

Faktoranalytische Validierung der der zweiten Trendstudie zur Berufstreue bei angehenden Notfallsanitäter:innen

Christian Eisenbast   ¹

Thomas Hofmann   ²

¹ Fraunhofer-IESE, Research Program Digital Healthcare, Kaiserslautern, Deutschland

² DGRe e. V., Deutschland

Zusammenfassung

Hintergrund, Fragestellung: Die zweite Berufstreuestudie (BTS 2) beschäftigt sich mit der Berufstreue von angehenden Notfallsanitäter:innen in Deutschland und untersucht die Gründe für eine mögliche Berufsabkehr sowie die Berufszufriedenheit. Das Ziel dieser Validierungsstudie ist, die Itembatterie "Zufriedenheit" der BTS 2 mittels konfirmatorischer Faktorenanalyse eingehend zu untersuchen, um datenbasierte Aussagen über die Validität treffen zu können und somit die Gültigkeit dieser Messung zu erhellen.

Methodik: Mit der kompletten Stichprobe der BTS 2 ($n = 1340$) werden nach Prüfung der Verteilung, internen Konsistenz und Korrelationen eine konfirmatorische Faktorenanalyse und Fehlspezifizierungen auf Itemebene berechnet.

Ergebnisse: Cronbachs Alpha wird mit $\alpha = .813$ angegeben. Alle Items weisen eine Schiefe nahe null auf, abgesehen von Item C001_06 (-1.39). Bei Kurtosis weisen alle Items ein Werte nahe 3 aus, abgesehen von C001_06 (4.95) und C001_08 (1.75). Es liegen keine Korrelationen > 0.7 und keine negativen Varianzen vor. Fit Indizes der CFA: RMSEA = 0.131, SRMR = 0.068, CFI = 0.915 und TLI = 0.854.

Abstract

Background: The second Professional Loyalty Study (BTS 2) focuses on the professional loyalty of prospective paramedics in Germany and explores the reasons for potential career changes as well as job satisfaction. The aim of this validation study is to thoroughly investigate the "Satisfaction" item battery of BTS 2 through confirmatory factor analysis in order to make data-based statements about its validity, thereby clarifying the accuracy of this measurement.

Methods: Using the complete sample of BTS 2 ($n = 1340$), after examining the distribution, internal consistency, and correlations, a confirmatory factor analysis and misspecifications at the item level are calculated.

Results: Cronbach's alpha is reported as $\alpha = .813$. All items show skewness close to zero, except for item C001_06 (-1.39). For kurtosis, all items have values close to 3, except for C001_06 (4.95) and C001_08 (1.75). There are no correlations > 0.7 and no negative variances. Fit indices of the CFA are: RMSEA = 0.131, SRMR = 0.068, CFI = 0.915, and TLI = 0.854.

Discussion: Significant misspecifications are evident, and the model cannot be considered

Diskussion: Es ist von erheblichen Fehlspezifizierungen auszugehen, dem Modell kann keine Validität unterstellt werden. Die Stärke dieser Validierungsstudie liegt in ihrer erheblichen Stichprobengröße, die entsprechende Teststärke erwarten lässt. Limitierend auf die Methode CFA wirkt sich bereits die Struktur der Itematterie aus, genauer gesagt, dass ein Faktor lediglich aus zwei Items besteht. Die analysierte Itematterie kann zwar deskriptiv auf Basis der Aussagekraft der Einzelitems Verwendung finden, ist jedoch nicht als psychometrisches Konstrukt zu verstehen.

Fazit: Es spricht nichts gegen die deskriptive Analyse der einzelnen Items im Rahmen ihrer jeweiligen Aussagekraft, die bei jedem Item dann für sich selbst steht. Die Itematterie ist jedoch nicht in der Lage, Zufriedenheit als psychometrische Skala zu reflektieren. Bei der Neuauflage der BTS ist eine iterative und Expert:innen gestützte Neukonzeption zu empfehlen.

valid. The strength of this validation study lies in its considerable sample size, which suggests adequate statistical power. The structure of the item battery itself, specifically that a factor consists of only two items, limits the methodological efficacy of the Confirmatory Factor Analysis (CFA). While the analyzed item battery can be used descriptively based on the informative value of individual items, it cannot be understood as a psychometric construct.

Conclusions: There is nothing against conducting a descriptive analysis of the individual items within the scope of their respective informative value, as each item stands on its own. However, the item battery is not capable of reflecting satisfaction as a psychometric scale. For the new edition of the BTS, an iterative redesign supported by experts is recommended.

Schlagwörter

Validierungsstudie, Zufriedenheit, Berufstreue, Faktorenanalyse, CFA

Keywords

Validation Study, Satisfaction, Professional Loyalty, Factor Analysis, CFA

Problem-/Ziel-/Fragestellung

Die Validierungsstudie ist im Kontext der zweiten Berufstreuestudie mit dem Titel „Deskriptive Analyse der zweiten Umfrage zur Berufstreue angehender Notfallsanitäter:innen (Berufstreuestudie 2): Eine Trendstudie“ (Elsenbast et al., 2024) zu verstehen.

Zwecks Orientierung zum Hintergrund wird im folgenden Abschnitt eine **Einführung in die Berufstreuestudie 2 (BTS 2)** gegeben. Im deutschen Rettungsdienst offenbart sich ein markantes Dilemma, charakterisiert durch eine steigende Anzahl an Einsätzen einerseits und einem akuten Mangel an Fachpersonal andererseits. Obwohl der steigende Arbeitsdruck innerhalb des Sektors in zunehmendem Maße Beachtung findet, bleibt die Problematik des Personalmangels weitestgehend im Schatten. Eine erhebliche Anzahl von Rettungsdienstmitarbeiter:innen hegt Überlegungen zum Berufswechsel. Ein Umstand, der bereits durch die erste Berufstreuestudie (BTS 1) aus dem Jahr 2020 dokumentiert wurde. Die BTS 2 knüpft als Trendstudie an diese Erkenntnisse an, indem sie die Gründe für die Berufsabkehr und die Berufszufriedenheit weiter erforscht, um eine fundierte Basis für zukünftige Untersuchungen sowie Ansätze zur Problemlösung zu bieten. Im Rahmen einer Online-Umfrage, die vom 01.11.22 bis zum 31.01.23 durchgeführt wurde, nahmen 1618 Auszubildende zur/zum Notfallsanitäter:in in Deutschland teil. Der Fragebogen, eine Erweiterung der vorherigen Berufstreuestudie, diente der Erhebung deskriptiver Statistiken bezüglich der Demografie der Teilnehmenden, ihrer Berufszufriedenheit, der Neigung, im Beruf zu

verbleiben, sowie der vermuteten Gründe für einen vorzeitigen Berufsausstieg, aufgeschlüsselt nach Bundesländern. Für Details verweisen wir auf die BTS 2 selbst.

Bezugnehmend auf das Thema der vorliegenden Studie ist zunächst wichtig, sich einen Überblick über den Charakter von sog. *latenten Merkmalen* zu schaffen. Einige Merkmale, die analysiert werden können, sind direkt beobachtbar, z.B. die Körpergröße von Menschen in Zentimetern. Andere Variablen sind nicht direkt beobachtbar, wie zum Beispiel Haltungen oder auch Zufriedenheit mit einer zu definierenden unabhängigen Variable. Um dieses nicht beobachtbare Merkmal zu analysieren, werden mehrere verwandte beobachtbare Variablen erstellt, um anschließend in einem Test „indirekt“ Schlussfolgerungen ziehen zu können. „Die mit dem Test zu messenden nicht direkt beobachtbaren Konstrukte [...] werden als latente Merkmale [...] aufgefasst, welche das manifeste Verhalten bestimmen“ (Döring et al., S. 483). Methodologisch ist dabei die Item Response Theory leitend, die innerhalb der Psychometrie, also der Disziplin, die sich allgemein mit Theorie und Methode des psychologischen Messens befasst und einen wichtigen Stellenwert bei der Fragebogenforschung einnimmt.

Sollen latente Merkmale analysiert werden, geschieht das mit einer Kombination von Items, die meistens als Fragen in einem Fragebogen formuliert sind. Um ein nicht beobachtbares Konstrukt (wie hier Zufriedenheit) zuverlässig darstellen zu können, müssen Belege erbracht werden, die auf die Validität schließen lassen.

Validität (auch: Gültigkeit) als psychometrisches Hauptgütekriterium gibt im

traditionellen Testverständnis an, ob ein (psychologischer) Test tatsächlich das misst, was er messen soll (Hartig, J., Frey, A. & Jude, N., 2012, S. 144). Eine hohe **Reliabilität** ist Voraussetzung für eine hohe Validität (Döring et al., S. 445), was sich dadurch erschließt, dass man Reliabilität als Zuverlässigkeit, Präzision oder Messgenauigkeit versteht (Döring et al., S. 441). Der Vollständigkeit halber, um alle drei Kerngütekriterien quantitativer Forschung angeführt zu haben, ist ebenso auf **Objektivität** hinzuweisen, die angibt, ob ein Test bzw. dessen Ergebnisse von der testanwendenden Person unabhängig sind.

Um die angesprochene Validität zu prüfen, stehen mehrere Möglichkeiten zu Verfügung, wie die **Faktorenanalyse**, die auf faktorielle Validität, die wiederum zur Familie der Konstruktvalidität gehört, prüft. Vorwegnehmend: Ein Faktor in der Faktorenanalyse repräsentiert eine latente Variable, die nicht direkt gemessen werden kann, aber aus den beobachteten Variablen abgeleitet wird und diese somit reflektiert. Diese latente Variable steht für ein zugrundeliegendes Konstrukt oder eine Dimension, die gemeinsame Muster oder Korrelationen zwischen mehreren beobachteten Variablen erklärt. So sollte beispielsweise die Variable C001_01 das Konstrukt „Zufriedenheit mit der *theoretischen* Ausbildung in der Berufsfachschule“ reflektieren, die in „Zufriedenheit mit der Berufsfachschule“ und diese wiederum in „(Berufs-)Zufriedenheit“ als übergeordnetes Modell mündet (siehe auch Abbildung 1 und Abbildung 4).

Das **Ziel** der vorliegenden Validierungsstudie besteht darin, die Itembatterie "Zufriedenheit" der BTS 2 mittels konfirmatorischer Faktorenanalyse eingehend zu untersuchen,

um datenbasierte Aussagen über die Validität der Zufriedenheitsmessung innerhalb der BTS 2 zu treffen und somit die Gültigkeit dieser Messung zu erhellen.

Studiendesign/Untersuchungsmethoden

Die **Erhebung der BTS 2** fand vom 01.11.22 bis zum 31.01.23 per Online-Fragebogen mit dem Online-Umfragewerkzeug SoSci-Survey statt. Alle im Internet recherchierbaren deutschen Ausbildungsstätten ($n = 143$) für Notfallsanitäter:innen wurden angeschrieben und gebeten, die Umfrage den Auszubildenden zugänglich zu machen. Weiterhin wurde die Umfrage über die Social-Media-Kanäle der DGR e beworben. Es handelt sich folglich um eine nicht-probabilistische Stichprobe. Bis auf drei Variablen waren fehlende Werte in allen Variablen zwischen 0% und maximal 3,6% vorhanden, durch fehlende Zwangseingaben bedingt und folgten keinem Muster (missing at random).

Diese Validierungsstudie fokussiert die **Itembatterie „Zufriedenheit“** (Abbildung 1), die mit einer sechsstufigen Likert-Skala (1 = stimme gar nicht zu, 2 = stimme voll zu) beantwortet wird. Diese Itembatterie wurde für die erste Durchführung der Berufstreuestudie von den Autor:innen der BTS als Ad-hoc-Skala auf Basis inhaltlicher Erwägungen erstellt, ohne tiefer gehende Anforderungen der Skalenbildung entsprechend der psychometrischen Testtheorie (Döring et al., S. 268) zu berücksichtigen. Andere Itemgruppen der BTS 2 werden nicht validiert, da diese nicht im Sinne eines psychometrischen Instruments auswertbar sind.

C001_01: Mit meiner theoretischen Ausbildung in der Berufsfachschule bin ich sehr zufrieden
 C001_02: Mit meiner theoretischen Ausbildung im Krankenhaus bin ich sehr zufrieden
 C001_03: Mit meiner theoretischen Ausbildung an der Rettungswache bin ich sehr zufrieden
 C001_04: Mit meiner praktischen Ausbildung in der Berufsfachschule bin ich sehr zufrieden
 C001_05: Mit meiner praktischen Ausbildung im Krankenhaus bin ich sehr zufrieden
 C001_06: Mit meiner praktischen Ausbildung an der Rettungswache bin ich sehr zufrieden
 C001_07: Mit meiner Ausbildung bin ich insgesamt sehr zufrieden
 C001_08: Die Ausbildung enttäuscht die Erwartungen, die ich zum Ausbildungsbeginn hatte
 C001_09: Meine Ausbildung hält alle Vorgaben ein (z.B. NotSanG, NotSan-APrV)
 C001_10: Mein Ausbildungsbetrieb engagiert sich sehr für meine qualitativ hochwertige Ausbildung

Abbildung 1: Infobox "Itembatterie Zufriedenheit"

Um Hinweise auf **interne Konsistenz** als Reliabilitätsparameter zu erhalten, wurde zunächst das unstandardisierte Cronbachs Alpha (α) berechnet. Dabei wurde der Empfehlung von Backhaus et al. (2015, S. 142) folgend mit der Berechnung von Cronbachs Alpha bewusst auf ein Prüfmaß der klassischen Testtheorie zurückgegriffen. Da Cronbachs Alpha jedoch weder ein Maß für Validität noch für die Eindimensionalität einer Skala darstellt, ist die Validität gesondert zu überprüfen (Knekta et al., 2019, S. 5–6). Dies kann als grundsätzliche Notwendigkeit bei der Nutzung eines psychometrischen Messinstruments verstanden werden (Kane, 2016; Knekta et al., 2019, S. 16).

Zur **Validitätsanalyse** wird nicht nur die Augenscheinvalidität herangezogen, sondern auch multivariate Analysemethoden wie die Faktorenanalyse. Faktorenanalysen sind statistische Verfahren, die die Beziehungen zwischen einer Reihe von Umfrageitems analysieren, um festzustellen, ob die Antworten

der Teilnehmenden mit verschiedene Untergruppen von Items enger miteinander zusammenhängen als mit andere Untergruppen. Die Faktorenanalyse ist also eine Analyse der Dimensionalität unter den Items (Knekta et al., 2019, S. 6–7). Die Faktorenanalyse kann in exploratorische (EFA) oder konfirmatorische Faktorenanalyse (Confirmatory Factor Analysis = CFA) unterschieden werden. Eine EFA eignet sich für die frühen Phasen der Instrumentenentwicklung (Knekta et al., 2019, S. 6). Da in der vorliegenden Studie ein bereits eingesetztes, aber noch nicht validiertes Instrument analysiert wird, bietet sich jedoch eine EFA nicht an, sondern eine CFA.

Die **CFA** ist der Spezialfall eines Strukturgleichungsmodells. Sie wird verwendet, um ein zuvor angegebenes theoretisches Modell zu analysieren. Die CFA ist dann geeignet, wenn Forschende eine bereits existierende Umfrage verwenden, die eine etablierte Struktur mit einer ähnlichen Population von Befragten einschließt (Knekta et al., 2019, S. 7). Diese Voraussetzung liegt bei der BTS 2 vor. Mathematisch gesehen beinhaltet die Faktorenanalyse die Analyse der Varianzen und Kovarianzen zwischen den Items.

Bezugnehmend die **eingesetzte Software**, wird die CFA mit dem Paket „lavaan“ (Rosseel, 2012) unter RStudio [R Version 4.3.3 (2024-02-29 ucrt), RStudio-Version 2023.12.1, Build 402] ausgeführt. Faktorladungen sind in dieser Studie promax-rotiert, d.h. zuerst orthogonal rotiert und dann oblique transformiert. Ziel ist, die absoluten Werte der Primärladungen zu maximieren und die Sekundärladungen gegen Null gehen zu lassen. Eine orthogonale Rotation scheint

aufgrund der Null-Korrelation bei der inhaltlichen Interpretation der latenten Variablen, die hier vorgenommen wird, nicht geeignet. In diesem Bezug kann es zum Missverständnis kommen, dass Faktorladungen nicht > 1 sein könnten und ein Messfehler vorliegen würde. Dieses Missverständnis rührt wahrscheinlich von der klassischen explorativen Faktorenanalyse her, bei der die Faktorladungen Korrelationen sind, wenn eine Korrelationsmatrix analysiert wird und die Faktoren standardisiert und unkorreliert (orthogonal) sind (Jöreskog, 1999). Wenn die Faktoren jedoch korreliert sind (schiefwinklig), sind die Faktorladungen Regressionskoeffizienten und keine Korrelationen und können daher > 1 sein.

Bei der **Auswertung der Faktorenanalyse** wird sich auf die Konventionen von Hu & Bentler (1999) sowie von Backhaus et al. (2015) bezogen. Demnach ist für den **Tucker-Lewis Index (TLI) und Comparative-Fit-Index (CFI)** ein Trennwert von 0.95 zu berücksichtigen (Hu & Bentler, 1999, S. 27). Für den **SRMR**, als deskriptives Maß, gilt der Cut-Off von 0.09. **SRMR** bedeutet „standardized root mean squared residual (SRMR)“ (Hu & Bentler, 1999, S. 27). Für den **RMSEA** gilt der Cut-Off von 0.06. RMSEA bedeutet „root mean squared error of approximation“ (Hu & Bentler, 1999, S. 28). Der RMSEA-Index ist ein inferenzstatistisches Maß und ein globales Gütekriterium der Modellprüfung. Er prüft, ob ein Modell die Realität gut schätzen kann. Wie von Backhaus et al. (2015, S. 148–149) erläutert, ist der RMSEA robuster als der Chi-Quadrat-Test (χ^2), der an eine Reihe von Annahmen geknüpft ist und auch von Bentler & Bonett (1980, S. 603–604) aufgrund seiner mangelnden Robustheit kritisiert wird. Der χ^2 -

Test wird hier aus diesem Grund nicht zur Analyse herangezogen.

Die vorgestellten Fit Indizes reflektieren Modellfehlspezifikationen. Jedoch sind nicht alle Modellfehlspezifikationen problematisch. Betrachten wir Ordnungseffekte, können zwei Items unabhängig von ihrem gemeinsamen Faktor korrelierte Fehler haben, weil eines dem anderen, im Sinne einer seriellen Korrelation, folgt. Das Fehlen dieses korrelierten Fehlers in einer CFA würde sich negativ auf jeden globalen Fit-Index auswirken. Außerdem sagen die globalen Fit-Indizes nichts über die spezifischen Fehlspezifikationen des Modells aus, sondern sind eher dichotome Instrumente (Saris et al., 2009). So schlagen Saris et al. vor, die **erwartete Parameteränderung** (Expected Parameter Change = EPC) in Kombination mit dem **Modifikationsindex (MI)** und der **Power** des MI-Tests zu verwenden (Saris et al., 2009, S. 20–21). Dafür wird dem **dargestellten Analysepfad von Uanhoro** (2017) gefolgt (Abbildung 2), d.h.: Zu Beginn wird die Power festgestellt. Ist die Power unter 75%, wird der MI ausgewertet. Ist dieser nicht signifikant, ist das Modell nicht aussagekräftig. Ist der MI signifikant, handelt es sich um ein fehlspezifiziertes Modell. Hinweisend auf Saris et al. ist der MI bei einem Signifikanzlevel von $\alpha = 0.05$ erst dann signifikant, wenn er auch Werte über 3.84 aufweist. Durch diese Vorgehensweise wurde die Abschätzung der inhaltlichen Größe und Richtung möglicher Abhängigkeiten ermöglicht. Der Zusammenhang zwischen EPC und MI wird von Uanhoro (2017) beschrieben: $MI = (EPC/\sigma)^2$. Ist der EPC über der Akzeptabilitätsschwelle, handelt es sich um eine Fehlspezifikation (epc:m wird indiziert). Unterhalb der Akzeptabilitätsschwelle liegt keine

Fehlspezifikation vor (epc:nm). Diesbezüglich empfehlen Saris et al. für Faktorladungen einen Wert > 0.4, für korrelierte Fehler, > 0.1 sowie die Berechnung eines Nicht-Zentralitäts-Parameter (ncp), $ncp = (\delta/\sigma)^2$. Dieser ncp folgt einer nicht-zentralen- χ^2 -Verteilung, die verwendet wird, um die statistische Power zu berechnen.

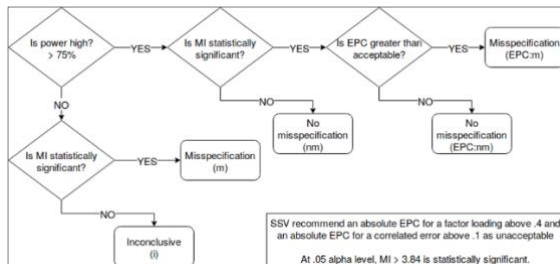


Abbildung 2: Entscheidungsbaum für Fehlspezifizierungen

In der **konkreten Auswertung** wird die Itembatterie Zufriedenheit isoliert ausgewertet, also C001_01 – C001_10 (Abbildung 1). Item C001_08 wurde aufgrund der inversen Polung rekodiert und C001_09 wurde aufgrund fehlender Augenscheinvalidität ausgeschlossen. Zunächst werden Cronbachs Alpha (α) und danach die Korrelationsmatrix zwischen den Items ausgewertet, um Hinweise auf Reliabilität zu und Multikollinearität zu erhalten. Anschließend wird durch Prüfung von Schiefe und Kurtosis auf Abweichung von der Normalverteilung analysiert, bevor das Modell spezifiziert, geschätzt und inklusive der Modifikationsindizes evaluiert wird.

Ergebnisse

Stichprobengröße: Analysiert wurden n = 1350 vollständig vorliegende Antworten der Itembatterie „Zufriedenheit“. Cronbachs Alpha (unstandardisiert): Die Itembatterie weist ein $\alpha = .813$ auf. Verteilung: Alle Items weisen eine Schiefe nahe null auf, abgesehen von Item C001_06 (-1.39). Bei Kurtosis weisen alle Items

ein Werte nahe 3 auf, abgesehen von C001_06 (4.95) und C001_08 (1.75). Korrelationen/Varianzen: Es liegen keine Korrelationen > 0.7 und keine negativen Varianzen vor. Abbildung 3 zeigt die Korrelationen als Heatplot-Grafik.

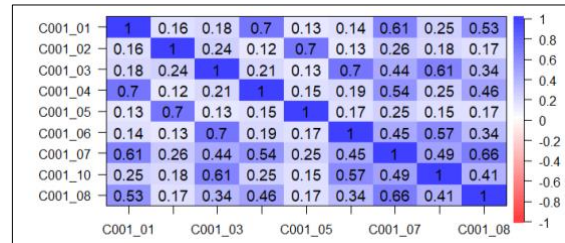


Abbildung 3: Korrelationen

CFA: Das Modell konvergiert nach 44 Iterationen normal (n=1350). Tabelle 1 zeigt die Fit Indizes.

Tabelle 1: Fit Indizes der CFA

Fit Index	Ausprägung	Trennwerte
RMSEA	0.131	< 0.06
SRMR	0.068	< 0.09
CFI	0.915	> 0.95
TLI	0.854	

Faktorladungen: Die Faktorladungen sind Abbildung 4 zu entnehmen. Dabei bezieht sich „KH“ auf den Ausbildungsort Krankenhaus, „RW“ auf die Rettungswache und „Ges. Ausb.“ erfasst die Items, die keinem spezifischen Lernort zugerechnet werden können und die gesamte Ausbildung betreffen.

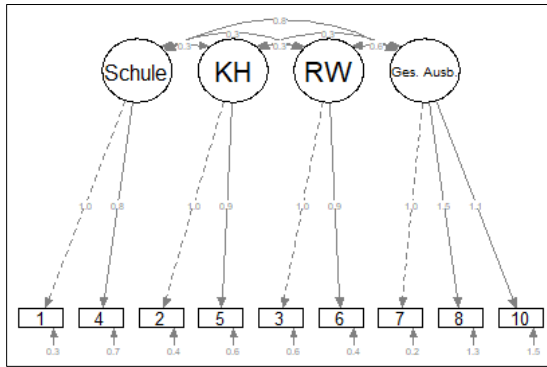


Abbildung 4: Ergebnisse der CFA

Fehlspezifizierungen: Es wurden Modellfehlspezifizierungen nach der von Uanhoro (2017) vorgestellten Methode berechnet (Abbildung 2). Tabelle 2 zeigt die fehlspezifizierten Itempaare auf Basis korrelierter Fehler.

Tabelle 2: Ergebnisse der Prüfung auf Fehlspezifikation - Fehlspezifizierte Items

Links	Rechts	MI	EPC	Power
C001_02	C001_03	87,44	0,23	0,99
C001_03	C001_10	83,58	0,30	0,87
C001_05	C001_03	61,95	-0,19	0,99
C001_02	C001_06	54,32	-0,15	1,00
C001_07	C001_08	37,58	0,22	0,80
C001_05	C001_06	33,58	0,12	1,00
C001_06	C001_10	31,40	0,15	0,96
C001_03	C001_07	28,98	-0,11	1,00
C001_01	C001_10	18,90	-0,13	0,91
C001_07	C001_10	16,94	-0,11	0,97

Links = Linke Seite der Gleichung, Rechts = Rechte Seite der Gleichung, MI = Mod Indizes, EPC = Expected Parameter Change

Diskussion

Zusammenfassend wurden die Itembatterie „Zufriedenheit“ der Berufsstreuestudie 2, die insgesamt aus zehn Items besteht, faktoranalytisch durch die Anwendung einer konfirmatorische Faktorenanalyse untersucht.

Bei der **Stichprobengröße** ist die statistisch Power zu berücksichtigen, also die Wahrscheinlichkeit, die Nullhypothese zurückzuweisen, wenn sie falsch ist. Verweisend auf die spezifischen Desiderate von Wolf et al. 2013, ist bei einer Stichprobengröße von $n = 1340$ bei dem hier vorliegenden Modell von einer ausreichenden Teststärke der CFA auszugehen.

Begreift man **Reliabilität** als Voraussetzung von Validität, liefert Cronbachs Alpha ($\alpha = 0.8$) als Indiz einer guten internen Konsistenz bereits eine Legitimation zu Durchführung einer Validitätsprüfung.

Die Ergebnisse der **Faktorenanalyse**, konkret gesagt die Fit Indizes, weisen auf Probleme bei der Modellanpassung hin. So zeigt der RMSEA (0.131) eine unzureichende Anpassungsgüte des vorliegenden Modells in Kontrast zu einem perfekten Modell. Der SRMR (0.7) liegt nur wenig unterhalb des Trennwerts. Geht man über die reine Dichotomie des Trennwerts hinaus, bemerken Backhaus et al. (2015, S. 151), dass Werte unter 0.5 als gut zu bezeichnen sind, was den vorliegenden SRMR auch in kritischerem Licht erscheinen lässt. Auch der TLI (0.92) und CFI (0.86) zeigen eine unzureichende Passung des Modells, denn der Trennwert wurde nicht erreicht.

Die Prüfung der **Fehlspezifizierungen** auf Itemebene (Tabelle 2) erlaubt eine Lokalisierung der ursächlichen Validitätsprobleme. Basierend auf dem Entscheidungsbaum aus Abbildung 2 werden

die fehlspezifizierten Itempaare identifiziert, die alle auf einen unzureichenden EPC zurückgehen. Diese sind als problembehaftet zu betrachten, wenn ein valides Modell zustande kommen soll. Aufgrund der Vielzahl der fehlspezifizierten Itempaare wird auf eine elaborierende Ausführung der Itemanalyse verzichtet.

Reflektiert man die Ergebnisse der CFA inklusive der Fehlspezifizierungen, kann dem vorliegenden Modell keine Validität unterstellt werden. Dies führt dazu, dass die Itembatterie sich nicht zu Analysen eignet, die über eine Beschreibung auf Itemebene hinausgehen. Bevor eine Verwendung als psychometrisches Instrument stattfinden kann, ist eine grundlegende Überarbeitung notwendig. Um den Charakter einer Ad-hoc-Skala zu überwinden, ist zunächst die Erstellung einer Grobstruktur erforderlich, indem Fragen- bzw. Itemblöcke erstellt werden, die jeweils ein latentes Merkmal messen (Döring et al., S. 405–408). Weitere Aspekte der Berufstreuestudie können ebenfalls psychometrisch erfasst werden, beispielweise Verlassintentionen. Im Rahmen dieser Grobstrukturierung können einzelne Itemvorschläge gesammelt und danach mit Expert:innen weiter operationalisiert werden. Diese Expert:innen sollten iterative Partizipation erfahren und zumindest bei wichtigen Meilensteinen der Entwicklung eingebunden werden, um Inhaltsvalidität zu gewährleisten (Boateng et al., 2018, S. 6). In der Feinkonzeption sind Art der Items und Antwortformate sowie Reihenfolge der Items, ggf. notwendige Filterführung sowie das Layout zu berücksichtigen. Diskutiert werden kann weiterhin, ob sich die Studie auch zukünftig ausschließlich an Auszubildende richten soll

oder das Scope um die bereits berufstätigen Notfallsanitäter:innen erweitert werden soll. Ein Pretest, der sog. kognitive Interviews umschließt (Boateng et al., 2018, S. 7–8), aber auch mit einer exploratorischen Faktorenanalyse ausgewertet werden kann, bevor er in der Erhebungsphase eingesetzt wird, ist empfehlenswert. Eine grundlegende iterative Anordnung der Prozessschritte ist außerdem zu erwägen.

Die **Stärke** dieser Validierungsstudie liegt in ihrer erheblichen Stichprobengröße, die entsprechende Teststärke erwarten lässt (Wolf et al., 2013). **Limitierend** wirkt sich die Struktur der Itembatterie an sich aus, genauer gesagt, dass ein Faktor lediglich aus zwei Items besteht. Obwohl es keine allgemeingültige Aussage gibt, wieviel Items pro Faktor der verbindliche Minimalstandard sein müssen, lässt sich generell sagen, dass mehr Faktoren mit weniger nicht übereinstimmenden Faktorergebnissen, weniger fehlerhaften Ergebnissen, besserer Interpretierbarkeit, genauerer und stabilerer Parameterschätzungen und zuverlässigeren Faktoren assoziiert sind (Marsh et al., 1998, S. 18). Aufgrund der nicht-probabilistischen Stichprobe kann eine Stichprobenverzerrung durch Selbstselektion trotz sorgfältiger und umfangreicher Rekrutierungsmaßnahmen nicht ausgeschlossen werden.

In diesem Kontext lässt sich festhalten, dass die analysierte Itembatterie zwar deskriptiv auf Basis der Aussagekraft der Einzelitems Verwendung finden kann, jedoch nicht als psychometrisches Konstrukt zu verstehen ist, welches Zufriedenheit bzw. Berufszufriedenheit erklären kann. Auf die BTS 2 übertragen bedeutet dies, dass die einzelnen Items zur deskriptiven Analyse auf ihre jeweilige

Fragestellung begrenzt herangezogen werden können. Das folgende Beispiel anhand Item C001_01 („Mit meiner theoretischen Ausbildung in der Berufsfachschule bin ich sehr zufrieden“) illustriert die Aussage: **Der Analyse des Einzelitems und der Ableitung der Zufriedenheit zur theoretischen Ausbildung anhand des Mittelwerts steht nichts entgegen.** Es wäre jedoch problematisch, die Items zu einer Aussage wie „Die Berufszufriedenheit der Auszubildenden im Beruf Notfallsanitäter ist gut“, zusammenzuführen, da die faktorielle Validität nicht angenommen werden kann, wie diese Studie gezeigt hat.

Fazit für die Praxis

Es spricht nichts gegen die deskriptive Analyse der einzelnen Items im Rahmen ihrer jeweiligen Aussagekraft, die bei jedem Item für sich selbst steht. Die Itematterie ist jedoch nicht in der Lage, Zufriedenheit als psychometrische Skala, die ein latentes Konstrukt repräsentiert, zu reflektieren. Es empfiehlt sich bei einer Neuauflage der Berufstreuestudie, dass der Charakter einer Ad-Hoc-Skala durch eine Neukonzeption überwunden wird, die als psychometrische Skala zu verstehen ist. Dazu ist ein iteratives Vorgehen zu berücksichtigen, das Expert:innen miteinbezieht und ausführlichen Raum zum Pretesting gibt.

Ethische Richtlinien:

Interessenskonflikt: Beide Autor:innen sind auch an der Publikation der BTS 2 als Autor:innen beteiligt. Autor 1 war für Analyse der Daten und die Manuskripterstellung zuständig. Autor 2 war für die Beantwortung spezifisch-inhaltlicher Fragen zur Berufstreuestudie zuständig.

Ethikvotum: Durch die Ethikkommission der HSD Hochschule Döpfer wurde der BTS 2 ethische Unbedenklichkeit bescheinigt (Aktenzeichen BEth_31_22 vom 11.08.2022).

Korrespondierende*r Autor*in:

Christian Elsenbast

Fraunhofer-IESE, Research Program Digital Healthcare, Kaiserslautern

christian.elsenbast@iese.fraunhofer.de

Artikel frei zugänglich unter
<https://doi.org/10.25974/gjops.v1i1.27>

Eingereicht: 2024-03-10

Angenommen: 2024-05-14

Veröffentlicht: 2024-08-09

Copyright

© 2024 Christian Elsenbast. Dieser Artikel ist ein Open Access-Beitrag und wird unter den Bedingungen der [Creative Commons-Lizenz BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) veröffentlicht.

Literatur

- Backhaus, K., Erichson, B. & Weiber, R. (2015). Fortgeschrittene Multivariate Analysemethoden. Springer Berlin Heidelberg.
- Bentler, P. M [P. M.] & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606. <https://doi.org/10.1037/0033-2909.88.3.588>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R. & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6. <https://doi.org/10.3389/fpubh.2018.00149>
- Döring, N., Bortz, J. & Pöschl, S. *For-schungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5., vollst. überarb., aktualisierte und erw. Aufl.). Springer.
- Eisenbast, C., Böhm, D. & Hofmann, T. (2024). Deskriptive Analyse der zweiten Umfrage zur Berufstreue angehender Notfallsanitäter:innen (Berufstreuestudie 2): Eine Trendstudie“ [under review].
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Springer Berlin Heidelberg.
- Hu, L. & Bentler, P. M [Peter M.] (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jöreskog, K. G. (1999). How Large Can a Standardized Coefficient be? <https://cscar.github.io/misc/HowLargeCanaStandardizedCoefficientbe.pdf>* [Zuletzt geprüft am 04.03.24].
- Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198–211. <https://doi.org/10.1080/0969594X.2015.1060192>
- Knekta, E., Runyon, C. & Eddy, S. (2019). One Size Doesn't Fit All: Using Factor Analysis to Gather Validity Evidence When Using Surveys in Your Research. *CBE life sciences education*, 18(1), rm1. <https://doi.org/10.1187/cbe.18-04-0064>
- Marsh, H. W., Hau, K. T., Balla, J. R. & Grayson, D. (1998). Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis. *Multivariate behavioral research*, 33(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1
- Rosseel, Y. (2012). lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Saris, W. E., Satorra, A. & van der Veld, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), 561–582. <https://doi.org/10.1080/10705510903203433>
- Uanhoro, J. (2017). Misspecification and fit indices in covariance-based SEM. <https://www.jamesuanhoro.com/post/2017/10/28/misspecification-and-fit-indices-in-covariance-based-sem/>* [Zuletzt geprüft am 03.03.2024].
- Wolf, E. J., Harrington, K. M., Clark, S. L. & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and psychological measurement*, 76(6), 913–934. <https://doi.org/10.1177/0013164413495237>