

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/213848>

Please be advised that this information was generated on 2020-04-03 and may be subject to change.

Ein Vergleich traditioneller und computergestützter Methoden zur Erstellung einer deutschsprachigen Need for Cognition Kurzskala

Hanna Beißert¹ , Meike Köhler², Marina Rempel³ und Peter M. Kruyen⁴

¹DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

²Comma Soft AG, Bonn

³Universität zu Köln

⁴Institute for Management Research, Radboud University Nijmegen

Zusammenfassung: Die vorliegende Arbeit dient der Entwicklung einer Kurzskala zur Messung von Need for Cognition (NFC). Neben traditionellen Verfahren der Itemreduktion auf Basis von Itemkennwerten der klassischen Testtheorie wurde in Studie 1 ($N = 282$) ein neues, computergestütztes Verfahren des „full information approach“ vorgestellt. Mithilfe der beiden Verfahren wurden 3 mögliche Skalen mit jeweils 5 Items selektiert, welche in einem unabhängigen Datensatz in Studie 2 ($N = 530$) Validierungskriterien unterzogen wurden. Aus den 3 Skalen mit ähnlichen Ergebnissen bezüglich der Gütekriterien wurde eine anhand der „full information approach“ erstellten Skalen als finale Kurzskala NFC-K ausgewählt. Diese NFC-K weist ein Cronbachs α von .69 auf, eine Korrelation mit der Langskala von .81 und keinen signifikanten Zusammenhang mit sozialer Erwünschtheit. Neben der systematischen und objektiven Selektion und Validierung der Kurzskala NFC-K stellt die vorliegende Arbeit auch eine Fallstudie zu den Herausforderungen der Kurzskalenentwicklung auf Basis klassischer und computergestützter Selektionsverfahren dar.

Schlüsselwörter: Need for Cognition, Kurzskala, Itemselektion, Full Information Approach

A Comparison of Traditional and Computer-Assisted Methods for the Development of a German Need for Cognition Short Scale

Abstract: The current study presents the development and validation of a Need for Cognition (NFC) short scale. In Study 1 ($N = 282$), traditional item selection procedures based on classic test theory were used as well as an innovative computational approach, our “full information approach.” The procedures led to three different short scales with comparable psychometric quality, which were validated in Study 2 ($N = 530$). Based on different validation criteria, one of the three scales obtained from the full information approach was selected as the final short scale NFC-K. This NFC-K achieved a Cronbach’s α of .69, a correlation of .81 with the original scale, and showed no significant correlation with social desirability. Besides presenting a systematic and objective selection and validation of the NFC-K, this article represents a case study of the challenges of developing a short scale comparing both traditional and computational approaches.

Keywords: Need for Cognition, short scale, item selection, full information approach

Das Konstrukt Need for Cognition (NFC) beschreibt interindividuelle Unterschiede im Engagement und in der Freude bei Denkaufgaben (Cacioppo & Petty, 1982). Dieses überdauernde Persönlichkeitsmerkmal ist von Relevanz in vielen Teildisziplinen der Psychologie. Eine ökonomische Erfassung des Konstrukts anhand einer Kurzskala ist deshalb von großem Interesse. Ziel der vorlie-

genden Studie war deshalb die Entwicklung und Validierung einer Fünf-Item-Kurzskala zur Messung von NFC. Dabei wurden zur Entwicklung klassische Verfahren der Itemreduktion einem computergestützten¹ automatisierten Reduktionsverfahren (*full information approach*) gegenübergestellt.

¹ Mit „computergestützt“ ist in diesem Zusammenhang ein automatisiertes Absuchen des Raumes aller möglichen Subskalen sowie das Berechnen verschiedener Optimierungskriterien pro Subskala gemeint.

Die Idee eines Persönlichkeitskonstrukts, das die Freude am Denken darstellt, entstand schon in der Sozialpsychologie der 40er- und 50er-Jahre des vergangenen Jahrhunderts und ist in Arbeiten von Maslow (1943), Murphy (1947), Asch (1952) sowie Sarnoff und Katz (1954) zu finden. In den 1980er Jahren setzten sich Petty, Cacioppo und Kollegen intensiv mit dem Konstrukt NFC auseinander (Cacioppo & Petty, 1982; Cacioppo, Petty, & Morris, 1983; Cacioppo, Petty & Kao, 1984; Haugtvedt, Petty, Cacioppo & Steidley, 1988) und prägten auch die bis heute gebräuchliche Definition von NFC: „Need for Cognition refers to an individual’s tendency to engage in and enjoy effortful cognitive endeavors.” (Cacioppo et al., 1984, S. 306).

Notwendigkeit einer NFC-Kurzskala

NFC nimmt in vielen Teildisziplinen der Psychologie wie der Sozial-, Motivations-, Personal-, Gesundheits- und Pädagogischen Psychologie, einen wichtigen Platz in der Erforschung und der Vorhersage von Verhalten ein (Cavazos & Campbell, 2008; Chen & Wu, 2012; Fleischhauer et al., 2010; Haugtvedt et al., 1988; Mussel, Spengler, Litman & Schuler, 2012). Große Relevanz hat das Konstrukt NFC im Bereich der Sozialpsychologie im Rahmen des Elaboration Likelihood Models (ELM; Petty & Cacioppo, 1986), in dem es als wichtiger Moderator fungiert. Dieses Modell der Verarbeitung persuasiver Informationen nimmt eine zentrale Rolle in der sozialpsychologischen Entscheidungs- und Einstellungsforschung der Medienwirkungsforschung ein. Damit kann davon ausgegangen werden, dass NFC Informationsverarbeitung, Werbewahrnehmung, oder Entscheidungsverhalten beeinflusst (vgl. Pechtl, 2009). Weiterhin ist das Konstrukt in der Gesundheitspsychologie von wesentlicher Bedeutung, da NFC mit geringerer Ängstlichkeit und niedrigerem negativen Affekt assoziiert ist, was wiederum mit Wohlbefinden und Lebenszufriedenheit in Zusammenhang gebracht werden kann (Cavazos & Campbell, 2008; Cacioppo, Petty, Feinstein & Jarvis, 1996). Durch den Zusammenhang von NFC mit leistungsrelevanten oder motivationalen Variablen hat das Konstrukt auch Relevanz in der pädagogisch-psychologischen Forschung. Beispielhaft seien hier der Zusammenhang von NFC mit Erfolgserwartungen (Reinhard & Dickhäuser, 2009), mit dem Umgang mit Feedback (Anseel, Lievens & Schollaert, 2009) und mit intrinsischer Motivation in kognitiven Bereichen (Olson, Camp & Fuller, 1984) sowie der Zusammenhang von NFC mit berufsbezogener Kreativität (Chen, Himsel, Kasof, Greenberger & Dmitrieva, 2006; Curşeu, 2011; Hahn & Lee, 2016) genannt. Doch auch außerhalb der psychologischen Forschung zeigen sich Anwendungs-

gebiete für eine Erfassung des Konstrukts NFC. Beispielsweise im Bereich der Management Studies ist das Konstrukt NFC von Interesse. Darüber hinaus könnten Mitarbeiterinnen und Mitarbeiter, die eine größere Freude an schwierigen Aufgaben haben und dazu neigen, ihre Urteile eher auf empirischen Informationen und rationalen Erwägungen zu gründen, im Fokus der Personalauswahl stehen (Cacioppo et al., 1996).

Da NFC in vielen verschiedensten Teilgebieten der Psychologie ein relevantes Konstrukt darstellt, ist eine reliable, valide und effiziente Erfassung des Konstrukts von großem Interesse. Zur Messung von NFC liegt im deutschsprachigen Raum die Skala von Bless, Wänke, Bohner, Fellhauer und Schwarz (1994) vor. Es handelt sich dabei um eine Übersetzung der Originalskala von Cacioppo und Petty (1982). Die deutsche Fassung liegt in einer Langform mit 33 Items vor sowie in einer kürzeren Variante, die aus 16 Items besteht. Beide Varianten wurden von Bless et al. (1994) erfolgreich validiert. Vor allem die 16-Item-Kurzform ist in der deutschsprachigen Forschung stark verbreitet. Wie oben geschildert kommt das Konstrukt NFC in den verschiedensten Forschungsfeldern zur Anwendung. Häufig handelt es sich um Studien, in denen eine ganze Reihe verschiedener Variablen erfasst werden. Für eine solche Verwendung weist aber selbst die effizientere 16-Item-Form der NFC-Skala von Bless et al. (1994) noch eine problematische Gesamtlänge im Hinblick auf die Zumutbarkeit für die Befragten sowie die Erhebungskosten auf. Deshalb erscheint es äußerst sinnvoll und notwendig, eine Kurzskala für NFC zu entwickeln, die aus nur wenigen, sorgfältig ausgewählten Items besteht, um das Konstrukt effizient und unter Sicherung der psychometrischen Gütekriterien zu erfassen. In einer Vorarbeit von Beißert, Köhler, Rempel und Beierlein (2014), wurde bereits ein erster Versuch unternommen, eine solche Kurzskala zu entwickeln. Da sich aber die dort zugrunde gelegte Zweifaktorlösung als fraglich erwies, ist es Ziel der vorliegenden Studien, an diese Vorarbeit anzuknüpfen und eine systematische Kurzskalenentwicklung auf Basis bewährter sowie innovativer Selektionsmethoden zu präsentieren.

Möglichkeiten der Itemreduktion

Zur Erstellung von Kurzskalen aus einem Pool bestehender Items stehen verschiedene Vorgehensweisen zur Verfügung. In traditionellen Ansätzen zur Itemselektion gehen neben inhaltlichen Aspekten vorrangig Itemkennwerte der klassischen Testtheorie mit ein. So ist die Auswahl von Items anhand von Kennwerten wie Itemschwierigkeit, Itemvarianz und Itemtrennschärfe (Moosbrugger & Kelava, 2012) oder auch Faktorladungen üblich (Beier-

lein, Kemper, Kovaleva & Rammstedt, 2013; Schipolowski et al., 2014; Rammstedt & John, 2005; Klingsieck & Fries, 2012). Zusätzlich gehen hier meist auch inhaltliche Kriterien bezüglich der Itemformulierung wie zum Beispiel das Vermeiden langer, komplexer Sätze oder Items mit multiplen Stimuli mit ein.

Diesem klassischen Ansatz stehen computergestützte Ansätze gegenüber, die in den vergangenen Jahren zunehmend Anklang gefunden haben. Grundidee ist hierbei, dass eine Vielzahl möglicher Subskalen erstellt wird und aus diesen eine optimale Kurzsкала auf Basis quantifizierbarer Gütekriterien selektiert wird. Als Kriterien zum Identifizieren einer optimalen Kurzsкала können hierbei neben Kennwerten der klassischen Itemanalyse und Korrelationen mit der entsprechenden Langform auch eine Vielzahl weiterer Kriterien hinzugezogen werden. Der Fokus der aktuellen Arbeit liegt auf der Abdeckung verschiedener Validitätskriterien, um das Konstrukt NFC möglichst breit abzudecken und eine optimale Kurzsкала – gegebenenfalls über die Güte der Langform hinaus – zu ermöglichen. Diese Kriterien umfassen zum einen Aspekte der Konstruktvalidität, wie sie durch Modellfitstatistiken aus Strukturgleichungsmodellen erfasst werden, und zum anderen Aspekte der konvergenten und divergenten Validität, in den Korrelationen mit verwandten Konstrukten berechnet werden. Die Korrelationen der reduzierten Kurzsкала mit den relevanten Konstrukten können anschließend mit den Korrelationen der Langform verglichen werden (Kruyen, Emons & Sijtsma, 2013). Je nach Konstrukt können darüber hinaus auch weitere Kriterien, wie z. B. die Messinvarianz, zum Finden einer optimalen Kurzsкала miteinfließen. Das Testen von Messinvarianz ist üblich, um die Vergleichbarkeit einer Skala in verschiedenen Gruppen, wie z. B. in interkulturellen Vergleichsstudien oder zu verschiedenen Messzeitpunkten, z. B. in Paneldaten, sicherzustellen (Van De Schoot, Schmidt, Beuckelaer, Lek & Zondervan-Zwijnenburg, 2015). Darüber hinaus sind weitere Kriterien denkbar. Die Auswahl von Optimierungskriterien stellt eine Abwägungsentscheidung und Schwerpunktsetzung dar, basierend auf theoretischen und empirischen Aspekten des Konstrukts.

Besteht die ursprüngliche Langform der Skala aus einer großen Anzahl von Items, ist eine exhaustive Suche, also das Berechnen aller möglichen Subskalen, aufgrund des hohen Rechenaufwands nicht möglich. Hier können Algorithmen der kombinatorischen Optimierung, wie zum Beispiel der Ameisenalgorithmus (Leite, Huang & Marcoulides, 2008) oder evolutionäre Algorithmen (z. B. Sandy,

Gosling & Koelkebeck, 2014) Abhilfe schaffen. In diesen Verfahren wird nur ein Teil aller möglichen Subskalen getestet. Zwar zielen diese Verfahren darauf ab, trotz der Einschränkung des Suchraums die optimale Kurzsкала zu finden, dennoch besteht die Unsicherheit, ob von allen möglichen Itemkombinationen auch wirklich die beste Kurzsкала selektiert wurde. In der vorliegenden Studie wurde deshalb ein *full information approach* genutzt, in dem alle möglichen Subsets von Items miteinander verglichen werden. Durch die Nutzung des reduzierten Itempools der etablierten 16-Item NFC Skala als Ausgangsbasis der Selektion war diese exhaustive Suche möglich.

Entwicklung der NFC-Kurzsкала

In der vorliegenden Arbeit wird die systematische und objektive Entwicklung einer deutschsprachigen NFC-Kurzsкала auf Basis bewährter und innovativer Selektionsmethoden präsentiert. Damit sollen mögliche Fallstricke der Kurzsкаlenentwicklung, wie beispielsweise das Verwenden unzureichender Validierungskriterien, Rechnung getragen werden (Smith, McCarthy & Anderson, 2000; Ziegler, Kemper & Kruyen, 2014). Hierbei wird zunächst in Studie 1 die Kurzsкаlenselektion anhand klassischer Verfahren der Itemselektion und des *full information approach* beschrieben. Die resultierenden, neu entwickelten Kurzsкаlen werden in Studie 2 validiert und bezüglich ihrer Güte bewertet. Darauf basierend wird in der abschließenden Diskussion eine finale Empfehlung einer besten NFC-Kurzsкала im Hinblick auf Reliabilitäts- und Validitätskriterien präsentiert.

Ziel der vorliegenden Arbeit ist somit neben der Präsentation einer neu entwickelten NFC-Kurzsкала auch ein Vergleich verschiedener Vorgehensweisen (klassisch vs. *full information approach*) bei der Kurzsкаlenerstellung. Der Fokus der Darstellung liegt hierbei auf der transparenten Präsentation aller relevanten Schritte in der Selektion, Validierung und Bewertung der Kurzsкала. All diese Schritte können reproduziert und nachvollzogen werden mit Hilfe des Code Supplements im elektronischen Supplement (ESM) 2, in dem sich die anonymisierten Daten von Studie 1 und 2 sowie alle notwendigen Analyse-Skripte befinden.²

² Das Code Supplement ist unter <https://github.com/Meike-K/nfc-k> verfügbar. Eine Verlinkung dorthin findet sich auch im ESM 2. Es beinhaltet neben den Rohdaten alle R-, SPSS- und Mplus-Skripte, um die berichteten Analysen und Ergebnisse nachzuvollziehen und bietet darüber hinaus die Möglichkeit, weitere Analysen durchzuführen.

Studie 1: Kurzskalenentwicklung

In Studie 1 wurden anhand eines Online-Fragebogens die validierte 16-Item-Skala von Bless et al. (1994) sowie verschiedene Konstrukte, die Zusammenhänge mit NFC aufweisen, erhoben. Basierend auf dieser Stichprobe wurden zum einen mit klassischen Verfahren zur Itemselektion und zum anderen anhand des von uns vorgeschlagenen full information approach verschiedene potenzielle Kurzskalen aus der 16-Item Form ausgewählt, die in Studie 2 einer Validierung unterzogen wurden.

Selektionskriterien

Die Itemselektion erfolgte zum einen anhand klassischer Itemkennwerte und zum anderen zur Optimierung von Reliabilitäts-, Modellanpassungs- sowie Validitätskriterien, um verschiedene potenzielle Kurzskalen zu ermitteln.

Itemselektion anhand klassischer Itemkennwerte

Zur Itemselektion auf Basis klassischer Itemkennwerte wurden Schwierigkeit, Varianz und Trennschärfe sowie eine exploratorische Faktorenanalyse berechnet. Ziel war die Auswahl von Items mit mittlerer Schwierigkeit, hoher Varianz, hoher Trennschärfe sowie einer hohen Ladung auf dem Hauptfaktor der unrotierten Faktorlösung. Daneben wurden inhaltliche Überlegungen herangezogen, um den bestehenden Itempool auf eine möglichst geringe Anzahl an Items zu reduzieren. Um eine Repräsentation der beiden Facetten von NFC, des Engagements und der Freude bei Denkaufgaben, zu gewährleisten, wurde die Auswahl von insgesamt mindestens vier Items angestrebt.

Full information approach

Im full information approach wurden für mögliche Kurzskalen aus der 16-Item Form Reliabilitäts-, Validitäts- und Modellfit-Kriterien berechnet. Konkret wurden aus dem Pool der 16 Items alle möglichen Kurzskalen berechnet, die zwischen drei und sieben Items enthielten und damit 26 196 potenzielle Kurzskalen gebildet. Für jede dieser Testskalen wurde Cronbachs α , der Greatest Lower Bound (Jackson & Agunwanba, 1977), ω (Brunner, Nagy & Wilhelm, 2012) sowie Fitindizes (χ^2 , CFI, TLI, RMSEA) aus einer konfirmatorischen Faktorenanalyse berechnet. In der konfirmatorischen Faktorenanalyse luden alle Items

auf einem Hauptfaktor. Neben der einfachen konfirmatorischen Faktorenanalyse (CFA_{unkorr}), wie sie häufig zur Darstellung des Konstrukts verwendet wird, wurde zudem ein Modell (CFA_{korrr}) berechnet, in dem die Fehlerterme der negativ formulierten Items korreliert wurden. Damit wurden neuere Ergebnisse zur Faktorstruktur der NFC und dem Einfluss negativ formulierter Items miteinbezogen (Forsterlee & Ho, 1999; Hevey et al., 2012). Dabei wurden alle Subskalen ausgeschlossen, deren CFA_{korrr} durch die Anzahl der korrelierten Fehlerterme nicht identifizierbar war. Daneben wurden für alle Testskalen die Korrelationen zu den erhobenen Konstrukten berechnet. Hierbei wurde eine hohe mittlere Korrelation zu den verwandten Konstrukten Präferenz für Deliberation, Typisches Intellektuelles Interesse, Lernzielorientierung und Offenheit für neue Erfahrung angestrebt sowie eine Minimierung des Zusammenhangs mit sozialer Erwünschtheit. Die mittlere Korrelation wurde als das arithmetische Mittel der Korrelationskoeffizienten berechnet. Zudem wurden diese Korrelationen mit der Höhe des Zusammenhangs zwischen Langskala und den verwandten Konstrukten verglichen.

Bei dieser Anwendung des full information approach standen die konvergente und divergente Validität im Vordergrund, um unter Gewährleistung akzeptabler Reliabilitäts-Maße eine breite inhaltliche Abdeckung des Konstrukts NFC mittels weniger Items der etablierten Skala zu erzielen.

Darüber hinaus wurden weitere Kennwerte berechnet, um eine umfangreiche Bewertung der Kurzskalen zu ermöglichen. So wurde die Korrelation mit der Langform berechnet, in die Selektion jedoch nur nachrangig zur Qualitätssicherung miteinbezogen. Zudem wurden in einer Post-Hoc Sensitivitätsanalyse konfigurale, metrische und skalare Invarianz getestet bezüglich des Bildungslevels (hoch, mittel, niedrig) und bezüglich des Geschlechts. Invarianz wurde bei nicht-signifikantem $\Delta\chi^2$ -Test sowie einem Δ CFI und Δ RMSEA < 0.01 (Chen, 2007) angenommen.

Zur Berechnung der konfirmatorischen Faktorenanalysen und Messinvarianz-Analysen kam die Software Mplus, Version 8 (Muthén & Muthén, 1998–2017) sowie die Software R, Version 3.5.1 (R Core Team, 2018) zum Einsatz. Mittels eines R-Skripts wurden alle Subskalen zur Modellschätzung mittels Maximum Likelihood (ML) Schätzung an Mplus übergeben und die Resultate innerhalb R verarbeitet. Als weitere Sensitivitätsanalyse wurde zudem ein Set der besten Kurzskalen mittels einer robusten ML-Schätzung (MLR) berechnet, um auch die Robustheit der Ergebnisse bezüglich des Schätzverfahrens zu untersuchen.

Methode

Stichprobe

An der Onlinestudie zur Selektion der Kurzsкаlen nahmen 344 Personen teil, die den Fragebogen bis zum Ende vollständig ausgefüllt haben. Die Daten von 62 Personen wurden von den Analysen ausgeschlossen, da auf Grund ihrer Bearbeitungszeit von unter drei Minuten Zweifel an der Ernsthaftigkeit der Antworten bestanden. Somit ergab sich eine Analysestichprobe von 282 Personen (63% weiblich). Die Teilnehmenden der Studie wurden über verschiedene deutschsprachige Onlineforen und Mailinglisten rekrutiert und waren durchschnittlich 27.7 Jahre alt ($SD = 9.5$). Signifikante Unterschiede zwischen männlichen ($M = 5.03$, $SD = 0.83$) und weiblichen Personen ($M = 4.96$, $SD = 0.91$) in Bezug auf NFC wurden nicht festgestellt ($t(280) = -.653$, $p = .514$). 4% der Personen hatten keinen Schulabschluss, 12% einen Volks-, Haupt- oder Realschulabschluss, 38% hatten (Fach-)Abitur, 45% einen (Fach-) Hochschulabschluss oder höher und 1% machte keine Angaben oder nannte einen nicht einzuordnenden Abschluss.

Ablauf und Inhalt der Erhebung

Die Studie wurde als Onlinefragebogen mit der Software Unipark von QuestBack programmiert und durchgeführt. Die Bearbeitung des Onlinefragebogens dauerte durchschnittlich 17 Minuten. Nach einer Begrüßung und Aufklärung der Teilnehmenden folgte eine kurze Einweisung und die Erhebung soziodemografischer Daten. Im Anschluss wurden relevante Konstrukte zur Erfassung der konvergenten und divergenten Validität potenzieller NFC-Kurzsкаlen bearbeitet. Alle verwendeten Skalen hatten ein siebenstufiges Antwortformat (von 1 = „trifft überhaupt nicht zu“ bis 7 = „trifft voll und ganz zu“). Die erfassten Konstrukte in der Reihenfolge ihrer Erhebung werden im Folgenden vorgestellt.

Offenheit

Da angenommen wird, dass Personen mit höherem NFC neuen Ideen und Informationen offener entgegen treten, weil diese mit vermehrter kognitiver Aktivität einhergehen, wurde der Zusammenhang zwischen Offenheit für neue Ideen und NFC untersucht. In mehreren Forschungsarbeiten wurden diesbezüglich bereits positive Korrelationen zwischen der ursprünglichen NFC-Skala und Offenheit dokumentiert (Bertrams & Dickhäuser, 2010; Berzonsky & Sullivan, 1992; Sadowski & Cogburn, 1997).

Zur Erfassung des Persönlichkeitsmerkmals Offenheit für neue Erfahrungen wurden die 12 Items der entsprechenden Subskala des NEO-Fünf-Faktoren-Inventar nach Costa und Mc Crae (NEO-FFI; Borkenau & Ostendorf, 2008) verwendet. Da die Items des NEO-FFI für die Teil-

nehmenden relativ leicht zugänglich und weniger abstrakt als die meisten anderen Konstrukte der Erhebung sind, wurden sie als Einstieg in den Fragebogen verwendet.

Präferenz für Deliberation

Die Tendenz zur Deliberation beschreibt einen reflektiven Entscheidungsmodus (Betsch, 2004). Personen, die eher zu deliberativen als zu intuitiven Entscheidungen tendieren, beziehen auch Kognitionen stärker in ihre Urteile mit ein. Ein positiver Zusammenhang zwischen NFC und eine Präferenz für Deliberation wurde daher bereits festgestellt (Epstein, Pacini, Denes-Raj & Heier, 1996). Zur Erfassung dieser Eigenschaft kamen die neun Items der Subskala Deliberation aus dem Fragebogen *Präferenz für Intuition und Deliberation* (PID) von Betsch (2004) zum Einsatz.

Typisches Intellektuelles Engagement

Weiterhin wurde ein Zusammenhang zwischen dem Konstrukt des Typical Intellectual Engagement (TIE) und NFC angenommen. TIE ist teilweise aus der ursprünglichen NFC-Skala abgeleitet und geht davon aus, dass sich Personen in ihrem üblichen Engagement bei intellektuellen Aktivitäten unterscheiden (Wilhelm, Schulze, Schmiedek & Süß, 2003). Personen, die eine hohe Ausprägung des TIE aufweisen, werden vor allem durch ein breites Interesse und den Wunsch nach tieferem Verständnis von Sachverhalten charakterisiert. Da dieses Konstrukt Freude an kognitiver Aktivität voraussetzt und aus dem Konstrukt NFC abgeleitet wurde, wird auch hierbei ein positiver Zusammenhang mit NFC erwartet. Das TIE der Teilnehmenden wurde mithilfe von 18 Items des Fragebogens TIE (Wilhelm et al., 2003) erfasst.

Motivation

In früheren Arbeiten wurde postuliert, dass NFC mit Motivations- und Leistungskomponenten zusammenhängt (Chen & Wu, 2012). Im Fokus der vorliegenden Studie stand darauf basierend das Konstrukt der Lernzielorientierung im Zentrum, also der Motivation, eigene Fähigkeiten zu erweitern. Dieses Konstrukt ist eng verwandt mit intrinsischer Motivation zum Lernen (Spinath, Stiensmeier-Pelster, Schöne & Dickhäuser, 2012). Da Lernen zumeist mit Denken und kognitiver Beschäftigung verbunden ist, wurde ein positiver Zusammenhang zwischen Lernzielorientierung und NFC vermutet. Um die Lernmotivation der Teilnehmenden zu messen, wurde die aus acht Items bestehende Subskala Lernzielorientierung aus den *Skalen zur Erfassung der Lern- und Leistungsmotivation* (SELLMO, Spinath et al., 2012) verwendet.

Soziale Erwünschtheit

Des Weiteren gilt es sicherzustellen, dass die neu entwickelten Kurzsкаlen tatsächlich die Freude und das Engagement bei Denkaufgaben abbilden und nicht etwa die Selbstdarstellung als viel denkende Person. Aus diesem Grund wurde ein möglicher Zusammenhang mit dem Konstrukt der sozialen Erwünschtheit untersucht. Erfasst wurde die soziale Erwünschtheit mit der deutschen Form des *Balanced Inventory of Desirable Responding (Inventar zur Erfassung zwei Faktoren Sozialer Erwünschtheit)*; Musch, Brockhaus & Bröder, 2002). Anhand von zwei Subskalen mit je 10 Items werden die beiden Teilaspekte Selbst- und Fremdtäuschung erfasst.

Need for Cognition

Need for Cognition wurde mit dem 16-Item-Fragebogen von Bless et al. (1994) erfasst. Diese bereits reduzierte Form der englischen NFC-Langskala beschreibt eine gängige Form der Erfassung von NFC im deutschsprachigen Raum. Die neu entwickelten Kurzsкаlen sollten substanzial mit dieser etablierten Form zusammenhängen. Darüber hinaus dient die Erfassung von NFC anhand der 16-Item-Version dazu, die gefundenen Zusammenhänge der neuen Kurzsкаlen mit den beschriebenen Korrelaten vergleichen und einordnen zu können.

Ergebnisse

Deskriptive Ergebnisse

Alle erfassten Skalen wiesen eine für Persönlichkeitsmerkmale akzeptable interne Konsistenz (Cronbachs α) von mindestens .60 auf (siehe Anhang 2 in ESM 1). Die 16-Item Form wies die erwarteten positiven Zusammenhänge mit den verwandten Konstrukten Offenheit ($r = .40$, $p < .001$), Präferenz für Deliberation ($r = .29$, $p < .001$) Typisches Intellektuelles Engagement ($r = .58$, $p < .001$) und Lernzielorientierung ($r = .41$, $p < .001$) auf. Weiterhin zeigte sich ein Zusammenhang mit sozialer Erwünschtheit ($r = .21$, $p < .001$).

Itemselektion anhand klassischer Itemkennwerte

Bei der klassischen Itemselektion wurden in einem ersten Schritt alle Items in Hinblick auf ihre Formulierung und sprachliche Komplexität hin begutachtet. Daraufhin wurde Item 2 ausgeschlossen, da dieses Item aufgrund seiner Länge und Komplexität als kritisch eingestuft wurde („Ich würde lieber eine Aufgabe lösen, die Intelligenz erfordert, schwierig und bedeutend ist, als eine Aufgabe, die zwar irgendwie wichtig ist, aber nicht viel Nachdenken erfordert.“).

Im nächsten Schritt wurde eine Itemanalyse durchgeführt sowie eine explorative Faktorenanalyse mit einem

Hauptfaktor gerechnet. Für die Selektion der Items für die Need for Cognition-Kurzsкаla-1 (NFC-K-1) wurden Items mit einer Faktorladung $> .55$ und einer Trennschärfe $> .50$ miteinbezogen. Zudem sollten nicht mehr als drei Items invertiert sein und beide Facetten des Konstrukts, Freude und Engagement bei Denkaufgaben, abgebildet werden. Beispielsweise repräsentiert das Item „Die Aufgabe, neue Lösungen für Probleme zu finden, macht mir wirklich Spaß.“ die Facette Freude, das Item „Ich setze mir eher solche Ziele, die nur mit erheblicher geistiger Anstrengung erreicht werden können.“ hingegen die Facette Engagement. Bei der Abwägung der verschiedenen Kriterien wurden insgesamt fünf Items für die NFC-K-1 ausgewählt, da diese vergleichbar gute Kennwerte erzielten. Anhang 1 in ESM 1 gibt eine Übersicht über die Itemkennwerte und Faktorladungen, auf denen die Auswahl der Items nach klassischer Vorgehensweise basiert.

Selektion nach dem full information approach

Ziel des full information approach war es, alle möglichen Subsets von Items miteinander zu vergleichen. Aus allen berechneten potenziellen Testskalen wurden zunächst Kurzsкаlen mit vier oder fünf Items ausgewählt, da Subskalen mit weniger Items kaum zufriedenstellende Reliabilitätsergebnisse erzielten. Zudem wurden Skalen mit einem Anteil negativ formulierter Items von mindestens 25% und maximal 75%, Cronbachs $\alpha \geq .70$, Greatest Lower Bound $\geq .70$ sowie CFI ≥ 0.95 (für CFA_{korr} und CFA_{unkorr}) ausgewählt. Aus diesem Set wurden alle Testskalen, die Item 2 enthielten, ausgeschlossen, da dieses Item auf Grund seiner Länge und Komplexität als kritisch eingestuft wurde (siehe Abschnitt *Itemselektion anhand klassischer Itemkennwerte*). Zur finalen Selektion einer Kurzsкаla wurden die verbleibenden 104 Skalen bezüglich der Kriterien (a) RMSEA minimal (für CFA_{korr} und CFA_{unkorr}), (b) mittlere Korrelation mit den Skalen TIE, Präferenz für Deliberation, Offenheit für neue Erfahrung und Lernzielorientierung maximal, sowie (c) soziale Erwünschtheit minimal verglichen. Unter Berücksichtigung dieser Kriterien wurden zwei Kurzsкаlen mit je fünf Items zur weiteren Validierung ausgewählt (NFC-K-2 und NFC-K-3) mit einem Cronbachs α von .72 und .75, einem Greatest Lower Bound von .76 und .78 sowie einem ω von 0.73 und 0.76. Basierend auf den konfirmatorischen Faktorenanalysen ergab sich für die NFC-K-2 $\chi^2_{\text{korr}}(\text{df}) = 0.42$ (2), $\chi^2_{\text{unkorr}}(\text{df}) = 3.72$ (5), $CFI_{\text{korr}} = 1$, $CFI_{\text{unkorr}} = 1$, $RMSEA_{\text{korr}} = 0.000$ sowie $RMSEA_{\text{unkorr}} = 0.000$. Für die NFC-K-3 ergab sich $\chi^2_{\text{korr}}(\text{df}) = 0.48$ (2), $\chi^2_{\text{unkorr}}(\text{df}) = 7.34$ (5), $CFI_{\text{korr}} = 1$, $CFI_{\text{unkorr}} = 0.992$, $RMSEA_{\text{korr}} = 0.000$, sowie $RMSEA_{\text{unkorr}} = 0.041$ (Faktorladungen und Residualvarianzen siehe Anhang 3 in ESM 1). Die mittlere Korrelation lag bei beiden Skalen bei .40 und die Korrelation mit sozialer Erwünschtheit bei .17 und .19. Aller-

Tabelle 1. Items der entwickelten NFC-Kurzskalen

Item	NFC-K-1	NFC-K-2	NFC-K-3
Die Aufgabe, neue Lösungen für Probleme zu finden, macht mir wirklich Spaß.	x		
Ich setze mir eher solche Ziele, die nur mit erheblicher geistiger Anstrengung erreicht werden können.	x	x	x
Ich finde es besonders befriedigend, eine bedeutende Aufgabe abzuschließen, die viel Denken und geistige Anstrengung erfordert hat.		x	
Ich würde lieber etwas tun, das wenig Denken erfordert, als etwas, das mit Sicherheit meine Denkfähigkeit herausfordert. ^a	x	x	x
Ich finde wenig Befriedigung darin, angestrengt und stundenlang nachzudenken. ^a	x	x	x
In erster Linie denke ich, weil ich muss. ^a		x	
Denken entspricht nicht dem, was ich unter Spaß verstehe. ^a			x
Ich habe es gern, wenn mein Leben voller kniffliger Aufgaben ist, die ich lösen muss.	x		x

Anmerkungen: NFC-K-1 = Need for Cognition-Kurzskala-1; NFC-K-2 = Need for Cognition-Kurzskala-2; NFC-K-3 = Need for Cognition-Kurzskala-3. ^aItem ist invertiert.

dings gilt hierbei anzumerken, dass viele der möglichen Kurzskalen sehr ähnliche Ergebnisse erzielten, sodass zur finalen Auswahl zusätzlich die Höhe von Cronbachs α sowie eine ausgewogene Korrelation zu den verwandten Konstrukten im Vergleich zur Langskala hinzugezogen wurde.

Als Sensitivitätsanalyse und zur umfangreichen Bewertung der Subskalen wurden, basierend auf CFI_{kor} (siehe auch Ergebnisse im Code Supplement in ESM 2), für die Langform sowie für die besten 104 Subskalen Messinvarianzmetriken bezüglich des Geschlechts und Bildungslevels berechnet. Hierbei waren für die Langform der Skala die Kriterien der metrischen und skalaren Messinvarianz sowohl für die Gruppierung nach Geschlecht als auch nach Bildung basierend auf dem $\Delta\chi^2$ Kriterium und dem Kriterium $\Delta CFI < 0.01$ nicht erfüllt. Im Vergleich hierzu schnitten beide selektierten Subskalen sowohl hinsichtlich des Gesamtfits im Basismodell als auch hinsichtlich der Messinvarianz besser ab (siehe Anhang 5 in ESM 1). Bezüglich der NFC-K-2 ergab sich Messinvarianz bezüglich des Geschlechts auf allen Metriken und bezüglich Bildung auf allen außer CFI ($\Delta CFI = 0.011$) bei metrischer Invarianz. Die NFC-K-3 erzielte Messinvarianz bezüglich des Geschlechts auf allen Metriken außer $RSMEA$ ($\Delta RMSEA = 0.033$) bei metrischer Invarianz und bezüglich Bildung auf allen Metriken. Die selektierten Subskalen zeigten zudem keine deutlich schlechteren Messinvarianzresultate im Vergleich zu den anderen 102 untersuchten Subskalen. Insgesamt erzielten nur 14% der selektierten Subskalen sowohl hinsichtlich Bildung als auch Geschlecht auf allen Metriken Invarianz.

Die weitere Sensitivitätsanalyse zeigte auch Robustheit bezüglich des Schätzverfahrens. Bei Schätzung der besten 104 Subskalen mit dem MLR Schätzer zeigten sich vergleichbare Ergebnisse in den Fit-Statistiken zur ML

Schätzung (siehe Code Supplement in ESM 2). So erzielten beispielsweise 99 der 104 Skalen CFI -Werte (korreliert und unkorreliert) von über 0.95 und alle Skalen, die in der ML-Schätzung einen $RMSEA_{\text{kor}}$ und $RSMEA_{\text{unkor}}$ unter 0.05 erzielt hatten, fielen auch in der MLR-Schätzung unter diesen Grenzwert.

Neu entwickelte Kurzskalen zur Messung von NFC

Basierend auf der klassischen Vorgehensweise einerseits und dem full information approach andererseits wurden wie in den vorhergehenden Abschnitten beschrieben insgesamt drei Kurzskalen zur Messung von NFC entwickelt (klassisch: NFC-K-1; full information approach: NFC-K-2, NFC-K-3), wobei sich eine große Überschneidung bezüglich der ausgewählten Items in allen drei Kurzskalen ergab (siehe Tabelle 1).

Studie 2: Kurzskalenvalidierung

Zur Validierung und Bewertung der drei Kurzskalen wurde eine weitere Online-Studie durchgeführt. Hierbei wurden in einem randomisierten Design die drei Kurzskalen an unterschiedlichen Probanden und Probandinnen erhoben. Zur Validierung wurden zusätzlich die 16-Item-NFC-Skala sowie verwandte Persönlichkeitskonstrukte erfasst.

Methode

Stichprobe

Die Teilnehmenden der Studie wurden über verschiedene deutschsprachige Onlineforen und Mailinglisten rekrui-

tiert. Insgesamt bearbeiteten 557 Personen den Fragebogen. Zur Sicherung der Datenqualität wurden 27 Personen ausgeschlossen, die den Fragebogen in weniger als 10 Minuten bearbeitet haben oder keine Varianz auf verschiedenen Variablen aufwiesen. Somit ergab sich eine Analytestichprobe von $N = 530$ Personen. Die Stichprobe enthielt deutlich mehr weibliche ($n = 398$; 75 %) als männliche ($n = 132$; 25 %) Personen. Das Durchschnittsalter betrug 31,9 Jahre ($SD = 10,7$). Die Verteilung des höchsten Bildungsabschlusses war ähnlich zu Studie 1. Es gaben 0,2 % der Personen an, keinen Schulabschluss zu haben, 17 % verfügten über einen Volks-, Haupt- oder Realschulabschluss, 39 % hatten (Fach-) Abitur, 43 % einen (Fach-) Hochschulabschluss oder höher. 1 % der Personen machte keine Angaben oder gab einen nicht einzuordnenden Abschluss an.

Ablauf und Inhalt der Erhebung

Die Durchführung der Validierungsstudie erfolgte analog zur ersten Studie. Die Studie wurde als Onlinefragebogen mit der Software Unipark von QuestBack programmiert und durchgeführt. Die Bearbeitung des Onlinefragebogens dauerte durchschnittlich 25 Minuten. Nach einer Begrüßung und Aufklärung der Teilnehmenden, folgte eine kurze Einweisung und es wurden einige soziodemografische Daten abgefragt. Anschließend wurden alle weiteren Skalen erhoben. Die Teilnehmenden erhielten über randomisierte Zuteilung eine der drei neu entwickelten Kurzskalen. Alle weiteren Skalen zur Überprüfung der Validität wurden von allen Teilnehmenden bearbeitet. Alle Skalen hatten ein siebenstufiges Antwortformat (von 1 = „trifft überhaupt nicht zu“ bis 7 = „trifft voll und ganz zu“). Es handelt sich größtenteils um dieselben Skalen wie in der Selektionsstudie, weshalb auf eine ausführliche Beschreibung dieser im Folgenden verzichtet wird. Für die Validierung zusätzlich kamen lediglich die Konstrukte Vermeidungsleistungszielorientierung und Gewissenhaftigkeit hinzu. Anhang 2 in ESM 1 gibt einen Überblick über die Reihenfolge der verwendeten Skalen sowie deren interne Konsistenzen.

Vermeidungsleistungszielorientierung

Unter Vermeidungsleistungszielorientierung wird die Motivation verstanden, mangelnde Fähigkeiten zu verbergen. Da Bless und Kollegen (1994) zeigen konnten, dass NFC negativ mit Vermeidungsleistungszielorientierung zusammenhängt, erwarteten wir im Rahmen der Validierung ebenfalls einen negativen Zusammenhang. Das Konstrukt wurde mit der Subskala *Vermeidungsleistungszielorientierung* aus den Skalen zur Erfassung der Lern- und Leistungsmotivation (SELLMO; Spinath et al., 2012) gemessen.

Gewissenhaftigkeit

Mit Gewissenhaftigkeit wird ein Persönlichkeitsmerkmal beschrieben, das mit Ordnungsliebe, Pflichtbewusstsein und Zuverlässigkeit einhergeht (Asendorpf, 2015). Studien zum Zusammenhang zwischen NFC und verschiedenen Persönlichkeitsmerkmalen konnten zeigen, dass NFC mit Gewissenhaftigkeit korreliert (z. B. Bless et al., 1994; Sadowski & Cogburn, 1997). Da Cacioppo und Petty (1982) NFC theoretisch als zweck- und aufgabengebundenes, organisiertes und engagiertes Denken beschreiben, wird angenommen, dass Gewissenhaftigkeit aufgrund seiner Facetten des akkuraten und gewissenhaften Denkens mit NFC korreliert. Zur Erfassung des Persönlichkeitsmerkmals Gewissenhaftigkeit wurden die 12 Items der entsprechenden Subskala des NEO-FFI (Borkenau & Ostendorf, 2008) verwendet.

Ergebnisse

Zur Auswahl der finalen Kurzskala wurden analoge Kriterien wie bei der Itemselektion herangezogen. Hierbei wurden Cronbachs α zur Erfassung der Reliabilität, Modellfit Kennwerte einer konfirmatorischen Faktorenanalyse (CFI, RMSEA), die mittlere Korrelation mit verwandten Konstrukten sowie die Korrelation mit sozialer Erwünschtheit ermittelt. Nachdem in der Selektion die konfirmatorische Faktorenanalyse mit korrelierten Fehlertermen der negativ formulierten Items einen deutlich besseren Fit erzielt hatte, analog zu den Ergebnissen von Hevey et al (2012), wurden für die finale Selektion nur die Ergebnisse der CFA_{kor} miteinbezogen.

Alle drei untersuchten Kurzskalen erzielten ähnliche Ergebnisse bezüglich der untersuchten Gütekriterien (siehe Tabelle 2). Es zeigten sich keine signifikanten Unterschiede für Cronbachs α zwischen den Kurzskalen ($p = .628$) basierend auf dem Test von Feldt, Woodruff und Salih (1987). Alle Skalen erzielten eine zufriedenstellende interne Konsistenz (Cronbachs $\alpha = .69-.72$). Auch die Fit-Statistiken der konfirmatorischen Faktorenanalyse mit Korrelation der negativ formulierten Items waren zufriedenstellend (Faktorladungen und Residualvarianzen siehe Anhang 4 in ESM 1). Dabei wies lediglich die dritte getestete Skala (NFC-K-3) einen nur akzeptablen Fit (RMSEA = 0,06) auf und schneidet in diesem Punkt somit etwas schlechter ab als die anderen beiden Skalen. Alle Kurzskalen korrelierten hoch mit der 16-Itemversion ($r = .80-.81$, $p < .001$). Die mittlere Korrelation aller Kurzskalen mit den erfassten Konstrukten Präferenz für Deliberation, Typisches intellektuelles Interesse, Lernzielorientierung und Offenheit für neue Erfahrung war zwischen allen Kurzskalen vergleichbar und zeigt zwischen den Skalen keine signifikanten Unterschiede ($p_{K1-K2} = .735$, $p_{K2-K3} = .665$, $p_{K1-K3} = .912$). Die

Tabelle 2. Kennwerte der Kurzskaen zur finalen Auswahl

	NFC-K-1 (klassisch)	NFC-K-2 (full information approach)	NFC-K-3
<i>M (SD)</i>	4.69 (1.07)	5.22 (1.03)	4.76 (1.11)
Cronbachs α	.71	.69	.72
CFI	1.00	1.00	.99
RMSEA	0.00	0.00	.06
Mittlere Korrelation	.38	.41	.37
Korrelation mit der NFC	.80**	.81**	.81**
Korrelation mit Sozialer Erwünschtheit	.19**	.08	.28**

Anmerkungen: NFC-K-1 = Need for Cognition-Kurzska-1; NFC-K-2 = Need for Cognition-Kurzska-2; NFC-K-3 = Need for Cognition-Kurzska-3. * $p < .05$, ** $p < .01$.

Tabelle 3. Korrelationen zwischen den erhobenen Konstrukten mit den NFC-Skaen

	NFC-K-1	NFC-K-2	NFC-K-3	16-NFC ¹
Konvergente Validität				
NFC ¹	.80**	.81**	.81**	1.00
CI (95 %)	[.78; .98]	[.79; .98]	[.78; .97]	530
<i>N</i>	184	171	175	
Offenheit	.22**	.42**	.29**	.39**
CI (95 %)	[.10; .47]	[.34; .67]	[.21; .61]	[.37; .56]
<i>N</i>	184	171	175	530
Gewissenhaftigkeit	.35**	.30**	.28**	.29**
CI (95 %)	[.23; .52]	[.18; .50]	[.16; .52]	[.22; .39]
<i>N</i>	184	171	175	530
Präferenz für Deliberation	.35**	.29**	.24**	.31**
CI (95 %)	[.21; .48]	[.17; .50]	[.12; .47]	[.24; .40]
<i>N</i>	184	171	175	530
TIE	.51**	.49**	.57**	.62**
CI (95 %)	[.46; .76]	[.43; .76]	[.55; .86]	[.61; .76]
<i>N</i>	184	171	175	530
Lernzielorientierung	.43**	.43**	.37**	.49**
CI (95 %)	[.35; .66]	[.34; .67]	[.26; .59]	[.44; .61]
<i>N</i>	180	167	170	517
Vermeidungsleistungszielorientierung	-.28**	-.17*	-.24**	-.29**
CI (95 %)	[-.30; -.06]	[-.20; -.01]	[-.28; -.07]	[-.23; -.13]
<i>N</i>	180	167	170	517
Diskriminante Validität				
Soziale Erwünschtheit	.19*	.08	.29**	.26**
CI (95 %)	[.07; .52]	[-.14; .42]	[.19; .57]	[.24; .47]
<i>N</i>	178	166	172	516

Anmerkungen: NFC = Need for Cognition; NFC-K-1 = Need for Cognition-Kurzska-1; NFC-K-2 = Need for Cognition-Kurzska-2; NFC-K-3 = Need for Cognition-Kurzska-3; TIE = Typisches Intellektuelles Interesse. * $p < .05$, ** $p < .01$.

¹16-Item-Version von Bless et al. (1994).

Kurzskaen NFC-K-1 und NFC-K-3 wiesen im Gegensatz zur Kurzska NFC-K-2 signifikante Korrelationen mit der sozialen Erwünschtheit auf.

Eine detaillierte Übersicht der erhobenen Konstrukte zur Validierung und ihre Korrelationen mit den drei NFC-Kurzskaen sowie der 16-Itemversion der NFC ist in Tabelle 3 dargestellt. Es zeigten sich keine deutlichen Un-

terschiede in den Korrelationsmustern zwischen den getesteten Kurzskaen und der Langform. Lediglich im Hinblick auf die soziale Erwünschtheit ist die NFC-K-2 den anderen beiden Kurzskaen überlegen.

Dementsprechend wurde die NFC-K-2 als finale Kurzska gewählt. Während alle Kurzskaen sehr ähnliche Ergebnisse bezüglich der internen Konsistenz, des CFI,

der Korrelation mit verwandten Konstrukten und der 16-Itemversion aufwies, zeigte die NFC-K-2 einen zufriedenstellenden RMSEA und wies keinen Zusammenhang zur sozialen Erwünschtheit auf.

Diskussion

Ziel der vorliegenden Arbeit war einerseits die Entwicklung einer Kurzskala, die das Konstrukt Need for Cognition möglichst effizient, reliabel und valide misst. Andererseits wurde ein Vergleich klassischer und innovativer Itemselektionsstrategien durchgeführt mit besonderem Fokus auf eine transparente und möglichst objektive Darstellung aller Schritte bei der Selektion, Validierung und Bewertung. Zur Selektion und Validierung der Skala wurden unabhängige Datensätze verwendet und umfassende Validierungskriterien angelegt.

In Studie 1 wurde einerseits anhand klassischer Itemselektion und andererseits anhand des von uns vorgeschlagenen full information approach drei mögliche Fünf-Item-Kurzskalen zur Messung von NFC selektiert. Bei dem klassischen Vorgehen standen inhaltliche und sprachliche Attribute sowie Kennwerte der klassischen Testtheorie (Itemschwierigkeit, Itemvarianz, Itemtrennschärfe) im Mittelpunkt. Beim full information approach wurden neben inhaltlichen und Reliabilitätskriterien auch die Konstruktvalidität und die konvergente Validität möglicher Subskalen herangezogen, um die besten Skalen zu ermitteln. In Studie 2 wurden die drei ausgewählten potenziellen Subskalen an einer zweiten, unabhängigen Stichprobe umfassend validiert. Basierend auf den Gütekriterien der internen Konsistenz, des Modellfits sowie der konvergenten und diskriminanten Validität wurde schließlich eine finale NFC-Kurzskala ausgewählt. Da die Repräsentation eines Konstruktes besser gelingen kann, wenn der Konstruktraum mit heterogenen Items gut abgedeckt ist (Little, Lindenberger & Nesselroade, 1999), dies aber zu einer relativ geringen internen Konsistenz führen kann, besonders bei einer kleinen Anzahl an Items, lag unser Fokus auf der Optimierung von Modellfit und prädiktiver Validität.

Trotz unterschiedlichen Vorgehens in den beiden Selektionsansätzen erzielten die resultierenden Kurzskalen ähnliche Ergebnisse in der Validierungsstudie, insbesondere bezüglich der konvergenten Validität. Ein solches Ergebnis erzielten auch Sandy und Kollegen (2014) in ihrem Vergleich eines genetischen Algorithmus und einer klassischen Kurzskalenentwicklung. Mögliche Gründe für dieses Ergebnis liegen zum einen in der Ähnlichkeit der Kurzskalen. So waren insgesamt drei Items in allen drei Kurzskalen enthalten. Zum anderen zeigte sich im Rah-

men der Durchführung des full information approach, dass viele potenzielle Subskalen ähnliche Ergebnisse bezüglich der Gütekriterien erzielten. Das legt den Schluss nahe, dass innerhalb des Itempools der 16-Item-NFC eine Reihe möglicher Itemkombinationen die Gütekriterien einer guten Kurzskala angemessen erfüllen. Während das gute Abschneiden verschiedener Subskalen die Auswahl einer finalen, eindeutig besten Kurzskala erschwerte, illustriert es einen Vorteil des computergestützten Vorgehens. Es konnte explizit und quantifizierbar aufgezeigt werden, dass viele Subskalen gute Kurzformen der Skala darstellen und in welchen Items diese Subskalen überlappen.

Als beste NFC-Kurzskala wurde in Studie 2 eine der beiden anhand des full information approach selektierten Kurzskalen, die NFC-K-2, bewertet. Die finale Auswahlentscheidung basierte vorrangig auf Modellfitkennwerten sowie dem Zusammenhang mit der sozialen Erwünschtheit. Damit wurde eine NFC-Kurzskala entwickelt, die das Konstrukt NFC valide und reliabel misst und gleichzeitig den unerwünschten substantiellen Zusammenhang zwischen der 16-Itemversion und dem Antwortstil der sozialen Erwünschtheit in der Kurzskala reduziert.

Die vorliegende Studie illustriert die Breite der Kriterien, mit denen potenzielle Kurzskalen in dieser Klasse von stark computergestützten Selektionsverfahren bewertet werden. Neben Aspekten der internen Konsistenz, wie sie in der klassischen Vorgehensweise dominant ist, sind dies auch Aspekte der Konstruktvalidität durch Modellpassungskennwerte sowie diskriminante und konvergente Validität. Als Vorteil dieses Verfahrens sind damit alle zentralen Kriterien einer Kurzskala bereits im Selektionsprozess mit abgedeckt. Diese Kriterien können problemlos erweitert werden, um weiteren Anforderungen an Kurzskalen im Hinblick auf unterschiedliche Aspekte Rechnung zu tragen, wie beispielsweise der Messäquivalenz in unterschiedlichen Gruppen bei interkulturellen Studien oder komplexeren Faktorstrukturen.

Eine Herausforderung innerhalb einer solch komplexen Form der Selektierung ist das Identifizieren einer optimalen Kurzskala über die verschiedenen Optimierungskriterien hinweg. In der vorliegenden Arbeit wurde keine a priori festgelegte mathematische Gewichtung der einzelnen Kriterien vorgenommen, da hierfür keine klare theoretische Fundierung vorliegt. Stattdessen wurde die Auswahl der optimalen Skala schrittweise durchgeführt über Untergrenzen der internen Konsistenz- sowie Modellfitmaße, dem Ausschluss eines problematischen Items und der Festlegung der Itemanzahl sowie der Anzahl invertierter Items. In einem letzten Schritt erfolgte eine Abwägung der verbleibenden Kriterien, da unterschiedliche Subskalen auf den jeweiligen Kriterien optimal waren.

Dies reduziert zwar die Objektivität des Vorgehens, hat jedoch den Vorteil, dass Forscherinnen und Forscher keine atheoretische Black Box bedienen müssen (Sandy et al., 2014), sondern eine informierte Auswahlentscheidung treffen können. Gerade bei einer solchen nicht-automatisierten Abwägung der einzelnen Kriterien ist es wichtig, einen klaren Fokus bezüglich der zu optimierenden Kriterien zu setzen, wie z.B. in unserem Fall die Konstrukt- und konvergente Validität für das Konstrukt NFC, um einen transparenten Selektionsprozess zu ermöglichen.

Bei der Auswahl der finalen Kurzsкала sowie dem direkten Vergleich der getesteten Kurzsкаlen sind gewisse Einschränkungen miteinzubeziehen. Alle Ergebnisse beruhen auf nicht-repräsentativen Stichproben. Die Stichprobengröße ist zudem am unteren Limit für sinnvolle Vergleiche zwischen Korrelationskoeffizienten und Subgruppen. Unterschiede in den Korrelationskoeffizienten zu den erhobenen Konstrukten sollten deshalb nicht überbewertet werden und repräsentative Aussagen über die Skalenverteilungen sind nicht möglich. Allerdings sind Kurzsкаlen im Allgemeinen in der Individualdiagnostik, wie beispielsweise in der Personalauswahl, nur eingeschränkt nutzbar (Kruyen, Emons, & Sijtsma, 2012) und so liegt auch der Fokus der Nutzung einer NFC-Kurzsкаla im Bereich der Forschung. Zudem beruhen alle erzielten Ergebnisse auf unabhängigen Datensätzen zur Selektion und zur Validierung, was mögliche Verzerrungen durch spezielle Stichproben reduziert. Des Weiteren zeigten sich zufriedenstellende Ergebnisse den Sensitivitätsanalysen bezüglich des Geschlechts und Bildung in den Messinvarianzanalysen der ersten Studie.

Generelle Einschränkungen des full information approach – unabhängig von unserer konkreten Studie – bestehen bezüglich der Auswahl möglicher Kriterien. Eine Vielzahl verschiedener Optimierungskriterien ist möglich und ihre Auswahl ist abhängig von der Art des Konstrukts, das in Kurzform erfasst werden soll. So lag der Fokus bezüglich der Kurzsкаla NFC auf der Konstruktvalidität und der konvergenten und divergenten Validität. Allgemein sollte die Auswahl zum einen inhaltlich getrieben sein bezüglich empirischer und theoretischer Erkenntnisse zum betreffenden Konstrukt. Zum andern sollten nur solche Kriterien einbezogen werden, deren Berechnung automatisiert möglich ist und bei denen klare Optima definiert werden können. Gerade bei einer Vielzahl von Kriterien ist zunehmend notwendig, eine klare Berechnungsregel für ein globales Optimum über alle Kriterien hinweg festzulegen, um eine objektive Skalenselektion zu gewährleisten. Zudem sollten Aspekte der Schätzung in der Subskalenselektion miteinbezogen werden. Im vorliegenden Manuskript wurden Skalen basierend auf den Fitindices von zwei verschiedenen Modellversionen (mit

und ohne korrelierte Fehlerterme) ausgewählt. Der Einfluss verschiedener Schätzer auf die Modellselektion wurde nur innerhalb einer Sensitivitätsanalyse untersucht. Auch wenn diese Analysen keine starken Unterschiede im Modellfit bezüglich anderer Schätzverfahren aufzeigten, wäre dies doch eine stringente Erweiterung des präsentierten Ansatzes, auch die Sensitivität bezüglich verschiedener Schätzer explizit miteinzubauen.

Weiterhin besteht eine gewisse Einschränkung darin, dass ein beträchtliches Maß an Programmierfähigkeiten notwendig ist, um alle möglichen Skalenkombinationen zu überprüfen. Durch die Bereitstellung der Daten und Skripte, mit denen die vorliegende Studie erstellt wurde, soll ein Nachvollziehen der präsentierten Ergebnisse erleichtert werden und können weitere Optimierungskriterien getestet werden. Darüber hinaus sind die Möglichkeiten des Ansatzes durch den rechnerischen Aufwand und die Anforderungen an den Arbeitsspeicher des genutzten Computers beschränkt. Die exhaustive Suche im full information approach ist eine effiziente Möglichkeit der Kurzsкаlenfindung bei Konstrukten mit einer Langform unter 25 Items. In diesem Fall ist die Berechnung auf einem handelsüblichen Computer möglich. Eine Selektion basierend auf größeren Itempools wäre zwar durch leistungsstärkere Rechner machbar, in diesem Fall sind allerdings die oben genannten Algorithmen der kombinatorischen Optimierung eine effizientere Option, da hierbei nur ein Teil aller möglichen Subskalen getestet wird.

Die vorliegende Arbeit präsentiert eine reliable und valide Kurzsкаla zur Erfassung des Konstrukts Need for Cognition mit nur fünf Items. Diese Skala schließt die Lücke einer effizienten Erfassung dieses viel beforschten Konstrukts. Am Fallbeispiel dieser Kurzsкаlenentwicklung werden Vorteile des innovativen full information approach in der Kurzsкаlenentwicklung demonstriert, aber auch praktische Herausforderungen und Abwägungsentscheidungen aufgezeigt, die mit einer Kurzsкаlenentwicklung einhergehen.

Elektronische Supplemente (ESM)

Die elektronischen Supplemente sind mit der Online-Version dieses Artikels verfügbar unter <https://doi.org/10.1026/0012-1924/a000242>

ESM 1. Ergänzende Kennwerte und Statistiken. Anhang 1: Itemkennwerte und Faktorladungen der Items der 16-Item-Version aus Studie 1. Anhang 2: Kennwerte aller verwendeten Skalen. Anhang 3: Faktorladungen und Residualvarianzen der CFA_{korrt} der selektierten Kurzsкаlen in der Selektionsstudie. Anhang 4: Faktorladungen und Residualvarianzen der CFA_{korrt} der getesteten Kurzsкаlen in

der Validierungsstudie. Anhang 5: Ergebnisse der Messinvarianz-Analysen auf den Subskalen und der Langform der Selektionsstudie

ESM 2. Code Supplement

Literatur

- Anseel, F., Lievens, F. & Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, 110, 23–35. <https://doi.org/10.1016/j.obhdp.2009.05.003>
- Asch, S. E. (1952). *Social Psychology*. New York, NY: Prentice-Hall. <https://doi.org/10.1037/10025-000>
- Asendorpf, J. B. (2015). Persönlichkeitsbereiche. In J. B. Asendorpf (Hrsg.), *Persönlichkeitspsychologie für Bachelor* (S. 65–120). Springer: Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-46454-0-4>
- Beierlein, C., Kemper, C. J., Kovaleva, A. & Rammstedt, B. (2013). Short scale for measuring general self-efficacy beliefs (ASKU). *Methoden, Daten, Analysen*, 7, 251–278. <https://doi.org/10.12758/mda.2013.014>
- Beißert, H., Köhler, M., Rempel, M. & Beierlein, C. (2014). *Eine deutschsprachige Kurzskala zur Messung des Konstrukts Need for Cognition*. GESIS-WorkingPapers, 32. <https://www.ssoar.info/ssoar/handle/document/40315>
- Bertrams, A. & Dickhäuser, O. (2010). University and school students' motivation for effortful thinking. *European Journal of Psychological Assessment*, 26, 263–268. <https://doi.org/10.1027/1015-5759/a000035>
- Berzonsky, M. D. & Sullivan, C. (1992). Social-cognitive aspects of identity style: Need for Cognition, experiential openness, and introspection. *Journal of Adolescent Research*, 7, 140–155. <https://doi.org/10.1177/074355489272002>
- Betsch, C. (2004). Präferenz für Intuition und Deliberation (PID). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 25, 179–197. <https://doi.org/10.1024/0170-1789.25.4.179>
- Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F. & Schwarz, N. (1994). Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben. *Zeitschrift für Sozialpsychologie*, 25, 147–154. <https://doi.org/10.1348/014466602321149858>
- Borkenau, P. & Ostendorf, F. (2008). *NEO-FFI: NEO-Fünf-Faktoren-Inventar nach Costa und McCrae*. Göttingen: Hogrefe.
- Brunner, M., Nagy, G. & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs: Hierarchically structured constructs. *Journal of Personality*, 80, 796–846. <https://doi.org/10.1111/j.1467-6494.2011.00749.x>
- Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A. & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119, 197–253. <https://doi.org/10.1037/0033-2909.119.2.197>
- Cacioppo, J. T., Petty, R. E. & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48, 306–307. https://doi.org/10.1207/s15327752jpa4803_13
- Cacioppo, J. T., Petty, R. E. & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, 45, 805–818. <https://doi.org/10.1037/0022-3514.45.4.805>
- Cavazos, J. T. & Campbell, N. J. (2008). Cognitive style revisited: The structure X cognition interaction. *Personality and Individual Differences*, 45, 498–502. <https://doi.org/10.1016/j.paid.2008.06.001>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464–504
- Chen, C., Himsel, A., Kasof, J., Greenberger, E. & Dmitrieva, J. (2006). Boundless creativity: Evidence for the domain generality of individual differences in creativity. *Journal of Creative Behavior*, 40, 179–199. <https://doi.org/10.1002/j.2162-6057.2006.tb01272.x>
- Chen, C.-H. & Wu, I.-A. (2012). The interplay between cognitive and motivational variables in a supportive online learning system for secondary physical education. *Computers and Education*, 58, 542–550. <https://doi.org/10.1016/j.compedu.2011.09.012>
- Curşeu, P. L. (2011). Need for cognition and active information search in small student groups. *Learning and Individual Differences*, 21, 415–418. <https://doi.org/10.1016/j.lindif.2011.02.005>
- Epstein, S., Pacini, R., Denes-Raj, V. & Heier, H. (1996). Individual Differences in Intuitive-Experiential and Analytical-Rational Thinking Styles. *Journal of Personality and Social Psychology*, 71, 390–405. <https://doi.org/10.1037/0022-3514.71.2.390>
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied psychological measurement*, 11, 93–103. <https://doi.org/10.1177/014662168701100107>
- Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A. & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin*, 36, 82–96. <https://doi.org/10.1177/0146167209351886>
- Forsterlee, R. & Ho, R. (1999). An examination of the short form of the need for cognition scale applied in an Australian sample. *Educational and Psychological Measurement*, 59, 471–480. <https://doi.org/10.1177/00131649921969983>
- Hahn, M. H. & Lee, K. C. (2016). Exploring the role of self-confidence, Need for Cognition, and the degree of IT support on individual creativity: Multilevel analysis approach. *Current Psychology*, 36, 565–576. <https://doi.org/10.1007/s12144-016-9445-z>
- Haugtvedt, C., Petty, R. E., Cacioppo, J. T. & Steidley, T. (1988). Personality and ad effectiveness: Exploring the utility of need for cognition. *Advances in Consumer Research*, 15, 209–213.
- Hevey, D., Thomas, K., Pertl, M., Maher, L., Craig, A. & Chuinagain, S.N. (2012). Method Effects and the Need for Cognition Scale. *The International Journal of Educational and Psychological Assessment*, 12, 20–33.
- Jackson, P. H. & Agunwanba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, 42, 567–578. <https://doi.org/10.1007/BF02295979>
- Klingsieck, K. B. & Fries, S. (2012). Allgemeine Prokrastination: Entwicklung und Validierung einer deutschsprachigen Kurzskala der General Procrastination Scale (Lay, 1986). *Diagnostica*, 58, 182–193. <https://doi.org/10.1026/0012-1924/a000060>
- Kruyen, P. M., Emons, W. H. & Sijtsma, K. (2012). Test length and decision quality in personnel selection: When is short too short? *International Journal of Testing*, 12, 32–1344. <https://doi.org/10.1080/15305058.2011.643517>
- Kruyen, P. M., Emons, W. H. & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, 13, 223–248. <https://doi.org/10.1080/15305058.2012.703734>
- Leite, W. L., Huang, I. C., & Marcoulides, G. A. (2008). Item selection for the development of short forms of scales using an ant colony optimization algorithm. *Multivariate Behavioral Research*, 43, 411–431. <https://doi.org/10.1080/00273170802285743>

- Little, T. D., Lindenberger, U. & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, 4, 192. <https://doi.org/10.1037/1082-989X.4.2.192>
- Maslow, A. H. (1943). Dynamics of personality organization. *Psychological Review*, 50, 514–539. <https://doi.org/10.1037/h0062222>
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer. <https://doi.org/10.1007/978-3-642-20072-4>
- Murphy, G. (1947). *Personality: A biosocial approach to origins and structure*. New York, NY: Harper & Brothers. <https://doi.org/10.1037/10759-000>
- Musch, J., Brockhaus, R. & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren Sozialer Erwünschtheit. *Diagnostica*, 48, 121–129. <https://doi.org/10.1026//0012-1924.48.3.121>
- Mussel, P., Spengler, M., Litman, J. A. & Schuler, H. (2012). Development and validation of the German work-related curiosity scale. *European Journal of Psychological Assessment*, 28, 109–117. <https://doi.org/10.1027/1015-5759/a000098>
- Muthén, L. K. & Muthén, B. O. (1998–2017). *Mplus User's Guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.
- Olson, K., Camp, C. & Fuller, D. (1984). Curiosity and need for cognition. *Psychological Reports*, 54, 71–74. <https://doi.org/10.2466/pr0.1984.54.1.71>
- Pechtl, H. (2009). *Anmerkungen zur Operationalisierung und Messung des Konstrukts „need for cognition“*. Diskussionspapier 05/09. Universität Greifswald.
- Petty, R. E. & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 123–205). New York, NY: Academic Press.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org>.
- Rammstedt, B. & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit. *Diagnostica*, 51, 195–206. <https://doi.org/10.1026/0012-1924.51.4.195>
- Reinhard, M.-A. & Dickhäuser, O. (2009). Need for cognition, task difficulty, and the formation of performance expectancies. *Journal of Personality and Social Psychology*, 96, 1062–1076. <https://doi.org/10.1037/a0014927>
- Sadowski, C. J. & Cogburn, H. E. (1997). Need for cognition in the big-five factor structure. *The Journal of Psychology*, 131, 307–312. <https://doi.org/10.1080/00223989709603517>
- Sandy, C. J., Gosling, S. D. & Koelkebeck, T. (2014). Psychometric comparison of automated versus rational methods of scale abbreviation: An illustration using a brief measure of values. *Journal of Individual Differences*, 35, 221–235. <https://doi.org/10.1027/1614-0001/a000144>
- Sarnoff, I. & Katz, D. (1954). The motivational bases of attitude change. *Journal of Abnormal and Social Psychology*, 49, 115–124.
- Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2014). Eine kurze Skala zur Messung kristalliner Intelligenz: Die Kurzskala gc des Berliner Tests zur Erfassung Fluidur und Kristalliner Intelligenz (BEFKI GC-K). *Gesis Working Papers*, 29. <https://www.ssoar.info/ssoar/handle/document/37786>
- Smith, G. T., McCarthy, D. M. & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12, 102–111. <https://doi.org/10.1037/1040-3590.12.1.102>
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2012). *Die Skalen zur Erfassung von Lern- und Leistungsmotivation (SELLMO)*. Göttingen: Hogrefe.
- Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K. & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in psychology*, 6, 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- Wilhelm, O., Schulze, R., Schmiedek, F. & Süß, H.-M. (2003). Interindividuelle Unterschiede im typischen intellektuellen Engagement. *Diagnostica*, 49, 49–60. <https://doi.org/10.1026//0012-1924.49.2.49>
- Ziegler, M., Kemper, C. J. & Krueger, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *Journal of Individual Differences*, 35, 185–189. <https://doi.org/10.1027/1614-0001/a000148>

Danksagung

Wir danken Frau Beatrice Rammstedt für die initiale Betreuung dieses Forschungsvorhabens, Frau Constanze Beierlein für die Beratung bei der Ausgestaltung und Erweiterung der Idee sowie für die Hilfe bei Planung und Umsetzung der Studien. Weiterhin danken wir Frau Silke Urso für die Unterstützung bei der Datenerhebung und allen Teilnehmerinnen und Teilnehmern, die an den Studien teilgenommen haben.

ORCID

Hanna Beißert

 <https://orcid.org/0000-0001-7767-3308>

Dr. Hanna Beißert

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Rostocker Straße 6
60323 Frankfurt am Main
beissert@dipf.de

Dr. Meike Köhler

Comma Soft AG
Pützchens Chaussee 202–204a
53229 Bonn

Marina Rempel, M. Sc.

Universität zu Köln
Albertus-Magnus-Platz
50923 Köln

Peter Krueger, Ph. D.

Radboud University Nijmegen
Institute for Management Research
Thomas van Aquinostraat 5
6525 GD Nijmegen
Niederlande