten sind sinnvoll eingeteilt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Es geht wenig Zeit für organisatorische Dinge (z.B. Absprachen, Materialbeschaffung) verloren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Es geht wenig Zeit für örtliche/räumliche Wechsel verloren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Seite 2


Anmerkung: Der vollständige Fragebogen findet sich, wie auch die Dozentenversion, im digitalen Anhang.

Anhang F

Evaluationsbogens (Studie III) – Auszug

WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Lehrgang: Gruppenführer



Liebe/r Lehrgangsteilnehmer/in,

wir schätzen es sehr, dass Sie an dieser Befragung teilnehmen. Diese findet im Rahmen eines Kooperationsprojektes zwischen dem Institut der Feuerwehr (IdF) Nordrhein-Westfalen und der Westfälischen Wilhelms-Universität (WWU) Münster statt. Das Projekt wurde im Zuge der Neuausrichtung der Führungsausbildung am IdF ins Leben gerufen, mit dem Ziel die Güte der alten und neuen Lehrgänge am IdF zu beurteilen.


Sie helfen uns die Qualität der Lehre am IdF zu beurteilen und gegebenenfalls zu verbessern, indem Sie diesen Lehrgang hinsichtlich verschiedener Aspekte bewerten. Hierzu werden wir Ihnen nachfolgend zunächst einige Aussagen präsentieren. Bitte teilen Sie uns für jede Aussage mit, wie sehr Sie dieser zustimmen. Kreuzen Sie für jede Aussage einen der vorgegebenen Kreise an, der den Grad Ihrer Zustimmung am besten wiedergibt. Anschließend präsentieren wir Ihnen auch einige Fragen, die Sie gerne stichpunktartig beantworten können. Generell gilt dabei: Es gibt bei dieser Befragung keine richtigen oder falschen Antworten. **Vielmehr interessieren wir uns für Ihre ganz persönliche Meinung.**

Bitte beachten Sie:

- Machen Sie hinter jeder Aussage jeweils **nur ein Kreuz in** einem der dafür vorgesehenen Kreise (und niemals zwischen oder neben Kreisen). Bitte achten Sie zudem darauf, keine Aussage auszulassen. Wenn eine Aussage für Sie nicht sinnvoll beantwortbar ist, können Sie uns dies durch ein Kreuz in dem hintersten Kreis kundtun.
- Aus Gründen der besseren Lesbarkeit wird nachfolgend jeweils nur die männliche Form (Dozenten/ Teilnehmer) verwendet, die weibliche Form ist dabei jeweils mit eingeschlossen.
- Ihre Teilnahme an dieser Befragung erfolgt **freiwillig**. Bitte geben Sie am Ende des Fragebogens an, ob wir Ihre Angaben für die Beurteilung der Lehre am IdF verwenden dürfen.
- Die Befragung erfolgt selbstverständlich **anonym**. Die Ergebnisse werden nur in gesamelter Form, d.h. z.B. in Form von Mittelwerten aller befragten Lehrgänge, an das IdF zurückgemeldet. Ein Rückschluss auf Ihre Person ist damit ausgeschlossen. Um Ihre Angaben jedoch gruppenstatistisch auswerten zu können, benötigen wir von Ihnen einige demografische Angaben.

Seite 1

WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER



Alter: _____ Jahre

Geschlecht: ☐ weiblich ☐ männlich ☐ keine Angabe

Was ist der höchste Schulabschluss, den Sie besitzen?

☐ Keinen Schulabschluss
☐ Haupt-/Volksschulabschluss
☐ Realschulabschluss/ Mittlere Reife
☐ Fachhochschulreife
☐ Abitur
☐ Sonstiger Schulabschluss: _____

Was ist Ihr derzeitiger beruflicher Ausbildungsstand?

☐ Keine abgeschlossene Berufsausbildung
☐ Noch in Ausbildung oder Student/in
☐ Abgeschlossene Lehre oder Meisterbrief
☐ Berufsakademie-, Fachhochschul- oder Universitätsabschluss
☐ Promotion oder höher
☐ Sonstiger beruflicher Ausbildungsstand: _____

Wie viel (Feuerwehr-) Einsatzerfahrung haben Sie?

☐ hauptberuflich: _____ Jahre _____ Einsätze/ Monat
☐ ehrenamtlich: _____ Jahre _____ Einsätze/ Monat

Sind Sie aktuell hauptberuflich und/oder ehrenamtlich bei der Feuerwehr tätig?

☐ hauptberuflich ☐ ehrenamtlich

Seite 2

In diesem Fragebogen geht es zunächst nur um die Beurteilung der Lehrveranstaltung. Nach den Prüfungen werden wir Sie morgen und übermorgen jeweils auch zu den verschiedenen Prüfungsformen befragen. Um Ihre Angaben einander zuordnen zu können und Ihnen dennoch Anonymität zu gewähren, **möchten wir Sie daher bitten einen persönlichen Code zu generieren.**

Bitte generieren Sie Ihren persönlichen Code nach dem von uns vorgegebenen Schema. Wir präsentieren Ihnen dieses Schema bei jeder Befragung. So müssen Sie sich Ihren Code nicht merken, sondern können diesen mit Hilfe der Anleitung jedes Mal wieder zusammenstellen.

Ihr persönlicher Code ergibt sich wie folgt:






Erster und letzter Buchstabe der Stadt, in der Sie geboren wurden:
(Angenommen Sie wurden in **M**ünster geboren, dann würden Sie nun ein **M** in das erste und ein **R** in das zweite rechte Kästchen schreiben)

Die beiden Ziffern des Monats, in dem Sie geboren wurden:
(Angenommen Sie wurden im **Juni** geboren, dann würden Sie nun die Ziffern **0** und **6** in die beiden Kästchen rechts eintragen (Juni = 6. Monat im Jahr))

Erster und zweiter Buchstabe des Vornamens Ihrer Mutter:
(Angenommen Ihre Mutter hieß **Birgit**, dann würden Sie nun ein **B** in das erste und ein **I** in das zweite rechte Kästchen schreiben)

Erster und zweiter Buchstabe des Vornamens Ihres Vaters:
(Angenommen Ihr Vater hieß **Jochen**, dann würden Sie nun ein **J** in das erste und ein **O** in das zweite rechte Kästchen schreiben)

Bevor Sie gleich mit der Beurteilung des Lehrgangs beginnen können interessieren wir uns noch für Ihre aktuelle Stimmungslage. **Wählen Sie hierzu bitte denjenigen Smiley aus, der Ihre momentane Stimmung am besten widerspiegelt:**

Wenn Sie Fragen oder Anregungen haben, wenden Sie sich gerne per Email an uns (s_babi02@uni-muenster.de). **Im Namen aller Beteiligten danken wir Ihnen ganz herzlich für Ihre Teilnahme,**

Stephanie Babel, B.Sc. (WWU Münster)
PD Dr. Meinold Thielsch (WWU Münster)

Seite 3

	Stimme gar nicht zu	Stimme nicht zu	Stimme eher nicht zu	Neutral	Stimme eher zu	Stimme zu	Stimme vollkommen zu	Nicht sinnvoll beantwortbar
1. Ich finde, die Lehrveranstaltung war klar strukturiert.	1	2	3	4	5	6	7	8
2. Ich konnte im Verlauf der Lehrveranstaltung die Gliederung immer nachvollziehen.	1	2	3	4	5	6	7	8
3. Ich finde, die Lehrveranstaltung gab einen guten Überblick über das Themengebiet.	1	2	3	4	5	6	7	8
4. Zu Beginn der Lehrveranstaltung wurden die Lernziele deutlich dargestellt.	1	2	3	4	5	6	7	8
5. Ich finde, die Dozenten teilten die zur Verfügung stehende Zeit gut ein.	1	2	3	4	5	6	7	8
6. Ich finde, Lern- und Pausenzeiten waren sinnvoll eingeteilt.	1	2	3	4	5	6	7	8
7. Es ging wenig Zeit für organisatorische Dinge (z.B. Absprachen, Materialbeschaffung) verloren.	1	2	3	4	5	6	7	8
8. Die Dozenten haben das Thema interessant aufgearbeitet.	1	2	3	4	5	6	7	8
9. Die Dozenten benutzten oft Beispiele, die zum Verständnis der Lehrinhalte beitrugen.	1	2	3	4	5	6	7	8
10. Die eingesetzten Lehrmethoden (z.B. Präsentation, Gruppenarbeit) trugen zum Verständnis der Inhalte bei.	1	2	3	4	5	6	7	8
11. Ich finde, theoretische und praktische Anteile standen in einem angemessenen Verhältnis zueinander.	1	2	3	4	5	6	7	8
12. Es gab ausreichend Gelegenheiten, das Gelernte praktisch zu erproben.	1	2	3	4	5	6	7	8
13. Ich finde, die Lehrmethoden waren gut auf unsere Teilnehmergruppe abgestimmt.	1	2	3	4	5	6	7	8
14. Ich profitierte von „Aha-Erlebnissen“ durch sehr eindrückliche Übungen oder Beispiele.	1	2	3	4	5	6	7	8
15. Ich finde, die Dozenten gingen auf Fragen und Anregungen der Teilnehmer angemessen ein.	1	2	3	4	5	6	7	8

Seite 4

Anmerkung: Der vollständige Fragebogen findet sich, wie auch die Versionen zur Evaluation anderer Lehrgangsformen findet sich im digitalen Anhang.

Anhang G

Gruppierung der Items für weitere Analysen

Gruppe 1: Kein 7-stufiges Antwortformat	Gruppe 2: Biasvariablen	Gruppe 3: EFA I	Gruppe 4: EFA II
Item 58	Item 37	Item 123	Item 6
Item 60	Item 38	Item 126	Item 11
Item 88	Item 57	Item 127	Item 12
Item 99	Item 84	Item 128	Item 13
Item 90	Item 86	Item 129	Item 15
Item 105	Item 99	Item 130	Item 16
	Item 100	Item 131	Item 17
	Item 103	Item 132	Item 18
	Item 104	Item 133	Item 19
	Item 111	Item 134	Item 20
	Item 112	Item 135	Item 22
	Item 113	Item 137	Item 23
	Item 119	Item 138	Item 26
	Item 121	Item 139	Item 27
	Item 122	Item 140	Item 28
	Item 125	Item 141	Item 29
		Item 142	Item 32
			Item 33
			Item 34
			Item 35
			Item 36
			Item 42
			Item 43
			Item 45
			Item 47
			Item 48
			Item 49
			Item 51
			Item 52
			Item 54
			Item 56
			Item 59
			Item 62
			Item 63
			Item 64
			Item 65
			Item 66
			Item 67
			Item 68
			Item 69
			Item 70
			Item 72
			Item 73
			Item 75
			Item 76
			Item 77
			Item 78
			Item 94
			Item 96
			Item 97

Anhang H

Ergebnisse der ersten exploratorischen Faktorenanalyse (Faktorladungen, MW und SD)

	Faktor I.1	Faktor I.2	MW (SD)
Item 123		0,70	6,18 (0,79)
Item 126		0,77	5,55 (1,08)
Item 127		0,80	5,53 (1,14)
Item 128		0,92	5,75 (1,13)
Item 129	0,68		5,84 (0,87)
Item 130	0,77		5,76 (0,90)
Item 131	0,55	0,32	5,81 (0,82)
Item 132	0,78		5,61 (1,09)
Item 133	0,83		5,46 (1,02)
Item 134	0,56		5,86 (0,97)
Item 135	0,85		5,96 (0,96)
Item 137	0,72		6,08 (0,87)
Item 138	0,90		5,86 (0,88)
Item 139	0,73		6,06 (0,98)
Item 140	0,50		5,64 (1,01)
Item 141		0,63	5,92 (0,79)
Item 142	0,41	0,45	5,67 (1,00)

Anhang I

Ergebnisse der zweiten exploratorischen Faktorenanalyse (Faktorladungen, *MW* und *SD*)

	Faktor II.1	Faktor II.2	Faktor II.3	Faktor II.4	Faktor II.5	Faktor II.6	Faktor II.7	Faktor II.8	<i>MW (SD)</i>
Item 6	0,62								6,25 (0,80)
Item 11	0,56			0,43					6,31 (0,77)
Item 12	0,67								6,37 (0,81)
Item 13	0,45			0,30					5,95 (1,03)
Item 15							0,71		5,69 (1,25)
Item 16							0,80		4,93 (1,64)
Item 17							0,52		5,57 (1,08)
Item 18	0,37								6,08 (0,93)
Item 19	0,79								6,50 (0,58)
Item 20	0,62	0,33							6,24 (0,77)
Item 22	0,77								6,07 (0,84)
Item 23	0,53						0,40		6,25 (0,83)
Item 26	0,70						0,31		6,34 (0,89)
Item 27	0,78								6,69 (0,52)
Item 28	0,78								6,20 (0,87)
Item 29	0,75								6,46 (0,72)
Item 32				0,73					6,00 (0,85)
Item 33				0,74					5,73 (0,95)
Item 34				0,73					5,92 (0,88)
Item 35				0,73					5,74 (1,14)
Item 36									6,13 (0,86)
Item 42	0,58								5,79 (1,23)
Item 43	0,58								5,67 (1,39)
Item 45		0,77							5,15 (1,47)

	Faktor II.1	Faktor II.2	Faktor II.3	Faktor II.4	Faktor II.5	Faktor II.6	Faktor II.7	Faktor II.8	MW (SD)
Item 47		0,36					0,59		3,07 (1,38)
Item 48		0,83							5,21 (1,45)
Item 49		0,46					0,35		5,04 (1,53)
Item 51		0,82							5,86 (1,13)
Item 52		0,62							5,56 (1,25)
Item 54		0,78							5,75 (1,22)
Item 56		0,38							5,89 (1,08)
Item 59						0,79			5,89 (1,17)
Item 62						0,81			5,57 (1,30)
Item 63						0,65			5,78 (1,07)
Item 64						0,58			5,71 (0,97)
Item 65						0,70			5,75 (0,93)
Item 66			0,44				0,34		3,75 (1,77)
Item 67			0,64						5,14 (1,32)
Item 68			0,65						5,53 (1,06)
Item 69			0,85						5,72 (1,02)
Item 70			0,59						5,62 (1,27)
Item 72			0,65						6,06 (1,05)
Item 73			0,55						5,86 (0,96)
Item 75					0,73				5,73 (1,06)
Item 76					0,74				5,83 (1,09)
Item 77					0,88				5,76 (1,04)
Item 78					0,76				5,84 (0,96)
Item 94								0,62	5,91 (0,79)
Item 96								0,88	6,39 (0,69)
Item 97								0,73	6,57 (0,68)

Anhang J

Korrelationen zwischen den Biasvariablen und den vier definierten Lehrerfolgskriterien

Biasvariablen	Kompetenzerwerb (Skala I.1)	Vorbereitung & Transfer (Skala I.2)	Gesamtbenotung (Item 145*)	Subjektiver Lerngewinn (Item 123)
Item 37	0,21**	0,15*	0,21**	0,17**
Item 38	0,13*	0,07	0,11	0,15*
Item 57	0,29**	0,24**	0,24**	0,25**
Item 84	0,33**	0,19**	0,31**	0,18**
Item 86	0,17**	0,16*	0,16*	0,09
Item 99	0,19**	0,09	0,08	0,14*
Item 100	0,11	0,07	0,06	0,08
Item 103	0,26**	0,18**	0,24**	0,25**
Item 104	0,27**	0,26**	0,25**	0,27**
Item 111	0,17*	0,22**	0,25**	0,18**
Item 112	0,28**	0,09	0,26**	0,21**
Item 113	0,40**	0,26**	0,34**	0,22**
Item 119	0,47**	0,21**	0,52**	0,38**
Item 121	0,38**	0,38**	0,44**	0,44**
Item 122	0,16*	0,12	0,18**	0,24**
Item 125	0,44**	0,37**	0,59**	0,42**

Anmerkung: * $p < 0,05$ (zweiseitig), ** $p < 0,01$ (zweiseitig).

Anhang K

Standardfragebogen zur Evaluation der Gruppen- und Zugführerlehrgänge am IdF NRW.

	Stimme gar nicht zu	Stimme nicht zu	Stimme eher nicht zu	Neutral	Stimme eher zu	Stimme zu	Stimme vollkommen zu	Nicht sinnvoll beantwortbar
1. Der sprachliche Ausdruck der Dozenten war verständlich.	1	2	3	4	5	6	7	0
2. Ich finde, die Dozenten gingen auf Fragen und Anregungen der Teilnehmer angemessen ein	1	2	3	4	5	6	7	0
3. Ich finde, die Dozenten gaben nützliches Feedback.	1	2	3	4	5	6	7	0
4. Die Dozenten motivierten mich, mich einzubringen.	1	2	3	4	5	6	7	0
5. Ich finde, die Dozenten haben ihr Vorgehen gut miteinander abgestimmt.	1	2	3	4	5	6	7	0
6. Ich finde, theoretische und praktische Anteile standen in einem angemessenen Verhältnis zueinander.	1	2	3	4	5	6	7	0
7. Es gab ausreichend Gelegenheiten, das Gelernte praktisch zu erproben.	1	2	3	4	5	6	7	0
8. Ich finde, die Lehrmethoden waren gut auf unsere Teilnehmergruppe abgestimmt.	1	2	3	4	5	6	7	0
9. Ich konnte im Verlauf der Lehrveranstaltung die Gliederung immer nachvollziehen.	1	2	3	4	5	6	7	0
10. Ich finde, die Lehrveranstaltung gab einen guten Überblick über das Themengebiet.	1	2	3	4	5	6	7	0
11. Zu Beginn der Lehrveranstaltung wurden die Lernziele deutlich dargestellt.	1	2	3	4	5	6	7	0
12. Der Umfang der zu lernenden Inhalte hat mich überfordert.	1	2	3	4	5	6	7	0
13. Das Tempo der Stoffvermittlung war zu hoch für mich.	1	2	3	4	5	6	7	0
14. Die Dozenten passten die vermittelten Inhalte zu wenig an den Wissensstand der Teilnehmer an.	1	2	3	4	5	6	7	0
15. Die anderen Teilnehmer brachten sich aktiv ein.	1	2	3	4	5	6	7	0
16. Die Teilnehmer unterstützen sich gegenseitig.	1	2	3	4	5	6	7	0
17. Ich finde, es herrschte ein guter Zusammenhalt im Lehrgang.	1	2	3	4	5	6	7	0



Liebe/r Lehrgangsteilnehmer/in,

wir schätzen es sehr, dass Sie an dieser Befragung teilnehmen. Diese findet im Rahmen eines Kooperationsprojektes zwischen dem Institut der Feuerwehr (IdF) Nordrhein-Westfalen und der Westfälischen Wilhelms-Universität (WWU) Münster statt.

Mit ihren Antworten helfen Sie uns die Qualität der Lehre am IdF zu beurteilen und gegebenenfalls zu verbessern, indem Sie diesen Lehrgang hinsichtlich verschiedener Aspekte bewerten. Hierzu werden wir Ihnen nachfolgend zunächst einige Aussagen präsentieren. Bitte teilen Sie uns für jede Aussage mit, wie sehr Sie dieser zustimmen. Kreuzen Sie für jede Aussage einen der vorgegebenen Kreise an, der den Grad Ihrer Zustimmung am besten wiedergibt. Anschließend präsentieren wir Ihnen auch einige Fragen, die Sie gerne stichpunktartig beantworten können. Generell gilt dabei: Es gibt bei dieser Befragung keine richtigen oder falschen Antworten. **Vielmehr interessieren wir uns für Ihre ganz persönliche Meinung.**

Bitte beachten Sie:

- Machen Sie hinter jeder Aussage jeweils **nur ein Kreuz** in einem der dafür vorgesehenen Kreise (und niemals zwischen oder neben Kreisen). Bitte achten Sie zudem darauf, keine Aussage auszulassen. Wenn eine Aussage für Sie nicht sinnvoll beantwortbar ist, können Sie uns dies durch ein Kreuz in dem hintersten Kreis kundtun.
- Aus Gründen der besseren Lesbarkeit wird nachfolgend jeweils nur die männliche Form (Dozenten/ Teilnehmer) verwendet, die weibliche Form ist dabei jeweils mit eingeschlossen.
- Ihre Teilnahme an dieser Befragung erfolgt **freiwillig**. Bitte geben Sie am Ende des Fragebogens an, ob wir Ihre Angaben für die Beurteilung der Lehre am IdF verwenden dürfen.
- Die Befragung erfolgt selbstverständlich **anonym**. Die Ergebnisse werden nur in gesammelter Form, d.h. z.B. in Form von Mittelwerten aller befragten Lehrgänge, an das IdF zurückgemeldet. Ein Rückschluss auf Ihre Person ist damit ausgeschlossen. Um Ihre Angaben jedoch gruppenstatistisch auswerten zu können, benötigen wir von Ihnen einige demografische Angaben:

Alter: _____ Jahre

Geschlecht: ☐ weiblich ☐ männlich ☐ keine Angabe

	Stimme gar nicht zu	Stimme nicht zu	Stimme eher nicht zu	Neutral	Stimme eher zu	Stimme zu	Stimme vollkommen zu	Nicht sinnvoll beantwortbar
18. Ich habe bei den Gruppenarbeiten viel von anderen Teilnehmern gelernt.	1	2	3	4	5	6	7	0
19. Alle Teilnehmer haben sich, entsprechend ihrer Kompetenzen, an den Gruppenarbeiten beteiligt.	1	2	3	4	5	6	7	0
20. Die Gruppenarbeiten haben den Austausch zwischen den Teilnehmern gefördert.	1	2	3	4	5	6	7	0
21. Ich habe bei der Planübung viel gelernt.	1	2	3	4	5	6	7	0
22. Die Dozenten gaben mir nützliches Feedback zu meinen Leistungen bei der Planübung.	1	2	3	4	5	6	7	0
23. In der Planübung konnte ich das neuerworbene Wissen anwenden.	1	2	3	4	5	6	7	0
24. Der Schwierigkeitsgrad der Planübung war angemessen.	1	2	3	4	5	6	7	0
25. Ich habe bei den Einsatzübungen viel gelernt.	1	2	3	4	5	6	7	0
26. Die Dozenten gaben mir nützliches Feedback zu meinen Leistungen in der Einsatzübung.	1	2	3	4	5	6	7	0
27. In den Einsatzübungen konnte ich das neuerworbene Wissen anwenden.	1	2	3	4	5	6	7	0
28. Der Schwierigkeitsgrad der Einsatzübung war angemessen.	1	2	3	4	5	6	7	0
29. Ich hatte ausreichend Zeit die EVA-Aufgaben zu bearbeiten.	1	2	3	4	5	6	7	0
30. Der Umfang der EVA-Aufgaben war angemessen.	1	2	3	4	5	6	7	0
31. Die EVA-Aufgaben haben mein Verständnis der Inhalte weiter vertieft.	1	2	3	4	5	6	7	0
32. Die Aufgabenstellungen der EVA-Aufgaben waren verständlich.	1	2	3	4	5	6	7	0
33. Der Schwierigkeitsgrad der EVA-Aufgaben war angemessen.	1	2	3	4	5	6	7	0

	Stimme gar nicht zu	Stimme nicht zu	Stimme eher nicht zu	Neutral	Stimme eher zu	Stimme zu	Stimme vollkommen zu	Nicht sinnvoll beantwortbar
34. Ich fühlte mich auf diesen Lehrgang sehr gut vorbereitet.	1	2	3	4	5	6	7	0
35. Der mit dem Lehrgang verbundene Zeitaufwand war passend für mich.	1	2	3	4	5	6	7	0
36. Die Größe der Teilnehmergruppe war angemessen.	1	2	3	4	5	6	7	0
37. Ich konnte mich voll und ganz auf den Lehrgang konzentrieren.	1	2	3	4	5	6	7	0
38. Ich fühle mich auf die Prüfung sehr gut vorbereitet.	1	2	3	4	5	6	7	0
39. Ich fühle mich auf meinen nächsten Einsatz, den ich als Gruppen-/Zugführer leiten werde, sehr gut vorbereitet.	1	2	3	4	5	6	7	0
40. Durch die praktischen Übungen im Lehrgang habe ich die nötige Sicherheit gewonnen, einen Einsatz als Gruppen-/Zugführer zu leiten.	1	2	3	4	5	6	7	0
41. Ich kann das Gelernte auf meine Heimatstelle übertragen.	1	2	3	4	5	6	7	0
42. Durch meine Teilnahme am Lehrgang gelingt es mir jetzt besser, Gefahrenlagen frühzeitig zu erkennen.	1	2	3	4	5	6	7	0
43. Durch meine Teilnahme am Lehrgang fällt es mir jetzt leichter, Entscheidungen in kritischen Situationen zu treffen.	1	2	3	4	5	6	7	0
44. Durch meine Teilnahme am Lehrgang kenne ich meine persönlichen Grenzen besser.	1	2	3	4	5	6	7	0
45. Durch meine Teilnahme am Lehrgang traue ich mir besser zu, in stressigen Situationen ruhig zu bleiben.	1	2	3	4	5	6	7	0
46. Durch meine Teilnahme am Lehrgang kann ich korrektere und verständlichere Anweisungen geben.	1	2	3	4	5	6	7	0
47. Ich habe in der Lehrveranstaltung sehr viel gelernt.	1	2	3	4	5	6	7	0

48. Ich würde dieser Lehrveranstaltung insgesamt folgende Schulnote geben:	<div>sehr gut gut befriedigend ausreichend mangelhaft ungenügend</div> <div>1 2 3 4 5 6</div>
49. Ich habe das Angebot vom IdF zur Vorbereitung auf diesen Lehrgang genutzt,	<input type="radio"/> Ja <input type="radio"/> Nein
50. Was war gut an dieser Veranstaltung? (-Bitte stichpunktartig antworten-)	
<hr/> <hr/> <hr/>	
51. Was hat Ihnen an dieser Veranstaltung nicht gefallen? Und welche Verbesserungen schlagen Sie vor?	
<hr/> <hr/> <hr/>	
Platz für weitere Anmerkungen:	
<hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>	
Herzlichen Dank für Ihre Teilnahme!	
Seite 5	

Anmerkung: Die Originalversionen finden sich im digitalen Anhang.

Plagiatserklärung

Hiermit versichere ich, dass die vorliegende Arbeit mit dem Titel „*Evaluation der Führungsausbildung am Institut der Feuerwehr NRW – Entwicklung eines Fragebogens zur Bewertung der Zug- und Gruppenführerlehrgänge*“ selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

(Datum, Unterschrift)

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in einer Datenbank einverstanden.

(Datum, Unterschrift)