

Die deutsche Version des Big Five Inventory 2 (BFI-2)

*Daniel Danner, Beatrice Rammstedt,
Matthias Bluemke, Lisa Treiber, Sabrina
Berres, Christopher Soto, & Oliver John*

Zitierung

Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C., & John, O. (2016). Die deutsche Version des Big Five Inventory 2 (BFI-2). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis247

1 Überblick

Zusammenfassung

Die deutsche Version des Big Five Inventory 2 (BFI-2) erfasst die Big Five Persönlichkeitseigenschaften Extraversion, Verträglichkeit, Gewissenhaftigkeit, Negative Emotionalität (früher: Neurotizismus) und Offenheit mit jeweils drei Facetten. Die Skala basiert auf dem Big Five Inventory 2 von Christopher Soto und Oliver John. Die Items wurden in einem mehrstufigen Verfahren vom Englischen ins Deutsche übertragen. Erste empirische Untersuchungen belegen die Reliabilität und Validität der Messungen.

Zentrale Kennwerte

Titel: Die deutsche Version des Big Five Inventory 2 (BFI-2)

Autor: Danner, Rammstedt, Bluemke, Treiber, Berres, Soto, & John

In ZIS seit: 2016

Anzahl der Items: 60

Reliabilität: .60 bis .88

Validität: Hinweise auf die Konstrukt- und Kriteriumsvalidität

Konstrukt: Extraversion, Verträglichkeit, Gewissenhaftigkeit, Neurotizismus, Negative Emotionalität, Emotionale Stabilität, Offenheit

Sprache der Dokumentation: Deutsch

Sprache Items: Deutsch, Englisch

Item(s) in Bevölkerungsumfrage eingesetzt: ja

Entwicklungsstand: validiert

2 Instrument

Instruktion

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. Würden Sie über sich z. B. sagen, dass Sie gerne Zeit mit anderen Menschen verbringen? Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.

Items

Tabelle 1

BFI-2 Items

Nr.	Deutsches Item	Polung	Domäne	Facette
1	Ich gehe aus mir heraus, bin gesellig.	+	E	Geselligkeit
2	Ich bin einfühlsam, warmherzig.	+	V	Mitgefühl
3	Ich bin eher unordentlich.	-	G	Ordnungsliebe
4	Ich bleibe auch in stressigen Situationen gelassen.	-	N	Ängstlichkeit
5	Ich bin nicht sonderlich kunstinteressiert.	-	O	Ästhetisches Empfinden
6	Ich bin durchsetzungsfähig, energisch.	+	E	Durchsetzungsfähigkeit
7	Ich begegne anderen mit Respekt.	+	V	Höflichkeit
8	Ich bin bequem, neige zu Faulheit.	-	G	Fleiß
9	Ich bleibe auch bei Rückschlägen zuversichtlich.	-	N	Niedergeschlagenheit
10	Ich bin vielseitig interessiert.	+	O	Intellektuelle Neugierde
11	Ich schäume selten vor Begeisterung über.	-	E	Aktivität
12	Ich neige dazu, andere zu kritisieren.	-	V	Vertrauen
13	Ich bin stetig, beständig.	+	G	Verlässlichkeit
14	Ich kann launisch sein, habe schwankende Stimmungen.	+	N	Unbeständigkeit der Gefühle
15	Ich bin erfinderisch, mir fallen raffinierte Lösungen ein.	+	O	Kreativer Einfallreichtum
16	Ich bin eher ruhig.	-	E	Geselligkeit
17	Ich habe mit anderen wenig Mitgefühl.	-	V	Mitgefühl
18	Ich bin systematisch, halte meine Sachen in Ordnung.	+	G	Ordnungsliebe
19	Ich reagiere leicht angespannt.	+	N	Ängstlichkeit
20	Ich kann mich für Kunst, Musik und Literatur begeistern.	+	O	Ästhetisches Empfinden
21	Ich neige dazu, die Führung zu übernehmen.	+	E	Durchsetzungsfähigkeit
22	Ich habe oft Streit mit anderen.	-	V	Höflichkeit
23	Ich neige dazu, Aufgaben vor mir	-	G	Fleiß

	herzuschieben.			
24	Ich bin selbstsicher, mit mir zufrieden.	-	N	Niedergeschlagenheit
25	Ich meide philosophische Diskussionen.	-	O	Intellektuelle Neugierde
26	Ich bin weniger aktiv und unternehmungslustig als andere.	-	E	Aktivität
27	Ich bin nachsichtig, vergebe anderen leicht.	+	V	Vertrauen
28	Ich bin manchmal ziemlich nachlässig.	-	G	Verlässlichkeit
29	Ich bin ausgeglichen, nicht leicht aus der Ruhe zu bringen.	-	N	Unbeständigkeit der Gefühle
30	Ich bin nicht besonders einfallsreich.	-	O	Kreativer Einfallsreichtum
31	Ich bin eher schüchtern.	-	E	Geselligkeit
32	Ich bin hilfsbereit und selbstlos.	+	V	Mitgefühl
33	Ich mag es sauber und aufgeräumt.	+	G	Ordnungsliebe
34	Ich mache mir oft Sorgen.	+	N	Ängstlichkeit
35	Ich weiß Kunst und Schönheit zu schätzen.	+	O	Ästhetisches Empfinden
36	Mir fällt es schwer, andere zu beeinflussen.	-	E	Durchsetzungsfähigkeit
37	Ich bin manchmal unhöflich und schroff.	-	V	Höflichkeit
38	Ich bin effizient, erledige Dinge schnell.	+	G	Fleiß
39	Ich fühle mich oft bedrückt, freudlos.	+	N	Niedergeschlagenheit
40	Es macht mir Spaß, gründlich über komplexe Dinge nachzudenken und sie zu verstehen.	+	O	Intellektuelle Neugierde
41	Ich bin voller Energie und Tatendrang.	+	E	Aktivität
42	Ich bin anderen gegenüber misstrauisch.	-	V	Vertrauen
43	Ich bin verlässlich, auf mich kann man zählen.	+	G	Verlässlichkeit
44	Ich habe meine Gefühle unter Kontrolle, werde selten wütend.	-	N	Unbeständigkeit der Gefühle
45	Ich bin nicht sonderlich fantasievoll.	-	O	Kreativer Einfallsreichtum
46	Ich bin gesprächig.	+	E	Geselligkeit
47	Andere sind mir eher gleichgültig, egal.	-	V	Mitgefühl
48	Ich bin eher der chaotische Typ, mache selten sauber.	-	G	Ordnungsliebe
49	Ich werde selten nervös und unsicher.	-	N	Ängstlichkeit
50	Ich finde Gedichte und Theaterstücke langweilig.	-	O	Ästhetisches Empfinden
51	In einer Gruppe überlasse ich lieber anderen die Entscheidung.	-	E	Durchsetzungsfähigkeit
52	Ich bin höflich und zuvorkommend.	+	V	Höflichkeit

53	Ich bleibe an einer Aufgabe dran, bis sie erledigt ist.	+	G	Fleiß
54	Ich bin oft deprimiert, niedergeschlagen.	+	N	Niedergeschlagenheit
55	Mich interessieren abstrakte Überlegungen wenig.	-	O	Intellektuelle Neugierde
56	Ich bin begeisterungsfähig und kann andere leicht mitreißen.	+	E	Aktivität
57	Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	+	V	Vertrauen
58	Manchmal verhalte ich mich verantwortungslos, leichtsinnig.	-	G	Verlässlichkeit
59	Ich reagiere schnell gereizt oder genervt.	+	N	Unbeständigkeit der Gefühle
60	Ich bin originell, entwickle neue Ideen.	+	O	Kreativer Einfallsreichtum

Anmerkung: E = Extraversion, V = Verträglichkeit, G = Gewissenhaftigkeit, N = Negative Emotionalität, O = Offenheit, die englischen Items sind in [Anhang 1](#) dargestellt.

Antwortvorgaben

Jedes Item kann auf einer fünfstufigen Skala mit den Abstufungen „stimme überhaupt nicht zu“, „stimme eher nicht zu“, teils, teils“, „stimme eher zu“ und „stimme voll und ganz zu“ beantwortet werden.

Auswertungshinweise

Mit dem BFI-2 können Skalenwerte für die Dimensionen Extraversion, Verträglichkeit, Gewissenhaftigkeit, Negative Emotionalität (früher: Neurotizismus) und Offenheit gebildet werden. Zusätzlich können Skalenwerte für insgesamt 15 Facetten gebildet werden. Jeweils drei Facetten lassen sich inhaltlich einer Dimension zuordnen. Dazu werden die Antwortkategorien von 1 („stimme überhaupt nicht zu“) bis 5 („stimme voll und ganz zu“) kodiert und anschließend aggregiert. Negativ kodierte Items müssen zuvor rekodiert werden (rekodierter Wert = 6 - Rohwert).

Tabelle 2

Auswertungsschlüssel

Skala	Items
Extraversion	1, 6, 11R, 16R, 21, 26R, 31R, 36R, 41, 46, 51R, 56
Geselligkeit	1, 16R, 31R, 46
Durchsetzungsfähigkeit	6, 21, 36R, 51R
Aktivität	11R, 26R, 41, 56
Verträglichkeit	2, 7, 12R, 17R, 22R, 27, 32, 37R, 42R, 47R, 52, 57
Mitgefühl	2, 17R, 32, 47R
Höflichkeit	7, 22R, 37R, 52
Vertrauen	12R, 27, 42R, 57

Gewissenhaftigkeit	3R, 8R, 13, 18, 23R, 28R, 33, 38, 43, 48R, 53, 58R
Ordnungsliebe	3R, 18, 33, 48R
Fleiß	8R, 23R, 38, 53
Verlässlichkeit	13, 28R, 43, 58R
Negative Emotionalität	4R, 9R, 14, 19, 24R, 29R, 34, 39, 44R, 49R, 54, 59
Ängstlichkeit	4R, 19, 34, 49R
Niedergeschlagenheit	9R, 24R, 39, 54
Unbeständigkeit der Gefühle	14, 29R, 44R, 59
Offenheit	5R, 10, 15, 20, 25R, 30R, 35, 40, 45R, 50R, 55R, 60
Ästhetisches Empfinden	5R, 20, 35, 50R
Intellektuelle Neugierde	10, 25R, 40, 55R
Kreativer Einfallsreichtum	15, 30R, 45R, 60

Anmerkung: negativ kodierte Items (R) müssen rekodiert werden (rekodierter Wert = 6 – Rohwert).

Anwendungsbereich

Die Skala ist für schriftliche Befragungsmodi entwickelt und kann als Paper & Pencil Version und im Rahmen von Online Studien eingesetzt werden (PASI, CASI). Prinzipiell kann die Skala auch in mündlichen Befragungen (wie PAPI, CAPI, CATI) eingesetzt werden. Die Messäquivalenz dieser Skala für unterschiedliche Modi wurde allerdings bisher noch nicht empirisch geprüft. Die Skala wurde für die deutsche Allgemeinbevölkerung zwischen 18 und 65 Jahren entwickelt und an dieser erprobt.

3 Theoretischer Hintergrund

Die dem BFI-2 und dem Vorgänger Big Five Inventory (BFI; John, Donahue, & Kentle, 1991) zugrundeliegende Theorie ist die der fundamentalen „Big Five“ Traits zur prototypischen Persönlichkeitsbeschreibung (John, 1989, 1990). Auf Basis des lexikalischen Ansatzes wurden Adjektive und Kurzphrasen genutzt, um rational-empirisch einen Test zur Messung von fünf weitestgehend unabhängigen Dimensionen zu konstruieren. Diese Dimensionen („Domänen“) sind geläufig unter den Namen Extraversion, Verträglichkeit, Gewissenhaftigkeit, Neurotizismus und Offenheit.

Jeder Persönlichkeitszug kann jedoch entweder eher breiter oder eher spezifischer erfasst werden (Bandbreiten-Dilemma; Cronbach & Gleser, 1957). Das BFI-2 versucht beiden Aspekten gerecht zu werden, indem es spezifische Facetten mit hoher Genauigkeit (aber nur eingeschränkter Vorhersagebreite) misst; diese werden zu breiten, hierarchisch höherliegenden Domänen zusammengefasst, welche zwar bei der Vorhersage von spezifischen Kriterien weniger treffgenau sind, dafür aber eine Vielzahl von Kriterien vorhersagen können. Während die Facetten möglichst spezifisch erfasst werden, sollen die Domänen - durch die gewählten Facetten - möglichst breit repräsentiert sein.

Anders als das originale BFI ist das BFI-2 hierarchisch angelegt, um die Überlappung bzw. Ähnlichkeit zwischen Facetten adäquat abzubilden (Soto & John, 2016). Es ordnet den fünf Domänen eine niedriger liegende, spezifische Facettenstruktur von jeweils drei Facetten zu. Die erste Facette sollte

immer der zentralste Aspekt einer Domäne sein und gleichzeitig am klarsten unterschieden (orthogonal) von den weiteren vier Domänen. Beispielsweise umfasst Extraversion die drei Facetten Geselligkeit, Durchsetzungsfähigkeit und Aktivität.

Im englischen BFI-2 erhielten zwei Domänen des BFI-2 neue Bezeichnungen: Neurotizismus (engl. neuroticism) wurde in Negative Emotionalität umbenannt, um Verwechslungen mit klinischen Symptombildern zu vermeiden; Offenheit (engl. openness) wurde als mentale Offenheit (nicht gemeint war soziale Offenheit) treffsicherer benannt (engl. openmindedness). Letzterem wurde der Vorzug vor verengenden Namen wie Intellekt, Kultur, Imagination gegeben. Im Deutschen werden die Namen Extraversion, Verträglichkeit, Gewissenhaftigkeit, Negative Emotionalität und Offenheit vergeben.

4 Skalenentwicklung

Itemkonstruktion und Itemselektion

Die *englische Version* des BFI-2 wurde von Christopher Soto und Oliver John entwickelt und stellt eine Weiterentwicklung des Big Five Inventory (BFI; John et al., 1991; Rammstedt, 1997) dar. Ziel der Neuentwicklung war ein Instrument, welches die Domänen Extraversion, Verträglichkeit, Gewissenhaftigkeit, Negative Emotionalität und Offenheit mit jeweils drei Facetten abbildet. Dabei sollen die Domänenwerte des BFI und des BFI-2 maximal vergleichbar sein. Zusätzlich sollten die Skalen der Domänen und Facetten balanciert sein, also eben so viel positiv gepolte wie negativ gepolte Items enthalten. Balancierte Skalen können verhindern, dass Messwerte von Akquieszenz verzerrt werden (Danner, Aichholzer, Rammstedt, 2015; Rammstedt, Goldberg, & Borg, 2010; Soto, John, Gosling, & Potter, 2008). Die Entwicklung der englischen Version ist ausführlich bei Soto und John (2016) beschrieben.

Die *deutsche Version* des BFI-2 wurde in einem mehrstufigen Verfahren in Anlehnung an Janet Harkness TRAPD Approach (Harkness, 2003) entwickelt. Der TRAPD Approach beschreibt ein mehrstufiges Vorgehen, bei dem Items übersetzt (translation), dann diskutiert (review, adjudication) und anschließend empirisch erprobt (pretesting) und dokumentiert (documentation) werden. In einem ersten Schritt wurden die Items der englischen Version von zwei unabhängigen Übersetzern vom Englischen ins Deutsche übersetzt. Anschließend wurden die zwei Übersetzungen von den zwei Übersetzern, zwei Experten aus dem Bereich Persönlichkeitspsychologie und Umfrageforschung und einem Moderator diskutiert. Im Rahmen der Diskussion wurden abweichende Übersetzungen erörtert und teilweise modifiziert. Am Ende der Diskussion standen für jedes Item entweder eine deutsche Übersetzung oder zwei alternative Übersetzungen zur Verfügung, die anschließend in einer ersten empirischen Untersuchung erprobt wurden.

Die Übersetzungen wurden an einer für die deutsche Allgemeinbevölkerung repräsentative Quotenstichprobe ([Stichprobe 1](#)) empirisch getestet. Basierend auf den Ergebnissen dieser ersten Untersuchung wurde für jedes englische Item eine deutsche Übersetzung ausgewählt. Entscheidungskriterien für die Itemauswahl waren die relativen Trennschärfen der Items auf

Domänenebene, die relativen Trennschärfen der Items auf Facettenebene und die inhaltliche Heterogenität der Items. Die relative Trennschärfe der Items auf Domänenebene wurde bestimmt, indem für jedes Item die Trennschärfe des Items auf die zugehörige Domäne relativ zur maximalen Trennschärfe des Items auf andere Domänen bestimmt wurde. Beispielweise wurde die Trennschärfe des Items „Ich gehe aus mir heraus, bin gesellig“ auf die Domäne Extraversion ($r_{it} = .75$) relativ zur Trennschärfe des Items auf die Domäne Verträglichkeit ($r_{it} = .45$) betrachtet ($rr_{it} = .75/.45 = 1.67$). Werte >1 indizieren, dass das Item die zugehörige Domäne besser abbildet als eine andere Domäne. Daher wurden nur Items mit relativen Trennschärfen > 1 ausgewählt. Analog wurden die relativen Trennschärfen auf Facettenebene gebildet. Die inhaltliche Heterogenität der Items wurde maximiert, indem darauf geachtet wurde, dass unterschiedliche Items einer Skala möglichst unterschiedliche inhaltliche Aspekte abdecken.¹ Da 10 der 60 Items entweder niedrige Trennschärfen auf die zugehörigen Domänen oder Facetten oder substantielle Trennschärfen auf andere Domänen oder Facetten aufwiesen, wurden für diese 10 Items revidiert und in einer zweiten empirischen Untersuchung erprobt.

Die ursprünglichen Übersetzungen und die revidierten Items wurden an einer zweiten, für die deutsche Allgemeinbevölkerung repräsentative Quotenstichprobe ([Stichprobe 2](#)) empirisch getestet. Basierend auf den Ergebnissen dieser zweiten Untersuchung wurden die finalen Items des deutschen BFI-2 ausgewählt. Wie zuvor wurden die relativen Trennschärfen auf Domänen- und Facettenebene und die inhaltliche Heterogenität der Items als Entscheidungskriterien herangezogen. Bei vergleichbarer Eignung wurden Items bevorzugt, die bereits im ursprünglichen BFI (Rammstedt, 1997) enthalten waren, um die Vergleichbarkeit der Domänenwerte des BFI und des BFI-2 zu maximieren.

Stichproben

Stichprobe 1 ($N = 552$) ist eine hinsichtlich Alter, Geschlecht und Bildungsabschluss repräsentative Quotenstichprobe für die deutsche Allgemeinbevölkerung. Der Quotenplan wurde nach dem Zensus 2011 (www.zensus2011.de) erstellt. Geschlecht wurde zweistufig, Alter dreistufig (14-29, 30-49, 49-75) und Bildungsabschluss ebenfalls dreistufig (ohne Bildungsabschluss/Hauptschule, Mittlerer Schulabschluss und gymnasiale Oberstufe, (Fach)Hochschulreife und höher) quotiert. Die Befragung als computer-assisted-self-interview (CASI) mit dem Online-Access-Panel der Respondi AG durchgeführt. Die Teilnehmer wurden für die Befragung finanziell entlohnt.

Stichprobe 2 ($N = 770$) ist ebenfalls eine hinsichtlich Alter, Geschlecht und Bildungsabschluss repräsentative Quotenstichprobe für die deutsche Allgemeinbevölkerung. Der Quotenplan wurde nach dem Zensus 2011 (www.zensus2011.de) erstellt. Geschlecht wurde zweistufig, Alter dreistufig (18-29, 30-49, 49-65) und Bildungsabschluss ebenfalls dreistufig (ohne Bildungsabschluss/Hauptschule, Mittlerer Schulabschluss und gymnasiale Oberstufe, (Fach)Hochschulreife und höher) quotiert. Die Befragung als computer-assisted-self-interview (CASI) mit dem Online-Access-Panel der Respondi

¹ Für einige Items lagen eine, für andere Items zwei Übersetzungen vor. Daher waren manche inhaltlichen Aspekte der Domänen und Facetten häufiger vertreten als andere. Um sicherzustellen, dass die Ergebnisse davon nicht verzerrt wurden, wurden die relativen Trennschärfen nach der ersten Itemauswahl neu berechnet. Die Ergebnismuster waren äquivalent.

AG durchgeführt. Teilnehmer, die eine Aufmerksamkeitsprüfung nicht bestanden („Bitte klicken Sie hier zur Überprüfung der Funktionsfähigkeit unseres Fragebogen »stimme eher nicht zu« an“) oder die Befragung in weniger als 5 oder mehr als 45 Minuten beendeten, wurden von der Analyse ausgeschlossen. Zusätzlich wurde Stichprobe 2 in einem experimentellen Design erhoben, wobei die Hälfte der Teilnehmer (zufällige Zuteilung) für Ihre Teilnahme an der Befragung finanziell belohnt wurden und die Hälfte der Teilnehmer nicht finanziell belohnt wurde, dafür aber eine Rückmeldung zu Ihrem Persönlichkeitsprofil erhielt. Da alle Ergebnisse für beide Bedingungen äquivalent waren, werden die Ergebnisse gepoolt über beide Bedingungen berichtet. Die Stichprobencharakteristika sind in Tabelle 2 dargestellt.

Tabelle 2

Stichprobencharakteristika

	Stichprobe 1	Stichprobe 2
Stichprobenart	Quotenstichprobe	Quotenstichprobe
Modus	CASI	CASI
<i>N</i>	552	770
Mittelwert Alter (SD)	47.22 (15.70)	44.47 (13.81)
Anteil Frauen in Prozent	52.17	51.43
Anteil ohne Bildungsabschluss/ Hauptschule in Prozent	41.85	35.71
Anteil Mittlerer Schulabschluss und gymnasiale Oberstufe in Prozent	29.89	32.99
Anteil (Fach)Hochschulreife und höher in Prozent	28.26	31.30

Itemanalysen

Die faktorielle Struktur der Items und die faktorielle Struktur der Facettenwerte wurde mit explorativen und konfirmatorischen Faktoranalysen in Stichprobe 2 untersucht. Die aus Platzgründen in [Anhang 2](#) dargestellte Tabelle deutet auf gut interpretierbare, fünf-faktorielle Struktur der Items hin. In gleicher Weise deutet das Ladungsmuster der Facettenwerte auf eine fünf-faktorielle Struktur der Facetten hin, wobei die Facetten auf die angenommenen Domänen laden (Tabelle 3).

Tabelle 3

Faktorladungen der Facettenwerte auf fünf Faktoren

Facette	F1	F2	F3	F4	F5
Geselligkeit	0.86	-0.03	-0.04	0.15	0.01
Durchsetzungsfähigkeit	0.76	0.14	-0.16	-0.23	0.16
Aktivität	0.75	0.17	-0.16	0.23	0.25
Mitgefühl	0.14	0.24	0.07	0.81	0.15
Höflichkeit	-0.02	0.35	-0.21	0.72	0.10
Vertrauen	0.04	-0.11	-0.19	0.75	0.03
Ordnungsliebe	0.05	0.86	0.00	0.00	0.07
Fleiß	0.26	0.80	-0.18	0.13	0.07
Verlässlichkeit	0.00	0.76	-0.28	0.26	0.05

Ängstlichkeit	-0.19	-0.07	0.88	0.00	-0.09
Niedergeschlagenheit	-0.44	-0.20	0.71	-0.09	-0.05
Unbeständigkeit der Gefühle	0.07	-0.20	0.81	-0.29	-0.11
Ästhetisches Empfinden	0.01	-0.01	0.03	0.17	0.84
Intellektuelle Neugierde	0.16	0.09	-0.22	0.03	0.80
Kreativer Einfallsreichtum	0.45	0.14	-0.08	0.04	0.60

Anmerkung: Hauptkomponentenanalyse, Varimax rotiert, $N = 770$.

Für die konfirmatorische Faktoranalyse wurde ein hierarchisches Strukturgleichungsmodell spezifiziert. Die Varianzen und Kovarianzen der 60 Items wurden durch 15 Facetten erklärt, von denen jeweils drei Facetten auf eine gemeinsame Domäne laden. Das Modell ist in Abbildung 1 skizziert und weist einen akzeptablen Fit auf, $RMSEA = .062$, $CFI = .744$, $SRMR = .082$. Diese Ergebnisse deuten darauf hin, dass der BFI-2 die fünf Domänen Extraversion, Gewissenhaftigkeit, Negative Emotionalität (früher: Neurotizismus) und Offenheit mit jeweils drei Facetten abbilden kann.

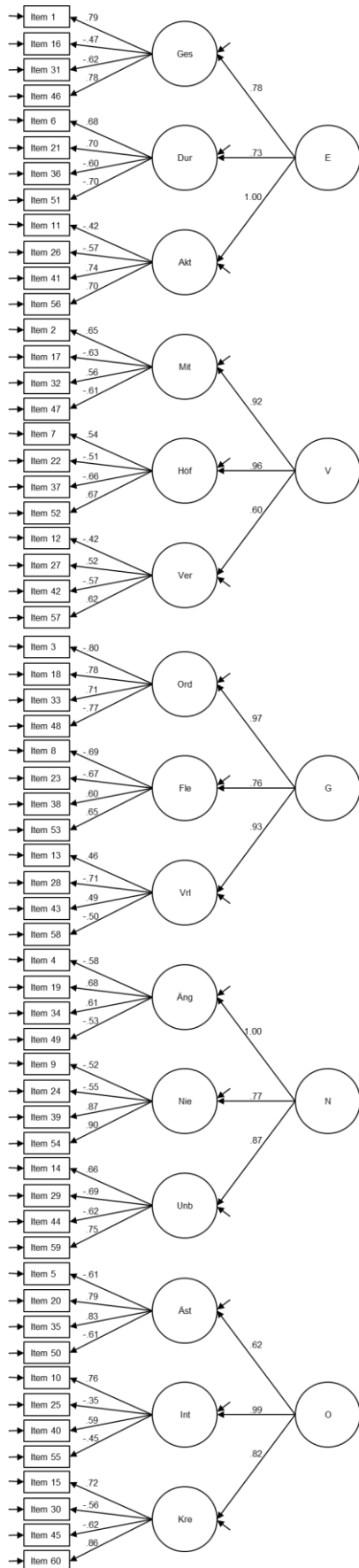


Abbildung 1. Hierarchisches Faktormodel des BFI-2 mit standardisierten Pfadkoeffizienten (Vergrößerung per Klick auf die Abbildung).

Itemkennwerte

In Tabelle 4 sind Mittelwert, Standardabweichung, Schiefe und Exzess aller 60 Items in Stichprobe 2 dargestellt. Zusätzlich werden die Trennschärfen der Items auf die Facetten und Domänen berichtet.

Tabelle 4

Mittelwert, Standardabweichung, Schiefe, Exzess und Trennschärfen der manifesten Items

Item	<i>M</i>	<i>SD</i>	Schiefe	Exzess	Trennschärfe (Facette)	Trennschärfe (Domäne)
1	3.54	1.02	-0.32	-0.40	0.61	0.65
2	4.05	0.80	-0.63	0.31	0.49	0.54
3	3.44	1.12	-0.22	-0.71	0.73	0.66
4	2.69	0.93	0.17	-0.14	0.48	0.53
5	2.78	1.30	0.18	-1.07	0.56	0.50
6	3.43	0.95	-0.17	-0.22	0.57	0.54
7	4.27	0.77	-1.21	2.35	0.40	0.43
8	3.36	1.11	-0.14	-0.69	0.55	0.61
9	2.48	0.91	0.26	-0.15	0.51	0.53
10	3.95	0.87	-0.69	0.44	0.43	0.60
11	2.92	1.03	0.13	-0.54	0.37	0.39
12	3.14	0.89	-0.03	-0.29	0.25	0.42
13	3.64	0.77	-0.16	0.03	0.37	0.39
14	2.82	1.06	0.17	-0.63	0.55	0.53
15	3.41	0.98	-0.33	-0.40	0.61	0.46
16	2.51	1.08	0.41	-0.47	0.47	0.44
17	4.00	1.01	-0.88	0.17	0.54	0.51
18	3.52	0.98	-0.23	-0.45	0.70	0.67
19	2.80	0.89	0.01	-0.21	0.48	0.60
20	3.26	1.21	-0.25	-0.78	0.66	0.59
21	3.00	1.09	0.00	-0.66	0.60	0.52
22	4.17	0.80	-0.91	0.88	0.44	0.41
23	3.23	1.07	-0.04	-0.59	0.58	0.58
24	2.72	0.98	0.23	-0.34	0.54	0.53
25	3.10	1.19	-0.18	-0.86	0.36	0.37
26	3.42	1.04	-0.28	-0.59	0.48	0.46
27	3.40	0.93	-0.40	-0.12	0.39	0.36
28	3.38	0.95	-0.22	-0.36	0.46	0.63
29	2.68	1.00	0.31	-0.39	0.55	0.64
30	3.76	0.98	-0.60	-0.08	0.55	0.41
31	3.24	1.15	-0.15	-0.82	0.59	0.58
32	3.86	0.79	-0.58	0.58	0.42	0.50
33	3.96	0.87	-0.64	0.25	0.65	0.60
34	3.40	1.04	-0.17	-0.67	0.49	0.57
35	3.38	1.09	-0.31	-0.48	0.69	0.63
36	3.26	0.90	-0.19	-0.14	0.51	0.47

37	3.72	1.02	-0.40	-0.63	0.53	0.54
38	3.54	0.87	-0.27	-0.06	0.49	0.54
39	2.60	1.15	0.35	-0.72	0.72	0.65
40	3.54	0.96	-0.30	-0.33	0.43	0.45
41	3.33	0.91	-0.09	-0.11	0.54	0.57
42	2.92	0.92	-0.16	-0.13	0.44	0.35
43	4.38	0.68	-0.89	0.71	0.39	0.38
44	2.57	0.97	0.26	-0.38	0.56	0.49
45	3.68	1.00	-0.46	-0.36	0.58	0.51
46	3.53	1.03	-0.27	-0.58	0.65	0.63
47	3.91	0.96	-0.52	-0.41	0.52	0.52
48	3.91	1.13	-0.83	-0.21	0.69	0.67
49	2.93	1.02	0.12	-0.49	0.45	0.50
50	3.00	1.26	-0.03	-1.02	0.55	0.49
51	3.09	0.97	-0.06	-0.19	0.57	0.58
52	4.18	0.69	-0.42	-0.26	0.50	0.54
53	3.99	0.83	-0.61	0.30	0.55	0.56
54	2.44	1.21	0.52	-0.71	0.75	0.68
55	3.18	1.05	-0.13	-0.43	0.45	0.48
56	3.41	0.92	-0.19	-0.23	0.51	0.61
57	3.39	1.04	-0.46	-0.28	0.45	0.40
58	3.65	1.04	-0.40	-0.59	0.41	0.41
59	2.47	1.02	0.34	-0.47	0.64	0.60
60	3.40	0.97	-0.16	-0.52	0.66	0.59

Anmerkung. Skala von 1 („stimme überhaupt nicht zu“) bis 5 („stimme voll und ganz zu“), negativ gepolte Items wurden zuvor rekodiert, $N = 770$.

5 Gütekriterien

Objektivität

Standardisierte Instruktionen, Auswertungsvorschriften und vorläufige Referenzwerte gewährleisten die Durchführungs-, Auswertungs-, und Interpretationsobjektivität des BFI-2.

Reliabilität

Im Folgenden werden die Reliabilitätsschätzer der Facetten- und Domänenwerte basierend auf Stichprobe 2 berichtet. Da momentan noch keine längsschnittlichen Daten zum BFI-2 vorliegen, können derzeit nur Schätzer der internen Konsistenz berichtet werden. Cronbach's Alpha wird berichtet, da es die gebräuchlichste Schätzmethode für die Reliabilität einer Skala ist. Wie in Tabelle 5 ersichtlich, liegen die Reliabilitätsschätzer für die Domänenwerte in einem guten und die Reliabilitätsschätzer der Facettenwerte in einem akzeptablen Bereich.

Tabelle 5

Reliabilitätsschätzungen der Skalen

Skala	Cronbach's Alpha
Extraversion	.86
Geselligkeit	.77
Durchsetzungsfähigkeit	.76
Aktivität	.69
Verträglichkeit	.81
Mitgefühl	.70
Höflichkeit	.68
Vertrauen	.60
Gewissenhaftigkeit	.87
Ordnungsliebe	.85
Fleiß	.74
Verlässlichkeit	.62
Negative Emotionalität	.88
Ängstlichkeit	.69
Niedergeschlagenheit	.81
Unbeständigkeit der Gefühle	.77
Offenheit	.84
Ästhetisches Empfinden	.80
Intellektuelle Neugierde	.63
Kreativer Einfallsreichtum	.79

Anmerkung: $N = 770$.

Validität

Die inhaltliche Validität der Skala kann als hoch bewertet werden, da es sich bei der deutschen Version des BFI-2 um eine Adaption der englischen Originalversion handelt. Die Entwicklung der englischen Originalversion orientierte sich am Big-Five Modell der Persönlichkeit (Goldberg, 1993; John, Naumann, & Soto, 2008; McCrae & Costa, 2008) und ist ausführlich bei Soto & John (2016) dargestellt. Die faktorielle Struktur der Items und Facetten (vgl. Abbildung 1 und Tabelle 3) können als erste Hinweise für die Konstruktvalidität der Skala gedeutet werden.

Zusätzlich wurde die Kriteriumsvalidität der Skala in Hinblick auf den Zusammenhang mit Bildung, Einkommen, Gesundheitszustand, Risikobereitschaft und Lebenszufriedenheit bewertet. Es wurde erwartet, dass Offenheit und Gewissenhaftigkeit einen positiven Zusammenhang mit Bildungserfolg aufweisen (Rammstedt, Kemper, Céline, Klein & Kovaleva, 2013; Poropat, 2009), Negative Emotionalität (früher: Neurotizismus) einen negativen Zusammenhang mit Einkommen aufweist, Gesundheit positiv mit Gewissenhaftigkeit korreliert (Bogg & Roberts, 2004), Risikobereitschaft positiv mit Extraversion und Offenheit korreliert (Beierlein, Kovaleva, Kemper & Rammstedt, 2014) und

Lebenszufriedenheit negativ mit Negativer Emotionalität (früher: Neurotizismus) korreliert (Rammstedt, 2007).

Die Zusammenhänge wurden in Stichprobe 2 untersucht und sind in Tabelle 6 dargestellt. Nach Cohen (1988) werden Korrelationen $r > .50$ als groß, $r > .30$ als mittel und $r > .10$ als klein interpretiert. Insgesamt bestätigen die Zusammenhänge mit den Außenkriterien die Kriteriumsvalidität der Skalen. Der stärkste Zusammenhang zeigt sich zwischen Lebenszufriedenheit und Negativer Emotionalität (früher: Neurotizismus, $r = -.52$). Emotional labile Personen beschreiben sich als besonders unzufrieden. Betrachtet man die Zusammenhänge auf Facettenebene zeigt sich, dass v.a. die Facette Niedergeschlagenheit zu diesem Zusammenhang beiträgt. Erwartungsgemäß zeigt sich ebenfalls ein stark positiver Zusammenhang zwischen Extraversion und Risikobereitschaft ($r = .41$), der v.a. durch die Facette Aktivität determiniert wird. Offenheit weist erwartungsgemäß einen mittleren positiven Zusammenhang mit Bildung auf, der v.a. durch die Facette Ästhetische Sensitivität erklärt werden kann. Kleine erwartungsgemäße Zusammenhänge zeigen sich auch zwischen Gesundheit und Gewissenhaftigkeit ($r = .15$) und zwischen Einkommen und Negativer Emotionalität ($r = -.12$).

Tabelle 6

Korrelationen zwischen Kriterien und Domänen- und Facettenwerten

Skala	Bildungsabschluss	Einkommen	Gesundheit	Risikobereitschaft	Lebenszufriedenheit
Extraversion	.10**	.07*	.25**	.41***	.34***
Geselligkeit	.03	.06	.18***	.30***	.25***
Durchsetzungsfähigkeit	.11**	.11**	.13***	.35***	.24***
Aktivität	.10**	.03	.31***	.38***	.39***
Verträglichkeit	-.03	-.06	.08*	.01	.18***
Mitgefühl	-.03	.07	-.01	.00	.09*
Höflichkeit	-.07*	.02	.11***	-.08*	.16***
Vertrauen	-.01	.04	.10***	.10***	.18***
Gewissenhaftigkeit	-.03	.12***	.15***	-.02	.25***
Ordnungsliebe	-.02	.15***	.07*	-.04	.16***
Fleiß	-.03	.08*	.15***	.10***	.25*
Verlässlichkeit	-.01	.08*	.16***	-.13***	.24***
Negative Emotionalität	-.06	-.12**	-.45***	-.21***	-.52***
Ängstlichkeit	-.05	.14***	-.38***	.23***	.40***
Niedergeschlagenheit	-.07	.09*	-.50***	-.24***	.61***
Unbeständigkeit der Gefühle	-.02	.07*	-.27***	-.07	.30***
Offenheit	.25***	.02	.13***	.21***	.15***
Ästhetisches Empfinden	.27***	.02	.06	.07	.03
Intellektuelle Neugierde	.19***	.02	.16***	.24***	.17***
Kreativer Einfallsreichtum	.11***	.03	.09*	.23***	.17***

Anmerkung: $N = 770$. * $p < .05$, ** $p < .01$, *** $p < .001$.

Deskriptive Statistiken (Normierung)

Stichprobe 2 liefert vorläufige Referenzdaten für die Domänen- und Facettenwerte. Mittelwerte und Standardabweichungen für die Gesamtstichprobe und für Männer und Frauen sind in Tabelle 7 dargestellt.

Tabelle 6

Referenzwerte für die Gesamtstichprobe und Frauen und Männer

	Gesamtstichprobe (N = 770)		Frauen (N = 396)		Männer (N = 374)	
	M	SD	M	SD	M	SD
Extraversion	3.22	0.63	3.26	0.65	3.18	0.61
Geselligkeit	3.20	0.83	3.32	0.85	3.08	0.79
Durchsetzungsfähigkeit	3.20	0.75	3.12	0.76	3.28	0.73
Aktivität	3.27	0.70	3.34	0.71	3.19	0.69
Verträglichkeit	3.76	0.51	3.87	0.49	3.65	0.49
Mitgefühl	3.95	0.65	4.13	0.61	3.77	0.64
Höflichkeit	4.08	0.59	4.17	0.59	3.99	0.58
Vertrauen	3.25	0.65	3.31	0.66	3.20	0.64
Gewissenhaftigkeit	3.67	0.62	3.73	0.61	3.60	0.62
Ordnungsliebe	3.71	0.85	3.73	0.84	3.68	0.87
Fleiß	3.53	0.73	3.62	0.73	3.43	0.72
Verlässlichkeit	3.76	0.60	3.84	0.57	3.68	0.62
Negative Emotionalität	2.72	0.67	2.79	0.68	2.64	0.65
Ängstlichkeit	2.96	0.70	3.07	0.70	2.83	0.68
Niedergeschlagenheit	2.56	0.85	2.59	0.87	2.53	0.83
Unbeständigkeit der Gefühle	2.63	0.78	2.70	0.80	2.56	0.76
Offenheit	3.38	0.64	3.42	0.64	3.34	0.62
Ästhetisches Empfinden	3.11	0.96	3.23	0.94	2.97	0.97
Intellektuelle Neugierde	3.48	0.68	3.47	0.68	3.50	0.67
Kreativer Einfallsreichtum	3.56	0.77	3.56	0.79	3.56	0.75

Nebengütekriterien

Die Messinvarianz des BFI-2 für Männer und Frauen wurde mit Multigruppen Strukturgleichungsmodellen untersucht (Breyer, Bluemke, & Danner, 2016; Chen, 2007; Vandenberg & Lance, 2000). Konfigurale Invarianz liegt vor, wenn die Faktorstruktur der Items vergleichbar ist, wobei sich die Faktorladungen und die Intercepts der Items zwischen den Gruppen unterscheiden dürfen. Liegt konfigurale Invarianz vor, weist das Konstrukt in allen Gruppen eine vergleichbare dimensionale Struktur auf. Metrische Invarianz liegt vor, wenn die Höhe der Faktorladungen zwischen den Gruppen gleich ist. Liegt metrische Invarianz vor, können Korrelationen zwischen den Gruppen verglichen werden. Skalare Invarianz liegt vor, wenn zusätzlich die Intercepts (Schwierigkeiten) der Items in allen Gruppen gleich ist. Liegt skalare Invarianz vor, können auch Mittelwerte zwischen den

Gruppen verglichen werden. Welche Form von Invarianz vorliegt, wurde anhand der Kriterien von Chen (2007) entschieden. Chen empfiehlt bei Differenzen von $\Delta\text{RMSEA} \leq .015$, $\Delta\text{CFI} \leq .010$, und $\Delta\text{SRMR} \leq .030$ zwischen konfiguralem und metrischem Modell das metrische Modell anzunehmen, und bei Differenzen von $\Delta\text{RMSEA} \leq .015$, $\Delta\text{CFI} \leq .010$ und $\Delta\text{SRMR} \leq .010$ zwischen metrischen und skalaren Modell das skalare Modell anzunehmen. Die Modelfitindices sind in Tabelle 7 dargestellt und deuten auf die skalare Invarianz des Modells für Frauen und Männer hin. Das bedeutet, dass Korrelationen und Mittelwerte der Domänen- und Facettenwerte zwischen Männern und Frauen verglichen werden können.

Tabelle 7

Messinvarianzprüfung

	RMSEA	CFI	SRMR	df
Konfigurale Invarianz	.064	.727	.088	3377
Metrische Invarianz	.064	.726	.089	3422
Skalare Invarianz	.065	.716	.090	3467

Anmerkung: $N = 770$.

6 Literatur und Datenquellen

Kontakt zu Autor(en)

Daniel Danner, GESIS – Leibniz-Institut für Sozialwissenschaften, Postfach 122155, 68072 Mannheim, daniel.danner@gesis.org

Literaturverzeichnis

- Beierlein, C., Kovaleva, A., Kemper, C.J. & Rammstedt, B. (2014). Eine Single-Item-Skala zur Erfassung von Risikobereitschaft: Die Kurzskala Risikobereitschaft-1 (R-1). *GESIS – Leibniz-Institut für Sozialwissenschaften Working Papers*, 34.
- Bogg, T. & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological bulletin*, 130(6), 887. doi: 10.1037/0033-2909.130.6.887
- Breyer, B., Bluemke, M. & Danner, D. (2016). *ZIS Publication Guide: Leitfaden für die Veröffentlichung von Erhebungsinstrumenten in der Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). Mannheim: GESIS. doi:<http://dx.doi.org/10.6102/pubguide1>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, 14(3), 464-504.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, 48(1), 26.

- Harkness, M. R. (2003). *U.S. Patent No. 6,589,776*. Washington, DC: U.S. Patent and Trademark Office.
- John, O. P. (1989). Towards a taxonomy of personality descriptors. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 261–271). New York, NY: Springer. doi:10.1007/978-1-4684-0634-4_20
- John, O. P. (1990). The “Big Five” factor taxonomy: Dimensions of personality in the natural language and questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York, NY: Guilford Press.
- John, O. P., Donahue, E. M. & Kentle, R. (1991). ‘The “Big Five. *inventory—version 4a and, 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
- John, O. P., Naumann, L. P. & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, 3, 114-158.
- Costa, P. T. & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, 2, 179-198.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates. doi:10.1111/j.2044-8317.1981.tb00621.x
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, 135(2), 322.
- Rammstedt, B. (1997). *Die deutsche Version des Big Five Inventory (BFI): Übersetzung und Validierung eines Fragebogens zur Erfassung des Fünf-Faktoren-Modells der Persönlichkeit*. Unpublished thesis. University of Bielefeld, Germany.
- Rammstedt, B., Goldberg, L. R. & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, 44(1), 53-61. doi:10.1016/j.jrp.2009.10.005
- Rammstedt, B., Kemper, C. J., Céline, M., Klein, C. B. & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. *Methoden, Daten, Analysen*, 7, 233-249. doi:10.12758/mda.2013.013
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173-184.
- Soto, C. J. & John, O. P. (2016). The Next Big.Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model with 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power. *Journal of Personality and Social Psychology*. doi: 10.1037/pspp0000096
- Soto, C. J., John, O. P., Gosling, S. D. & Potter, J. (2008). The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of personality and social psychology*, 94(4), 718. doi:10.1037/0022-3514.94.4.718
- Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4-70.