

Authentische Führung

Entwicklung und Validierung einer modifizierten deutschen Fassung des Authentic Leadership Inventory von Neider und Schriesheim (2011)

Luise Franke-Bartholdt, Dirk Frömmer, Jürgen Wegge und Anja Strobel

Zusammenfassung: Zur Erfassung authentischer Führung fehlt im deutschen Sprachraum ein transparent validiertes Messverfahren. Der vorliegende Beitrag soll mit einer deutschen Adaption des *Authentic Leadership Inventory* von Neider und Schriesheim (2011) diese Lücke schließen. Nach der Übersetzung des Originals wurde das *Deutsche Inventar Authentischer Führung* (DIAF) in drei Stichproben ($N_{\text{gesamt}} = 705$) geprüft und modifiziert. In konfirmatorischen Faktorenanalysen konnte die theoretisch postulierte Binnenstruktur mit vier Komponenten bestätigt werden: Selbstbewusstheit, Transparenz in Beziehung zu anderen, verinnerlichte moralische Perspektive und ausgewogene Informationsverarbeitung. Die internen Konsistenzen der Gesamtskala (16 Items) und Einzelkomponenten lagen im guten bis sehr guten Bereich. Es wurden erwartungskonforme Zusammenhänge zu anderen Führungsskalen gefunden (positive Korrelationen zu ethischer Führung und Leader Member Exchange, negative Korrelationen zu destruktiver Führung). Das Instrument zeigte bedeutsame Zusammenhänge zu zentralen organisationalen Ergebniskriterien (Wohlbefinden, Arbeitsengagement, individuelle Leistung) und inkrementelle Validität über andere Führungsskalen hinaus. Insgesamt kann das DIAF als ökonomisches und valides Verfahren zur Erfassung authentischer Führung eingesetzt werden.

Schlüsselwörter: Authentische Führung, Adaption, Messinstrument, Validität, Authentic Leadership Inventory

Authentic Leadership: Validation of a Modified German Version of the Authentic Leadership Inventory by Neider and Schriesheim (2011)

Abstract: The German-speaking research landscape lacks a verifiably validated measure of authentic leadership. We aimed at closing this gap with a German adaptation of the *Authentic Leadership Inventory* provided by Neider and Schriesheim (2011). We tested and adapted the translated DIAF (*Deutsches Inventar Authentischer Führung*) in three samples ($N_{\text{total}} = 705$). Confirmatory factor analysis confirmed the expected four component-structure: self-awareness, relational transparency, internalized moral perspective, and balanced processing. The overall scale (16 items) and the subscales had good to excellent internal consistencies. As expected, the DIAF was significantly related to other leadership scales (i.e., positive correlations with ethical leadership and leader member exchange, negative correlations with destructive leadership). The measure correlated significantly with criteria of organizational performance (i.e., well-being, work engagement, individual performance) and showed incremental validity above and beyond other leadership scales. In sum, the DIAF provides a viable and valid measure for assessing authentic leadership.

Keywords: authentic leadership, adaption, measurement, validity, Authentic Leadership Inventory

Unmoralisches Verhalten in Organisationen belastet die Gesellschaft jährlich mit Kosten in Milliardenhöhe (Eigensetter, Strobel & Stumpf, 2012). Wirtschaftsskandale der jüngeren Vergangenheit zeigen, dass Führungskräfte dabei häufig eine unrühmliche Rolle spielen. Vor dem Hintergrund derartiger gesellschaftlicher Ereignisse wird in der neueren arbeits- und organisationspsychologischen Forschung die Bedeutung (un)moralischen Verhaltens in Organisationen intensiv diskutiert. Eine steigende Zahl konzeptueller Arbeiten zu positiven bzw. moralischen Formen der Führung rückt zunehmend den Einfluss von Führungskräften in den Fokus (Banks, McCauley, Gardner & Guler, 2016; Brown &

Mitchell, 2010; Treviño, den Nieuwenboer & Kish-Gephart, 2014). Als wesentliche Gestalter des Arbeitsumfeldes haben Vorgesetzte aufgrund ihrer Vorbildfunktion maßgeblichen Einfluss auf Einstellungen und Verhalten von Mitarbeitenden (Treviño et al., 2014). Moralische Führungsformen sind mit Konzepten wie Vertrauenswürdigkeit, Echtheit und hohen Wertmaßstäben assoziiert (Brown & Mitchell, 2010) und sollten daher das moralische Verhalten von Mitarbeitenden positiv beeinflussen (Frömmer, Wegge & Strobel, 2014).

Als vielversprechendes Konzept unter den moralischen Führungsformen gilt authentische Führung (Brown &

Luise Franke-Bartholdt und Dirk Frömmer teilen sich die Erstautorenschaft.

Mitchell, 2010). Es bestehen u. a. positive Zusammenhänge von authentischer Führung mit Arbeitsleistung und -engagement, Arbeitszufriedenheit, Vertrauen und Hilfeverhalten sowie negative Zusammenhänge mit Burnout und kontraproduktivem Arbeitsverhalten (Banks et al., 2016; Hoch, Bommer, Dulebohn & Wu, 2016), d. h. das Konzept ist von hoher Relevanz im Rahmen des Human Resource Managements. Grundlegend für den Erkenntnisfortschritt ist es, wie gut es gelingt, Dimensionen und Schlüsselverhaltensweisen dieses Führungskonzeptes zu operationalisieren (Cooper, Scandura & Schriesheim, 2005). Während im angloamerikanischen Sprachraum verschiedene Instrumente zur Messung authentischer Führung zur Verfügung stehen (Avolio, Gardner & Walumbwa, 2007; Neider & Schriesheim, 2011), fehlt im deutschen Sprachraum ein umfassend und transparent validiertes Messinstrument. Das war die zentrale Motivation für die Entwicklung und erste Validierung des *Deutschen Inventars Authentischer Führung* (DIAF), einer deutschen Adaption des *Authentic Leadership Inventory* von Neider und Schriesheim (2011). Im Folgenden wird zunächst das Konzept der authentischen Führung theoretisch eingeordnet. Daran anschließend wird die hypothesengeleitete Prüfung des DIAF dargestellt.

Theoretische Grundlagen

Authentizität und authentische Führung

Authentische Führung wurde erstmals von Luthans und Avolio (2003) als eigenständige Führungsform vorgeschlagen und hat seither verschiedene theoretische Weiterentwicklungen erfahren (vgl. Review von Gardner, Cogliser, Davis & Dickens, 2011). Die wohl am weitesten akzeptierte Definition authentischer Führung von Walumbwa, Avolio, Gardner, Wernsing und Peterson (2008) spiegelt das aktuelle Verständnis wider:

(...) a pattern of leader behavior that draws upon and promotes both positive psychological capacities and a positive ethical climate, to foster greater self-awareness, an internalized moral perspective, balanced processing of information, and relational transparency on the part of leaders working with followers, fostering positive self-development. (S. 94)

Nach Walumbwa et al. (2008) umfasst authentische Führung vier Verhaltenskomponenten (vgl. Frömmer et al., 2014). Die Komponente (1) *Selbstbewusstheit* (SB) beschreibt die Selbstreflexion der Führungskraft, ihr vertieftes Verständnis für eigene Stärken und Schwächen so-

wie für die eigene Wirkung auf andere. Die Komponente (2) *Transparenz in Beziehung zu anderen* (TiB) umfasst im Kern, anderen Personen sein wahres Selbst zu zeigen, ohne Gedanken oder Gefühle zu verbergen oder sich zu verstellen. Der Aspekt (3) *Ausgewogene Informationsverarbeitung* (AIV) meint die objektive Analyse relevanter Informationen vor einer Entscheidung. Führungskräfte erbitten Stellungnahmen von Mitarbeitenden und stellen eigene (auch moralische) Grundansichten zur Diskussion. Schließlich beschreibt die Komponente (4) *Verinnerlichte moralische Perspektive* (VMP) eine Form der Selbstregulation, die über innere moralische Normen und Werte gesteuert wird und die zu Entscheidungen und Verhaltensweisen führt, die mit diesen Normen und Werten übereinstimmen.

Erfassung authentischer Führung

Zur Erfassung authentischer Führung kommen im angloamerikanischen Sprachraum vor allem der *Authentic Leadership Questionnaire* (ALQ; Avolio et al., 2007; Walumbwa et al., 2008) und das *Authentic Leadership Inventory* (ALI; Neider & Schriesheim, 2011) zum Einsatz (Banks et al., 2016). Walumbwa et al. (2008) stellten mit dem ALQ erstmals ein Instrument vor, um die vier Komponenten authentischer Führung zu erfassen. In der Originalpublikation von Walumbwa et al. (2008) wurden jedoch nur 8 der insgesamt 16 Items des ALQ veröffentlicht. Die Gesamtskala ist urheberrechtlich geschützt (Avolio et al., 2007) und für Forschungszwecke nur dann kostenfrei erhältlich, wenn eine Registrierung auf der vertreibenden Online-Plattform erfolgt und ausführliche Informationen zur geplanten Studie angegeben werden.

Neben der eingeschränkten Verfügbarkeit kritisierten Neider und Schriesheim (2011) Aspekte des von Walumbwa et al. (2008) gewählten Vorgehens zur Konstruktvalidierung des ALQ, d. h. die gewählten Modellspezifikationen in den CFA sowie die eingesetzten Skalen zur Bestimmung der divergenten Validität. Sie merkten zudem an, dass die Itemauswahl durch eine vergleichsweise kleine Stichprobe von Personen erfolgte, was die Inhaltsvalidität der Skala einschränke. Mit dem Ziel, die genannten Kritikpunkte zu verbessern, entwickelten Neider und Schriesheim (2011) mit dem ALI ein alternatives Instrument zur Erfassung authentischer Führung. Sie orientierten sich dabei eng an den Definitionen des Konzepts von Walumbwa et al. (2008) und formulierten vier Items pro Komponente, wobei sie die veröffentlichten ALQ-Items (Walumbwa et al., 2008) paraphrasierten. Nach einer ersten Prüfung der Inhaltsvalidität, u. a. über Mittelwertsvergleiche und Faktorenanalysen, wurden 14 Items in die finale Version des ALI aufgenommen. Die Kompo-

nenten des ALI wiesen mit internen Konsistenzen von $.70 (SB) \leq \alpha \leq .85 (AIV)$ eine befriedigende bis gute Reliabilität auf. In CFA wiesen ein Modell mit vier korrelierten Faktoren erster Ordnung ($.59 \leq r \leq .89$), welche inhaltlich den vier Komponenten authentischer Führung nach Walumbwa et al. (2008) entsprachen, und ein Modell mit vier Faktoren erster Ordnung und einem Faktor zweiter Ordnung (Gesamtwert authentische Führung) den besten Fit auf. Letzteres wurde dem Grundsatz der Sparsamkeit folgend verworfen. In einer Multitrait-Multimethod-Analyse konnte die konvergente Validität des ALI zu den veröffentlichten Items des ALQ (Walumbwa et al., 2008) gezeigt werden. In CFA ließen sich die ALI-Komponenten von den Facetten transformationaler Führung (Bass, 1985) abgrenzen, was für die divergente Validität des Verfahrens spricht. Ferner zeigten sich klare Zusammenhänge mit organisationalen Ergebniskriterien, z. B. zu $\bar{r} = .42$ mit allgemeiner Arbeitszufriedenheit und $\bar{r} = .30$ mit organisationalem Commitment (Neider & Schriesheim, 2011).

Mit dem ALI (Neider & Schriesheim, 2011) liegt für den angloamerikanischen Sprachraum ein validiertes, frei verfügbares Instrument zur Messung authentischer Führung vor. Im deutschsprachigen Raum besteht hingegen ein deutlicher Bedarf für Messinstrumente. Es liegt zwar eine deutsche Version des ALQ (Avolio et al., 2007) vor, die bereits in Studien zum Einsatz kam (z. B. Peus, Wesche, Streicher, Braun & Frey, 2012). Diese ist jedoch urheberrechtlich geschützt und wurde zudem nicht umfassend und transparent validiert. Als Adaption unterliegt sie grundsätzlich ähnlichen Kritikpunkten wie die amerikanische Originalversion des ALQ (Walumbwa et al., 2008). Aus den genannten Gründen entschieden wir uns dafür, eine deutsche Adaption des ALI von Neider und Schriesheim (2011) zu entwickeln und zu prüfen: das *Deutsche Inventar Authentischer Führung* (kurz: DIAF).

Herleitung der Hypothesen

Nach der Übersetzung der amerikanischen Originalversion des ALI (Neider & Schriesheim, 2011) wurden die psychometrischen Eigenschaften des DIAF untersucht. Dazu zählte in erster Linie die faktorielle Binnenstruktur der Skala, wobei in CFA überprüft wurde, ob sich die von Walumbwa et al. (2008) definierten Komponenten authentischer Führung (SB, AIV, TiB, VMP) im DIAF abbilden ließen. Neben der Bestimmung der Reliabilität der Gesamtskala und ihrer Komponenten erfolgten eine weiterführende Prüfung der Konstruktvalidität sowie eine erste Betrachtung der Kriteriumsvalidität.

Zur Prüfung der Konstruktvalidität wurde das DIAF in Beziehung zu anderen Führungskonzepten gesetzt, die im

nomologischen Netz eine unterschiedlich starke Nähe zu authentischer Führung aufweisen. Grundlegend bildet ein nomologisches Netz ein Theoriegerüst ab, spezifiziert relevante Konstrukte, deren Operationalisierungen und Verknüpfungen untereinander (Gardner et al., 2011). Gardner et al. (2011) stellten in ihrem Review ihren Entwurf zum nomologischen Netz authentischer Führung vor. Darin werden vor allem konzeptuelle Bezüge zu ethischer Führung (Brown, Treviño & Harrison, 2005) und transformationaler Führung (Judge & Piccolo, 2004) aufgezeigt. Transformationale Führung betont stark die Leistung und Motivation der Mitarbeitenden zur Erreichung organisationaler Ziele (Hoch et al., 2016). Ethische und authentische Führung fokussieren im Unterschied dazu stärker die moralischen Werte und das positive, moralische Verhalten der Führungskraft auch im Sinne der Vorbildwirkung gegenüber den Mitarbeitenden (Banks et al., 2016; Hoch et al. 2016). Sie bilden somit deutlicher den moralischen Kern als wesentliche Grundbedingung positiver Führung ab (Gardner et al., 2011). Für die erste Validierung des DIAF konzentrierten wir uns somit zunächst auf die ethische Führung. Führungskräfte werden dann als ethisch führend wahrgenommen (Brown et al., 2005; Brown & Mitchell, 2010; Rowold, Borgmann & Heinitz, 2009), wenn sie (1) über Eigenschaften wie Prinzipienorientierung, Vertrauenswürdigkeit oder Ehrlichkeit verfügen und insofern als Vorbilder fungieren (Ethisches Rollenmodell; ERM) und (2) in ihrem Einflussbereich normativ angemessenes Verhalten fördern, indem sie beispielsweise ethische Standards setzen (Ethische Mitarbeiterführung; EMF). Im Vergleich zu authentischer Führung wird bei ethischer Führung weniger stark die Beziehungsebene zu Mitarbeitenden und der offene Umgang mit Informationen, Gedanken und Gefühlen betont (Cropanzano & Walumbwa, 2010), da dem Konzept eher transaktionale Prozesse wie Klarheit über Belohnung und Bestrafung zugrunde liegen. Ferner stehen bei authentischer Führung die Selbstreflexion und der konkrete Diskurs mit den Mitarbeitenden stärker im Fokus als bei ethischer Führung. Dennoch werden inhaltliche Überschneidungen beider Führungsformen sichtbar: Beide Konzepte betonen einerseits die Vorbildwirkung der Führungskraft, die aus ihrer Funktion erwächst, und schreiben ihr eine positive moralische Orientierung zu (ERM bei ethischer Führung; VMP bei authentischer Führung). Andererseits ist in beiden Ansätzen zentral, auf welche Art und Weise die Führungskraft ihre Wertvorstellungen vertritt und in der Zusammenarbeit mit Mitarbeitenden agiert, diese einbindet und Rückmeldungen einholt (EMF bei ethischer Führung; AIV, TiB und SB bei authentischer Führung). Aufgrund der konzeptuellen Überlappung sollten hohe positive Zusammenhänge zwischen authentischer und ethischer Führung bestehen (Hypothese 1a).

Ferner sollten die Komponenten AIV, TiB und SB von authentischer Führung stärker mit der Facette EMF korrelieren als mit der Facette ERM von ethischer Führung, da sie die Einflussnahme von Führung sowie die Beziehung zu Mitarbeitenden in den Vordergrund stellen (Hypothese 1b). Demgegenüber sollte die Komponente VMP von authentischer Führung stärker mit der Komponente ERM als mit der Komponente EMF von ethischer Führung korrelieren, da die Orientierung an moralischen Grundsätzen und Werten bei beiden Teilaspekten im Vordergrund steht (Hypothese 1c).

In der bisherigen Forschung wurden positive Führungsformen weitgehend isoliert betrachtet (vgl. Übersichtsarbeiten von Brown & Mitchell, 2010; Schyns & Schilling, 2013). Für ein umfassendes Abbild von Führungsprozessen sollten jedoch auch negative, unmoralische Führungsweisen einbezogen werden. Dieser Aspekt fand weder im nomologischen Netz authentischer Führung nach Gardner et al. (2011) noch bei der Validierung des ALI durch Neider und Schriesheim (2011) Berücksichtigung. Um diese Lücke zu schließen, griffen wir auf das Konzept der *Abusive Supervision* (Tepper, 2000) zurück, welches unter den negativen Führungskonzepten das breiteste theoretische und empirische Fundament aufweist (vgl. Brown & Mitchell, 2010; Schyns & Schilling, 2013). Bei *Abusive Supervision* stehen im Wesentlichen destruktive Verhaltensweisen der Führungskraft im Vordergrund, die einer positiven Mitarbeiterorientierung widersprechen. Im nomologischen Netz ist *Abusive Supervision* daher gegensätzlich zur authentischen Führung ausgerichtet. Der Fokus auf die Person der Führungskraft, ihre Überzeugungen, Werte oder Fähigkeiten zur Selbstreflexion fehlt allerdings weitgehend. Die Konzeptnähe zur authentischen Führung ist daher geringer ausgeprägt als bei ethischer Führung. *Abusive Supervision* (Tepper, 2000) wird definiert als das von Mitarbeitenden wahrgenommene Ausmaß, in dem Führungskräfte „(...) engage in the sustained display of hostile verbal and nonverbal behaviors, excluding physical contact“ (S. 178). Zeigt die Führungskraft *Abusive Supervision*, fehlt Mitarbeitenden die Vertrauensbasis für einen offenen Austausch mit ihr (Tepper, Moss, Lockhart & Carr, 2007). Der direkte Kontakt mit der Führungskraft wird als risikoreich erlebt. Vor diesem Hintergrund sollten zwischen *Abusive Supervision* und authentischer Führung sowie ihren vier Komponenten negative Zusammenhänge bestehen (Hypothese 2).

Um die Kriteriumsvalidität des DIAF zu prüfen, wurden zunächst Zusammenhänge zu demografischen Variablen bestimmt. Weiterhin wurden sowohl behaviorale (d.h. Leistung) als auch affektive (Engagement, Wohlbefinden) und relationale Kriterien (Beziehungsqualität zwischen Mitarbeitenden und Führungskraft) berücksichtigt, um die verschiedenen Wirkrichtungen von Führung zu be-

rücksichtigen. Des Weiteren sollten durch die Auswahl wesentliche organisationale Ergebniskriterien abgebildet werden (Leistung, Wohlbefinden), die durch Führung letztlich positiv beeinflusst werden sollen (Wegge, 2004) und somit auch in das nomologische Netz authentischer Führung einfließen (Gardner et al., 2011). Es wird argumentiert (Hoch et al., 2016; Cropanzano & Walumbwa, 2010; Walumbwa et al., 2008), dass authentische Führung insbesondere die zwischenmenschliche Beziehung und den offenen Umgang mit anderen stärkt, dabei auch die Gefühlsebene berücksichtigt und sich diesbezüglich auch von anderen Führungsstilen abhebt. Um diesen Aspekt des nomologischen Netzes in die Validierung zu integrieren, wurden insbesondere affektive Kriterien und der Leader-Member-Exchange (Graen & Scandura, 1987) als relationales Kriterium erfasst. Auf Basis theoretischer Überlegungen (Gardner et al., 2011) und metaanalytischer Befunde (Banks et al., 2016) waren individuelle Unterschiede in demografischen Merkmalen (z. B. Alter, Geschlecht) nicht als bedeutsame Antezedenzen authentischer Führung zu erwarten. Daher sollte das DIAF weitgehend unabhängig von demografischen Merkmalen der Führungskräfte und Mitarbeitenden sein (Hypothese 3).

Authentische Führungskräfte fördern durch ihre Vorbildwirkung und die stetige Einbindung der Mitarbeitenden deren Selbstbestimmung und interne Regulationsprozesse (Ilies, Morgeson & Nahrgang, 2005). Auf diese Weise wird die intrinsische Motivation gesteigert (Deci, Connell & Ryan, 1989), was zu einer erhöhten Leistung führt. Empirisch zeigten Metaanalysen einen Zusammenhang zwischen authentischer Führung und Leistung von $\rho = .12$ (Hoch et al., 2016) bzw. $\rho = .14$ (Banks et al., 2016). Daher erwarteten wir einen niedrigen positiven Zusammenhang zwischen dem DIAF und Leistung (Hypothese 4).

Authentische Führungskräfte verdeutlichen ihren Mitarbeitenden, dass Offenheit und Austausch gewollt sind. Sie fördern auf diese Weise ein positives Arbeitsklima und die Mitarbeiterbeteiligung (Frömmner et al., 2014). Das sollte sich positiv auf das Arbeitsengagement und Wohlbefinden der Mitarbeitenden auswirken (vgl. Wegge et al., 2010). So zeigte sich metaanalytisch ein positiver Zusammenhang zu Engagement von $\rho = .47$ (Hoch et al., 2016), während zu Burnout/Stress ein negativer Zusammenhang von $\rho = -.27$ (Banks et al., 2016) gefunden wurde. Daher sollte sich ein mittlerer positiver Zusammenhang zwischen dem DIAF und Arbeitsengagement (Hypothese 5) und Wohlbefinden (Hypothese 6) zeigen.

Authentische Führung sollte auch die Beziehungsqualität zwischen Führungskraft und Mitarbeitenden beeinflussen, die sich anhand des Leader-Member-Exchange (LMX; Graen & Scandura, 1987) umschreiben lässt. Die positive Form der Führung sollte dabei durch qualitativ hochwertige Austauschbeziehungen geprägt sein (vgl.

Bishop, Scott & Burroughs, 2000), da der transparente und vertrauensvolle Umgang untereinander sowie die aktive Einbindung anderer Personen in Entscheidungsprozesse betont werden. In Metaanalysen zeigten sich Zusammenhänge von $\rho = .67$ (Hoch et al., 2016) bzw. $\rho = .65$ (Banks et al., 2016) zwischen authentischer Führung und LMX. Wir vermuteten demnach einen mittleren positiven Zusammenhang zwischen dem DIAF und LMX (Hypothese 7).

Für eine methodisch strenge Prüfung der Kriteriumsvalidität wurde zudem die inkrementelle Validität des DIAF bestimmt. Für die Kriterien Leistung, Arbeitsengagement, Wohlbefinden und LMX (Graen & Scandura, 1987) vermuteten wir, dass authentische Führung Varianz in den Kriterien über die Führungskonzepte ethische Führung und Abusive Supervision hinaus aufklären kann (Hypothese 8a). Der Anteil zusätzlich aufgeklärter Varianz sollte bei Abusive Supervision größer sein als bei ethischer Führung (Hypothese 8b), da geringere konzeptuelle Überlappungen zwischen authentischer Führung und Abusive Supervision bestehen.

Methoden

Übersetzung

Die Übertragung des ALI (Neider & Schriesheim, 2011) aus dem Englischen ins Deutsche erfolgte in einem mehrstufigen Prozess durch die Kombination aus Forward- und Back-Translation (vgl. Hambleton, 2005). Es gingen alle Ursprungsisems von Neider und Schriesheim (2011) in die Übersetzung ein, um ihre Anwendbarkeit im deutschen Sprachraum zu prüfen. Zunächst übersetzte eine Person aus dem Autorenteam die Items, wobei die inhaltliche Passung zum Konstrukt sowie die Angemessenheit für den deutschen Sprachraum bei der Übersetzung eine größere Rolle spielten als eine möglichst wortgetreue Abbildung der originalen Formulierungen. Der entstandene Entwurf wurde von den anderen Autorinnen und Autoren kommentiert und anschließend in einer Gruppendiskussion überarbeitet (vgl. Geisinger, 1994). Nach dieser Überarbeitung wurde der Entwurf von einer bilingualen Person in Englische rückübersetzt (vgl. Brislin, 1970). Mangelnde Übereinstimmungen der Rückübersetzungen mit den originalen Items wurden im Autorenteam diskutiert und die betroffenen deutschen Übersetzungen gegebenenfalls modifiziert. Die vorläufige Fassung des DIAF umfasste 16 Items zur Einschätzung des Verhaltens der direkten Führungskraft, die wie bei der Originalversion auf einer fünf-stufigen Likertskala (von

1 = *stimme gar nicht zu* bis 5 = *stimme vollkommen zu*) beantwortet werden sollten.

Vorgehen zur Prüfung

Die Überprüfung des DIAF erfolgte in drei Onlinestudien. Ziel war zunächst die Entwicklung einer Skala, die vier Komponenten authentischer Führung mit Items erfasst, deren Ladungsmuster eine Einfachstruktur aufweisen. So erfolgte anhand der Daten aus Studie 1 eine Überprüfung der faktoriellen Validität. Die Ergebnisse waren Anlass für eine grundlegende Überarbeitung der Skala, wobei einige Items umformuliert und andere neu aufgenommen wurden. Die überarbeitete Fassung des DIAF wurde in Studie 2 erneut hinsichtlich ihrer faktoriellen Validität geprüft. Durch den Ausschluss ungeeigneter Items und die sprachliche Schärfung weniger weiterer Items konnte die finale Version des DIAF erstellt werden. Mit dieser finalen Version konnten bereits in Studie 2 Analysen in Bezug auf Item- und Skalenkennwerte, Reliabilität und Validität der Skala durchgeführt werden. In Studie 3 erfolgte schließlich eine Kreuzvalidierung der deutschen Adaption, wobei auch weitere Befunde zu Konstrukt- und Kriteriumsvalidität ergänzt wurden.

Stichproben

An den drei Studien nahmen jeweils aktuell berufstätige Personen teil, die einer direkten Führungskraft unterstellt waren. Die Rekrutierung der Teilnehmenden für Studie 1 und Studie 2 erfolgte via Internet, beispielsweise über soziale Medien. Die Teilnehmenden erhielten keine Aufwandsentschädigung. Die Teilnehmenden an Studie 3 wurden über das SoSci Panel (Leiner, 2012) gewonnen und konnten jeweils einen von sechs Gutscheinen für einen Internethandel gewinnen.

In Studie 1 riefen $N = 473$ Personen die Startseite der Befragung auf, wovon 72.3% die Bearbeitung begannen und 42.3% die Befragung abschlossen. Das Alter dieser $n = 200$ Personen (81.5% weiblich) lag zwischen 18 und 63 Jahren ($M = 37.67$, $SD = 11.12$). Von den $N = 477$ Personen, welche die Startseite der Befragung in Studie 2 aufgerufen hatten, begannen 53.9% die Bearbeitung und schlossen 28.7% die Befragung ab. Das Alter dieser $n = 137$ Personen (67.9% weiblich) lag zwischen 19 und 63 Jahren ($M = 34.77$, $SD = 10.82$). In Studie 3 riefen $N = 538$ Personen die Startseite der Befragung auf, wovon 88.1% die Bearbeitung begannen und 69.3% die Befragung abschlossen. Das Alter dieser $n = 373$ Personen (61.7% weiblich) lag zwischen 19 und 65 Jahren ($M = 41.14$,

Tabelle 1. Demografische Variablen der Teilnehmenden in Studie 1 bis 3

Merkmale	Relative Häufigkeit (in %)		
	Studie 1 (n = 200)	Studie 2 (n = 137)	Studie 3 (n = 373)
<i>Dauer Tätigkeit im Arbeitsfeld</i>			
≤ 5 Jahre	33.5	38.7	31.1
6 bis 20 Jahre	45.0	48.9	46.6
> 20 Jahre	21.5	12.4	22.3
<i>Art der Beschäftigung</i>			
Vollzeit	71.0	65.0	70.8
Teilzeit	29.0	33.5	29.0
keine Angabe	0.0	1.5	0.3
<i>Dauer Zusammenarbeit mit Führungskraft</i>			
< 1 Jahr		21.2	15.3
1 bis 2 Jahre		25.5	20.6
3 bis 4 Jahre		25.5	22.0
5 bis 6 Jahre	^a	4.4	14.7
> 6 Jahre	31.0	23.4	27.3
<i>Geschlecht Führungskraft</i>			
weiblich	57.0	42.3	32.7

Anmerkung: ^a In Studie 1 wurde die Dauer der Zusammenarbeit mit der Führungskraft weniger differenziert erfasst. Von den Teilnehmenden arbeiteten 69.0% bis zu 6 Jahre mit ihrer Führungskraft zusammen.

SD = 11.23). Tabelle 1 zeigt weitere demografische Merkmale der Teilnehmenden in allen drei Studien.

Instrumente

Zur Bestimmung der Konstruktvalidität des DIAF wurden weitere Führungskonzepte erfasst. Zur Erhebung von ethischer Führung kam in Studie 3 die deutsche Adaptation der Ethical Leadership Scale (Rowold et al., 2009) zum Einsatz. Der Fragebogen erfasst die Facetten EMF (fünf Items, z. B. „Mein direkter Vorgesetzter denkt an die Interessen der Mitarbeiter.“) und ERM (vier Items, z. B. „Mein direkter Vorgesetzter gibt Beispiele, wie Dinge aus ethischer Sicht richtig gemacht werden sollten.“). Die Items werden auf einer fünf-stufigen Likertskala (von 1 = *nie* bis 5 = *immer*) beantwortet.

Abusive Supervision wurde in den Studien 2 und 3 mit der Skala von Tepper (2000) in der deutschen Fassung von Loock und Schilling (2010) erhoben. Der Fragebogen umfasst 15 Items (z. B. „Mein direkter Vorgesetzter macht sich über mich lustig.“), die auf einer fünf-stufigen Likertskala (von 0 = *nie* bis 4 = *sehr oft*) beantwortet werden sollen.

Zur Bestimmung der Kriteriumsvalidität wurden Maße zur Erfassung von Leistung, Arbeitsengagement, Wohlbefinden und der Beziehungsqualität zwischen Mitarbei-

tenden und Führungskraft eingesetzt. Leistung wurde über eine Selbsteinschätzung auf zwei Items (Robinson, 1996) in Studie 2 und 3 erfasst. Dabei bewerteten die Teilnehmenden auf einer fünf-stufigen Notenskala (von 5 = *mangelhaft* bis 1 = *sehr gut*) ihre Leistung aus eigener Perspektive („Wie schätzen Sie Ihre eigene berufliche Leistung ein?“) und gaben an, wie ihre Führungskraft ihre Leistung beurteilen würde. Da beide Items hinreichend hoch zusammenhingen (Studie 2 und 3: Cronbachs $\alpha = .59$), wurden sie im Rahmen der Auswertung zu einem Gesamtwert für die Arbeitsleistung gemittelt.

Arbeitsengagement wurde in Studie 2 mit der Utrecht Work Engagement Scale (UWES; Schaufeli & Bakker, 2003) erfasst, welche neun Items (z. B. „Ich gehe völlig in meiner Arbeit auf.“) umfasst, die auf einer sieben-stufigen Likertskala (von 0 = *nie* bis 6 = *immer*) eingeschätzt werden sollen.

Zur Erfassung des Wohlbefindens kam in den Studien 2 und 3 der Wohlbefindens-Index der WHO (WHO-5; Brähler, Mühlan, Albani & Schmidt, 2007) zum Einsatz. Dieser erfasst mit fünf Items (z. B. „In den letzten zwei Wochen war ich froh und guter Laune.“) das Wohlbefinden innerhalb der letzten zwei Wochen. Die Items werden auf einer sechs-stufigen Likertskala (von 0 = *zu keinem Zeitpunkt* bis 5 = *die ganze Zeit*) beantwortet.

Die Beziehungsqualität zwischen Führungskraft und Mitarbeitendem wurde in Studie 3 mit der Leader-Mem-

ber-Exchange Skala (LMX7; Schyns & Paul, 2014) erfasst. Die sieben Items (z. B. „Wie würden Sie das Arbeitsverhältnis mit Ihrem Vorgesetzten beschreiben?“) werden auf fünf-stufigen Likertskalen mit unterschiedlichen verbalen Verankerungen (z. B. von 1 = *sehr ineffektiv* bis 5 = *sehr effektiv*) beantwortet.

Die ermittelten internen Konsistenzen aller Instrumente sind in den Tabellen 5 und 6 dargestellt.

Analysen

In Studie 1 traten aufgrund technischer Fehlspezifikationen vereinzelt Fehlwerte auf. Aus diesem Grund lag die Stichprobengröße in den Analysen für Studie 1 bei $n = 197$. In keiner der Studien lagen univariate Ausreißer mit Werten von $z \geq 3.29$ (Field, 2009) vor. Den Empfehlungen von Tabachnick und Fidell (2007) folgend wurde nach einer Analyse anhand von Mahalanobis-Distanzen in Studie 2 und 3 jeweils eine Person als multivariater Ausreißer ausgeschlossen, sodass die Stichprobengröße in Studie 2 bei $n = 136$ und $n = 372$ in Studie 3 lag.

Die Prüfung der faktoriellen Validität erfolgte in CFA mit dem Programm IBM SPSS AMOS (Version 23.0; Arbuckle, 2014). Unter Verwendung der ML-Schätzmethode diente der χ^2 -Test zur Prüfung des globalen Modell-Fits. Da der Test mit steigender Stichprobengröße bereits bei kleineren Abweichungen vom postulierten Modell signifikant wird, betrachteten wir auch andere Fit-Indizes: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) und Standardized Root Mean Square Residual (SRMR). Zur Interpretation der Indizes orientierten wir uns an den vorgeschlagenen Cutoff-Werten von Hu und Bentler (1999): Indizes CFI und TLI jeweils $\geq .95$, RMSEA $< .06$ (bei $N > 250$) bzw. $< .08$ (bei $N \leq 250$) und SRMR $< .11$. Wir betrachteten die Fit-Indizes für unterschiedliche theoretisch abgeleitete Modelle, deren Fit wir zudem mit einem χ^2 -Differenz-Test verglichen.

Um die faktorielle Binnenstruktur des DIAF noch differenzierter abzubilden, berechneten wir zudem in allen drei Studien explorative Faktorenanalysen (EFA) in Form von Hauptachsenanalysen mit obliquer Rotation (Promax), da von korrelierten Faktoren auszugehen war. Als Abbruchkriterium für die Faktorenextraktion verwendeten wir den Scree-Test. Ziel war, eine Einfachstruktur der Faktorladungen zu erzielen (d. h. Nebenladungen $\lambda < .30$; vgl. Kline, 1994).

Zur weiteren Prüfung der Konstruktvalidität des DIAF wurden Produkt-Moment-Korrelationen zu Variablen mit unterschiedlicher Konstruktnähe berechnet. In Studie 3 konnten diese Zusammenhänge mit einem Test für abhängige Korrelationen von Lenhard und Lenhard (2014)

verglichen werden, um die Hypothesen 1b und 1c zu prüfen. Ferner betrachteten wir die korrelativen Zusammenhänge zwischen dem DIAF und demografischen Variablen der Teilnehmenden. Die Prüfung der Kriteriumsvalidität erfolgte über die Berechnung von Produkt-Moment-Korrelationen zu relevanten Ergebnismaßen. Zudem erfolgte eine Bestimmung der inkrementellen Validität in multiplen hierarchischen Regressionsanalysen. Dabei wurde untersucht, ob das DIAF (Schritt 3) über etablierte Führungsinstrumente (Schritt 2) und Kontrollvariablen (Schritt 1) hinaus zusätzliche Varianz in relevanten Kriterien aufklärte.

Ergebnisse

Faktorielle Validität

Ziel von Studie 1 war es, die faktorielle Binnenstruktur des DIAF zu prüfen. In CFA wurden drei Modelle gegeneinander getestet: (1) Ein Modell mit vier interkorrelierten Faktoren, die die Komponenten authentischer Führung widerspiegeln (vgl. Walumbwa et al., 2008), (2) ein Modell mit vier Faktoren erster Ordnung und einem Faktor zweiter Ordnung (Gesamtwert authentische Führung) und (3) ein Modell mit nur einem Faktor, auf dem alle Items luden. Wie Tabelle 2 zeigt, wies Modell 1 mit vier interkorrelierten Faktoren (mittlere Korrelation: $\bar{r} = .89$) im Vergleich einen signifikant besseren Fit auf als die Modelle 2 und 3, wenngleich die Fit-Indizes absolut gesehen unter den empfohlenen Cutoff-Werten von Hu und Bentler (1999) zurückblieben.

Da die angenommene Faktorenstruktur die Daten nur unzureichend abbildete, führten wir eine EFA durch, um die Binnenstruktur der Daten zu erfassen. Diese erbrachte zwei Faktoren (Eigenwertverlauf: Faktor 1 = 7.18; Faktor 2 = 1.54; Faktor 3 = 0.98; Faktor 4 = 0.81), die 48.0 % der Varianz aufklärten und zu $r = .68$ korrelierten. Wie Tabelle 3 zeigt, luden die Items der Komponenten SB und AIV überwiegend auf dem ersten Faktor, während die Items der Komponente VMP auf dem zweiten Faktor luden. Die Items der Komponente TiB konnten keinem Faktor eindeutig zugeordnet werden.

Aufgrund der unzureichenden faktoriellen Validität der Skala wurden einzelne Items sprachlich überarbeitet (vgl. Tabelle 3). Darüber hinaus wurden auf Basis inhaltlich-theoretischer Überlegungen zwei neue Items formuliert, um vor allem die Komponenten TiB sowie SB, deren Items vergleichsweise niedrige Hauptladungen aufwiesen, inhaltlich zu schärfen.

Eine EFA der Daten aus Studie 2 ergab anhand des Scree-Kriteriums erneut zwei Faktoren (Eigenwertverlauf:

Tabelle 2. Ergebnisse der konfirmatorischen Faktorenanalysen (Studie 1 bis 3)

Modell	χ^2 Test ^a			CFI	TLI	RMSEA	SRMR	χ^2 Differenz-Test		
	χ^2	df	p					$\Delta\chi^2$	Δdf	p
Studie 1 (n = 197) ^b										
1	246.08	98	.001	.90	.88	.09	.06			
2	253.11	100	.001	.89	.87	.09	.06	7.03	2	.030
3	309.93	104	.001	.86	.84	.10	.07	63.85	6	<.001
Studie 2 (n = 136)										
1	157.83	98	.067	.96	.95	.07	.04			
2	174.33	100	.028	.95	.94	.07	.05	16.50	2	<.001
3	300.29	104	.001	.86	.84	.12	.07	142.46	6	<.001
Studie 3 (n = 372)										
1	292.79	98	.001	.95	.93	.07	.04			
2	309.40	100	.001	.94	.93	.08	.05	16.61	2	<.001
3	708.98	104	.001	.83	.80	.13	.08	399.58	6	<.001

Anmerkungen: Modell 1 = vier interkorrelierte Faktoren; Modell 2 = vier Faktoren erster Ordnung, ein Faktor zweiter Ordnung; Modell 3 = ein Faktor. ^a p-Wert nach Bollen-Stine-Bootstrap-Korrektur. ^b Einsatz einer Vorläuferversion des DIAF. Aufgrund einzelner fehlender Werte lag der Stichprobenumfang in Studie 1 bei n = 197. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Faktor 1 = 9.69; Faktor 2 = 1.44; Faktor 3 = 0.97; Faktor 4 = 0.84), die 57.0 % der Varianz aufklärten und zu $r = .73$ korrelierten. Die Items zeigten ein vergleichbares Ladungsmuster wie in Studie 1 (vgl. Tabelle 3). Alle ursprünglichen Items der Skala SB luden nun höher auf dem ersten Faktor, weshalb das zusätzlich aufgenommene Item für diese Komponente wieder ausgeschlossen wurde. Auf Faktor 1 luden erneut die Items der Komponente AIV, während die Items der Komponente VMP auf dem zweiten Faktor luden. Wie bereits in Studie 1 wiesen die Items der Komponente TiB keine Einfachstruktur auf. Besonders auffällig war das Item „(...) spricht mit anderen darüber, wenn ihm/ihr Fehler unterlaufen sind.“, das erneut eine starke Ladung auf dem ersten Faktor aufwies, während die anderen Items dieser Komponente mehrheitlich auf dem zweiten Faktor luden. Da sich dieses Item bereits in der amerikanischen Originalversion der Skala (Neider & Schriesheim, 2011) als problematisch erwiesen hatte, wurde es auch in der deutschen Adaption ausgeschlossen. Als Ersatz wurde das neu formulierte Item „(...) sagt offen, wie er/sie sich wirklich fühlt, und verstellt sich nicht“ final aufgenommen. Ein anderes Item, das im amerikanischen Original ausgeschlossen worden war (Item 1; vgl. Tabelle 3), wies in der deutschen Adaption akzeptable Faktorladungen auf und wurde deshalb beibehalten.

Wie Tabelle 2 zeigt, wirkten sich die sprachlichen Umformulierungen und Substitutionen positiv auf die faktorielle Validität der Skala aus. In CFA mit den finalen 16 Items des DIAF erwies sich das Modell 1 mit vier inter-

korrelierten Faktoren ($\bar{r} = .83$) als gut geeignet, die gewonnenen Daten aus Studie 2 abzubilden. Alle Fit-Indizes lagen unter den von Hu und Bentler (1999) empfohlenen Cutoff-Werten und der χ^2 -Test war nicht signifikant. Modell 1 erklärte die Datenstruktur zudem signifikant besser als die beiden Alternativmodelle 2 und 3, die wir testeten, wobei auch Modell 2 mit vier Faktoren erster Ordnung und einem übergeordneten Faktor zweiter Ordnung zufriedenstellende Fit-Indizes aufwies.

Die Ergebnisse von Studie 2 bildeten die Basis, um die finale Version des DIAF zu erstellen. Dabei wurden für Studie 3 noch einmal drei Items minimal sprachlich überarbeitet (z. B. Ersetzung des Begriffs „Grundüberzeugung“ durch „Überzeugung“). Die finale Version der Skala erfasste vier Komponenten von authentischer Führung mit jeweils vier Items, wobei ein Originalitem der Komponente TiB durch eine Neuformulierung ersetzt wurde.

Die finale Version wurde in Studie 3 noch einmal hinsichtlich ihrer faktoriellen Validität geprüft. Erneut zeigten die CFA, dass ein Modell mit vier interkorrelierten Faktoren ($\bar{r} = .82$) einen signifikant besseren Fit als die zwei Alternativmodelle aufwies (vgl. Tabelle 2), wengleich die Fit-Indizes TLI und RMSEA leicht unter den Cutoff-Werten von Hu und Bentler (1999) zurückblieben. Der Modell-Fit war dennoch insgesamt als befriedigend einzuschätzen. Eine EFA ergab erneut eine zweifaktorielle Lösung (Eigenwertverlauf: Faktor 1 = 8.16; Faktor 2 = 1.50; Faktor 3 = 0.86; Faktor 4 = 0.78), wobei die Faktoren 55.1 % der Varianz aufklärten und zu $r = .67$ korrelierten. Von einer Ausnahme abgesehen (Item 13;

Tabelle 3. Faktorladungen der Items auf Basis der exploratorischen Faktorenanalysen (Studie 1 bis 3)

Nr.	Itemwortlaut	Faktor 1			Faktor 2		
		λ_{11}	λ_{12}	λ_{13}	λ_{21}	λ_{22}	λ_{23}
<i>Selbstbewusstheit (SB)</i>							
1	erbittet Rückmeldung, um seinen/ihren Umgang mit anderen zu verbessern.	.67	.73	.62	-.10	-.08	-.06
5	kann genau beschreiben, wie seine/ihre Fähigkeiten durch andere beurteilt werden.	.55	.64	.75	.16	.14	-.02
8	macht deutlich, dass er/sie versteht, wo seine/ihre Stärken und Schwächen liegen.	.63	.73	.78	.16	.13	-.06
12	ist sich selbst klar bewusst, welche Wirkung er/sie auf andere hat. ¹	-.09	.46	.46	.55	.17	.23
<i>Transparenz in Beziehungen zu anderen (TiB)</i>							
2	sagt offen das, was er/sie wirklich denkt, sodass man weiß, woran man ist. ¹	.30	.30	.57	.39	.48	.17
9	schaft Transparenz und Klarheit, indem er/sie andere offen über Anliegen informiert, die ihm/ihr wichtig sind. ¹²	.36	.39	.70	.38	.49	.13
13	bringt anderen gegenüber klar zum Ausdruck, welche Gedanken und Ideen er/sie hat. ²	.04	.48	.53	.66	.26	.30
16	sagt offen, wie er/sie sich wirklich fühlt, und verstellt sich nicht. ^{a2}	–	.32	.59	–	.43	.15
<i>Verinnerlichte moralische Perspektive (VMP)</i>							
3	handelt in Übereinstimmung mit seinen/ihren Überzeugungen und Werten. ¹	.19	.02	.08	.53	.83	.74
6	stützt sich auf seine/ihre Überzeugungen und Werte, um wichtige Entscheidungen zu treffen. ¹	.06	-.03	-.02	.57	.82	.77
10	widersetzt sich, wenn er/sie zu Dingen gedrängt wird, die im Widerspruch zu seinen/ihren Überzeugungen und Werten stehen. ¹	-.26	-.08	-.10	.76	.57	.69
14	richtet seine/ihre Handlungen nach seinen/ihren moralischen Überzeugungen und Werten aus. ¹	.29	-.11	-.01	.46	.89	.87
<i>Ausgewogene Informationsverarbeitung (AIV)</i>							
4	fragt aktiv auch nach Standpunkten, die seine/ihre Überzeugungen in Frage stellen. ²	.90	.79	.91	-.19	.07	-.12
7	hört sich alternative Standpunkte sorgfältig an, bevor er/sie Schlüsse zieht.	.85	.86	.80	-.06	-.04	-.05
11	analysiert erst objektiv alle relevanten Informationen, bevor er/sie eine Entscheidung trifft.	.63	.60	.56	.12	.05	.17
15	ermutigt andere, offen auch gegensätzliche Sichtweisen zu äußern.	.85	.84	.89	-.02	.03	-.08

Anmerkungen: Die Items beginnen mit „Mein direkter Vorgesetzter / Meine direkte Vorgesetzte ...“. λ_{ij} = Ladung auf dem ersten Faktor in Studie $j = 1, 2$ oder 3 ; λ_{2j} = Ladung auf dem zweiten Faktor in Studie $j = 1, 2$ oder 3 . Fett hervorgehobene Werte kennzeichnen Hauptladungen. ¹ Nach Studie 1 sprachlich überarbeitet. ² Nach Studie 2 sprachlich überarbeitet. ^a Item wurde nach Studie 1 aufgenommen.

vgl. Tabelle 3), wiesen alle Items eine Einfachstruktur auf. Während die Items der Komponenten SB und AIV wie zuvor auf dem ersten Faktor luden, traf das nun auch vollständig auf die Items der Komponente TiB zu. Der Item-Ausschluss hatte demnach einen deutlich schärfenden Effekt auf die inhaltliche Bedeutung der Komponenten. Die Items der Komponente VMP luden wie in den beiden Studien zuvor auf dem zweiten Faktor.

Deskriptive Statistik der Items, Komponenten und Gesamtskala

Tabelle 4 zeigt relevante Kennwerte (deskriptive Statistik, Trennschärfen) der einzelnen Items in den Studien 2 und 3. Die deskriptiven Statistiken für die Komponenten bzw. Gesamtskala sind in den Tabellen 5 und 6 dargestellt. Die Schwierigkeiten der Items, Komponenten und Gesamtskala lagen in beiden Studien überwiegend im anzustrebenden mittleren Bereich. Einzig die Komponente VMP

hatte eine etwas geringere Schwierigkeit ($3.45 \leq M \leq 3.92$). In beiden Studien wiesen alle Items eine Schiefe $< |2|$ und Kurtosis $< |7|$ auf, was nach West, Finch und Curran (1995) für eine wünschenswerte Normalverteilung bei allen Items spricht. Wie Tabelle 4 zeigt, lagen die Trennschärfen der Items in Bezug auf die Gesamtskala mit $.42 \leq r_{it-G} \leq .80$ (Studie 2) bzw. $.44 \leq r_{it-G} \leq .77$ (Studie 3) in einem sehr guten Bereich (vgl. Kelava & Moosbrugger, 2012). Auch in Bezug auf die einzelnen Komponenten erwiesen sich die Trennschärfen mit Werten von $r_{it-K} \geq .47$ in beiden Studien als sehr gut.

Reliabilität

Die interne Konsistenz der Gesamtskala lag mit einem Cronbachs Alpha von .94 (Studie 2) bzw. .93 (Studie 3) in einem sehr guten Bereich. Auch die einzelnen Komponenten wiesen mit $.83$ (SB) $\leq \alpha \leq .88$ (AIV) in Studie 2 bzw. $.77$ (SB) $\leq \alpha \leq .87$ (AIV) in Studie 3 unter Berücksichtigung

Tabelle 4. Itemkennwerte des DIAF in den Studien 2 und 3

Studie 2 (n = 136)									Studie 3 (n = 372)							
Item	Min	Max	M	SD	r_{it-G}	r_{it-K}	Skew	Kurt	Min	Max	M	SD	r_{it-G}	r_{it-K}	Skew	Kurt
<i>Selbstbewusstheit (SB)</i>																
1	1.00	5.00	2.77	1.45	.60	.66	0.20	-1.30	1.00	5.00	2.53	1.25	.53	.47	0.45	-0.80
5	1.00	5.00	2.86	1.18	.72	.71	-0.11	-0.92	1.00	5.00	2.67	1.02	.70	.67	-0.01	-0.71
8	1.00	5.00	3.16	1.17	.80	.71	-0.18	-0.86	1.00	5.00	2.97	1.13	.68	.63	-0.09	-0.81
12	1.00	5.00	3.33	1.27	.59	.60	-0.25	-0.97	1.00	5.00	3.06	1.11	.62	.52	-0.21	-0.79
<i>Transparenz in Beziehungen zu anderen (TiB)</i>																
2	1.00	5.00	3.61	1.26	.69	.70	-0.53	-0.76	1.00	5.00	3.37	1.17	.68	.71	-0.39	-0.65
9	1.00	5.00	3.54	1.24	.79	.74	-0.49	-0.77	1.00	5.00	3.15	1.31	.76	.73	-0.23	-1.11
13	1.00	5.00	3.68	1.10	.67	.66	-0.41	-0.65	1.00	5.00	3.42	1.10	.74	.71	-0.42	-0.54
16	1.00	5.00	3.34	1.35	.67	.70	-0.27	-1.13	1.00	5.00	3.05	1.20	.67	.70	-0.14	-0.85
<i>Verinnerlichte moralische Perspektive (VMP)</i>																
3	1.00	5.00	3.92	1.06	.74	.71	-0.78	-0.15	1.00	5.00	3.77	1.02	.66	.69	-0.73	0.23
6	1.00	5.00	3.81	1.11	.69	.73	-0.83	0.01	1.00	5.00	3.68	1.01	.58	.70	-0.55	-0.21
10	1.00	5.00	3.61	1.06	.42	.49	-0.44	-0.19	1.00	5.00	3.45	1.12	.44	.56	-0.38	-0.55
14	1.00	5.00	3.79	1.06	.66	.75	-0.71	-0.11	1.00	5.00	3.56	1.04	.66	.77	-0.56	-0.22
<i>Ausgewogene Informationsverarbeitung (AIV)</i>																
4	1.00	5.00	2.88	1.32	.78	.76	0.06	-1.14	1.00	5.00	2.73	1.24	.75	.72	0.24	-0.97
7	1.00	5.00	3.44	1.21	.75	.80	-0.32	-0.85	1.00	5.00	3.23	1.15	.71	.78	-0.29	-0.74
11	1.00	5.00	3.39	1.14	.60	.64	-0.33	-0.71	1.00	5.00	3.10	1.12	.66	.63	-0.23	-0.63
15	1.00	5.00	3.40	1.29	.79	.78	-0.38	-0.99	1.00	5.00	3.02	1.27	.77	.78	-0.06	-1.05

Anmerkungen: Die Items werden auf einer fünf-stufigen Likertskala beantwortet (1 = *stimme gar nicht zu*, 2 = *stimme eher nicht zu*, 3 = *teils/teils*, 4 = *stimme eher zu*, 5 = *stimme vollkommen zu*). r_{it-G} = Trennschärfe in Bezug auf die Gesamtskala. r_{it-K} = Trennschärfe in Bezug auf die Komponente, aus der ein Item stammt. Skew = Schiefe. Kurt = Exzess.

der geringen Itemzahl pro Komponente gute bis sehr gute interne Konsistenzen auf (vgl. Tabelle 5 bzw. Tabelle 6).

Zusammenhänge zu anderen Führungsformen

Die Korrelationen zu anderen Führungsformen sind in Tabelle 5 für Studie 2 und Tabelle 6 für Studie 3 dargestellt. Der Gesamtwert authentischer Führung und der Gesamtwert ethischer Führung korrelierten in Studie 3 zu $r = .84$ miteinander, was erwartungskonform für große Überschneidungen zwischen beiden Führungskonzepten sprach (Hypothese 1a). Wie in Hypothese 1b angenommen, wiesen die Komponenten SB ($r = .75$), TiB ($r = .76$) und AIV ($r = .82$) signifikant höhere Zusammenhänge zur Facette EMF als zur Facette ERM ($.60 \leq r \leq .69$) von ethischer Führung auf, $5.12 \leq z \leq 6.76$ mit allen $p < .001$. Im Gegensatz zu den Erwartungen in Hypothese 1c korrelierte die Komponente VMP nicht höher mit der Facette ERM ($r = .60$) als mit der Facette EMF ($r = .61$), $z = 0.38$, $p = .354$.

Der Gesamtwert des DIAF korrelierte in den Studien 2 bzw. 3 zu $r = -.61$ bzw. $r = -.62$ mit Abusive Supervision,

was Hypothese 2 bestätigte. Auf Ebene der Komponenten korrelierte AIV mit Koeffizienten von $r = -.62$ (Studie 2) bzw. $r = -.65$ (Studie 3) am stärksten mit Abusive Supervision, während die übrigen Komponenten mit Werten zwischen $-.47 \leq r \leq -.53$ in Studie 2 bzw. $-.41 \leq r \leq -.55$ in Studie 3 niedriger mit diesem Führungskonzept zusammenhängen. Die Korrelationen verdeutlichten, dass beide Konzepte Überschneidungen aufwiesen, aber keinesfalls deckungsgleich waren.

Zusammenhänge zu demografischen Variablen

Die Korrelationen zwischen den Kontrollvariablen und den Führungsskalen sowie den erhobenen Kriterien sind in Tabelle 5 für Studie 2 und Tabelle 6 für Studie 3 dargestellt. In Studie 3 korrelierte das Geschlecht der Teilnehmenden positiv mit der Bewertung der Komponente VMP ($r = .14$, $p = .007$), was tendenziell, allerdings nicht signifikant auch in Studie 2 der Fall war ($r = .16$, $p = .071$): Frauen bewerteten diese Komponente bei ihren Führungskräften demnach höher als Männer. Das Geschlecht der Führungskraft korrelierte in beiden Studien nicht si-

Tabelle 5. Deskriptive Statistiken, Reliabilität und Interkorrelationen der eingesetzten Skalen in Studie 2 (*n* = 136)

	Min	Max	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Geschlecht	1.00	2.00	1.68	0.47	–												
2. Alter	19.00	63.00	34.87	10.79	-.18	–											
3. Geschlecht Führungskraft	1.00	2.00	1.43	0.50	.20	-.16	–										
4. Arbeitszeit ^a	1.00	2.00	1.34	0.48	.36	-.24	.32	–									
5. SB	1.00	5.00	3.03	1.04	.06	-.15	.17	.15	(.83)								
6. TIB	1.00	5.00	3.54	1.04	.09	-.20	.16	.24	.69	(.86)							
7. VMP	1.00	5.00	3.78	0.88	.16	-.15	.02	.21	.59	.70	(.84)						
8. AIV	1.00	5.00	3.28	1.07	-.01	-.09	.06	.17	.76	.72	.62	(.88)					
9. DIAF Gesamtwert	1.38	4.94	3.41	0.88	.08	-.17	.12	.22	.88	.89	.82	.90	(.94)				
10. Abusive Supervision	1.00	3.87	1.55	0.66	.07	.01	-.01	-.15	-.53	-.51	-.47	-.62	-.61	(.93)			
11. Wohlbefinden	1.60	5.80	3.83	0.97	.03	.06	.05	.02	.32	.46	.34	.41	.44	-.34	(.87)		
12. Arbeitsengagement	1.33	7.00	4.73	1.17	.05	.01	.10	.00	.34	.44	.37	.42	.45	-.22	.60	(.93)	
13. Leistung	2.00	5.00	4.05	0.55	.04	.04	-.01	.04	.16	.27	.21	.16	.23	-.13	.33	.39	(.59)

Anmerkungen: In der Diagonalen ist die interne Konsistenz (Cronbachs Alpha) der (Sub-)Skalen angegeben. Kodierung Geschlecht und Geschlecht Führungskraft: 1 = männlich, 2 = weiblich; Kodierung Arbeitszeit: 1 = Vollzeit, 2 = nicht Vollzeit. ^a *n* = 134 aufgrund einzelner fehlender Angaben. SB = Selbstbewusstheit; TIB = Transparenz in Beziehungen zu anderen; VMP = Verinnerlichte moralische Perspektive; AIV = Ausgewogene Informationsverarbeitung; DIAF = Deutsches Inventar Authentischer Führung. Korrelationen von *r* ≥ |.18| sind auf dem 5%-Niveau (*p* < .05), von *r* ≥ |.23| auf dem 1 %-Niveau (*p* < .01) und von *r* ≥ |.31| auf dem 0.1 %-Niveau (*p* < .001) signifikant (2-seitig).

Tabelle 6. Deskriptive Statistiken, Reliabilität und Interkorrelationen der eingesetzten Skalen in Studie 3 ($n = 372$)

	Min	Max	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Geschlecht	1.00	2.00	1.62	0.49	–															
2. Alter	19.00	65.00	41.12	11.24	-.12	–														
3. Geschlecht Führungskraft	1.00	2.00	1.33	0.47	.16	-.02	–													
4. Arbeitszeit ^a	1.00	2.00	1.29	0.45	.28	-.11	.16	–												
5. SB	1.00	5.00	2.81	0.87	.04	-.01	-.05	.01	(.77)											
6. TiB	1.00	5.00	3.25	1.01	.02	-.06	-.03	.11	.72	(.86)										
7. VMP	1.00	5.00	3.61	0.86	.14	-.03	.08	.06	.54	.61	(.84)									
8. AIV	1.00	5.00	3.02	1.02	.02	.01	-.05	.06	.79	.74	.54	(.87)								
9. DIAF Gesamtwert	1.00	4.94	3.17	0.81	.06	-.03	-.01	.07	.88	.90	.77	.90	(.93)							
10. EMF	1.00	5.00	3.28	0.90	.02	-.03	.02	.08	.75	.76	.61	.82	.85	(.90)						
11. ERM	1.00	5.00	2.88	0.95	.14	.03	.08	.09	.63	.60	.60	.69	.73	.78	(.84)					
12. Ethische Führung Gesamtwert	1.00	4.89	3.10	0.87	.08	.00	.05	.09	.74	.73	.64	.80	.84	.95	.93	(.92)				
13. Abusive Supervision	1.00	4.60	1.61	0.64	.02	-.11	.07	-.06	-.55	-.51	-.41	-.65	-.62	-.69	-.53	-.66	(.93)			
14. Wohlbefinden	1.00	6.00	3.75	1.09	.05	.25	-.06	-.03	.36	.28	.23	.34	.35	.41	.34	.40	-.41	(.88)		
15. Leistung	2.50	5.00	4.08	0.61	.08	.11	.09	.03	.32	.25	.20	.32	.32	.30	.29	.31	-.33	.37	(.59)	
16. LMX	1.00	5.00	3.40	0.85	-.01	.02	-.03	.04	.77	.74	.59	.80	.84	.83	.66	.79	-.69	.46	.36	(.91)

Anmerkungen: In der Diagonalen ist die interne Konsistenz (Cronbachs Alpha) der (Sub-)Skalen angegeben. Kodierung Geschlecht und Geschlecht Führungskraft: 1 = männlich, 2 = weiblich; Kodierung Arbeitszeit: 1 = Vollzeit, 2 = nicht Vollzeit. ^a $n = 371$ aufgrund einzelner fehlender Angaben. SB = Selbstbewusstheit; TiB = Transparenz in Beziehungen zu anderen; VMP = Verinnerlichte moralische Perspektive; AIV = Ausgewogene Informationsverarbeitung; DIAF = Deutsches Inventar Authentischer Führung; EMF = Ethische Mitarbeiterführung; ERM = Ethisches Rollenmodell; LMX = Leader-Member-Exchange. Korrelationen von $r \geq .11$ sind auf dem 5%-Niveau ($p < .05$), von $r \geq .14$ auf dem 1%-Niveau ($p < .01$) und von $r \geq .20$ auf dem 0.1%-Niveau ($p < .001$) signifikant (2-seitig).

gnifikant mit den Komponenten oder dem Gesamtwert des DIAF, wobei in Studie 2 zu den Komponenten SB ($r = .17, p = .051$) und TiB ($r = .16, p = .065$) die höchsten Zusammenhänge auftraten. Das Alter der Teilnehmenden korrelierte in Studie 2 negativ mit der Bewertung authentischer Führung ($-.20 \leq r \leq -.09$), wobei einzig der Zusammenhang zur Komponente TiB signifikant war ($p = .022$, andere $p \geq .053$). In Studie 3 traten keine Zusammenhänge zum Alter der Teilnehmenden auf ($-.06 \leq r \leq .01, p \geq .215$). Deutliche Zusammenhänge ergaben sich in Studie 2 zwischen der Arbeitszeit der Teilnehmenden und ihren Bewertungen: Personen, die nicht Vollzeit arbeiteten, bewerteten ihre Führungskraft signifikant höher bezüglich der Komponenten TiB ($r = .24$) und VMP ($r = .21$) sowie in Bezug auf den Gesamtwert ($r = .22$, alle $p \leq .015$). In Studie 3 trat ein signifikanter Zusammenhang nur zur Komponente TiB auf ($r = .11, p = .036$). Hypothese 3 konnte daher weitgehend, aber nicht vollständig bestätigt werden. Aufgrund der teilweise beobachteten Zusammenhänge wurden Alter, Geschlecht und Arbeitszeit der Teilnehmenden sowie das Geschlecht der Führungskraft in den Regressionsanalysen zur Bestimmung der Kriteriumsvalidität als Kontrollvariablen berücksichtigt.

Zusammenhänge zu Kriterien

Wie Tabelle 5 zeigt, korrelierten in Studie 2 die Komponenten TiB ($r = .27, p = .002$) und VMP ($r = .21, p = .014$) ebenso wie der Gesamtwert des DIAF ($r = .23, p = .008$) signifikant positiv mit der selbsteingeschätzten Leistung, während die Komponenten SB ($r = .16, p = .056$) und AIV ($r = .16, p = .072$) ebenso positiv, aber nicht signifikant mit dem Kriterium korrelierten. In Studie 3 (vgl. Tabelle 6) bestanden zwischen allen Komponenten sowie dem Gesamtwert des DIAF signifikante positive Korrelationen zur selbsteingeschätzten Leistung ($.20 \leq r \leq .32, p < .001$). Die Ergebnisse unterstützten überwiegend Hypothese 4.

In Studie 2 traten signifikante positive Zusammenhänge der Komponenten und des Gesamtwertes des DIAF zum Arbeitsengagement auf ($.34 \leq r \leq .45, p < .001$), was Hypothese 5 bestätigte. Die Komponenten und der Gesamtwert des DIAF korrelierten in beiden Studien zudem signifikant positiv mit dem Wohlbefinden der Teilnehmenden (Studie 2: $.32 \leq r \leq .44, p < .001$; Studie 3: $.23 \leq r \leq .36, p < .001$), was Hypothese 6 unterstützte.

Zum Kriterium der Beziehungsqualität zwischen Führungskraft und Mitarbeitenden bestanden in Studie 3 signifikante positive Korrelationen im hohen Bereich ($.59 \leq r \leq .84, p < .001$), was Hypothese 7 bestätigte.

In multiplen hierarchischen Regressionsanalysen wurde die inkrementelle Validität des DIAF über etablierte

Führungsskalen hinaus untersucht. Der Gesamtwert des DIAF erklärte bei allen eingesetzten Kriterien in Studie 2 und 3 einen statistisch bedeutsamen Varianzanteil (zwischen 2% und 28%) über Abusive Supervision hinaus (vgl. Tabelle 7). In Studie 3 erklärte der Gesamtwert des DIAF zudem bei den Kriterien Leistung und LMX einen signifikanten Varianzanteil über ethische Führung hinaus (vgl. Tabelle 8). Einzig beim Kriterium Wohlbefinden zeigte das DIAF keine inkrementelle Validität über ethische Führung. Hypothese 8a konnte somit überwiegend, aber nicht vollständig bestätigt werden. Wie in Hypothese 8b angenommen, war der zusätzliche Erklärungsbeitrag des DIAF über andere Führungskonzepte hinaus in Studie 3 sowohl beim Kriterium Leistung (2% vs. 1%) als auch beim Kriterium LMX (28% vs. 10%) für Abusive Supervision größer als für ethische Führung.

Diskussion

Ziel des vorliegenden Beitrages war es, mit einer deutschen Adaption des *Authentic Leadership Inventory* (ALI) von Neider und Schriesheim (2011) für den deutschen Sprachraum ein umfassend geprüftes Instrument zur Erfassung authentischer Führung zur Verfügung zu stellen. Das *Deutsche Inventar Authentischer Führung* (DIAF) erfasst mit 16 Items vier Komponenten authentischer Führung als Fremdeinschätzung der direkten Führungskraft. Die vier Komponenten *Selbstbewusstheit* (SB), *Transparenz in Beziehungen zu anderen* (TiB), *verinnerlichte moralische Perspektive* (VMP) und *ausgewogene Informationsverarbeitung* (AIV) entsprechen den von Walumbwa et al. (2008) theoretisch postulierten Aspekten authentischer Führung.

Das DIAF wurde in drei Studien geprüft und wies zufriedenstellende psychometrische Eigenschaften auf. Die Reliabilität der Gesamtskala und ihrer Komponenten lag im guten bis sehr guten Bereich. Die Items erwiesen sich bei hervorragenden Trennschärfen und überwiegend mittlerer Schwierigkeit als gut geeignet, um adäquat zwischen Personen zu differenzieren. Die Itemanzahl des DIAF ist im Vergleich zur finalen ALI-Skala mit 16 vs. 14 Items etwas größer. Das DIAF zeigte erwartungskonform positive Zusammenhänge zu ethischer Führung und negative Zusammenhänge zu Abusive Supervision, was für die Konstruktvalidität der Skala spricht. Die Berücksichtigung von Abusive Supervision als negative Seite von Führung ist zudem eine wichtige inhaltliche Erweiterung im Vergleich zum Validierungsansatz von Neider und Schriesheim (2011). In Bezug auf die Kriteriumsvalidität ergaben sich erwartungskonforme positive Zusammenhänge zu behavioralen (Arbeitsleistung), affektiven (Arbeitsengagement, Wohlbefinden) und relationalen (LMX)

Tabelle 7. Inkrementelle Validität des DIAF gegenüber Abuseive Supervision bei verschiedenen Kriterien (Studie 2 und 3)

	Studie 2						Studie 3					
	Engagement		Wohlbefinden		Leistung		Wohlbefinden		Leistung		Beziehung	
	β	$R^2/\Delta R^2$	β	$R^2/\Delta R^2$	β	$R^2/\Delta R^2$	β	$R^2/\Delta R^2$	β	$R^2/\Delta R^2$	β	$R^2/\Delta R^2$
Schritt 1												
Geschlecht	.08	.01/.01	.05	.01/.01	.05	.01/.01	.10	.08/.08***	.08	.03/.03	-.01	.00/.00
Alter	-.00		.06		.04		.26***		.12*		.03	
Geschlecht Führungskraft	.07		.02		-.04		-.07		.08		-.03	
Arbeitszeit	-.05		.01		.04		-.02		.00		.05	
Schritt 2												
Geschlecht	.10	.04/.03	.09	.11/.10***	.06	.02/.01***	.11*	.23/.15***	.09	.14/.11***	.00	.48/.48***
Alter	.00		.06		.04		.21***		.08		-.06	
Geschlecht Führungskraft	.08		.03		-.04		-.04		.11*		.03	
Arbeitszeit	-.08		-.05		.02		-.05		-.03		-.02	
Abusive Supervision	-.17		-.33***		-.11		-.39***		-.34***		-.70***	
Schritt 3												
Geschlecht	.06	.21/.18***	.06	.20/.10***		.05/.04*	.10*	.25/.02**	.08	.16/.02**	-.04	.76/.78***
Alter	.07		.11		.07		.23***		.10		.00	
Geschlecht Führungskraft	.05		.01		-.05		-.05		.10*		.01	
Arbeitszeit	-.12		-.08		.01		-.05		-.03		-.02	
Abusive Supervision	.14		-.10		.04		-.28***		-.22***		-.28***	
DIAF	.53***		.40***		.25*		.19**		.19**		.67***	

Anmerkungen: Kodierung Geschlecht und Geschlecht Führungskraft: 1 = männlich, 2 = weiblich; Kodierung Arbeitszeit: 1 = Vollzeit, 2 = nicht Vollzeit. Studie 2: $n = 134$, Studie 3: $n = 371$ (aufgrund einzelner fehlender Angaben bezüglich der Arbeitszeit). *** $p < .001$, ** $p < .01$, * $p < .05$ (2-seitig).

Tabelle 8. Inkrementelle Validität des DIAF gegenüber ethischer Führung bei verschiedenen Kriterien (Studie 3)

	Wohlbefinden		Leistung		Beziehung	
	β	$R^2/\Delta R^2$	β	$R^2/\Delta R^2$	β	$R^2/\Delta R^2$
<i>Schritt 1</i>						
Geschlecht	.10	.08/.08***	.08	.02/.02*	-.01	.00/.00
Alter	.26***		.12*		.03	
Geschlecht Führungskraft	-.07		.08		-.03	
Arbeitszeit	-.02		.00		.05	
<i>Schritt 2</i>						
Geschlecht	.08	.24/.16***	.07	.12/.10***	-.06	.64/.64***
Alter	.25***		.11*		.02	
Geschlecht Führungskraft	-.09		.07		-.05	
Arbeitszeit	-.05		-.02		-.01	
Ethische Führung	.41***		.32***		.80***	
<i>Schritt 3</i>						
Geschlecht	.08	.24/.00	.07	.13/.01*	-.06	.74/.10***
Alter	.26***		.12*		.03	
Geschlecht Führungskraft	-.08		.08		-.02	
Arbeitszeit	-.05		-.02		-.01	
Ethische Führung	.38***		.14		.30***	
DIAF	.04		.21*		.59***	

Anmerkungen: Kodierung Geschlecht Führungskraft: 1 = männlich, 2 = weiblich; Kodierung Arbeitszeit: 1 = Vollzeit, 2 = nicht Vollzeit. $n = 371$ (aufgrund einzelner fehlender Angaben bezüglich der Arbeitszeit). *** $p < .001$, ** $p < .01$, * $p < .05$ (2-seitig).

Kriterien. Das DIAF wies zudem inkrementelle Validität über Abusive Supervision (für Leistung, Arbeitsengagement, Wohlbefinden, LMX) und ethische Führung (für Leistung, LMX) hinaus auf, wobei die zusätzliche Varianzaufklärung gegenüber Abusive Supervision erwartungskonform größer ausgeprägt war als gegenüber ethischer Führung. Insbesondere bei der Beziehungsqualität zwischen Führungskraft und Mitarbeitenden (LMX) wurde der zusätzliche Erklärungsbeitrag des DIAF deutlich (28 % über Abusive Supervision, 10 % über ethische Führung). Dieses Ergebnis deckt sich mit theoretischen Annahmen, dass authentische Führung im Vergleich zu ethischer Führung stärker den offenen Umgang mit Gedanken und Gefühlen im Austausch mit Mitarbeitenden betont und die Beziehungsebene stärker berücksichtigt (Cropanzano & Walumbwa, 2010). Die inkrementelle Validität gegenüber Abusive Supervision erklärt sich möglicherweise darüber, dass authentische Führung die Konsistenz zwischen Überzeugungen und Handeln umfasst, während sich Abusive Supervision vorwiegend auf die reine Verhaltensebene beschränkt. Es ist plausibel, dass derart konsistentes Agieren wichtig für die Beziehungsgestaltung ist, um beispielsweise als verlässlicher Partner wahrgenommen zu werden. Zu den untersuchten

demografischen Variablen zeigte das DIAF bis auf wenige Ausnahmen keine bedeutsamen Zusammenhänge. Insgesamt sind die ersten Befunde zur Validität des DIAF als vielversprechend zu bewerten.

Bei der Prüfung der faktoriellen Binnenstruktur des DIAF wurden wie schon bei Neider und Schriesheim (2011) per CFA drei Modelle geprüft (ein Faktor vs. vier korrelierte Faktoren vs. vier Faktoren erster Ordnung mit einem Faktor zweiter Ordnung). Den besten Fit wies das Modell mit vier interkorrelierten Faktoren auf, welches auch einem Modell mit vier Faktoren erster Ordnung und einem übergeordneten Faktor zweiter Ordnung überlegen war. Bei Neider und Schriesheim (2011) erreichten beide Modelle hingegen ähnlich gute Kennwerte, wobei die Autoren das strukturell sparsamere Modell ohne übergeordneten Faktor präferierten. Die für den DIAF gefundene Struktur entsprach damit nicht vollständig den Annahmen und Befunden von Walumbwa et al. (2008), stützt aber die Befunde und das Vorgehen zur Originalskala von Neider und Schriesheim (2011). Aus unserer Sicht wäre es auf Basis des aktuellen Erkenntnisstandes verfrüht, das Vorhandensein eines übergeordneten Faktors für authentische Führung auszuschließen. Hohe Interkorrelationen zwischen den Komponenten des DIAF ($r \geq .54$) bzw. er-

mittelten Faktoren in CFA ($r \geq .63$) und EFA ($r \geq .67$) lassen die Bildung eines Gesamtwertes für das DIAF gerechtfertigt erscheinen. Die Befunde sollten als Ausgangspunkt für eine weitere theoretisch-inhaltliche Schärfung des Konzepts authentischer Führung betrachtet werden.

Hinweise dafür können die Ergebnisse aus den EFA liefern, die eine zweifaktorielle Struktur des Instrumentes nahelegen: Die drei Komponenten SB, AIV und TiB luden auf einem ersten Faktor, die Komponente VMP auf einem zweiten. In der theoretischen Konzeption steht SB inhaltlich für das vertiefte Selbstverständnis der Führungskraft, während AIV und TiB als Prozesskomponenten umschreiben, wie ein vertieftes Selbstverständnis entstehen kann. Die Ergebnisse der explorativen Analysen weisen darauf hin, dass sich die drei Komponenten in der derzeitigen Operationalisierung empirisch weniger trennscharf voneinander abgrenzen lassen. Die gefundene Struktur ähnelt dem von Knoll, Meyer, Kroemer und Schröder-Abé (2015) vorgeschlagenen zweidimensionalen Modell zur Authentizität. Der erste DIAF-Faktor mit den Komponenten SB, AIV und TiB spiegelt die Dimension *Authentic Self-Awareness* wider, welche die ganzheitliche Wahrnehmung des eigenen Selbst sowie die Motivation meint, weiter mehr über sich selbst zu erfahren. Der zweite Faktor des DIAF mit der Komponente VMP ist mit der Dimension *Authentic Self-Expression* vergleichbar, die Verhaltensweisen und Entscheidungen umfasst, die das eigene Selbst nach außen konsistent repräsentieren. Eine intensivere Verknüpfung des Konzepts der Authentizität (Knoll et al., 2015) mit der Führungsrolle und der täglichen Führungsarbeit könnte die Weiterentwicklung des Konzepts authentischer Führung unterstützen.

Erwartungskonform korrelierte der Gesamtwert des DIAF sehr hoch mit ethischer Führung ($r = .84$). Die Komponente VMP korrelierte dabei entgegen unserer Erwartung in vergleichbarer Höhe mit den Komponenten Ethisches Rollenmodell (ERM) und Ethische Mitarbeiterführung (EMF). Dieser Befund mag auf die Operationalisierung von ERM in dem von uns eingesetzten Instrument von Rowold et al. (2009) zurückgehen. Die dort verwendeten Items spiegeln zwar ein Vorleben ethischer Verhaltensweisen wider, bringen aber nicht zum Ausdruck, dass diese zwingend auf den *eigenen* moralischen Überzeugungen beruhen müssen. Hier liegt ein zentraler Unterschied zur Komponente VMP des DIAF, die auf eine Konsistenz zwischen *eigenen* Werten und *eigenem* Handeln fokussiert. Dass diese Übereinstimmung von Werten und Handeln im Konzept der ethischen Führung nicht abgebildet wird, ist zudem ein möglicher Grund dafür, dass VMP von allen DIAF-Komponenten die niedrigsten Korrelationen zu ethischer Führung aufwies. Ähnliches gilt in Bezug auf das Führungskonzept Abusive Supervision: Auch hier wies

VMP die niedrigsten Zusammenhänge auf, was darauf zurückgehen mag, dass auch bei Abusive Supervision die Kongruenz zwischen Werten und Handeln eine untergeordnete Rolle spielt.

Im Gegensatz dazu waren die Zusammenhänge der Komponente AIV zu anderen Führungsformen unter allen Komponenten am höchsten ausgeprägt. Es ist anzunehmen, dass die AIV-Items am ehesten dem Konzept der Mitarbeiterorientierung (vgl. Fleishman, 1953) entsprechen. Dieses bildet eine wichtige Basis für ethische Führung und weist dementsprechend hohe Zusammenhänge mit dieser Führungsform auf (Rowold et al., 2009). Abusive Supervision umfasst hingegen destruktive Verhaltensweisen, welche Mitarbeiterbeteiligung verhindern oder sanktionieren und somit quasi das Gegenstück zu Mitarbeiterorientierung abbilden.

Obwohl anders erwartet, traten vereinzelt signifikante Zusammenhänge geringer Höhe zwischen dem DIAF und soziodemografischen Variablen auf. Die deutlichsten Zusammenhänge zeigten sich zur Arbeitszeit der Teilnehmenden (Teilzeit vs. Vollzeit). Teilzeitbeschäftigte nahmen beispielsweise mehr Transparenz in Beziehungen bei ihren Führungskräften wahr als Vollzeitbeschäftigte. Möglicherweise investieren Führungskräfte mehr in die Beziehungsarbeit zu Mitarbeitenden in Teilzeit, um ein stabiles Vertrauensverhältnis aufzubauen, die Mitarbeitenden ins Team zu integrieren und längerfristig an die Organisation zu binden. In bisherigen Studien wurden Zusammenhänge zwischen authentischer Führung und demografischen Merkmalen noch nicht systematisch betrachtet (vgl. Banks et al., 2016). Hierzu besteht weiterer Forschungsbedarf.

Einschränkungen und Empfehlungen für zukünftige Forschung

Aus methodischer Sicht ist darauf hinzuweisen, dass alle drei Studien als Onlinebefragungen durchgeführt wurden. Die verwendeten Stichproben waren sehr heterogen, Verzerrungen durch Selektionseffekte sind nicht auszuschließen (für weitere Ausführungen zu Online-Studien siehe Kraut et al., 2004). Des Weiteren wurden sowohl das DIAF als auch die relevanten Außenkriterien in einem konkurrenten Design zu jeweils einem Messzeitpunkt erhoben, weshalb wir keine Aussagen zur prädiktiven Validität und zu möglichen Kausalitäten treffen können. Zukünftig sollten hierzu Längsschnittuntersuchungen umgesetzt werden. Dabei könnte auch eine Prüfung der Retest-Reliabilität des DIAF erfolgen. Die gefundenen Zusammenhänge zu den Außenkriterien wurden in unseren Studien möglicherweise überschätzt, da alle Variablen mit der gleichen Methode erfasst wurden. Für Folgestudien

ist zu empfehlen, auch objektive Maße (z. B. AU-Tage, Anzahl Verbesserungsvorschläge) in die Validierung einzubeziehen. Darüber hinaus beruhte die Einschätzung des Führungsverhaltens nur auf den Fremdeinschätzungen der Mitarbeitenden. Eine Selbsteinschätzung der Führungskräfte sollte zukünftig ebenfalls berücksichtigt werden. Eine DIAF-Variante zur Selbsteinschätzung wird derzeit durch die Autorinnen und Autoren dieses Beitrages erprobt.

Aus inhaltlicher Sicht sollten zur weiteren Prüfung der Konstruktvalidität des DIAF und seiner Komponenten die Beziehungen zu weiteren positiven Führungskonzepten wie transformationaler Führung (Bass, 1985) und Servant Leadership (van Dierendonck, 2011) untersucht werden, um Überlappungen und Unterschiede einzelner Konzepte abzubilden. Schließlich sollte auch theoretische Arbeiten zum Konzept der authentischen Führung und dessen Binnenstruktur vorangetrieben werden. Im Sinne der inhaltlichen Schärfung des übergeordneten Konzepts könnte klarer herausgearbeitet werden, welche gemeinsamen Bestimmungsstücke sich definitorisch in allen Teilkomponenten authentischer Führung widerspiegeln sollten. Mögliche Ansatzpunkte stellen die Kernaspekte Selbstreflexion sowie Kongruenz im Führungsverhalten dar. Insgesamt ist anzustreben, die Unterschiede authentischer Führung zu anderen Führungskonzepten stärker zu verdeutlichen. Aktuelle Metaanalysen (Banks et al., 2016; Hoch et al., 2016) belegen, dass eine klarere Differenzierung verschiedener Konzepte große Relevanz für den Erkenntnisfortschritt im Bereich von positiver Führung hätte. Die inhaltliche Präzisierung des Konstruktes kann zukünftig die Grundlage für eine verbesserte Operationalisierung sein, welche die einzelnen Komponenten authentischer Führung noch trennschärfer abbildet.

Implikationen für die Anwendung

Mit dem DIAF steht ein frei erhältliches, deutschsprachiges, ökonomisch einsetzbares Instrument zur Erfassung authentischer Führung zur Verfügung. Verschiedene Kernthemen der Führungskräfteentwicklung können damit unterstützt werden. So könnte die Skala als Teil von Mitarbeiterbefragungen in Organisationen eingesetzt werden, um den Status Quo der bestehenden Führungskultur zu bestimmen. Ferner kann das DIAF Mitarbeitenden ermöglichen, ihre Führungskräfte im Rahmen eines Führungskräfte-Feedbacks einzuschätzen. Die gemeinsame Auswertung kann helfen, Selbst- und Fremdbild abzugleichen, gegenseitige Erwartungen zum Führungsverhalten zu klären und gemeinsam Verbesserungsmaßnahmen abzuleiten. Darüber hinaus empfiehlt sich ein Einsatz der Skala im individuellen Führungskräfte-Coaching,

um beispielsweise das eigene Rollenverständnis sowie die Passung von Wertvorstellungen und eigenem Führungsverhalten zu reflektieren oder die Vorbildrolle und den souveränen Umgang mit eigenen Schwächen und Fehlern zu stärken. Das Verfahren misst vier Komponenten authentischer Führung und erlaubt daher ein differenziertes Screening des Führungsverhaltens.

Insgesamt stellt das Konzept der authentischen Führung ein empfehlenswertes Orientierungsmodell zur Entwicklung einer ausgewogenen Feedback- und Vertrauenskultur in Organisationen dar, um den offenen Diskurs auch zu moralischen Bedenken und kritischen Sachverhalten bei der Arbeit zu fördern (vgl. Knoll, Wegge, Unterrainer, Silva & Jönsson, 2016). Die positiven Auswirkungen authentischer Führung auf organisationale Erfolgskriterien wurden schon verschiedentlich nachgewiesen (vgl. Banks et al., 2016; Hoch et al., 2016) und sollten weiter beforscht werden. Der Mangel an psychometrisch abgesicherten Verfahren ist eines der bedeutendsten Probleme der Führungsforschung (Neider & Schriesheim, 2011). Ohne den Einsatz valider Verfahren kann es keinen echten Erkenntnisfortschritt geben, sondern vielmehr werden Fehlinterpretationen und nicht tragfähige Befunde wahrscheinlich. Mit dem DIAF steht nun für den deutschen Sprachraum ein valides Instrument zur Erfassung authentischer Führung zur Verfügung, das hoffentlich dazu beitragen wird, den Erkenntnisfortschritt weiter voranzutreiben.

Literatur

- Arbuckle, J. L. (2014). *Amos (Version 23.0)* [Computer Program]. Chicago, IL: IBM SPSS.
- Avolio, B. J., Gardner, W. L. & Walumbwa, F. O. (2007). *Authentic leadership questionnaire*. Menlo Park, CA: Mind Garden.
- Banks, G. C., McCauley, K. D., Gardner, W. L. & Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *The Leadership Quarterly*, 27, 634–652. <https://doi.org/10.1016/j.leaqua.2016.02.006>
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York, NY: Free Press.
- Bishop, J. W., Scott, K. D. & Burroughs, S. M. (2000). Support, commitment, and employee outcomes in a team environment. *Journal of Management*, 26, 1113–1132. <https://doi.org/10.1177/014920630002600603>
- Brähler, E., Mühlen, H., Albani, C. & Schmidt, S. (2007). Teststatistische Prüfung und Normierung der deutschen Versionen des EUROHIS-QOL Lebensqualität-Index und des WHO-5 Wohlbefindens-Index. *Diagnostica*, 53, 83–96. <https://doi.org/10.1026/0012-1924.53.2.83>
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216. <https://doi.org/10.1177/135910457000100301>
- Brown, M. E. & Mitchell, M. S. (2010). Ethical and unethical leadership: Exploring new avenues for future research. *Business*

- Ethics Quarterly*, 20, 583–616. <https://doi.org/10.5840/beq201020439>
- Brown, M. E., Treviño, L. K. & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, 97, 117–134. <https://doi.org/10.1016/j.obhdp.2005.03.002>
- Cooper, C., Scandura, T. A. & Schriesheim, C. A. (2005). Looking forward but learning from our past: Potential challenges to developing authentic leadership theory and authentic leaders. *The Leadership Quarterly*, 16, 475–493. <https://doi.org/10.1016/j.leaqua.2005.03.008>
- Cropanzano, R. & Walumbwa, F. O. (2010). Moral leadership: A short primer on competing perspectives. In M. Schminke (Ed.), *Managerial ethics: Managing the psychology of morality* (pp. 21–52). New York, NY: Psychology Press.
- Deci, E. L., Connell, J. P. & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, 74, 580–590. <https://doi.org/10.1037/0021-9010.74.4.580>
- Dierendonck, D. van (2011). Servant leadership: A review and synthesis. *Journal of Management*, 37(4), 1228–1261. <https://doi.org/10.1177/0149206310380462>
- Eigenstetter, M., Strobel, A. & Stumpf, S. (2012). Diagnostik ethischer Kompetenz. In S. Kaiser & A. Kozica (Hrsg.), *Ethik im Personalmanagement. Zentrale Konzepte, Ansätze und Fragestellungen* (S. 225–246). München: Rainer Hampp.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage.
- Fleishman, E. A. (1953). The description of supervisory behavior. *Journal of Applied Psychology*, 37, 1–6. <https://doi.org/10.1037/h0056314>
- Frömmer, D., Wegge, J. & Strobel, A. (2014). Risiko durch Verschlossenheit? Das Zusammenspiel von Führung, Mitarbeiter-schweigen und Managerversagen. *Wirtschaftspsychologie*, 16, 39–44.
- Gardner, W. L., Cogliser, C. C., Davis, K. M. & Dickens, M. P. (2011). Authentic leadership: A review of the literature and research agenda. *The Leadership Quarterly*, 22, 1120–1145. <https://doi.org/10.1016/j.leaqua.2011.09.007>
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6, 304–312. <https://doi.org/10.1037/1040-3590.6.4.304>
- Graen, G. B. & Scandura, T. A. (1987). Toward a psychology of dyadic organizing. *Research in Organizational Behavior*, 9, 175–208.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum.
- Hoch, J. E., Bommer, W. H., Dulebohn, J. H. & Wu, D. (2016). Do ethical, authentic, and servant leadership explain variance above and beyond transformational leadership? A meta-analysis. *Journal of Management*, 1–29. <https://doi.org/10.1177/0149206316665461>
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- Ilies, R., Morgeson, F. P. & Nahrgang, J. D. (2005). Authentic leadership and eudaemonic well-being: Understanding leader-follower outcomes. *The Leadership Quarterly*, 16, 373–394. <https://doi.org/10.1016/j.leaqua.2005.03.002>
- Judge, T. A. & Piccolo, R. F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology*, 89, 755–768. <https://doi.org/10.1037/0021-9010.89.5.755>
- Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktual. u. überarb. Aufl., S. 75–102). Berlin: Springer.
- Kline, P. (1994). *An easy guide to factor analysis*. London, UK: Routledge.
- Knoll, M., Meyer, B., Kroemer, N. B. & Schröder-Abé, M. (2015). It takes two to be yourself. An integrated model of authenticity, its measurement, and its relationship to work-related variables. *Journal of Individual Differences*, 36, 38–53. <https://doi.org/10.1027/1614-0001/a000153>
- Knoll, M., Wegge, J., Unterrainer, C., Silva, S. & Jönsson, T. (2016). Is our knowledge of voice and silence in organizations growing? Building bridges and (re)discovering opportunities. *German Journal of Human Resource Management*, 30, 161–194. <https://doi.org/10.1177/2397002216649857>
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J. & Couper, M. (2004). Psychological research online – Report of board of scientific affairs' advisory group on the conduct of research on the internet. *American Psychologist*, 59(2), 105–117. <https://doi.org/10.1037/0003-066X.59.2.105>
- Leiner, D. J. (2012, März). SoSci Panel: The Noncommercial Online Access Panel. Poster presented at the GÖR 2012, Mannheim. Verfügbar unter <https://www.sosicisurvey.de/panel/download/SoSciPanel.GÖR2012.pdf>.
- Lenhard, W. & Lenhard, A. (2014). *Signifikanztests bei Korrelationen*. Bibergau: Psychometrica. Verfügbar unter <https://www.psychometrica.de/korrelation.html>. <https://doi.org/10.13140/RG.2.1.2954.1367>
- Loock, C.-M. & Schilling, J. (2010, September). *Despostische Führung: Evaluation einer deutschen Fassung der „Abusive Supervision Scale“ von Tepper (2000)*. Poster, präsentiert auf dem 47. Kongress der Deutschen Gesellschaft für Psychologie, Bremen.
- Luthans, F. & Avolio, B. J. (2003). Authentic leadership development. In K. S. Cameron, J. E. Dutton & R. E. Quinn (Eds.), *Positive organizational scholarship* (pp. 241–258). San Francisco, CA: Berrett-Koehler.
- Neider, L. L. & Schriesheim, C. A. (2011). The authentic leadership inventory (ALI): Development and empirical tests. *The Leadership Quarterly*, 22, 1146–1164. <https://doi.org/10.1016/j.leaqua.2011.09.008>
- Peus, C., Wesche, J. S., Streicher, B., Braun, S. & Frey, D. (2012). Authentic leadership: An empirical test of its antecedents, consequences, and mediating mechanisms. *Journal of Business Ethics*, 107, 331–348. <https://doi.org/10.1007/s10551-011-1042-3>
- Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, 41, 574–599. <https://doi.org/10.2307/2393868>
- Rowold, J., Borgmann, L. & Heinitz, K. (2009). Ethische Führung – Gütekriterien einer deutschen Adaptation der Ethical Leadership Scale (ELS-D) von Brown et al. (2005). *Zeitschrift für Arbeits- und Organisationspsychologie*, 53, 57–69. <https://doi.org/10.1026/0932-4089.53.2.57>
- Schaufeli, W. B. & Bakker, A. B. (2003). *Test manual for the Utrecht Work Engagement Scale*. Unpublished manuscript, Utrecht University, Netherlands.
- Schyns, B. & Paul, T. (2014). Skala zur Erfassung des Leader-Member-Exchange (LMX7 nach Graen & Uhl-Bien, 1995) –

- Übersetzung. *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. <https://doi.org/10.6102/zis23>
- Schyns, B. & Schilling, J. (2013). How bad are the effects of bad leaders? A meta-analysis of destructive leadership and its outcomes. *The Leadership Quarterly*, 24, 138–158. <https://doi.org/10.1016/j.leaqua.2012.09.001>
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
- Tepper, B. J. (2000). Consequences of abusive supervision. *Academy of Management Journal*, 43, 178–190. <https://doi.org/10.2307/1556375>
- Tepper, B. J., Moss, S. E., Lockhart, D. E. & Carr, J. C. (2007). Abusive supervision, upward maintenance communication, and subordinates' psychological distress. *Academy of Management Journal*, 50, 1169–1180. <https://doi.org/10.2307/AMJ.2007.20159918>
- Treviño, L. K., den Nieuwenboer, N. A. & Kish-Gephart, J. J. (2014). (Un)Ethical behavior in organizations. *Annual Review of Psychology*, 65, 635–660. <https://doi.org/10.1146/annurev-psych-113011-143745>
- Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S. & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory-based measure. *Journal of Management*, 34, 89–126. <https://doi.org/10.1177/0149206307308913>
- Wegge, J. (2004). *Führung von Arbeitsgruppen*. Göttingen: Hogrefe.
- Wegge, J., Jeppesen, H.-J., Weber, W. G., Pearce, C. L., Silvia, S. A., Pundt, A., ... & Piecha, A. (2010). Promoting work motivation in organizations: Should employee involvement in organizational leadership become a new tool in the organizational psychologists' kit? *Journal of Personnel Psychology*, 9, 154–171. <https://doi.org/10.1027/1866-5888/a000025>
- West, S. G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Hrsg.), *Structural equation modeling: Concepts, issues, and applications* (S. 56–75). Thousand Oaks, CA: Sage.

Eingegangen: 27.09.2017

Revision eingegangen: 22.12.2017

Luise Franke-Bartholdt

Prof. Dr. Anja Strobel

Technische Universität Chemnitz

Institut für Psychologie

Professur für Persönlichkeitspsychologie und Diagnostik

Wilhelm-Raabe-Straße 43

09120 Chemnitz

anja.strobel@psychologie.tu-chemnitz.de

Dirk Frömmer

Prof. Dr. Jürgen Wegge

Technische Universität Dresden

Professur für Arbeits- und Organisationspsychologie

Zellescher Weg 17

01069 Dresden