

**43.651 Tests
ausgewertet:
So zuverlässig ist
der IQ-Test**



Analyse von 43.651 Testdurchführungen

Interne Konsistenz, Aufgabenschwierigkeit, Trennschärfe, Faktorstruktur und zeitliche Robustheit

Stichprobe

43.651

norm_valid = 1

Items

46

q40 ausgeschlossen

Alpha

0,876

interne Konsistenz

Omega

0,875

Pearson, 1 Faktor

Kurzfazit

Die 46 auswertbaren Aufgaben zeigen eine hohe interne Konsistenz. Die meisten Aufgaben tragen sinnvoll zum Gesamtscore bei, und ein deutlicher gemeinsamer Leistungsfaktor prägt die Antworten. Die Analyse ersetzt keine externe Validierung gegenüber professionellen Intelligenztests.

1. Ergebnis auf einen Blick

Die beiden zentralen Reliabilitätsmaße führen praktisch zur gleichen Einschätzung: Cronbachs Alpha beträgt 0,876, McDonalds Omega auf Basis der beobachteten Pearson-Korrelationen 0,875. Das ist ein starkes Ergebnis für die interne Konsistenz der 46-Item-Skala. Die Kennzahlen sagen jedoch nicht, ob ein ausgegebener IQ-Wert exakt mit einer professionellen Diagnostik übereinstimmt. Reliabilität und Validität sind unterschiedliche Anforderungen an einen Test. [1]

Bereich	Ergebnis	Einordnung
Reliabilität	$\alpha = 0,876$; $\omega = 0,875$	Hohe interne Konsistenz
Schwierigkeit	2,19 % bis 98,13 % richtig	Breiter Leistungsbereich
Trennschärfe	Median 0,3615	32 von 46 Aufgaben mindestens 0,30
Faktorstruktur	1. Eigenwert 12,98	Deutlicher gemeinsamer Faktor
Zeitstabilität	$r = 0,986$ / $0,989$	Sehr ähnliche Aufgabenrangfolge

Welche Daten wurden untersucht?

Ausgewertet wurden alle 43.651 im SQL-Export enthaltenen historischen Durchführungen mit `norm_valid = 1`. Der Zeitraum reicht vom 10. Mai 2010 bis zum 1. September 2026. Analysiert wurden 46 binär kodierte Aufgaben mit den Werten richtig = 1 und falsch = 0. Fehlende Itemwerte lagen nicht vor.

Festlegung	Umsetzung
Stichprobe	43.651 als normierungsg geeignet markierte Fälle
Analysierte Aufgaben	q1 bis q47 ohne q40
Ausschluss q40	Vollständig aus Skalen-, Item- und Faktorenanalysen entfernt
Personenbezogene Daten	Nicht Bestandteil dieses Berichts oder der veröffentlichten Excel-Datei

Warum q40 ausgeschlossen wurde

Bei q40 war historisch zeitweise die falsche Lösung 115 hinterlegt. Gespeichert wurde nur das damalige Richtig-falsch-Ergebnis, nicht die ursprüngliche Eingabe. Eine seriöse nachträgliche Korrektur ist deshalb nicht möglich.

2. Reliabilität der Gesamtskala

Cronbachs Alpha beschreibt, wie konsistent die Aufgaben als Skala zusammenarbeiten. Für binäre Aufgaben entspricht die Berechnung inhaltlich der Kuder-Richardson-Logik. [2,3] Der Wert darf nicht als Beleg für Eindimensionalität oder Validität missverstanden werden. [4]

Hauptbefund

Alpha = 0,8761 und Omega = 0,8751 liegen nahezu gleichauf. Die Übereinstimmung stärkt die Schlussfolgerung, dass die 46 Aufgaben als Gesamtskala zuverlässig zusammenarbeiten.

Ergebniswerte

Kennzahl	Wert	Verwendung
Cronbachs Alpha	0,876134	Hauptwert
Standardisiertes Alpha	0,872617	Methodischer Zusatz
McDonalds Omega, Pearson	0,875077	Direkt mit Alpha vergleichbare Ergänzung
Ordinales Omega, tetrachorisch	0,937743	Modellabhängige Sensitivitätsanalyse
Mittelwert Rohscore	24,4468	Keine IQ-Norm
Standardabweichung Rohscore	7,9982	Streuung der 46-Item-Summe
Median Rohscore	25	Mittlerer beobachteter Summenscore

Berechnungsprinzip

Cronbachs Alpha: $\alpha = k/(k - 1) \times [1 - \sum \sigma_i^2 / \sigma_x^2]$

Dabei ist k die Anzahl der Aufgaben, σ_i^2 die Varianz einer einzelnen Aufgabe und σ_x^2 die Varianz des Gesamtscores. Die ergänzende Excel-Datei dokumentiert die verwendeten Formeln und die veröffentlichten Ergebniswerte; vollständige Item- und Antwortdaten bleiben intern.

McDonalds Omega: $\omega = (\sum \lambda_i)^2 / [(\sum \lambda_i)^2 + \sum \theta_i]$

λ_i bezeichnet die Faktorladung einer Aufgabe, θ_i ihre nicht durch den gemeinsamen Faktor erklärte Varianz. Omega beruht auf einem Faktorenmodell und ergänzt Alpha sinnvoll. [5,6] Für kategoriale Summenscores sind modellbasierte Alternativen besonders relevant. [7]

Einordnung

Nach einer häufig verwendeten Faustregel liegt Alpha = 0,876 im Bereich 'gut'. [11] Solche Grenzen sind keine verbindlichen wissenschaftlichen Gütesiegel; entscheidend sind Testzweck, Itemstruktur und weitere Analysen. [4]

3. Aufgabenschwierigkeit

Die Aufgabenschwierigkeit wurde als Anteil richtiger Antworten definiert: $p_i = \text{Anzahl richtiger Antworten auf Aufgabe } i / N$. Ein hoher Wert steht damit für eine leichte Aufgabe. Der Median beträgt 0,5376; die typische Aufgabe wurde also von ungefähr 54 Prozent der Teilnehmenden gelöst.

Kennzahl	Ergebnis
Höchster Lösungsanteil	98,13 % richtig
Niedrigster Lösungsanteil	2,19 % richtig
Median	53,76 % richtig
Aufgaben zwischen 20 % und 80 %	37 von 46
Sehr leichte Aufgaben über 90 %	4 von 46
Sehr schwere Aufgaben bis 10 %	2 von 46

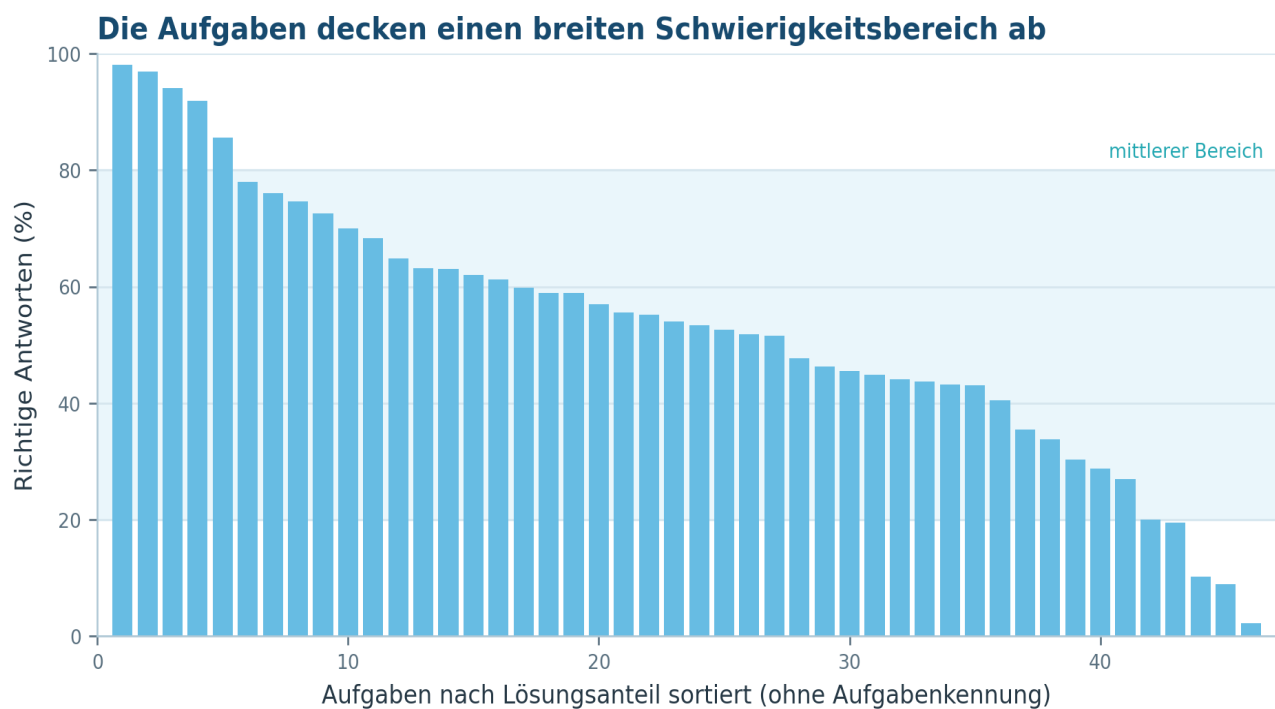


Abbildung 1: Anteil richtiger Antworten, sortiert von leicht nach schwer. Aufgabenkennungen werden nicht veröffentlicht; q40 ist ausgeschlossen.

Bewertung

Die unterschiedlichen Schwierigkeitsgrade sind ein positiver Befund: Leichte Aufgaben erleichtern den Einstieg, mittlere Aufgaben differenzieren im Zentrum und schwere Aufgaben erweitern die Messung am oberen Leistungsende. Problematisch wäre eine starke Häufung auf nur einem Schwierigkeitsniveau. Das ist hier nicht der Fall.

4. Korrigierte Trennschärfe

Die korrigierte Item-Gesamtkorrelation ist die Pearson-Korrelation zwischen einer Aufgabe und dem Summenscore der übrigen 45 Aufgaben: $r_{it(i)} = \text{cor}(X_i, X_{\text{gesamt}} - X_i)$. Das Entfernen der betrachteten Aufgabe verhindert, dass sie teilweise mit sich selbst korreliert. [8]

Eine Pearson-Korrelation kann zwischen -1 und +1 liegen. Ein positiver Wert bedeutet hier: Wer die Aufgabe richtig löst, erzielt tendenziell auch bei den übrigen Aufgaben einen höheren Score. Bei einer binären Aufgabe ist diese Kennzahl mathematisch eine punktbiseriale Korrelation, also ein Sonderfall der Pearson-Korrelation.

Kennzahl	Ergebnis
Median	0,3615
Mindestens 0,30	32 von 46 Aufgaben
Unter 0,20	11 von 46 Aufgaben
Höchster Wert	0,562
Niedrigster Wert	$q_{35} \cdot 0,017$
Weitere Aufgaben unter 0,20	10 von 46

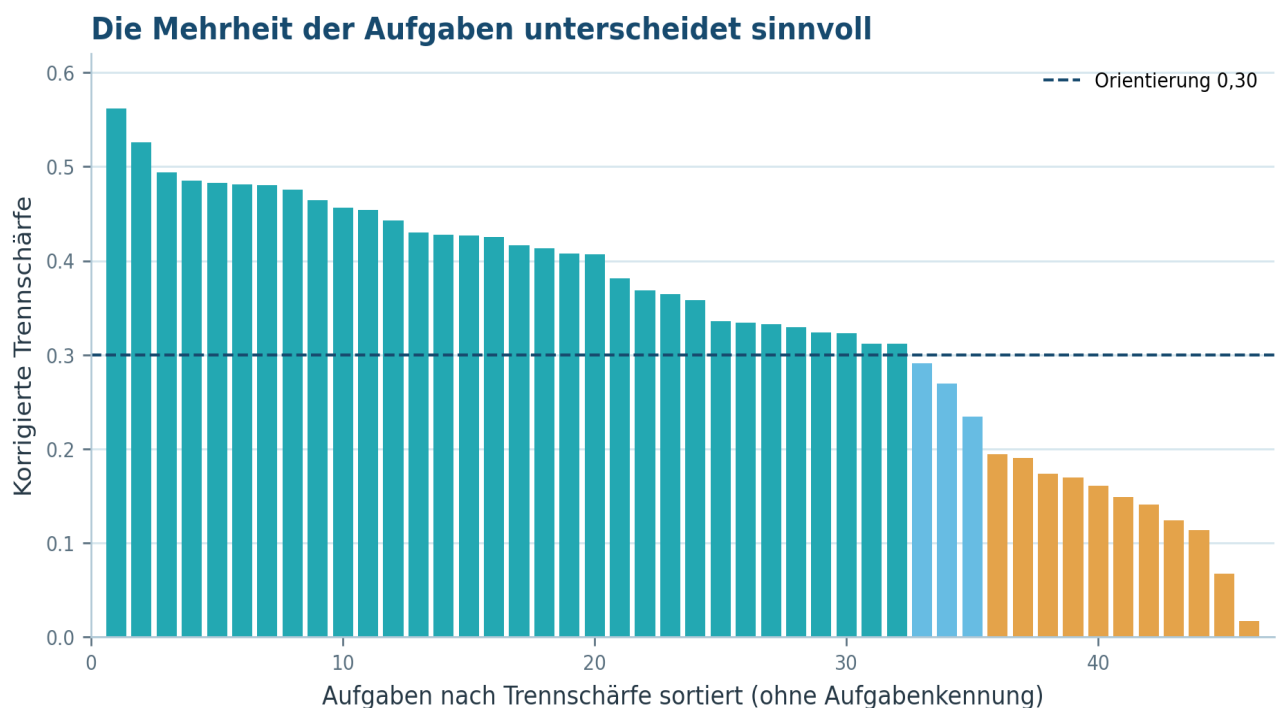


Abbildung 2: Korrigierte Trennschärfen, unabhängig sortiert und ohne Aufgabenkennungen. Höhere Werte zeigen einen stärkeren Zusammenhang mit dem übrigen Test.

Bewertung

Der Median von 0,3615 ist insgesamt solide. Die Mehrheit der Aufgaben unterscheidet sinnvoll zwischen schwächeren und stärkeren Gesamtergebnissen. q_{35} ist dagegen ein klarer Prüfpunkt; ohne q_{35} steigt Alpha leicht von 0,876 auf 0,879. Eine Aufgabe sollte dennoch nicht allein zur Maximierung von Alpha entfernt werden.

5. Faktorstruktur

Für jedes der 1.035 Aufgabenpaare wurde eine tetrachorische Korrelation geschätzt. Das Verfahren behandelt die beobachteten 0/1-Antworten als Schwellenwerte einer zugrunde liegenden stufenlosen Fähigkeit oder Lösungsneigung und wird häufig für Faktorenanalysen binärer und ordinaler Aufgaben eingesetzt. [9]

Kennzahl	Ergebnis
Erster Eigenwert	12,9766
Zweiter Eigenwert	2,4518
Dritter Eigenwert	1,9117
Vierter Eigenwert	1,3756
Verhältnis erster zu zweitem Eigenwert	5,29
Varianzanteil des ersten Faktors	28,21 %
Ein-Faktor-Ladungen mindestens 0,30	39 von 46 Aufgaben
Höchste Ladung	0,746
Niedrigste Ladung	q35 · 0,016

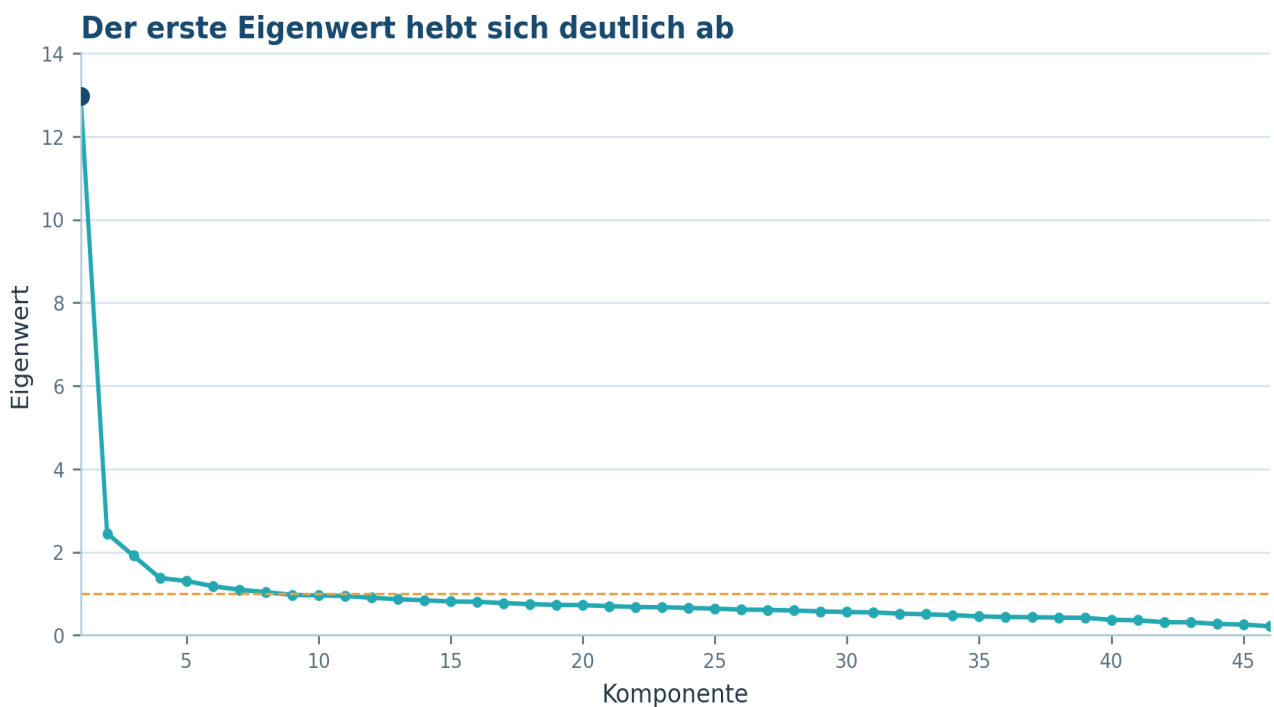


Abbildung 3: Scree-Plot der tetrachorischen Korrelationsmatrix. Der große Abstand zwischen erstem und zweitem Eigenwert spricht für einen dominanten gemeinsamen Faktor.

Fazit zur Struktur

Ein gemeinsamer Faktor ist für einen Test mit Gesamtscore grundsätzlich positiv: Die meisten Aufgaben besitzen eine gemeinsame Leistungsgrundlage. Gleichzeitig ist die Skala nicht vollkommen eindimensional. Mehrere Eigenwerte über 1 werden nicht automatisch als getrennte Fähigkeiten interpretiert, da diese Regel bei vielen Aufgaben zu viele kleine Faktoren nahelegen kann. [10]

6. Zeitliche Robustheit und Prüfpunkte

Als Robustheitscheck wurde die Stichprobe chronologisch in zwei nahezu gleich große Hälften geteilt. Die zentralen Kennwerte sind in beiden Zeitabschnitten sehr ähnlich.

Kennzahl	Frühere Hälfte	Spätere Hälfte / Vergleich
Fälle	21.825	21.826
Grenzdatum	bis 28.03.2014, 14:00	ab 28.03.2014, 14:41
Cronbachs Alpha	0,873	0,880
Korrelation der Itemschwierigkeiten	-	0,986
Korrelation der Trennschärfen	-	0,989

Offengelegte Prüfpunkte

Bereich	Befund	Nächster sinnvoller Schritt
q35	Trennschärfe 0,017; Faktorladung 0,016; in beiden Zeitabschnitten schwach	Aufgabeneinhalt, Lösung und Darstellung priorisiert prüfen
q40	Historische Auswertung nicht rekonstruierbar	Dauerhaft aus dieser Analyse ausgeschlossen lassen
Weitere Auffälligkeiten	Einzelne Aufgaben liegen unter den üblichen Orientierungswerten oder verändern sich über die Zeit	Intern inhaltlich und technisch prüfen

Gesamtbewertung der Robustheit

Die Itemschwierigkeiten und Trennschärfen behalten über die Zeit fast dieselbe Rangfolge. Das spricht dafür, dass die zentralen Ergebnisse nicht nur durch einen einzelnen historischen Zeitraum entstehen.

Umfang der öffentlichen Excel-Datei

Die ergänzende Excel-Datei enthält die zentralen Ergebniswerte, getrennt sortierte und anonymisierte Verteilungen der Aufgabenschwierigkeiten und Trennschärfen, das Eigenwertspektrum, die Methodik und die Quellen. Nicht veröffentlicht werden Aufgabenkennungen zu den Verteilungswerten, vollständige Itemkennwerte, Korrelationsmatrizen, zeitliche Einzelwerte, Antwortdaten und die Rohwert-IQ-Umrechnung.

7. Methodik und reproduzierbare Ergebniswerte

Analyseschritte

Analyse	Umsetzung
Stichprobe	Alle im Export enthaltenen Fälle mit norm_valid = 1
Items	46 Aufgaben; q40 vollständig ausgeschlossen
Reliabilität	Cronbachs Alpha; standardisiertes Alpha; Ein-Faktor-Omega
Itemschwierigkeit	Anteil richtiger Antworten je Aufgabe
Trennschärfe	Pearson-Korrelation mit dem Summenscore der übrigen 45 Aufgaben
Faktorstruktur	Tetrachorische Korrelationsmatrix; Eigenwerte; exploratives MinRes-Modell
Robustheit	Chronologische Halbierung und Vergleich der Itemkennwerte

Exakte Ergebniswerte

Analyse	Wert
Fälle	43.651
Items	46
Mittelwert Summenscore	24,4468
Standardabweichung Summenscore	7,9982
Median Summenscore	25
Cronbachs Alpha	0,876134
Standardisiertes Alpha	0,872617
McDonalds Omega, Pearson	0,875077
Ordinales Omega, tetrachorisch	0,937743
Median Itemschwierigkeit	0,537605
Median korrigierte Trennschärfe	0,361502
Erster tetrachorischer Eigenwert	12,9766
Zweiter tetrachorischer Eigenwert	2,4518
Verhältnis erster zu zweitem Eigenwert	5,2926
Anteil des ersten Faktors	28,2100 %
Ladungen auf Faktor 1 mindestens 0,30	39 von 46

Technischer Hinweis: Die tetrachorische Korrelationsmatrix war ohne Korrektur positiv definit; keine 2 x 2-Häufigkeitstabelle enthielt eine leere Zelle. Der erste Faktor erklärt 28,21 Prozent der Gesamtvarianz der Korrelationsmatrix. Dieser Wert ist keine Reliabilitätsquote.

8. Grenzen und Schlussfolgerung

Was diese Analyse nicht zeigt

Nicht untersucht	Konsequenz
Externe Validität	Keine Aussage zur Übereinstimmung mit WAIS, IST, Raven oder anderen etablierten Tests
Test-Retest-Reliabilität	Keine Aussage zur Stabilität derselben Person über wiederholte Messungen
Repräsentative Normierung	Die historische Online-Stichprobe ist keine repräsentative Bevölkerungsstichprobe
Individuelle Messunsicherheit	Kein Konfidenzintervall und keine diagnostische Genauigkeit für einen einzelnen IQ-Wert
q40	Keine nachträgliche Korrektur der historischen Antworten möglich

Seriöse Kernaussage

Die 46 historisch auswertbaren Aufgaben erreichten in 43.651 bereinigten Testdurchführungen eine hohe interne Konsistenz. Die meisten Aufgaben zeigen sinnvolle Zusammenhänge mit dem übrigen Testergebnis. Eine Faktorenanalyse auf Basis tetrachorischer Korrelationen ergibt einen deutlichen gemeinsamen Leistungsfaktor, daneben aber kleinere spezifische Strukturen.

Damit fällt das interne Ergebnis für einen frei zugänglichen Online-Test insgesamt stark aus. Besonders wertvoll ist, dass Alpha und Omega fast identische Werte liefern, die Aufgaben einen breiten Schwierigkeitsbereich abdecken und die Rangfolge der Itemkennwerte über die Zeit sehr stabil bleibt.

Zur wissenschaftlich sauberen Einordnung gehört zugleich die Begrenzung der Aussage: Die Analyse belegt die interne Funktionsweise der Skala. Sie ersetzt keine externe Validierung und keine psychologische Diagnostik. [1]

9. Quellen und weiterführende Literatur

Die Nummern entsprechen den Verweisen im Bericht. Links sind angegeben, soweit ein geeigneter öffentlicher Nachweis verfügbar ist.

- [1] American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014): Standards for Educational and Psychological Testing.
<https://www.aera.net/publications/books/standards-for-educational-psychological-testing-2014-edition>
- [2] Cronbach, L. J. (1951): Coefficient Alpha and the Internal Structure of Tests. Psychometrika, 16, 297-334.
<https://doi.org/10.1007/BF02310555>
- [3] Kuder, G. F. & Richardson, M. W. (1937): The Theory of the Estimation of Test Reliability. Psychometrika, 2, 151-160.
<https://doi.org/10.1007/BF02288391>
- [4] Sijtsma, K. (2009): On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. Psychometrika, 74, 107-120.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC2792363/>
- [5] McDonald, R. P. (1978): Generalizability in Factorable Domains: Domain Validity and Generalizability. Educational and Psychological Measurement, 38, 75-79.
<https://doi.org/10.1177/001316447803800111>
- [6] Dunn, T. J., Baguley, T. & Brunsden, V. (2014): From Alpha to Omega: A Practical Solution to the Pervasive Problem of Internal Consistency Estimation. British Journal of Psychology, 105, 399-412.
<https://doi.org/10.1111/bjop.12046>
- [7] Green, S. B. & Yang, Y. (2009): Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. Psychometrika, 74, 155-167.
<https://doi.org/10.1007/s11336-008-9099-3>
- [8] Henrysson, S. (1963): Correction of Item-Total Correlations in Item Analysis. Psychometrika, 28, 211-218.
<https://doi.org/10.1007/BF02289618>
- [9] Flora, D. B. & Curran, P. J. (2004): An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. Psychological Methods, 9, 466-491.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC3153362/>
- [10] Zwick, W. R. & Velicer, W. F. (1986): Comparison of Five Rules for Determining the Number of Components to Retain. Psychological Bulletin, 99, 432-442.
<https://doi.org/10.1037/0033-2909.99.3.432>
- [11] George, D. & Mallery, P. (2003): SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon.
<https://books.google.com/books?id=AghHAAAAMAAJ>

Dateien zum Bericht

Die ergänzende Excel-Arbeitsmappe iq-test-analyse-43651-oeffentliche-daten.xlsx enthält zentrale Ergebnismerte, anonymisierte Verteilungen, das Eigenwertspektrum, die Methodik und die Quellen. Itemzuordnungen, vollständige Korrelationsmatrizen, zeitliche Einzelwerte, Antwortdaten und die Rohwert-IQ-Umrechnung sind nicht enthalten.