

Feedback-Instrument zur Rettungs- kräfte-Entwicklung (FIRE)

*Niklas Schulte, Stephanie Babel,
Marius Messinger, Meinold T. Thielsch*

Zitierung

Schulte, N., Babel, S., Messinger, M., & Thielsch, M. T. (2019).
Feedback-Instrument zur Rettungs-kräfte-Entwicklung (FIRE).
*Zusammenstellung sozialwissenschaftlicher Items und Skalen
(ZIS)*. <https://doi.org/10.6102/zis268>

Nutzungsbedingungen

Dieses Erhebungsinstrument darf kostenfrei für nichtkommerzielle Forschungszwecke verwendet werden. Bei einem Einsatz für andere Zwecke oder in einer anderen als der hier dokumentierten Form ist das Einverständnis der Autoren bzw. Autorinnen einzuholen. In allen resultierenden Arbeiten und Publikationen ist diese Dokumentation als Quelle anzugeben.

1 Überblick

Zusammenfassung

Das Feedback-Instrument zur Rettungskräfte-Entwicklung (FIRE) erfasst die Qualität der Ausbildung von Rettungskräften. Das Instrument basiert auf qualitativen Analysen zu Merkmalen einer guten Ausbildung im Bereich Rettungswesen. Die auf Basis dieser Merkmale zusammengestellten Items wurden in zwei Studien am Institut der Feuerwehr Nordrhein-Westfalen (IdF NRW) hinsichtlich Relevanz und Verständlichkeit geprüft. Mithilfe einer explorativen Faktorenanalyse konnte das Instrument auf 21 Items reduziert und die Faktoren *Dozentenverhalten*, *Anforderungsniveau*, *Struktur* und *Gruppe* auf Ebene der Lernprozesse sowie die Faktoren *Kompetenzerwerb* und *Transfer* auf Ebene der Lernoutcomes identifiziert werden. Das Instrument wurde mithilfe von Lehrgangsteilnehmenden am IdF NRW validiert. Eine konfirmatorische Faktorenanalyse konnte die Faktorstruktur bestätigen; die Skalen zeigen eine akzeptable bis gute interne Konsistenz und es liegen eindeutige Hinweise auf Konstrukt- und Kriteriumsvalidität vor.

Abstract

The Feedback Instrument for Rescue Force Development (FIRE) measures the quality of the training of rescue forces. The instrument is based on qualitative analyses of characteristics of good training in the field of rescue. The items compiled on the basis of these characteristics were tested for relevance and comprehensibility in two studies at the Institute of the North Rhine-Westphalia Fire Brigade (IdF NRW). With the help of an explorative factor analysis the instrument could be reduced to 21 items and the factors lecturer behaviour, requirement level, structure and group at the level of learning processes as well as the factors acquisition of competence and transfer at the level of learning outcomes could be identified. The instrument was validated with the help of course participants at the IdF NRW. A confirmatory factor analysis was able to confirm the factor structure; the scales show an acceptable to good internal consistency and there are clear indications of construct and criterion validity.

Zentrale Kennwerte

Titel: Feedback-Instrument zur Rettungskräfte-Entwicklung (FIRE)

Autor/in: Schulte, Babel, Messinger, Thielsch

In ZIS seit: 2019

Anzahl der Items: 21

Reliabilität: Cronbachs Alpha $\geq .73$, McDonalds Omega $h \geq .76$

Validität: Hinweise auf Konstrukt- und Kriteriumsvalidität

Konstrukt: Feedback zur Rettungskräfteentwicklung (Dozentenverhalten, Anforderungsniveau, Struktur, Gruppe, Kompetenzerwerb und Transfer)

Schlagwörter: Feuerwehr, Evaluation | fire brigade, evaluation

Sprache Dokumentation: deutsch

Sprache Items: deutsch

Item(s) in Bevölkerungsumfrage eingesetzt: nein

Entwicklungsstand: validiert

2 Instrument

Instruktion

Liebe/r Lehrgangsteilnehmer/in,

wir schätzen es sehr, dass Sie an unserer Befragung teilnehmen. Indem Sie den Lehrgang hinsichtlich verschiedener Aspekte bewerten, helfen Sie uns, die Qualität der Lehre zu beurteilen und gegebenenfalls zu verbessern. Bitte kreuzen Sie hierzu bei jeder der folgenden Aussagen das Kästchen an, das den Grad Ihrer Zustimmung am besten beschreibt. 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 Kästchen (und niemals zwischen oder neben den Kästchen).
- Bitte lassen Sie keine Aussagen aus. Wenn eine Aussage für Sie nicht sinnvoll beantwortbar ist, können Sie uns dies durch ein Kreuz in dem entsprechenden Kästchen mitteilen.
- Aus Gründen der besseren Lesbarkeit wird an einigen Stellen nur die männliche Form (Dozent/Teilnehmer) verwendet. Die weibliche Form ist dabei jeweils miteingeschlossen.
- Ihre Teilnahme an der Befragung ist freiwillig und anonym.

Bitte teilen Sie uns bei den folgenden Aussagen mit, wie sehr Sie diesen zustimmen:

Items

Tabelle 1

FIRE-Items: Kernfragebogen

Nr.	Item	Subskala
1	Die Dozenten fassten schwierige Sachverhalte prägnant zusammen.	Dozentenverhalten
2	Ich finde, die Dozenten gaben nützliches Feedback.	Dozentenverhalten
3	Die Dozenten motivierten mich, mich einzubringen.	Dozentenverhalten
4	Ich finde, die Dozenten waren am Lernerfolg der Teilnehmer interessiert.	Dozentenverhalten
5	Der Umfang der zu lernenden Inhalte hat mich überfordert.	Anforderungsniveau
6	Das Tempo der Stoffvermittlung war zu hoch für mich.	Anforderungsniveau
7	Die Inhalte der Lehrveranstaltung waren zu schwierig für mich.	Anforderungsniveau
8	Ich finde, die Lehrveranstaltung war klar strukturiert.	Struktur
9	Ich konnte im Verlauf der Lehrveranstaltung die Gliederung immer nachvollziehen.	Struktur
10	Ich finde, die Lehrveranstaltung gab einen guten Überblick über das Themengebiet.	Struktur
11	Die anderen Teilnehmer brachten sich aktiv ein.	Gruppe
12	Die Teilnehmer unterstützten sich gegenseitig.	Gruppe
13	Ich finde, es herrschte ein guter Zusammenhalt im Lehrgang.	Gruppe
14	Durch meine Teilnahme am Lehrgang gelingt es mir jetzt besser, Gefahrenlagen frühzeitig zu erkennen.	Kompetenzerwerb
15	Durch meine Teilnahme am Lehrgang fällt es mir jetzt leichter, Entscheidungen in kritischen Situationen zu treffen.	Kompetenzerwerb
16	Durch meine Teilnahme am Lehrgang kenne ich meine persönlichen Grenzen besser.	Kompetenzerwerb
17	Durch meine Teilnahme am Lehrgang traue ich mir besser zu, in stressigen Situationen ruhig zu bleiben.	Kompetenzerwerb
18	Durch meine Teilnahme am Lehrgang kann ich korrektere und verständlichere Anweisungen geben.	Kompetenzerwerb
19	Ich fühle mich auf meinen nächsten Einsatz, den ich leiten werde, sehr gut vorbereitet.	Transfer
20	Durch die praktischen Übungen im Lehrgang habe ich die nötige Sicherheit gewonnen, einen Einsatz zu leiten.	Transfer
21	Ich kann das Gelernte auf meine Heimatstelle übertragen.	Transfer
22	Ich habe im Lehrgang sehr viel gelernt.	Globalurteil
23	Ich würde dem Lehrgang folgende Schulnote geben:	Globalurteil
24	Ich würde diesen Lehrgang weiterempfehlen.	Globalurteil
25	Anmerkungen für Dozenten (Vorschläge/Lob/Kritik)	Feedback

Anmerkung. Alle Items sind positiv gepolt. Bei den Items der Skala *Anforderungsniveau* gehen höhere Werte mit höherer Überforderung und damit einer schlechteren Evaluation des Lehrgangs einher. Für die Interpretation durch den Anwender kann es hilfreich sein, die numerischen Ergebnisse dieser Skala zu invertieren. Items 22 bis 25 sind optional und bilden keine eigenständigen psychometrischen Skalen.

Die FIRE-Items teilen sich in sechs Subskalen auf. Der hier eingeführte FIRE-Kernfragebogen (Tabelle 1) mit den Subskalen *Dozentenverhalten*, *Anforderungsniveau*, *Struktur*, *Gruppe*, *Kompetenzerwerb* und *Transfer* erfasst Qualitätsmerkmale, die unabhängig von konkreten Lehrgangsformaten relevant sind. Items 22 bis 25 (*Globalurteil* und *Feedback*) gehören nicht zum FIRE-Kernfragebogen. Item 22 und 23 orientieren sich an vergleichbaren Items aus etablierten Evaluationsinstrumenten (bspw. FEVOR, Staufenbiel, 2000; KIEL, Gediga et al., 2000; [MFE-Sr](#), Thielsch & Hirschfeld, 2012a). Die Items zum Globalurteil sollten nicht zu einem Gesamtmaß zusammengefasst werden. Ihr Einsatz ist dennoch empfehlenswert, da sie die Akzeptanz und die Kommunizierbarkeit der Evaluationsergebnisse innerhalb und außerhalb der Ausbildungsorganisation erleichtern.

Die Items sind auch in englischer Sprache verfügbar (vgl. Schulte & Thielsch, 2019).

Antwortvorgaben

Item 1-22:

Siebenstufiges Antwortformat mit den Optionen 1 = *stimme gar nicht zu*, 2 = *stimme nicht zu*, 3 = *stimme eher nicht zu*, 4 = *neutral*, 5 = *stimme eher zu*, 6 = *stimme zu*, 7 = *stimme vollkommen zu*. Ein zusätzliches Feld bietet die Möglichkeit anzugeben, dass das jeweilige Item *nicht sinnvoll beantwortbar* ist.

Item 23:

Sechsstufiges Antwortformat mit den Optionen 1 = *sehr gut*, 2 = *gut*, 3 = *befriedigend*, 4 = *ausreichend*, 5 = *mangelhaft* und 6 = *ungenügend*.

Item 24:

Dichotomes Antwortformat mit den Optionen *ja/nein*.

Item 25:

Offenes Antwortformat.

Auswertungshinweise

Die Auswertung des FIRE erfolgt in mehreren Schritten:

Schritt 1: Punkte für einzelne Antworten vergeben

Für jede Antwort einer Person auf ein Item werden Punkte vergeben. Es wird 1 Punkt vergeben für *stimme gar nicht zu*, 2 Punkte für *stimme nicht zu*, 3 Punkte für *stimme eher nicht zu*, 4 Punkte für *neutral*, 5 Punkte für *stimme eher zu*, 6 Punkte für *stimme zu* und 7 Punkte für *stimme vollkommen zu*. Bei der Antwort *nicht sinnvoll beantwortbar* werden keine Punkte vergeben und das betreffende Item wird in der weiteren Auswertung für diese Person ignoriert. Für alle Skalen sind hohe Werte erstrebenswert, ausgenommen die Items der Skala *Anforderungsniveau*. Bei letzterer sind eher niedrige bis mittlere Werte erstrebenswert, da hohe Werte auf dieser Skala für eine Überforderung der Lehrgangsteilnehmenden sprechen würden. Zur Vereinheitlichung der Rückmeldung können die Angaben auf der Skala *Anforderungsniveau* negativ kodiert werden (1 Punkt für *stimme vollkommen zu*, [...], 7

Punkte für *stimme gar nicht zu*), sodass auf allen Skalen eine positive Evaluation des Lehrgangs mit hohen Punktwerten einhergeht.

Schritt 2: Itemmittelwerte berechnen

In diesem Schritt werden für alle Items die Mittelwerte berechnet, indem die Punkte jedes Items addiert werden (Punktsomme eines Items über alle Antwortenden hinweg) und durch die Anzahl der Lehrgangsteilnehmenden, die auf das jeweilige Item geantwortet haben, dividiert wird. An dieser Stelle ist darauf hinzuweisen, dass die Anzahl der Antworten nicht notwendigerweise der Anzahl der Lehrgangsteilnehmenden entsprechen muss. Dies ist dann der Fall, wenn eine oder mehrere Personen Items als *nicht sinnvoll beantwortbar* klassifiziert haben. Diese Angaben sind als fehlende Werte zu behandeln. Für eine Online-Befragung empfehlen wir die Items 1-21 als Pflichtfragen vorzugeben. Für eine Papierbefragung empfehlen wir, Personen, die drei Items oder mehr komplett auslassen, aus der Auswertung herauszunehmen. Eine hohe Anzahl von fehlenden Werten bei vielen Personen kann in einer mangelnden Passung des Fragebogens in dem jeweiligen Evaluationskontext begründet sein.

Schritt 3: Skalenmittelwerte berechnen

Zur Berechnung der Mittelwerte der einzelnen Skalen des FIRE werden die Mittelwerte der Items auf der betreffenden Skala addiert und durch die Anzahl der Items auf dieser Skala dividiert (siehe auch Tabelle 2).

Tabelle 2

Berechnung der Mittelwerte der 6 FIRE-Skalen des Kernfragebogens

Skala	Itemnummern	Berechnung des Skalenmittelwerts
Dozentenverhalten	1, 2, 3, 4	$(\bar{y}_{item\ 1} + \bar{y}_{item\ 2} + \bar{y}_{item\ 3} + \bar{y}_{item\ 4})/4$
Anforderungsniveau	5, 6, 7	$(\bar{y}_{item\ 5} + \bar{y}_{item\ 6} + \bar{y}_{item\ 7})/3$
Struktur	8, 9, 10	$(\bar{y}_{item\ 8} + \bar{y}_{item\ 9} + \bar{y}_{item\ 10})/3$
Gruppe	11, 12, 13	$(\bar{y}_{item\ 11} + \bar{y}_{item\ 12} + \bar{y}_{item\ 13})/3$
Kompetenzerwerb	14, 15, 16, 17, 18	$(\bar{y}_{item\ 14} + \bar{y}_{item\ 15} + \bar{y}_{item\ 16} + \bar{y}_{item\ 17} + \bar{y}_{item\ 18})/5$
Transfer	19, 20, 21	$(\bar{y}_{item\ 19} + \bar{y}_{item\ 20} + \bar{y}_{item\ 21})/3$

Anmerkung. \bar{y} bezeichnet die Itemmittelwerte aus Schritt 2

Hinsichtlich der Interpretation der Globalurteile (Items Nr. 22-24) sowie des offenen Feedback-Items (Nr. 25) schlagen wir vor, diese lediglich als allgemeine Rückmeldung im Sinne einer Checkliste zu nutzen, beziehungsweise die offenen Anmerkungen erst bei einer hohen Zahl von vergleichbaren Nennungen als handlungsleitend heranzuziehen. Hinsichtlich der Schulnote (Item 23) empfehlen wir statt des Mittelwertes den Median zu bestimmen oder eine Häufigkeitstabelle zu erstellen, da Schulnoten typischerweise Ordinalskalenniveau aufweisen, d.h. nicht gleichabständig sind. Während die Items 1 bis 21 vollständig vorgegeben werden sollten, sind die Fragen 22 bis 25 optional und können je nach Bedarf eingesetzt oder ausgelassen werden. Eine spezifische Validierung dieser zusätzlichen Items erfolgte nicht, derartige Items sind allerdings in etablierten Evaluationsinstrumenten typisch (vgl.

z.B. FEVOR, Staufenbiel, 2000; KIEL, Gediga et al., 2000 oder [MFE-Sr](#), Thielsch & Hirschfeld, 2012a).

Generell möchten wir empfehlen, eine Auswertung erst vorzunehmen, wenn mindestens acht ausgefüllte Fragebögen vorliegen (bzw. bei Veranstaltungen mit weniger als 15 TeilnehmerInnen mindestens 50% an der Befragung teilgenommen haben). Dabei ist darauf zu achten, dass die Auswertung anonym erfolgt. Dementsprechend sollten keine demographischen Variablen erhoben werden, anhand derer Rückschlüsse auf einzelne Personen gezogen werden können. Bei Papierbefragungen sollten die offenen Angaben in Item 25 zunächst durch eine neutrale Person digitalisiert werden, damit Befragte nicht an der Handschrift erkannt werden können. Ist dies nicht möglich, so muss in der Befragung auf dieses Problem hingewiesen werden.

Schritt 4: Rückmeldung und Nutzung der Ergebnisse

Grundsätzlich sollte der subjektive Charakter einer solchen Evaluation berücksichtigt werden. Ein Feedback aus der FIRE-Evaluation ist für die AusbilderInnen eine Chance, wichtige Informationen über die eigene Lehrtätigkeit aus Sicht der VeranstaltungsteilnehmerInnen zu erhalten. Es empfiehlt sich eine Evaluationsbesprechung mit den Befragten abzuhalten. In diesem Rahmen können Ergebnisse des FIRE besprochen und mögliche Missverständnisse geklärt werden. Die Fragebögen eignen sich dabei auch zum Vergleich mit früheren gleichartigen Veranstaltungen oder zwischen verschiedenen Lehrkonzepten. Wenn keine eigenen Vergleichsdaten vorliegen, können die deskriptiven Statistiken in Tabelle 7 (siehe unten) als Orientierung dienen. Von Seiten der verantwortlichen Organisation sollte die Evaluation technisch, aber auch inhaltlich unterstützt werden. Dies umfasst insbesondere Hilfestellungen, sollten Evaluationen wiederholt Verbesserungspotentiale aufzeigen sowie Belobigungen für qualitativ gute Unterrichtstätigkeit.

Anwendungsbereich

Der FIRE ist ein schriftlicher Fragebogen zur Bewertung und Evaluation der Ausbildung von Rettungskräften der Feuerwehr und ist aus einer Kooperation des Instituts der Feuerwehr Nordrhein-Westfalen (IdF NRW) und der Organisations- und Wirtschaftspsychologie der Westfälischen Wilhelms-Universität Münster entstanden. Er ist für papierbasierte und digitale Erhebungsformen gleichermaßen geeignet. Bei der Entwicklung des Fragebogens wurde darauf geachtet, dass die Items keine rein feuerwehrspezifischen Inhalte erfassen, sondern prinzipiell auch bei Ausbildungen in anderen Organisationen, wie Rettungsdiensten, Polizei und Militär, eingesetzt werden können. Theoriegeleitet ist davon auszugehen, dass die Qualitätsmerkmale guter Ausbildungen in diesen Organisationen vergleichbar sind. Dennoch ist es unabdingbar, die Inhaltsvalidität der Skalen im Hinblick auf die konkret zu evaluierenden Lehrgänge zu beurteilen. Die Einführung sollte durch eine Validierung der Messungen im neuen Einsatzkontext begleitet werden. Beim FIRE handelt es sich um ein Evaluationsinstrument, bei dem die Bewertung durch die Lehrgangsteilnehmenden am Ende der entsprechenden Ausbildungsmodule erfolgt. Wird der zu evaluierende Lehrgang mit einer Prüfung abgeschlossen, empfiehlt sich die Durchführung der Evaluation am Ende der Ausbildungsphase, aber vor der Prüfung. Der FIRE ist an Ausbildungen für werdende und aktive Führungskräfte im Rettungswesen orientiert. Eine Reihe von

Ergänzungsmodulen für den FIRE befindet sich derzeit in Vorbereitung. Schriftliche Prüfungen im Rettungswesen können mittels des [FIRE-P](#) (Thielsch, Busjan, & Frerichs, 2018) bewertet werden. Für eine Evaluation der Basis-Ausbildungen im Rettungswesen für Personen ohne Führungsverantwortung steht der [FIRE-B](#) (Thielsch, Kläpker, & Streppel, 2019) zur Verfügung.

3 Theoretischer Hintergrund

Die Tätigkeit von Rettungskräften, insbesondere die Tätigkeit von Führungskräften in diesem Bereich, erfordert ein breitgefächertes Spektrum von verschiedensten Fertigkeiten, um herausfordernde Einsatzsituationen sicher zu meistern (Schulte & Thielsch, 2019). Ob Verkehrsunfälle, Brände oder technische Hilfeleistungen – die Arbeit der Feuerwehr ist durch hohe Anforderungen und teilweise extreme Bedingungen geprägt. Oftmals treffen an der Notfallstelle die miteinander arbeitenden Einsatzkräfte der Feuerwehr, Polizei, Luftrettung und weitere Helfer zum ersten Mal zusammen. Die Teams der Feuerwehr müssen dabei sofort funktionieren und eine effektive Kooperation und Kommunikation sicherstellen. Damit kommt insbesondere den Führungskräften eine entscheidende Rolle zu. Sie müssen die Einsatzlagen vor Ort einschätzen und oft unter hohem Zeitdruck Entscheidungen treffen. Führungskräfte der Feuerwehr müssen dabei zügig reagieren, die richtigen Befehle erteilen und dabei die Sicherheit der Bevölkerung und der eigenen Einsatzkräfte im Auge behalten.

Die Führungskräfteausbildung der Feuerwehr erfolgt in einem Zusammenspiel aus Theorieunterricht und praktischen Übungen. Das vorliegende Evaluationsinstrument wurde in Zusammenarbeit mit dem Institut der Feuerwehr Nordrhein-Westfalen für die dortigen Lehrgänge entwickelt. Sie sind folgendermaßen aufgebaut: In verschiedenen Unterrichtseinheiten werden den Lehrgangsteilnehmenden am IdF NRW zunächst explizite theoretische Inhalte meist im Frontalunterricht anhand von Präsentationen, Videosequenzen oder anderen Lehrmaterialien vermittelt. Weitere Methoden der theoretischen Wissensvermittlung sind unter anderem Gruppenarbeiten, das eigenverantwortliche Erarbeiten und Vertiefen von Inhalten sowie die Präsentation von Ergebnissen und kritische Diskussionen. Im praktischen Teil des Unterrichts üben die Lehrgangsteilnehmenden das erworbene theoretische Wissen in realitätsnahen Einsatzszenarien auf dem Außengelände ein und wenden es wiederholt praktisch an (IdF NRW Ausbildungsplan – Musterplan BIII, 2012; IdF NRW Ausbildungsplan – Musterplan FIII, 2013; IdF NRW Ausbildungsplan – Musterplan FIV, 2016). Im Rahmen dieser Szenarien erhält jeder der Lehrgangsteilnehmenden die Möglichkeit, einen Einsatz mit einer spezifischen Gefahrensituation als Gruppen- oder Zugführer zu leiten und erhält im Anschluss daran ausführliches Feedback von den Dozentinnen und Dozenten.

Unter welchen Kriterien lässt sich eine so geartete Ausbildung bewerten und was macht einen guten Lehrgang überhaupt aus? Das Evaluationsmodell von Kirkpatrick (1979) mit den vier Ebenen *Reaktion*, *Lernen*, *Verhalten* und *Ergebnis*, auf denen die Evaluation erfolgen soll, bietet einen konzeptuellen Rahmen, um die Qualität von Ausbildungsprogrammen oder Trainings zu beurteilen. Auf der Ebene der Reaktion wird untersucht, wie sehr das Training beziehungsweise Ausbildungsprogramm den Teilnehmenden gefällt. Eine positive Reaktion der Teilnehmenden ist ein Prädiktor für Lernprozesse (Blanchard & Thacker, 2010; Kirkpatrick, 1998), da sie Aufmerksamkeit und Motivation – wichtige kognitive Prozesse für effektives soziales Lernen – fördert (Bandura, 1977). Auf der Ebene des Ler-

nens werden erworbenes Wissen und Handlungskompetenzen der Teilnehmenden erfasst. Eine Evaluation auf Verhaltensebene erfasst das Ausmaß, in dem erlernte Trainings- oder Ausbildungsinhalte im Berufsalltag angewandt werden. Die Evaluation auf Ergebnisebene erlaubt eine Untersuchung des Trainings im Hinblick auf Unternehmensziele oder monetäre Größen. Kirkpatrick (1998) schlägt vor, ein Ausbildungsprogramm erst auf den Ebenen Verhalten und Ergebnis zu evaluieren, wenn sich die Evaluation auf den Ebenen Reaktion und Lernen als erfolgreich erwiesen hat. Ausbildungsprogramme lassen sich zudem im Hinblick auf die *Prozesse* während des Ausbildungsverlaufs (z.B. Ausbilderverhalten, Tagesablauf etc.) und die erzielten *Outcomes* (z.B. Lernerfolg) evaluieren (Blanchard & Thacker, 2010).

Der theoretische Teil der Lehre am IdF NRW und dessen methodische Gestaltung ähnelt in vielen Punkten durch seinen Vorlesungs- und vor allem Seminarcharakter der Lehre im universitären Kontext. Diese Ähnlichkeit ist für die Evaluation von Rettungskräfteausbildungen insofern relevant, als dass es eine Vielzahl von Forschungsarbeiten zur Evaluation universitärer Lehre gibt, welche eine Identifikation der kritischen Merkmale guter Lehre erlauben (bspw. Braun & Gusy, 2006; Souvignier & Gold, 2002). Im *multidimensionalen Bedingungsmodell des Lehrerfolgs* (Rindermann, 2001) ist der Erfolg der Lehre abhängig von einem Zusammenspiel aus Dozenten-/Studierendenvariablen und den Rahmenbedingungen der Lehre. Erfolgskritische Variablen auf der Seite des Dozenten sind hierbei vor allem Strukturiertheit und Klarheit, während auf der Seite der Studierenden hauptsächlich Vorwissen, Fähigkeiten, Interesse und Fleiß/Arbeitsbelastung den Lehrerfolg beziehungsweise Kompetenzerwerb bedingen.

Da wir trotz intensiver Literaturrecherche kein Evaluationsinstrument finden konnten, das sich auf den Ausbildungskontext der Feuerwehr anwenden ließ, entwickelten wir das hier vorgestellte Feedback-Instrument zur Rettungskräfte-Entwicklung. Der FIRE-Fragebogen vereint das bekannte Evaluationsmodell nach Kirkpatrick (1979) mit den Bedingungen guter Lehre aus dem universitären Kontext (Rindermann, 2001), in dem er auf Reaktions- und Lernebene Dozenten- und Teilnehmendenmerkmale sowie Rahmenbedingungen der Lehrgänge im IdF NRW erfasst, wodurch Prozess- und Outcomevariablen identifiziert werden, die Aussagen über die Lehrqualität und den Lehrerfolg erlauben.

4 Skalenentwicklung

Das IdF NRW als zentrale Ausbildungs- und Weiterbildungsstätte für Feuerwehrangehörige und Mitglieder von Krisenstäben des Landes NRW gab im Rahmen einer Neukonzeption der Lehre den Impuls, den FIRE-Fragebogen als Evaluationsinstrument der Gruppen- und Zugführerausbildungslehrgänge zu entwickeln. Der FIRE-Fragebogen wurde in einem multimodalen Studiendesign anhand von Stichproben entwickelt, die vorwiegend aus Lehrgangsteilnehmenden, aber auch aus Dozenten der Gruppen- oder Zugführerausbildungslehrgänge des IdF NRW bestanden. Am IdF NRW werden jährlich etwa 1.500 haupt- und ehrenamtliche Führungskräfte auf Gruppen- und Zugführerebene ausgebildet (Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalen, 2014; Sievers, 2015¹). Ein Gruppenführer ist befähigt, im Einsatz eine maximal acht Personen starke Mannschaft zu

¹ Diese Arbeit kann bei Bedarf durch die Autoren zur Verfügung gestellt werden.

leiten. Die Gruppe ist die taktische Grundeinheit bei Einsätzen der Feuerwehr. Bestehend aus Einsatzkräften und Einsatzmitteln übernimmt sie meistens die ersten Maßnahmen an der Einsatzstelle. Der Gruppenführer leitet die Einheit bis zum Einsatzende oder bis zur Übernahme der Einsatzleitung durch eine höhergestellte Führungskraft. Ein Zugführer ist dazu berechtigt, die nächstgrößte taktische Einheit der Feuerwehr, einen Zug, zu kommandieren. Ein Zug besteht aus bis zu 22 Feuerwehrleuten und dazugehörigen Einsatzmitteln (FwDV 3, 2008). In mehrwöchigen Lehrgängen werden die angehenden Führungskräfte befähigt, Einsätze erfolgreich zu leiten (AG-B III, 2007; AG-F III, 2005; AG-F IV, 2007).

Itemkonstruktion und Itemselektion

Studie I

Zunächst wurden halbstrukturierte Interviews mit $n = 5$ Dozenten und $n = 3$ Lehrgangsteilnehmenden am IdF NRW zu den Qualitätskriterien guter Lehre geführt (Babiel, 2016²; Schulte & Thielsch, 2019). Die Erkenntnisse aus diesen Interviews wurden mit den Angaben aus einem standardisierten Fragebogen mit $N = 56$ Lehrgangsteilnehmenden ergänzt. Durch den Einsatz des standardisierten Fragebogens sollten die verschiedenen Erfahrungen und Ansichten einer größeren Anzahl von Teilnehmern auf eine zeitlich flexiblere und ökonomische Weise erhoben werden (Howard, 2018; Schmidt-Atzert & Amelang, 2012). Im Folgenden wird auf die getrennte inhaltliche Darstellung der Ergebnisse der Interviews und Fragebogenmethode verzichtet, da für beide Methoden derselbe Fragebogen verwendet wurde.

Die Antworten wurden mittels qualitativer Inhaltsanalyse (Mayring, 2000) ausgewertet. Ausgehend von der von Mayring (2000) vorgeschlagenen induktiven Kategorienbildung konnten fünf Kategorien erfolgskritischer Lehre ermittelt werden. Um adäquate Interrater-Reliabilitäten (d. h. Beurteilerübereinstimmungen) für die inhaltsanalytischen Auswertungen sicherzustellen, wurden die Aussagen durch eine unabhängige zweite Person kategorisiert. Die zufallsbereinigten Übereinstimmungen der von den beiden Beurteilern unabhängig voneinander vorgenommenen Zuordnungen lagen hierbei im Bereich von $.85 \leq \kappa \leq .96$. Gemäß der Klassifikation von Greve und Wentura (1997, S. 111) spricht dies für eine gute bis ausgezeichnete Interrater-Reliabilität (da $\kappa > .75$). Die Kategorie *Dozent* umfasst verschiedene Verhaltensweisen und Merkmale der Dozenten, die aus Sicht der Befragten den Erfolg der Lehre beeinflussen. Dazu zählen vor allem didaktische Maßnahmen, wie beispielsweise eine gute Strukturierung der Lehrinhalte, aber auch sozial-interaktive Verhaltensweisen, wie die Gestaltung einer angenehmen Lernatmosphäre. Die *Rahmenbedingungen* als zweite Kategorie befassen sich mit der Ausstattung der Räumlichkeiten, der Organisation der Lehre und der Qualität und Verfügbarkeit der Lehrmaterialien. Die dritte und die vierte Kategorie – *Gruppe* und *Teilnehmer* – umfassen jeweils erfolgskritische Determinanten der Lehrgangsteilnehmenden, wobei unter *Gruppe* Aspekte des gemeinschaftlichen Miteinanders der Teilnehmenden zusammengefasst wurden und unter *Teilnehmer* Merkmale und Verhaltensweisen des einzelnen Lehrgangsteilnehmers, wie beispielsweise die Bereitschaft zur aktiven Mitarbeit oder die Motivation zum selbstständigen Nacharbeiten von Inhalten. In der fünften Kategorie – *Lehrerfolg* – finden sich vor allem Aspekte, die die Outcomes eines guten Lehr-

² Diese Arbeit kann bei Bedarf durch die Autoren zur Verfügung gestellt werden.

gangs beschreiben. Hierunter fallen beispielsweise der individuelle Wissenszuwachs oder der Kompetenzerwerb.

Aufbauend auf der beschriebenen Kategorisierung sollte eine möglichst große Itemmenge generiert werden, um alle bisher erfassten und angenommenen Determinanten guter Lehre umfassend zu berücksichtigen. Dieses Vorgehen führte zu einer Liste von 116 Items. Für jede Kategorie wurde eine unterschiedliche Itemanzahl generiert, da die Kategorien unterschiedlich umfangreich waren und für eine vollständige Beschreibung jeweils mehr oder weniger Aspekte berücksichtigt werden mussten. Bei der Itemkonstruktion konnten zum Teil auf Items aus bestehenden Lehrevaluationsinventaren mit guten Testgütekriterien zurückgegriffen werden (siehe auch Tabelle 3). Um eine optimale Passung an den Feuerwehrkontext sicherzustellen, wurden aber auch zahlreiche Items neu generiert.

Studie II

Eine wesentliche Voraussetzung für das Gelingen von Evaluationsprogrammen ist das frühzeitige Einbinden der Betroffenen (Wolf, Spiel & Pellert, 2001). Daher wurden die Items vor Erstellung der Fragebogenerstversion sowohl den Verantwortlichen als auch den Betroffenen (d. h. Dozenten sowie Lehrgangsteilnehmenden) am IdF NRW zur kritischen Begutachtung vorgelegt.

Die 116 Items wurden in Studie II $n = 7$ Dozenten und $n = 26$ Lehrgangsteilnehmenden präsentiert und von diesen hinsichtlich Verständlichkeit und Relevanz bewertet (vgl. Schulte & Thielsch, 2019). Außerdem sollten die Befragten die aus ihrer Sicht noch fehlenden Aspekte ergänzen. Die Items konnten anhand dieser Bewertung in zwei Gruppen eingeteilt werden. 65 Items zielten auf generelle Aspekte in der Ausbildung von Feuerwehrmännern und -frauen ab, die auf jede Ausbildungssituation übertragbar sind (44 Items zum Lernprozess, 21 Items zu Lernoutcomes). Weitere 51 Items zielten auf spezifische Lehrmethoden am IdF NRW ab und wurden deswegen aus der weiteren Analyse im Rahmen der Konstruktion des FIRE-Kernmoduls ausgeschlossen, gingen aber in die Konstruktion von hier nicht weiter besprochenen Erweiterungsmodulen für das FIRE ein.

Itemanalysen

Studie III

Basierend auf den Bewertungen aus Studie II konnte ein erster Fragebogen erstellt werden, dessen Items in je eine explorative Faktorenanalyse (EFA) für die Items zu Lernprozessen und zu Lernoutcomes eingingen. Von der Aufnahme in die EFAs ausgeschlossen wurden Items, die von weniger als 95% der Teilnehmenden beantwortet wurden (traf auf keines der Items zu), eine absolute Schiefe > 2 oder einen absoluten Exzess > 7 zeigten (keines der Items), Iteminterkorrelationen $> .75$ aufwiesen (5 Items) oder einen Mittelwert ≥ 6 und eine Standardabweichung ≤ 1 aufwiesen (3 Items). Ein weiteres Item wurde aufgrund mangelnder inhaltlicher Relevanz bei gleichzeitiger ungünstiger Verteilung der Itemantworten ausgeschlossen. Das Itemset für die EFAs bestand abschließend aus 39 Prozessitems und aus 17 Outcomeitems.

Es wurde je eine EFA für Prozess- und eine für Lernoutcomeitems berechnet. Das Ergebnis des Scree-tests für die Prozessitems legte die Extraktion von zwei oder vier Faktoren nahe, die Parallelanalyse die Extraktion von vier Faktoren, weshalb vier distinkte Faktoren auf der Prozessebene

extrahiert wurden. Aufgrund von Faktorladungen und Reliabilitätsindizes (Cronbachs α ohne das entsprechende Item) wurden im weiteren Verlauf insgesamt 26 Items ausgeschlossen.

Screeplot und Parallelanalyse auf Outcomeebene legten eine Extraktion von zwei Faktoren nahe, weswegen zwei distinkte Faktoren extrahiert wurden. Auf Outcomeebene wurden fünf Items aufgrund von Doppelladungen und vier Items aufgrund zu geringer Ladungen und inhaltlichen Überlegungen ausgeschlossen. Die Trennschärfe aller Items war zufriedenstellend ($\geq .58$) und die Reliabilitätsindizes differenzierten nicht zwischen den Items.

Das Resultat von Studie III war ein kurzer Fragebogen mit 21 Items auf 6 Skalen, die mit den in Studie I identifizierten Qualitätskriterien einer guten Lehre korrespondierten. Tabelle 3 zeigt eine Übersicht der 21 finalen Items inklusive der Quellen, nach denen sie adaptiert wurden.

Tabelle 3

Übersicht der finalen Itematterie

Nr.	Item	Quelle
1	Die Dozenten fassten schwierige Sachverhalte prägnant zusammen.	MFE-ZMo (Thielsch & Stegemöller, 2012)
2	Ich finde, die Dozenten gaben nützliches Feedback.	MFE-ZMo (Thielsch & Stegemöller, 2012)
3	Die Dozenten motivierten mich, mich einzubringen.	*
4	Ich finde, die Dozenten waren am Lernerfolg der Teilnehmer interessiert.	FB zur LVE (Wolf et al., 2001)
5	Der Umfang der zu lernenden Inhalte hat mich überfordert.	*
6	Das Tempo der Stoffvermittlung war zu hoch für mich.	MFE-Sr (Thielsch & Hirschfeld, 2012a)
7	Die Inhalte der Lehrveranstaltung waren zu schwierig für mich.	MFE-Sr (Thielsch & Hirschfeld, 2012a)
8	Ich finde, die Lehrveranstaltung war klar strukturiert.	Kurzskala zur Lehrevaluation (Zumbach, Spinath, Schahn, Friedrich, & Kögel, 2007)
9	Ich konnte im Verlauf der Lehrveranstaltung die Gliederung immer nachvollziehen.	MFE-Sr (Thielsch & Hirschfeld, 2012a)
10	Ich finde, die Lehrveranstaltung gab einen guten Überblick über das Themengebiet.	MFE-Sr (Thielsch & Hirschfeld, 2012a)
11	Die anderen Teilnehmer brachten sich aktiv ein.	MFE-Sr (Thielsch & Hirschfeld, 2012a)
12	Die Teilnehmer unterstützten sich gegenseitig.	*
13	Ich finde, es herrschte ein guter Zusammenhalt im Lehrgang.	*
14	Durch meine Teilnahme am Lehrgang gelingt es mir jetzt besser, Gefahrenlagen frühzeitig zu erkennen.	* (in Anlehnung an Mistele, 2007)
15	Durch meine Teilnahme am Lehrgang fällt es mir jetzt leichter, Entscheidungen in kritischen Situationen zu treffen.	* (in Anlehnung an Mistele, 2007)
16	Durch meine Teilnahme am Lehrgang kenne ich meine persönlichen Grenzen besser.	* (in Anlehnung an Mistele, 2007)
17	Durch meine Teilnahme am Lehrgang traue ich mir besser zu, in stressigen Situationen ruhig zu bleiben.	* (in Anlehnung an Mistele, 2007)
18	Durch meine Teilnahme am Lehrgang kann ich korrektere und verständlichere Anweisungen geben.	* (in Anlehnung an Mistele, 2007)
19	Ich fühle mich auf meinen nächsten Einsatz, den ich leiten werde, sehr gut vorbereitet.	*
20	Durch die praktischen Übungen im Lehrgang habe ich die nötige Sicherheit gewonnen, einen Einsatz zu leiten.	*
21	Ich kann das Gelernte auf meine Heimatstelle übertragen.	*

Anmerkung. Mit * markierte Items wurden für den FIRE-Fragebogen auf Basis der Ergebnisse aus Studie I neu generiert

Studie IV

In Studie IV wurde die in Studie III erhaltene interne Struktur des Fragebogens mithilfe von konfirmatorischen Faktorenanalysen (confirmatory factor analysis, CFAs) an einer Stichprobe aus Lehrgangs-

teilnehmenden am IdF NRW kreuzvalidiert (vgl. Schulte & Thielsch, 2019). Hierzu wurde das R-Paket lavaan (Rosseel, 2012) mit Maximum-Likelihood-Schätzer verwendet. Die CFAs konnten die in Studie III gefundene Faktorenstruktur bestätigen. Auf Ebene des Lernprozesses lassen sich die vier distinkten Faktoren *Dozentenverhalten*, *Anforderungsniveau*, *Struktur* und *Gruppe* unterscheiden, die Ergebnisse der CFA finden sich in Abbildung 1. Dieses Modell zeigte einen akzeptablen Fit (CFI = .96, RMSEA = 0.06 [0.05, 0.08], SRMR = 0.06). Lediglich der TLI erreichte mit 0.94 kein akzeptables Niveau. Der χ^2 -Test fiel signifikant aus mit $\chi^2(59) = 146.37$, $p < .001$, was typisch für große Stichproben ist (Schermelleh-Engel, Moosbrugger & Müller, 2003).

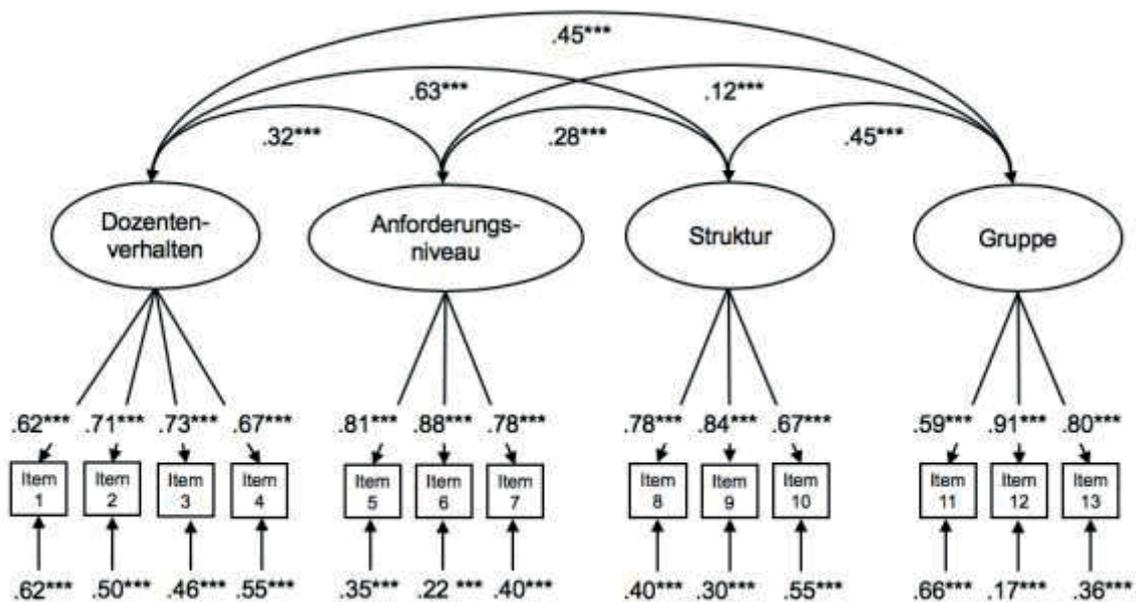


Abbildung 1. Ergebnisse der konfirmatorischen Faktorenanalyse der Prozessskalen, standardisierte Koeffizienten. *** $p < .001$. $N = 382$.

Auf Ebene des Lernoutcomes wurde ein Modell mit den Faktoren *Kompetenzerwerb* und *Transfer* getestet. Für dieses ursprüngliche Modell ($\chi^2(19) = 74.80$, $p < .001$, $\chi^2/df = 3.94$, RMSEA = 0.09 [0.07, 0.11], SRMR = 0.04, CFI = .95, TLI = .93) legten Modifikationsindizes das Hinzufügen einer Korrelation zwischen den Items 14 und 15 Item nahe. Der sich überschneidende Inhalt kann als plausible Ursache für einen solchen über die durch den Faktor *Kompetenzerwerb* erklärte Kovarianz hinausgehenden Zusammenhang angesehen werden. Das resultierende Modell ist in Abbildung 2 dargestellt. Das zweifaktorielle Modell zeigt einen guten (SRMR = 0.03, CFI = .97) bis akzeptablen Fit (TLI = .96, RMSEA = 0.07 [0.045, 0.09]). Der χ^2 -Test fiel ebenfalls signifikant aus mit $\chi^2(18) = 48.38$, $p < .001$.

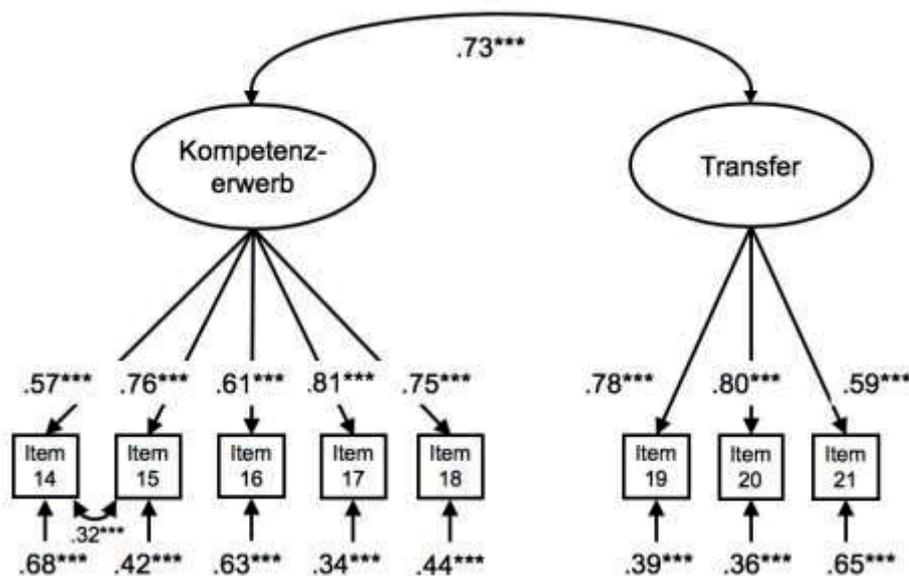


Abbildung 2. Ergebnisse der konfirmatorischen Faktorenanalyse der Outcomeskalen, standardisierte Koeffizienten. *** $p < .001$. $N = 382$.

Stichproben

Alle hier beschriebenen Studien fanden am IdF NRW statt. Bei den Befragten handelt es sich um Dozenten und Teilnehmende der dortigen Gruppen- und Zugführerlehrgänge. Die Teilnehmenden wurden im Rahmen ihrer Lehrgänge zur Teilnahme aufgefordert und diese wurde nicht vergütet. Die Teilnahme erfolgte auf freiwilliger und anonymer Basis. Die Erhebungen fanden im Zeitraum von 2015 bis 2017 statt.

Stichprobe in Studie I

In Studie I wurden zunächst drei Teilnehmer eines Zugführerlehrgangs (100% männlich) sowie fünf Dozenten und Prüfer (100% männlich) interviewt. Die interviewten Lehrgangsteilnehmer waren zwischen 40 und 50 Jahre alt ($M = 43.7$; $SD = 5.5$). Das Alter der interviewten Dozenten und Prüfer variierte zwischen 33 und 56 Jahren ($M = 42.2$; $SD = 8.8$).

Zusätzlich nahmen 56 Lehrgangsteilnehmer (100% männlich) an der Befragung mit dem standardisierten Fragebogen teil. Von ihnen absolvierten 55% eine Gruppen- und 45% einen Zugführerlehrgang. Das Alter der Teilnehmer lag dabei zwischen 23 und 50 Jahren ($M = 33.9$; $SD = 7.4$).

Stichprobe in Studie II

An der Bewertung der Items in Studie II nahmen sieben Dozenten (86% männlich) sowie 26 Teilnehmende eines Gruppenführerlehrgangs (96% männlich) teil, sodass insgesamt $N = 33$ Personen befragt wurden. Die befragten Dozenten waren zwischen 27 und 45 Jahre alt ($M = 38.9$; $SD = 6.0$). In der Gruppe der Teilnehmenden variierte das Alter zwischen 21 und 43 Jahren ($M = 30.5$; $SD = 6.9$).

Stichprobe in Studie III

In Studie III wurden $N = 263$ Lehrgangsteilnehmende befragt. Aus der Analyse mussten die Daten von 20 Teilnehmenden ausgeschlossen werden, da sie nicht mit der Nutzung ihrer Daten einverstanden waren ($n = 10$), mehr als 10% der Items nicht beantworteten ($n = 7$) oder einen monotonen oder unrealistischen Antwortstil aufwiesen ($n = 3$). Die Stichprobengröße für die EFA bestand demnach aus $n = 243$ Lehrgangsteilnehmenden.

Die Befragten (96% männlich) waren zwischen 21 und 55 Jahren alt ($M = 31.8$, $SD = 6.5$). Sie verfügten über eine mittlere Arbeitserfahrung von 13.8 Jahren ($SD = 8.2$) und absolvierten im Mittel 9.00 ($SD = 15.5$) Einsätze im Monat. 14% der Stichprobe arbeiteten bei der Berufsfeuerwehr, 70% bei der freiwilligen Feuerwehr und 16% gaben an, sowohl bei der Berufsfeuerwehr als auch bei der freiwilligen Feuerwehr tätig zu sein.

Stichprobe in Studie IV

Die für die CFAs in Studie IV genutzte Stichprobe bestand aus $N = 382$ Lehrgangsteilnehmenden, von denen zwei Personen wegen fehlender Antworten und eine Person wegen monotonem Antwortverhalten aus der Auswertung ausgeschlossen wurden. Das Alter der Lehrgangsteilnehmenden lag zwischen 20 und 55 Jahren ($M = 33.3$, $SD = 6.9$) und 95% der Stichprobe waren männlich. 54% der Lehrgangsteilnehmenden gehören der Berufsfeuerwehr an, 25% der freiwilligen Feuerwehr und 21% gaben an, in beiden Bereichen tätig zu sein. Die mittlere Berufserfahrung lag bei 14.42 Jahren ($SD = 6.6$) mit 6.35 ($SD = 12.0$) Einsätzen im Monat bei den Teilnehmenden der freiwilligen Feuerwehr und bei 9.35 Jahren ($SD = 7.3$) mit 24.62 ($SD = 23.9$) Einsätzen im Monat bei den Teilnehmenden der Berufsfeuerwehr.

Der hohe Männeranteil in allen beschriebenen Stichproben ist repräsentativ für die Population der Teilnehmenden an Führungskräftelehrgängen deutscher Feuerweherschulen.

Itemkennwerte

Eine Übersicht der Kennwerte der finalen Items (berechnet anhand der Stichprobe aus Studie IV) findet sich in Tabelle 4.

Tabelle 4

Mittelwerte, Standardabweichungen, Trennschärfen, Schiefe und Kurtosis der manifesten Items

Itemnummer	Mittelwert	Standardabweichung	Trennschärfe	Schiefe	Kurtosis	Subskala
1	5.95	0.78	.52	-1.08	2.52	Dozentenverhalten
2	6.22	0.80	.61	-1.21	2.09	Dozentenverhalten
3	5.78	0.94	.63	-0.87	0.67	Dozentenverhalten
4	6.29	0.83	.57	-1.66	4.17	Dozentenverhalten
5	5.44	1.21	.73	-0.54	-0.03	Anforderungsniveau
6	5.54	1.22	.77	-0.64	-0.26	Anforderungsniveau
7	5.86	1.06	.70	-1.14	2.02	Anforderungsniveau
8	5.78	1.12	.67	-1.48	2.78	Struktur
9	5.56	1.01	.72	-0.94	1.32	Struktur
10	5.92	0.85	.56	-0.80	0.78	Struktur
11	6.04	0.80	.50	-0.84	1.13	Gruppe
12	6.38	0.81	.76	-1.54	2.84	Gruppe
13	6.41	0.80	.65	-1.47	2.11	Gruppe
14	6.17	0.78	.55	-1.68	6.04	Kompetenzerwerb
15	6.06	0.79	.71	-1.48	6.03	Kompetenzerwerb
16	5.53	1.20	.55	-1.01	1.21	Kompetenzerwerb
17	5.58	0.91	.71	-0.98	1.27	Kompetenzerwerb
18	5.72	0.81	.62	-1.09	2.81	Kompetenzerwerb
19	5.69	0.83	.60	-0.95	1.86	Transfer
20	5.62	1.04	.65	-0.98	1.33	Transfer
21	5.53	1.08	.52	-0.98	1.44	Transfer

Anmerkung. Skala von 1 = *stimme gar nicht zu* bis 7 = *stimme vollkommen zu*. $N = 382$. Die Zahlen der Subskala *Anforderungsniveau* beruhen auf invertierten Itemrohwerten. Die Trennschärfen sind um das betreffende Item korrigiert.

5 Gütekriterien

Objektivität

Bei der Konstruktion des FIRE-Fragebogens wurden keine expliziten Objektivitätskriterien überprüft. Es kann davon ausgegangen werden, dass durch die Standardisierung des Fragebogens, das gebundene Antwortformat und die oben beschriebene Auswertungsmethode Durchführungs-, Auswertungs- und Interpretationsobjektivität gegeben sind.

Reliabilität

Reliabilitätskoeffizienten für alle FIRE-Skalen in Studie IV sind Tabelle 5 zu entnehmen. Für jede Skala ist ω_h der passende Koeffizient, da der Modellfit des kongenerischen Messmodells jeweils besser ausfällt als der des essentiell tau-äquivalenten Messmodells. Die Reliabilität ist nach den Kriterien von

Evers (2001) für Lernerfolg und Lehrprogrammevaluation insgesamt als akzeptabel (*Dozentenverhalten, Transfer*) bis gut (*Anforderungsniveau, Struktur, Gruppe, Kompetenzerwerb*) anzusehen.

Tabelle 5

Reliabilitätskoeffizienten und Messmodelltests für die FIRE-Skalen in Studie IV

FIRE-Skala	Cronbachs α	ω_h	<i>df</i>	$\Delta\chi^2$	<i>p</i>
Dozentenverhalten	.73	.78	4	156.50	<.001
Anforderungsniveau	.86	.86	3	22.44	<.001
Struktur	.80	.81	3	89.80	<.001
Gruppe	.79	.81	3	138.04	<.001
Kompetenzerwerb	.82	.82	5	127.24	<.001
Transfer	.75	.76	3	73.83	<.001

Anmerkung. Die χ^2 -Differenzentests vergleichen essentiell tau-äquivalente mit kongenerischen Messmodellen. $N = 382$.

Validität

In Ermangelung anderer Instrumente zur Evaluation der Rettungskräfteausbildung gestaltete sich die Validierung des FIRE nicht einfach, nichtsdestotrotz ließen sich einige Validitätshinweise finden.

Inhaltsvalidität. Die Qualitätskriterien zur Ausbildungsevaluation im Bereich der Feuerwehr wurden in semistrukturierten Interviews mit Dozenten und Lehrgangsteilnehmenden des IdF NRW ermittelt, daraus wurden Items abgeleitet, die zudem von anderen Lehrgangsteilnehmenden im Hinblick auf ihre Vollständigkeit und Relevanz bewertet wurden. Aufgrund dieser Entwicklungsschritte wird angenommen, dass die Items das zu messende Konstrukt angemessen repräsentieren und eine inhaltliche Validität gegeben ist.

Konstruktvalidität. Im Hinblick auf die Konstruktvalidität wurden andere Lehrevaluationsinstrumente als konvergente sowie Biasvariablen als divergente Maße erhoben. Biasvariablen sind solche Variablen, die Einfluss auf die Einschätzung der Lehrqualität durch die Urteiler nehmen, tatsächlich aber nicht mit ihr in Zusammenhang stehen (Spiel, 2001). Abschließend wurde die faktorielle Struktur überprüft.

Konvergente Validität. Die strukturell-didaktische Nähe der Führungskräfteausbildung der Feuerwehr zur Hochschullehre legt eine konvergente Validierung der FIRE-Skalen anhand von Evaluationsinstrumenten für die Hochschullehre nahe. Hierzu wurden das Heidelberger Inventar zur Lehrveranstaltungsevaluation (HILVE I; Rindermann, 2001) und das Trierer Inventar zur Lehrevaluation (TRIL; Gläßer et al., 2002) verwendet. Zur Validierung der FIRE-Skala *Dozentenverhalten* wurden die HILVE-Subskalen *Lehrkompetenz* und *Dozentenengagement* eingesetzt. Die *Struktur*-Skala des TRIL fungierte als konvergentes Validitätsmaß für die gleichnamige Skala des FIRE, zur Validierung der *Anforderungsniveau*-Unterskala des FIRE wurde die gleichnamige Unterskala des HILVE I benutzt. Da die Gruppe der Lehrgangsteilnehmenden in anderen Trainings- und Lehrkontexten oft irrelevant ist, kann-

te in der verfügbaren Literatur keine adäquate Validierungsskala gefunden werden. Aus diesem Grund wurden gruppenrelevante Aktivitäten als konvergentes Validitätskriterium für die FIRE-Skala *Gruppe* anhand von drei selbstgenerierten Items erfragt („Wie viele Abende des Lehrgangs haben Sie mit den anderen Lehrgangsteilnehmenden in einer Gaststätte verbracht?“, „An wie vielen Abenden des Lehrgangs haben Sie mehr als zwei Stunden zusammengesessen?“, und „Ich würde jederzeit wieder einen Lehrgang mit den anderen Teilnehmenden besuchen.“). Zur Validierung der Skala *Kompetenzwerb* wurden die Unterskalen *Lernen-qualitativ* und *Lernen-quantitativ* des HILVE II benutzt. Tabelle 6 zeigt eine Übersicht der Korrelationen der FIRE-Skalen mit den anderen Evaluationsinstrumenten. Zusammenfassend lassen sich konsistent positive Korrelationen zwischen konvergenten Evaluationsmaßen und dem FIRE-Fragebogen feststellen.

Divergente Validität. Zur divergenten Validierung sind im Kontext der Lehrveranstaltungsevaluation Biasvariablen von besonderem Interesse. Erfasst wurden in diesem Zusammenhang (1) das Ausmaß, in dem sich die Teilnehmenden gut auf den Lehrgang vorbereitet fühlten, (2) die wahrgenommene Angemessenheit der Gruppengröße, (3) der mit dem Lehrgang verbundene Zeitaufwand, (4) das Ausmaß, in dem sich die Teilnehmenden ohne äußere Störungen auf den Lehrgang konzentrieren konnten, (5) wie gut sie sich auf die Abschlussprüfung vorbereitet fühlten und (6) ihre Stimmung zum Zeitpunkt der Evaluation. Einige dieser Biasvariablen können als theoretisch nicht völlig unabhängig von einzelnen FIRE-Skalen erachtet werden. So ist zu erwarten, dass der Zeitaufwand und das wahrgenommene Ausmaß an Überforderung zusammenhängen. Auch mag sich ein großer Lernerfolg positiv auf die Stimmung auswirken. Kleine oder mittlere Zusammenhänge sprechen in diesem Zusammenhang daher nicht gegen die Validität der Skalen. Größere Zusammenhänge würden die Konstruktvalidität der Skalen jedoch in Frage ziehen.

Die Ergebnisse zeigen überwiegend kleine bis mittlere Zusammenhänge (zwischen $r = .03$, n.s., und $.45$, $p < .001$). Die größten unter diesen Zusammenhängen lassen sich meist gut theoriegeleitet erklären. So zeigt sich erwartungskonform ein größerer Effekt für den Zusammenhang von *Anforderungsniveau* und *Zeitaufwand* ($r = .43$, $p < .001$). Insgesamt korrelieren die untersuchten Biasvariablen zwar systematisch mit den FIRE-Skalen, die Größe der Zusammenhänge geht aber nicht über das unter der Annahme konstruktvalider Messungen der FIRE-Dimensionen zu erwartende Maß hinaus.

Faktorielle Validität. Zudem legen die Ergebnisse der CFA aus Studie IV nahe, dass der FIRE-Fragebogen verschiedene distinkte Facetten misst und sie zeigen, dass angehende Zug- und Gruppenführer ihre Einschätzung der Lehre am IdF zwischen einzelnen Aspekten angemessen differenzieren (Schulte & Thielsch, 2019). Dies impliziert die Abwesenheit von großen Halo-Effekten und bedeutet, dass die Bewertung nicht maßgeblich durch die Stimmung der Lehrgangsteilnehmenden oder äußere Faktoren verzerrt wurde (Marsh, 1987).

Tabelle 6

Validitätsmaße des FIRE

	Dozenten- verhalten	Anforde- rungs- niveau (r)	Struktur	Gruppe	Kom- petenz- erwerb	Transfer
Lehrkompetenz (HILVE)	.67***	.23***	.60***	.29***	.47***	.49***
Dozentenengagement (HILVE)	.59***	.18***	.44***	.37***	.42***	.41***
Struktur (TRIL)	.66***	.34***	.66***	.30***	.46***	.49***
Anforderungsniveau (HILVE)	.26***	.66***	.30***	.13*	.21***	.33***
Gruppe (neue Items)	.25***	.17**	.21***	.36***	.19***	.30***
Lernen qualitativ (HILVE)	.47***	.17**	.33***	.32***	.55***	.44***
Lernen quantitativ (HILVE)	.43***	.06 [†]	.33***	.27***	.50***	.43***
Benotung des Lehr- gangs	.43***	.30***	.40***	.21***	.34***	.41***
Gesamtzufriedenheit	.56***	.37***	.50***	.31***	.46***	.53***
Vorbereitung auf Lehrgang	.16**	.25***	.23***	.03 [†]	.14**	.27***
Gruppengröße	.17***	.28***	.23***	.21***	.18***	.21***
Zeitaufwand	.27***	.43***	.30***	.17***	.27***	.35***
Konzentration	.45***	.30***	.38***	.27***	.28***	.32***
Prüfungsvorbereitung	.38***	.35***	.29***	.11*	.20***	.37***
Stimmung	.37***	.27***	.27***	.24***	.18***	.28***

Anmerkung. $N = 129-382$. [†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$. r = negativ kodiert

Kriteriumsvalidität. Theoriegeleitet sollten hohe Werte auf den verschiedenen FIRE-Dimensionen mit einer hohen Gesamtzufriedenheit der Teilnehmenden mit dem Lehrgang einhergehen. Zur Validierung benutzten wir daher eine Gruppe von fünf Items (selbstgeneriert und entnommen aus Gläßer et al., 2002), welche die Gesamtzufriedenheit mit dem Lehrgang erfragten. Zudem wurde der Lehrgang anhand einer Schulnotenskala (siehe oben, Item Nr. 23) von den Teilnehmenden bewertet. Alle FIRE-Skalen zeigen mittlere bis große Korrelationen mit der Gesamtzufriedenheit der Lehrgangsteilnehmenden und ihrer Benotung des Lehrgangs (siehe Tabelle 6).

Deskriptive Statistiken

Tabelle 7 zeigt einen Überblick der deskriptiven Statistiken der Items und der Skalen, die anhand der Stichprobe aus Studie IV erhoben wurden.

Tabelle 7

Mittelwerte, Standardabweichungen, Schiefe und Kurtosis der FIRE-Skalen

FIRE-Skala	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis
Dozentenverhalten	6.06	0.65	-1.07	1.82
Anforderungsniveau*	5.61	1.04	-0.56	0.13
Struktur	5.76	0.84	-1.23	2.26
Gruppe	6.27	0.67	-1.33	2.47
Kompetenzerwerb	5.82	0.69	-1.17	3.77
Transfer	5.61	0.81	-0.88	1.54

Anmerkung. Die Beantwortung der Items erfolgte mit einem 7-stufigen Antwortformat mit 1 = *stimme gar nicht zu* bis 7 = *stimme vollkommen zu*. * = Skala wurde invertiert. *N* = 382.

Nebengütekriterien

Die FIRE-Skalen erlauben eine sehr zeit- und kostenökonomische Erfassung der Teilnehmendenurteile über die Qualität der Ausbildungen von Führungskräften im Rettungsdienst. Die Durchführungsdauer darf mit etwa 5 Minuten (Schätzwert der Autoren) als zumutbar sowohl für Dozentinnen und Dozenten als auch für Teilnehmende gelten. Es besteht die mit der Verwendung von Rating-/Likertskalen stets einhergehende Möglichkeit zur Verfälschung der Antworten. Um die Motivation zum ehrlichen Antworten zu erhöhen, sollte den Teilnehmenden in jedem Falle Anonymität zugesichert werden.

6 Literatur und Datenquellen

Weiterführende Literatur

- Schulte, N., & Thielsch, M. T. (2019). Evaluation of firefighter leadership trainings. *International Journal of Emergency Services*, 8, 34–49. doi:10.1108/IJES-03-2018-0020
- Thielsch, M. T., Busjan, J. N., & Frerichs, K. (2018). Feedback-Instrument zur Rettungskräfte-Entwicklung – Prüfungen (FIRE-P) [Feedback instrument for rescue forces education – Examinations (FIRE-P)]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis260

Danksagung

Die Autoren möchten sich herzlich beim Institut der Feuerwehr Nordrhein-Westfalen, insbesondere bei Thomas Löchteken, Stephanie Vöge, Yannick Ngatchou, Matthias Wegener und Bertholt Penkert für die umfassende Unterstützung dieser Instrumentenerstellung bedanken. Dank gebührt ebenso den weiteren Beteiligten in der Psychologie: Mona Olde lütke Beverborg, Jessica Busjan, Linda Loberg, Heinz Holling und Guido Hertel.

Kontakt zu Autoren

- Niklas Schulte, M.Sc., Westfälische Wilhelms-Universität Münster, Institut für Psychologie, n.schulte@uni-muenster.de

- Stephanie Babel, M.Sc., Psychologie, Trainerin und Beraterin für Führungskräfteentwicklung, stephanie.babel@profi-m.de
- Marius Messinger, B.Sc., Westfälische Wilhelms-Universität Münster, Institut für Psychologie, mmessinger@uni-muenster.de
- Meinald Thielsch, Prof. Dr. Dipl.-Psych., Westfälische Wilhelms-Universität Münster, Institut für Psychologie, thielsch@uni-muenster.de

Literaturverzeichnis

- AG-B III – LFV NRW, AGBF NRW, WFV NRW, & IdF NRW (2007). *Lernziele für die Ausbildung zum Gruppenführer in der Berufsfeuerwehr* [Learning objectives for the training of group leaders in professional fire departments]. Retrieved from http://www.idf.nrw.de/service/downloads/pdf/biii_lzk_2008.pdf
- AG-F III – LFV NRW, AGBF NRW, WFV NRW, & IdF NRW (2005). Gruppenführer-Ausbildung und Truppmann- / Truppführer-Aus- und Fortbildung der Freiwilligen Feuerwehren in Nordrhein-Westfalen [Group leader training and troop member/troop leader training for voluntary fire departments in North Rhine-Westphalia]. Retrieved from http://www.idf.nrw.de/service/downloads/pdf/rderlass_im_20051221_lernziele.pdf
- AG-F IV – LFV NRW, AGBF NRW, WFV NRW, & IdF NRW (2007). *Lernziele für die Ausbildung zum Zugführer (Freiwillige Feuerwehr)* [Learning objectives for training of platoon leaders (voluntary fire department)]. Retrieved from http://www.idf.nrw.de/service/downloads/pdf/lernziele_zugfuehrer_20070925.pdf
- Babel, S. (2016). *Evaluation der Führungsausbildung am Institut der Feuerwehr NRW* [Evaluation of the leader training at the Institute of Fire Brigade North Rhine-Westphalia] (Unpublished master's thesis). Westfälische Wilhelms-Universität, Münster.
- Bandura, A. (1977). *Social learning theory*. Oxford: Prentice-Hall.
- Blanchard, P. N., & Thacker, J. W. (2010). *Effective training: Systems, strategies, and practices* (4th ed.). Upper Saddle River, NJ: Pearson Education.
- Braun, E., & Gusy, B. (2006). Perspektiven der Lehrevaluation [Perspectives of teaching evaluation]. In G. Krampen & H. Zayer (Eds.), *Didaktik und Evaluation in der Psychologie [Didactics and evaluation in psychology]* (pp. 152-166). Göttingen: Hogrefe.
- Evers, A. (2001). The revised Dutch rating system for test quality. *International Journal of Testing*, 1, 155–182. doi:10.1207/S15327574IJT0102_4
- Feuerwehr-Dienstvorschrift (FwDV) 3 [German fire service regulation 3] (2008). *Einheiten im Löschen und Hilfeleistungseinsatz [Units in extinguishing and assistance incidents]*. Retrieved from http://www.idf.nrw.de/service/downloads/pdf/fwdv3_200802.pdf
- Gediga, G., von Kannen, K., Schnieder, F., Köhne, S., Luck, H., & Schneider, B. (2000). *KIEL. Ein Kommunikations-Instrument für die Evaluation von Lehrveranstaltungen [KIEL: An communication instrument for the evaluation of courses]*. Bissendorf: Methodos.

- Gläßer, E., Gollwitzer, M., Kranz, D., Meiniger, C., Schlotz, W., Schnell, T., & Voß, A. (2002). *Trierer Inventar zur Lehrevaluation [Trier inventory for teaching evaluation]*. Retrieved from www.zpid.de/pub/tests/PT_9004523_TRIL_weibl_Doiz_Fragebogen.pdf
- Greve, W., & Wentura, D. (1997). *Wissenschaftliche Beobachtung: Eine Einführung [Scientific observation: An introduction]*. Weinheim: Beltz.
- Howard, M. C. (2018). Scale Pretesting. *Practical Assessment, Research & Evaluation*, 23, 1–14. Retrieved from: <https://pareonline.net/getvn.asp?v=23&n=5>
- IdF NRW (2012). *Ausbildungsplan – Musterplan B III [IdF NRW exemplary syllabus for group leaders in professional fire departments]*. Retrieved from http://www.idf.nrw.de/ausbildung/katalog/dokumente/musterausbildungsplan_biii.pdf
- IdF NRW (2013). *Ausbildungsplan – Musterplan FIII. [IdF NRW exemplary syllabus for group leaders in voluntary fire departments]*. Retrieved from https://www.idf.nrw.de/ausbildung/katalog/dokumente/musterausbildungsplan_fiii_26.pdf
- IdF NRW (2016). *Ausbildungsplan – Musterplan FIV [IdF NRW exemplary syllabus for platoon leaders in voluntary fire departments]*.
- Kirkpatrick, D. L. (1979). Techniques for evaluating training programs. *Training and Development Journal*, 33, 78–92. Retrieved from [http://iptde.boisestate.edu/FileRepository.nsf/bf25ab0f47ba5dd785256499006b15a4/693b43c6386707fc872578150059c1f3/\\$FILE/Kirkpatrick_79.pdf](http://iptde.boisestate.edu/FileRepository.nsf/bf25ab0f47ba5dd785256499006b15a4/693b43c6386707fc872578150059c1f3/$FILE/Kirkpatrick_79.pdf)
- Kirkpatrick, D. L. (1998). *Evaluating training programs – The four levels* (2nd ed.). San Francisco, CA: Berrett-Koehler Publishers.
- 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. doi:10.1016/0883-0355(87)90001-2
- Mayring, P. (2000). Qualitative Inhaltsanalyse [Qualitative content analysis]. *Forum Qualitative Sozialforschung*, 1, article 20. Retrieved from www.qualitative-research.net/index.php/fqs/article/download/1089/2384
- Ministerium für Inneres und Kommunales des Landes Nordrhein-Westfalen [Ministry of the interior North Rhine Westphalia] (2014). *Gefahrenabwehr in Nordrhein-Westfalen: Jahresbericht 2013 [Emergency response in North Rhine Westphalia: Annual report 2013]*. Retrieved from <http://www.mik.nrw.de/themen-aufgaben/ Gefahrenabwehr-feuerwehr-katastrophenschutz/jahresbericht.html>
- Mistele, P. (2007). *Faktoren des verlässlichen Handelns: Leistungspotenziale von Organisationen in Hochrisikoumwelten [Factors of reliable acting: Performance potentials of organizations in high-risk environments]*. Wiesbaden: Deutscher Universitäts-Verlag.
- Rindermann, H. (2001). *Lehrevaluation: Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungs-evaluation an Hochschulen mit einem Beitrag zur Evaluation computer-basierter Unterrichts [Teaching evaluation: Introduction and overview of research and practice of course evaluation at colleges with an article on the evaluation of computer-based teaching]*. Landau: Verlag Empirische Pädagogik.

- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23–74. Retrieved from https://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf
- Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik* [Psychological diagnostics] (5th ed.). Berlin: Springer.
- Schulte, N., & Thielsch, M. (2019). Evaluation of firefighter leadership trainings. *International Journal of Emergency Services*, 8, 34–49.: doi 10.1108/IJES-03-2018-0020
- Sievers, S. C. (2015). *Effektiv führen, wenn es brennt! Eine Untersuchung des feuerwehr-spezifischen Führungsprozesses „Führungsvorgang“ mittels Critical Incident Technique [Leading effectively when the world is on fire! And investigation of a fire service-specific leadership process „leadership procedure“ with Critical Incident Technique]* (Unpublished master's thesis). Westfälische Wilhelms-Universität, Münster.
- Souvignier, E., & Gold, A. (2002). Fragebögen zur Lehrevaluation: Was können sie leisten [Questionnaires for teaching evaluation: What can they contribute?]. *Zeitschrift für Evaluation*, 1, 265–280.
- Spiel, C. (2001). Der differentielle Einfluß von Biasvariablen auf studentische Lehrveranstaltungsbeurteilungen [The differential influence of bias variables on students' course evaluations]. In U. Engel (Ed.), *Hochschul-Ranking: Zur Qualitätsbewertung von Studium und Lehre [College rankings: On the quality appraisal of studying and teaching]* (pp. 61–82). Frankfurt: Campus.
- Staufenbiel, T. (2000). Fragebogen zur Evaluation von universitären Lehrveranstaltungen durch Studierende und Lehrende [Questionnaire for the evaluation of college lectures by students and teachers]. *Diagnostica*, 46, 169–181. doi:10.1026//0012-1924.46.4.169
- Thielsch, M. T., Busjan, J. N., & Frerichs, K. (2018). Feedback-Instrument zur Rettungskräfte-Entwicklung – Prüfungen (FIRE-P) [Feedback instrument for rescue forces education – Examinations (FIRE-P)]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis260
- Thielsch, M. T., & Hirschfeld, G. (2012a). Münsteraner Fragebogen zur Evaluation von Seminaren – revidiert (MFE-Sr) [Münster questionnaire for the evaluation of seminars – revised (MFE-Sr)]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis86
- Thielsch, M. T., Kläpker, L., & Streppel, L. (2019). Feedback-Instrument zur Rettungskräfte-Entwicklung – Basisausbildung (FIRE-B) [Feedback instrument for rescue forces education – Basic training (FIRE-B)]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis266
- Thielsch, M. T., & Stegemöller, I. (2012). Münsteraner Fragebogen zur Evaluation – Zusatzmodul Moderation (MFE-ZMo) [Münster questionnaire for the evaluation – Additional module moderation (MFE-ZMo)]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis101

- Wolf, P., Spiel, C., & Pellert, A. (2001). Entwicklung eines Fragebogens zur globalen Lehrveranstaltungsevaluation – ein Balanceakt zwischen theoretischem Anspruch, Praktikabilität und Akzeptanz [Development of a questionnaire for global course evaluation – a balancing act between theoretical aspiration, practicability, and acceptance. In C. Spiel (Ed.), *Evaluation universitärer Lehre. Zwischen Qualitätsmanagement und Selbstzweck* [Evaluation of college teaching: Between quality management and end in itself] (pp.89–109). Münster: Waxmann.
- Zumbach, J., Spinath, B., Schahn, J., Friedrich, M., & Kögel, M. (2007). Entwicklung einer Kurzskaala zur Lehrevaluation [Development of a short scale for teaching evaluation]. In M. Krämer, S. Preiser, & K. Brusdeylins (Eds.), *Psychologiedidaktik und Evaluation VI* [Psychological didactics and evaluation IV] (pp. 317–325). Göttingen: V & R unipress.