

RESEARCH

Open Access



Reliability generalization meta-analysis of the internal consistency of the Big Five Inventory (BFI) by comparing BFI (44 items) and BFI-2 (60 items) versions controlling for age, sex, language factors

Waqar Husain^{1*} , Areen Jamal Haddad², Muhammad Ahmad Husain¹, Hadeel Ghazzawi² , Khaled Trabelsi^{3,4} , Achraf Ammar^{5,6} , Zahra Saif⁷ , Amir Pakpour⁸ and Haitham Jahrami^{7,9}

Abstract

Introduction The Big Five Inventory (BFI) is a popular measure that evaluates personality on the Big-Five model. Apart from its utilization across cultures, the literature did not reveal any meta-analysis for the reliability of the different versions of the BFI and its translations. The current study carried out a reliability generalization meta-analysis (REGEMA) to establish the reliability of the BFI across cultures and languages.

Methods We searched 30 databases for the relevant studies from 1991 to mid-November 2024. The studies that we intended to include in our meta-analysis required to have utilized the BFI (44 items) and the BFI-2 (60 items) and have reported Cronbach's alpha or McDonald's omega reliability estimates. Our coded variables included BFI version, sample size, population type, age, gender, clinical state, and reliability. A total of 57 studies (datapoints) published in 34 research articles (involving 43,715 participants; 60.24% women; Mean age = 30.08) from various cultures and languages were finally included. These studies used BFI and BFI-2 in Arabic, Chinese, Croatian, Czech, Danish, Dutch, English, French, German, Indonesian, Italian, Japanese, Malay, Norwegian, Polish, Portuguese, Russian, Serbian, Spanish, Swahili, and Turkish. Data analysis was conducted using the metafor and meta packages in R. The average correlation was computed using a random-effects model and reliability coefficients indicated effect size. I^2 and Cochran's Q tests were used to examine heterogeneity, with prediction intervals suggesting genuine influences around the pooled estimate. Using funnel plots, regression-based tests (e.g., Egger's regression, rank correlation), and trim-and-fill imputation, publication bias was adjusted to estimate unbiased effects.

Results We calculated the individual and combined reliability of the BFI and BFI-2 across languages and cultures. The results revealed the reliability of all five factors used in the BFI/BFI-2. The BFI estimates provide the following results: openness is estimated at 0.77 (95% CI: 0.75; 0.80); conscientiousness is estimated at 0.80 (95% CI: 0.78; 0.82); extraversion is also estimated at 0.80 (95% CI: 0.79; 0.82); agreeableness is estimated at 0.73 (95% CI: 0.71; 0.76); and neuroticism is estimated at 0.80 (95% CI: 0.79; 0.82). The BFI-2 estimates are as follows: openness is estimated at 0.83 (95% CI:

*Correspondence:

Waqar Husain
drsuikoon@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

0.82; 0.84); conscientiousness is estimated at 0.86 (95% CI: 0.85; 0.87); extraversion is estimated at 0.85 (95% CI: 0.84; 0.86); agreeableness is also estimated at 0.80 (95% CI: 0.79; 81); and neuroticism is estimated at 0.89 (95% CI: 0.88; 0.89).

Conclusion The current meta-analysis represents the first reliability analysis of the BFI and the first comparison between its two different versions, the BFI (44 items) and the BFI-2 (60 items). The generalized reliability of both the BFI and BFI-2 were established. The findings confirm that the BFI and BFI-2 have good reliability across all five factors.

Keywords Big-Five Inventory, Cultures, Generalizability, Meta-analysis, Reliability

Introduction

Personality, influenced by genetic and environmental factors [1], is a complicated construct that refers to the constant patterns of behavior, thoughts, and emotions that help explain individuals [2, 3]. Understanding personality is important as it influences various aspects of human life. Interpersonal relationships, occupational success, and the overall well-being of a person mostly relies on significant personality traits. One of the foundational theories of personality is Freud's psychoanalytic theory. This theory links unconscious impulses and conflicts to the development of personality [4–7]. Social cognitive theory views personality as a combination of behavioral and cognitive skills. These skills are shaped by an individual's interactions with the environment [8, 9]. Trait theories of personality focus on identifying and categorizing stable traits that can be observed across different situations and over time [10–12]. Modern psychobiological theories emphasize the interplay of biological, psychological, and social elements in personality development [13–15].

One of the most renowned and well-researched frameworks in personality psychology is the Big Five Model [16]. This model includes five primary personality traits *i.e.* openness, conscientiousness, extraversion, agreeableness, and neuroticism. These five traits or factors capture a broad range of human behavior and attitudes. These traits form the acronym OCEAN. Openness to experience reflects differences in creativity, curiosity, attractiveness, and social attitudes. Individuals with high openness tend to be imaginative, artistic, and open-minded. They often accept new ideas and experiences. Individuals with low openness may prefer routine and traditional values. They demonstrate a preference for familiar experiences. Conscientiousness incorporates adherence to social norms, persistence, and goal orientation. This results in hard work and diligence. Individuals with high conscientiousness are known for being reliable, organized, and disciplined. They are goal-oriented and often excel in achieving their objectives. Those with low conscientiousness may display more spontaneous and adaptable behaviors. They prioritize flexibility over strict adherence to rules and plans. Extraversion refers to social dominance, friendliness, enthusiasm, and activity. Highly

extroverted individuals are often outgoing and lively. They enjoy engaging with others in social settings. They may seek out excitement and thrive in environments that provide opportunities for interaction. Introverted individuals often prefer solitary and quieter surroundings. They find fulfillment in more introspective activities. Agreeableness reflects sympathy and collaboration rather than skepticism. Individuals with high agreeableness are typically empathic, friendly, and cooperative. They prioritize balanced relationships with others. Those with low agreeableness may exhibit more competitive or skeptical tendencies. They prioritize their own interests over those of others. Neuroticism is characterized by frequent negative emotions, fear, despair, anxiety, impatience, and low self-esteem. Individuals with high neuroticism may experience frequent mood swings and unpleasant emotions. This makes them more susceptible to stress and emotional challenges. Individuals with low neuroticism tend to be more emotionally durable and relaxed. They manage stress and adversity more effectively [17–19].

The Big Five Model has been widely studied and validated across many groups and cultures. This makes it a valuable tool for understanding human personality. Researchers have developed various psychological measures to evaluate these traits and gain insights into individuals' unique qualities [20]. In 1991, John, Donahue, and Kentle created the 44-item Big Five Inventory (BFI-44) to measure the five primary traits [21]. Participants respond to each statement using a Likert scale, ranging from "strongly disagree" to "strongly agree," to indicate the extent to which the statement applies to them. BFI's reliability and validity make it a popular instrument. It has been shown to predict a range of outcomes, including academic performance, career success, and relationship satisfaction. The BFI's adaptability for cross-cultural studies is demonstrated by its translation and validation in many languages and cultures. In 2017, Soto and John introduced the updated 60-item Big Five Inventory-2 (BFI-2) [22].

Cultural and geographic influences have been shown to affect personality traits [23–26]. Consequently, while the Big Five Model is widely recognized and used [27–29], it may not universally apply across all societies [30, 31]. The current study was the first reliability analysis of

the BFI through reliability generalization meta-analysis (REGEMA). This method assesses the reliability of an instrument across different cultural contexts, samples, and languages. REGEMA has been recognized as an effective method for strengthening the psychometric properties of an instrument [32]. Another key feature of the current study is its focus beyond a single version of the BFI. It also examines two distinct variants of the BFI, evaluating their reliability across diverse populations. This approach will enhance the trust of researchers in utilizing the BFI across diverse cultures enhancing understanding of personality dimensions in various fields such as educational and human resources management.

Methods

Literature search/Inclusion criteria

This meta-analysis searched over 30 databases and sources, including 19 major indexing and abstracting services: BibCnrs, CNKI, CNPIEC, Digital Science, DOAJ, EBSCO, Scopus (Elsevier Databases), Gale, PubMed/MEDLINE, PMC (National Library of Medicine), OpenAIRE, ProQuest, PSYINDEX, SafetyLit, SSCI (Web of Science), Current Contents—Social & Behavioral Sciences, and 5 journal ranking lists: ANVU, 4 journal directories: EZB, JournalGuide, Hinari (Research4Life), OARE; 2 repository delivery services: DeepGreen, Jisc; 2 digital preservation services: CLOCKSS, Helveticat; and 3 content aggregators: Google Scholar, Scilit, WorldCat.

The search included publications from 1991 to mid-November 2024. We searched for "Big Five Personality Inventory", "BFI", and "reliability" as keywords to identify possible publications. We included those studies that gave a reliability estimate (*e.g.*, Cronbach's alpha, McDonald's omega) for at least one of the BFI versions.

Coding variables and procedures

The version of BFI used in each study, sample size, population type, age, gender composition, clinical status, and reliability type were coded. Two researchers independently coded all studies. We calculated inter-rater agreement statistics and resolved disagreements by consensus.

Included studies

A total of 57 studies (reported in 34 papers) were finally included in the current meta-analysis. These studies used the BFI and BFI-2 in diverse cultures and languages *i.e.* Arabic [33], Chinese [34, 35], Croatian [36], Czech [37], Danish [38, 39], Dutch [40], English [22, 41–45], French [44, 46, 47], German [44, 48], Indonesian [49], Italian [50, 51], Japanese [44, 52], Malay [53, 54], Norwegian [55, 56], Polish [44], Portuguese [57], Russian [58], Serbian [59, 60], Slovak [61], Spanish [44, 62, 63], Swahili [64], and

Turkish [65]. We did not include the short versions of the BFI in the current meta-analysis.

Meta-analysis

Effect size was measured using reliability coefficients. To estimate pooled correlation and account for study heterogeneity, a random-effects model was used. The I^2 statistic and Cochran's Q test analyzed heterogeneity. Heterogeneity impact was assessed using prediction intervals to estimate variability around the pooled estimate. The DerSimonian-Laird estimation estimated heterogeneity variance (tau-squared). Knapp-Hartung adjustments calibrated confidence intervals and significance tests. Outliers and influential examples were detected using studentized residuals and Cook's distance values. Publication bias was assessed using funnel plots and regression-based tests (Egger's regression, rank correlation). Publication bias modifications included trim-and-fill imputation to estimate an unbiased effect and p-uniform publication bias testing. A fail-safe N (the number of null studies needed to invalidate results) was developed to determine the number of missed null experiments that would negate results. Excess significance testing assessed if significant findings surpassed expectations. P-curve analysis estimated effect sizes adjusted for p-hacking to assess robustness. Unless otherwise stated, all analyses employed two-tailed $\alpha = 0.05$. The reporting requirements for a meta-analysis were strictly adhered [66].

Statistical analyses

All analyses were conducted using the R statistical software (version 4.2.1) and the metafor (version 3.0–2) and meta (version 5.1–1) packages.

The data from the 57 included studies were compiled into a single dataset, each row represented a sample with its reliability coefficients for the five BFI dimensions (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism abbreviated as OCEAN). Additional variables included in the dataset were the study reference, clinical state, language, sample size, mean age, percentage of female participants, and COSMIN quality rating [67].

Reliability generalization meta-analyses

Separate random-effects meta-analyses were conducted for each of the five BFI dimensions using the metaphor package. Cronbach's alpha reliability coefficients were transformed using the Fisher's z transformation prior to analysis. The pooled reliability, 95% confidence intervals, and 95% prediction intervals were computed for each dimension.

Heterogeneity was assessed using the I^2 statistic, which represents the percentage of total variability in the effect sizes due to between-study heterogeneity. The

Tau² statistic was used to estimate the amount of true heterogeneity in the reliability coefficients.

Publication bias was evaluated using the Rank Correlation Test (Begg and Mazumdar), Regression Test (Egger's method), and the Test of Excess Significance. The Fail-Safe N was also calculated to estimate the number of unpublished or missing studies that would be needed to render the meta-analytic results non-significant. Additionally, the p-uniform method was used to provide an alternative effect size estimate that is less influenced by publication bias.

Sensitivity analyses

Sensitivity analyses were conducted by removing one study at a time from the meta-analyses and re-running the analyses to assess the impact on the pooled reliability estimates. The changes in the pooled reliability coefficients were evaluated to ensure no single study had a disproportionate influence on the results.

Meta-regression analyses

To explore potential sources of heterogeneity, meta-regression analyses were performed. Separate models were fitted for each BFI dimension, with participant age and sex (percentage of females) as the predictor variables. A combined model including both age and sex was also tested. The significance of the regression coefficients was used to determine if these study-level characteristics could explain the observed variability in the reliability estimates.

All statistical tests were conducted at a significance level of $\alpha=0.05$. The R codes and outputs for the meta-analyses and supplementary analyses are available upon request. We used Traffic light and summary charts for visualizing [68]. These visualizations provide a clear and concise representation of the findings.

Results

The reliability generalization meta-analysis included 57 studies examining the internal consistency of the BFI and BFI-2 across 22 different languages and cultural contexts, totaling 43,715 participants (Table 1, Fig. 1). The 57 studies (datapoints) are published in 34 papers [22, 33–65]. The samples were diverse in age range, ranging from adolescents (13 years old) to older adults (51 years old), with a median age of 30.08 years across studies. There was also variation in gender representation, with the percentage of female participants ranging from 40 to 86% (mean 60.24%). All studies met rigorous "Very Good" quality standards based on COSMIN criteria. Figure 2 shows traffic light plot of the risk of bias, and Fig. 3 shows summary plot analysis of the risk of bias.

For the BFI, the median Cronbach's alpha reliability coefficients across studies were 0.77 for Openness, 0.80 for Conscientiousness, 0.80 for Extraversion, 0.73 for Agreeableness, and 0.80 for Neuroticism. Reliability tended to be lowest for the Agreeableness dimension. For the newer BFI-2, median reliability was higher across dimensions: 0.83 for Openness, 0.86 for Conscientiousness, 0.85 for Extraversion, 0.80 for Agreeableness, and 0.88 for Neuroticism.

The studies represented a wide range of sample sizes, from 30 participants in the smallest study [54] to 3882 in the largest study [49]. All the studies utilized healthy populations, *i.e.* without adding any clinical population. Detailed descriptive results are shown in Table 1.

The reliability generalization meta-analysis conducted on the BFI provided a comprehensive overview of the internal consistency reliability across its five personality dimensions. The meta-analytic results for each domain are summarized in Table 2, which presents detailed information on the estimated pooled reliability coefficients, 95% confidence and predicted intervals, heterogeneity statistics, and tests for publication bias and excess significance.

Dimension O: Openness to experience

The meta-analysis of 57 studies with a total sample size of 43,715 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.79, 0.82) for the openness to experience dimension of the combined BFI and BFI-2 (Table 2, Dimension O). The 95% predicted interval for the reliability coefficient across different studies ranges from 0.72 to 0.89, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2=92.59\%$, $H^2=13.49$) confirm the presence of significant variability across studies. The rank correlation test ($p<0.001$) and regression test ($p<0.001$) indicate potential publication bias, with smaller studies reporting higher reliability estimates. The fail-safe N suggests a very large number of missing studies would be needed to render the overall effect insignificant (Table 2, Dimension O).

For the BFI, the subgroup meta-analysis of 29 studies with a total sample size of 20,243 participants yielded an estimated internal consistency reliability coefficient of 0.77 (95% CI: 0.75, 0.80) for the openness to experience dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.65 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=94.51\%$, $H^2=18.23$) confirm the presence of significant variability across studies (Table 3, BFI).

For the BFI-2, the subgroup meta-analysis of 28 studies with a total sample size of 23,472 participants yielded an estimated internal consistency reliability coefficient of 0.83 (95% CI: 0.82, 0.84) for the openness to experience

Table 1 Descriptive statistics of the included studies in the reliability generalization meta-analysis of the internal consistency of the Big Five Inventory (BFI)

SN	Label	Ref	Scale	Language	Sample	O	C	E	A	N	Mean age	% Female	COSMIN
1	Rammstedt B., 2024, English, BFI-2	[44]	BFI-2	English	794	0.82	0.88	0.84	0.79	0.91	41.68	42	Very Good
2	Rammstedt B., 2024, Polish, BFI-2	[44]	BFI-2	Polish	882	0.79	0.87	0.86	0.77	0.87	41.28	47	Very Good
3	Rammstedt B., 2024 [French, BFI-2]	[44]	BFI-2	French	988	0.82	0.86	0.79	0.77	0.89	42.66	40	Very Good
4	Rammstedt B., 2024, Spanish, BFI-2	[44]	BFI-2	Spanish	964	0.85	0.87	0.81	0.77	0.85	42	48	Very Good
5	Rammstedt B., 2024, Japanese, BFI-2	[44]	BFI-2	Japanese	718	0.82	0.85	0.84	0.77	0.88	44.64	46	Very Good
6	Rammstedt B., 2024, German, BFI-2	[44]	BFI-2	German	1118	0.86	0.89	0.87	0.80	0.90	44.6	49	Very Good
7	Snezana S., 2024, Serbian, BFI-2	[60]	BFI-2	Serbian	1061	0.82	0.86	0.84	0.79	0.89	32.48	58.6	Very Good
8	Garrashi H. H., 2023, Swahili, BFI	[64]	BFI	Swahili	200	0.57	0.73	0.56	0.54	0.49	20.16	57	Very Good
9	Hausding M., 2023, Croatian, BFI-2	[36]	BFI-2	Croatian	320	0.78	0.89	0.79	0.75	0.83	39.86	68	Very Good
10	Junior P.R.S.R., 2023, Portuguese, BFI	[57]	BFI	Portuguese	490	0.74	0.76	0.80	0.63	0.82	44.5	75	Very Good
11	Lignier B., 2023, French, BFI-2	[47]	BFI-2	French	1003	0.82	0.85	0.79	0.76	0.92	19.5	53.84	Very Good
12	Tucaković L., 2023, Serbian, BFI	[59]	BFI	Serbian	263	0.83	0.84	0.82	0.70	0.87	24.02	83.7	Very Good
13	Tucaković L., 2023, Serbian, BFI	[59]	BFI	Serbian	1350	0.78	0.82	0.80	0.71	0.84	27.48	82.4	Very Good
14	Tucaković L., 2023, Serbian, BFI	[59]	BFI	Serbian	215	0.81	0.84	0.80	0.75	0.85	35.59	67.3	Very Good
15	Dionigi A., 2022, Italian, BFI	[50]	BFI	Italian	607	0.84	0.83	0.82	0.70	0.81	36.5	64.7	Very Good
16	Follesdal H., 2022, Norwegian, BFI-2	[56]	BFI-2	Norwegian	606	0.81	0.83	0.82	0.79	0.86	33.79	72.2	Very Good
17	Follesdal H., 2022, Norwegian, BFI-2	[56]	BFI-2	Norwegian	409	0.83	0.85	0.85	0.81	0.90	26.77	61.12	Very Good
18	Gallardo-Pujol D., 2022, Spanish, BFI-2	[63]	BFI-2	Spanish	1000	0.84	0.86	0.77	0.77	0.88	23.1	81.8	Very Good
19	Rammstedt B., 2022, English, BFI-2	[45]	BFI-2	English	2234	0.84	0.89	0.86	0.81	0.92	*	57.9	Very Good
20	Yoshino S., 2022, Japanese, BFI-2	[52]	BFI-2	Japanese	487	0.81	0.82	0.90	0.74	0.89	20.01	58.93	Very Good
21	Yoshino S., 2022, Japanese, BFI-2	[52]	BFI-2	Japanese	500	0.8	0.84	0.88	0.79	0.89	44.94	50	Very Good
22	Zhang B., 2022, Chinese, BFI-2	[35]	BFI-2	Chinese	1194	0.84	0.85	0.87	0.79	0.87	19.46	60.6	Very Good
23	Zhang B., 2022, Chinese, BFI-2	[35]	BFI-2	Chinese	470	0.85	0.85	0.87	0.83	0.89	21.68	66.5	Very Good
24	Zhang B., 2022, Chinese, BFI-2	[35]	BFI-2	Chinese	1000	0.83	0.88	0.88	0.82	0.90	29.25	50	Very Good
25	Zhang B., 2022, Chinese, BFI-2	[35]	BFI-2	Chinese	486	0.83	0.86	0.85	0.81	0.86	29.38	59.8	Very Good
26	Kalugin A.Yu., 2021, Russian, BFI-2	[58]	BFI-2	Russian	1787	0.82	0.85	0.87	0.84	0.89	26.31	68.1	Very Good
27	Vedel A., 2021, Danish, BFI-2	[39]	BFI-2	Danish	101	0.88	0.92	0.88	0.85	0.92	22.8	81.2	Very Good
28	Vedel A., 2021, Danish, BFI-2	[39]	BFI-2	Danish	287	0.82	0.85	0.83	0.81	0.89	25.75	79.6	Very Good
29	Denissen J.J.A., 2020, Dutch, BFI-2	[40]	BFI-2	Dutch	827	0.85	0.87	0.86	0.84	0.88	50.5	50	Very Good
30	Halama P., 2020, Slovak, BFI-2	[61]	BFI-2	Slovak	542	0.79	0.83	0.80	0.83	0.83	41.79	50.5	Very Good
31	Thorvaldur S. P., 2020, Danish, BFI	[38]	BFI	Danish	100	0.77	0.75	0.84	0.66	0.84	41.7	64	Very Good
32	Danner D., 2019, German, BFI-2	[48]	BFI-2	German	1224	0.85	0.86	0.85	0.80	0.89	43.34	50	Very Good
33	Haslina M., 2018, Malay, BFI	[53]	BFI	Malay	437	0.60	0.70	0.74	0.77	0.70	21.15	55.8	Very Good
34	Ørnfjord M., 2018, Norwegian, BFI	[55]	BFI	Norwegian	484	0.80	0.78	0.79	0.71	0.83	24.98	81.8	Very Good
35	Alansari B., 2016, Arabic, BFI	[33]	BFI	Arabic	685	0.79	0.9	0.82	0.82	0.83	21.19	55.47	Very Good
36	Carciofo R., 2016, Chinese, BFI	[34]	BFI	Chinese	214	0.74	0.73	0.74	0.70	0.74	24.64	43.4	Very Good
37	Carciofo R., 2016, Chinese, BFI	[34]	BFI	Chinese	235	0.79	0.73	0.78	0.74	0.72	25.5	76.1	Very Good
38	Carciofo R., 2016, Chinese, BFI	[34]	BFI	Chinese	172	0.77	0.81	0.77	0.73	0.75	27.75	60.4	Very Good
39	Carciofo R., 2016, Chinese, BFI	[34]	BFI	Chinese	186	0.81	0.78	0.72	0.71	0.78	31.36	61.2	Very Good
40	Hřebíčková M., 2016, Czech, BFI	[37]	BFI	Czech	494	0.72	0.70	0.65	0.67	0.76	13	43	Very Good
41	Hřebíčková M., 2016, Czech, BFI	[37]	BFI	Czech	1109	0.70	0.74	0.79	0.68	0.81	15	54.6	Very Good
42	Hřebíčková M., 2016, Czech, BFI	[37]	BFI	Czech	245	0.75	0.79	0.81	0.68	0.83	51	63.7	Very Good
43	Muhammad R.F.W., 2016, Indonesian, BFI	[49]	BFI	Indonesian	3882	0.68	0.78	0.84	0.69	0.81	17.83	42.4	Very Good
44	Soto C.J., 2016, English, BFI-2	[22]	BFI-2	English	1000	0.82	0.84	0.87	0.82	0.86	29.25	50	Very Good
45	Soto C.J., 2016, English, BFI-2	[22]	BFI-2	English	1000	0.84	0.88	0.88	0.83	0.90	28.73	50	Very Good
46	Soto C.J., 2016, English, BFI-2	[22]	BFI-2	English	470	0.85	0.86	0.88	0.85	0.90	21.68	66.59	Very Good
47	Soto C.J., 2016, English, BFI	[22]	BFI	English	1000	0.81	0.85	0.88	0.79	0.85	28.73	50	Very Good
48	Soto C.J., 2016, English, BFI	[22]	BFI	English	470	0.80	0.81	0.88	0.80	0.84	21.68	66.59	Very Good
49	Ong C. H., 2014, Malay, BFI	[54]	BFI	Malay	30	0.85	0.88	0.77	0.90	0.73	*	86	Very Good
50	Reyes Zamorano E., 2014, Spanish, BFI	[62]	BFI	Spanish	472	0.77	0.78	0.76	0.62	0.74	31.35	54.03	Very Good

Table 1 (continued)

SN	Label	Ref	Scale	Language	Sample	O	C	E	A	N	Mean age	% Female	COSMIN
51	Ubbiali A., 2013, Italian, BFI	[51]	BFI	Italian	1023	0.80	0.83	0.80	0.69	0.80	34.8	56.9	Very Good
52	Neslihan G.K., 2010, Turkish, BFI	[65]	BFI	Turkish	1153	0.86	0.84	0.77	0.81	0.75	*	*	Very Good
53	Plaisant O., 2010, French, BFI	[46]	BFI	French	2499	0.74	0.80	0.82	0.75	0.82	20.2	69	Very Good
54	John O.P., 2008, English, BFI	[41]	BFI	English	829	0.83	0.82	0.86	0.79	0.87	*	*	Very Good
55	Zheng L., 2008, English, BFI	[42]	BFI	English	633	0.86	0.89	0.90	0.83	0.93	23	66	Very Good
56	Zheng L., 2008, English, BFI	[42]	BFI	English	437	0.85	0.87	0.89	0.81	0.91	23.5	49.6	Very Good
57	Worrell F.C., 2004, English, BFI	[43]	BFI	English	329	0.72	0.79	0.83	0.70	0.77	20.56	66	Very Good

O openness to experience, C conscientiousness, E extraversion, A agreeableness, N neuroticism

* Not Reported

dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.80 to 0.86, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2=66.23\%$, $H^2=2.96$) indicate lower variability across studies compared to the BFI (Table 3, BFI-2).

The comparison of the reliability coefficients between the BFI and BFI-2 for the openness to experience dimension showed a statistically significant difference ($p<0.01$), with the BFI-2 demonstrating higher internal consistency reliability (Table 3, Comparison p-value). The meta-analysis of 27 studies with a total sample size of 17,589 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.78, 0.82) for the openness to experience dimension of the combined BFI and BFI-2 (Fig. 4). The predicted interval for the reliability coefficient across different studies ranges from 0.70 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=91.72\%$, $H^2=12.08$) confirm the presence of significant variability across studies. The rank correlation test ($p=0.02$) and regression test ($p=0.02$) indicate potential publication bias (Fig. 5), with smaller studies reporting higher reliability estimates. The fail-safe N suggests a substantial number of missing studies would be needed to render the overall effect insignificant (Table 2).

For the BFI, the subgroup meta-analysis of 18 studies with a total sample size of 10,434 participants yielded an estimated internal consistency reliability coefficient of 0.78 (95% CI: 0.76, 0.81) for the openness to experience dimension of the BFI. The predicted interval for the reliability coefficient across different studies ranges from 0.69 to 0.87, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=88.84\%$, $H^2=8.96$) confirm the presence of significant variability across studies (Table 3).

For the BFI-2, the subgroup meta-analysis of 9 studies with a total sample size of 7,155 participants yielded an estimated internal consistency reliability coefficient of 0.83 (95% CI: 0.82, 0.84) for the openness to experience dimension of the BFI-2. The predicted interval for the reliability coefficient across different studies ranges

from 0.81 to 0.86, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2=41.97\%$, $H^2=1.72$) indicate lower variability across studies compared to the BFI (Table 3).

The comparison of the reliability coefficients between the BFI and BFI-2 for the openness to experience dimension showed a statistically significant difference ($p<0.01$), with the BFI-2 demonstrating higher internal consistency reliability.

Dimension C: Conscientiousness

The meta-analysis of 57 studies with a total sample size of 43,715 participants yielded an estimated internal consistency reliability coefficient of 0.83 (95% CI: 0.82, 0.84) for the conscientiousness dimension of the combined BFI and BFI-2 (Table 2, Dimension C). The 95% predicted interval for the reliability coefficient across different studies ranges from 0.76 to 0.91, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=92.89\%$, $H^2=14.06$) confirm the presence of significant variability across studies. The rank correlation test ($p<0.001$) and regression test ($p<0.001$) indicate potential publication bias, with smaller studies reporting higher reliability estimates. The fail-safe N suggests a very large number of missing studies would be needed to render the overall effect insignificant (Table 2, Dimension C).

For the BFI, the subgroup meta-analysis of 29 studies with a total sample size of 20,243 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.78, 0.82) for the conscientiousness dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.71 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=93.31\%$, $H^2=14.96$) confirm the presence of significant variability across studies (Table 3, BFI).

For the BFI-2, the subgroup meta-analysis of 28 studies with a total sample size of 23,472 participants yielded an estimated internal consistency reliability coefficient of 0.86 (95% CI: 0.85, 0.87) for the conscientiousness

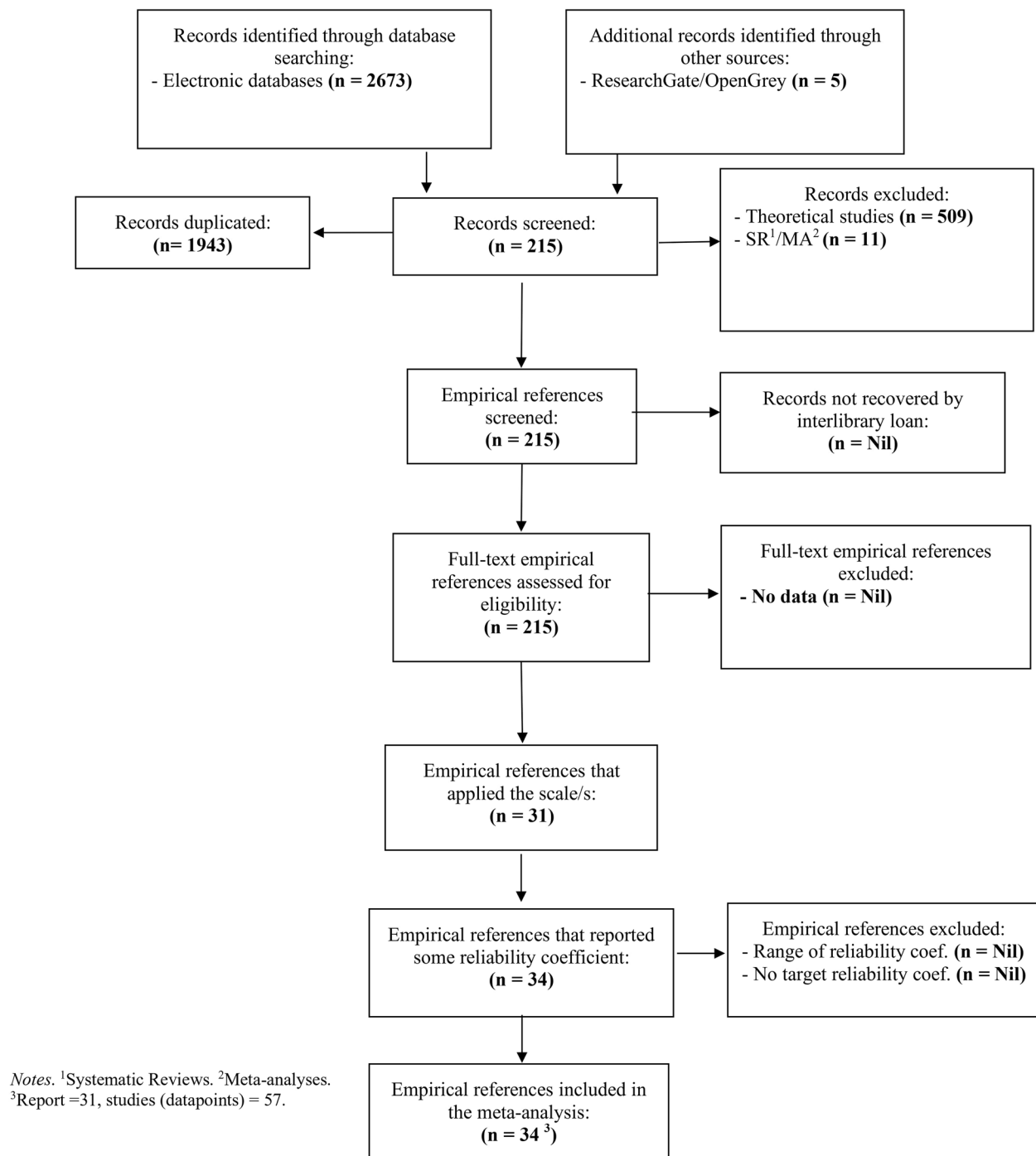


Fig. 1 REGEMA flow diagram for study selection

dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.83 to 0.90, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2=79.40\%$, $H^2=4.85$) indicate lower variability across studies compared to the BFI (Table 3, BFI-2).

The comparison of the reliability coefficients between the BFI and BFI-2 for the conscientiousness dimension showed a statistically significant difference ($p < 0.01$), with the BFI-2 demonstrating higher internal consistency reliability (Table 3, Comparison p-value). The meta-analysis of 27 studies with a total sample size of 17,589 participants yielded an estimated internal

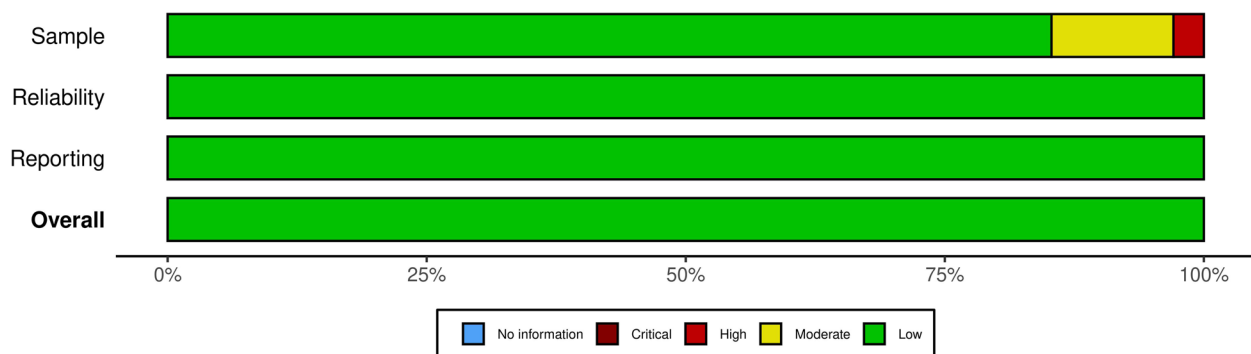


Fig. 2 Summary plot of the assessment of the risk of bias

consistency reliability coefficient of 0.82 (95% CI: 0.80, 0.84) for the conscientiousness dimension of the combined BFI and BFI-2 (Fig. 6). The predicted interval for the reliability coefficient across different studies ranges from 0.70 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 94.73\%$, $H^2 = 18.99$) confirm the presence of significant variability across studies. The rank correlation test ($p < 0.01$) and regression test ($p < 0.01$) indicate potential publication bias (Fig. 7), with smaller studies reporting higher reliability estimates. The fail-safe N suggests a substantial number of missing studies would be needed to render the overall effect insignificant (Table 2).

For the BFI, the subgroup meta-analysis of 18 studies with a total sample size of 10,434 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.77, 0.82) for the conscientiousness dimension of the BFI. The predicted interval for the reliability coefficient across different studies ranges from 0.70 to 0.89, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 91.29\%$, $H^2 = 11.48$) confirm the presence of significant variability across studies (Table 3).

For the BFI-2, the subgroup meta-analysis of 9 studies with a total sample size of 7,155 participants yielded an estimated internal consistency reliability coefficient of 0.86 (95% CI: 0.85, 0.87) for the conscientiousness dimension of the BFI-2. The predicted interval for the reliability coefficient across different studies ranges from 0.83 to 0.90, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2 = 77.25\%$, $H^2 = 4.39$) indicate lower variability across studies compared to the BFI (Table 3).

The comparison of the reliability coefficients between the BFI and BFI-2 for the conscientiousness dimension showed a statistically significant difference ($p < 0.01$), with the BFI-2 demonstrating higher internal consistency reliability.

Dimension E: Extraversion

The meta-analysis of 57 studies with a total sample size of 43,715 participants yielded an estimated internal consistency reliability coefficient of 0.83 (95% CI: 0.82, 0.84) for the extraversion dimension of the combined BFI and BFI-2 (Table 2, Dimension E). The 95% predicted interval for the reliability coefficient across different studies ranges from 0.76 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 92.2\%$, $H^2 = 12.82$) confirm the presence of significant variability across studies. The rank correlation test ($p < 0.001$) and regression test ($p < 0.001$) indicate potential publication bias, with smaller studies reporting higher reliability estimates. The fail-safe N suggests a very large number of missing studies would be needed to render the overall effect insignificant (Table 2, Dimension E).

For the BFI, the subgroup meta-analysis of 29 studies with a total sample size of 20,243 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.79, 0.82) for the extraversion dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.69 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 93.43\%$, $H^2 = 15.22$) confirm the presence of significant variability across studies (Table 3, BFI).

For the BFI-2, the subgroup meta-analysis of 28 studies with a total sample size of 23,472 participants yielded an estimated internal consistency reliability coefficient of 0.85 (95% CI: 0.84, 0.86) for the extraversion dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.80 to 0.90, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2 = 88.48\%$, $H^2 = 8.68$) indicate lower variability across studies compared to the BFI (Table 3, BFI-2).

The comparison of the reliability coefficients between the BFI and BFI-2 for the extraversion dimension showed a statistically significant difference ($p < 0.01$), with the

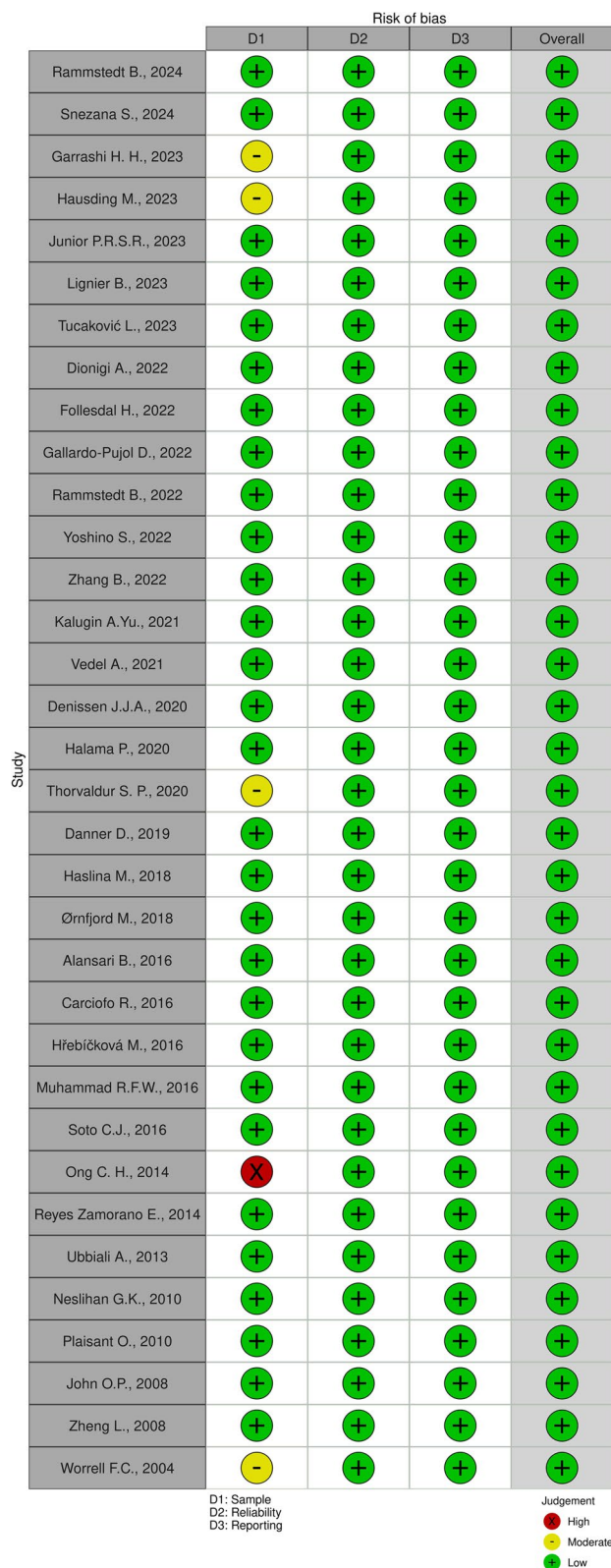


Fig. 3 Traffic light plot of the assessment of the risk of bias

BFI-2 demonstrating higher internal consistency reliability (Table 3, Comparison p-value).

The BFI-2 showed significantly higher internal consistency reliability than the original BFI across all five dimensions (openness, conscientiousness, extraversion, agreeableness, and neuroticism), with lower heterogeneity across studies for each dimension. This suggests an improvement in the psychometric properties of the revised BFI-2 scale. The meta-analysis of 27 studies with a total sample size of 17,589 participants yielded an estimated internal consistency reliability coefficient of 0.82 (95% CI: 0.80, 0.84) for the extraversion dimension of the combined BFI and BFI-2 (Fig. 8). The predicted interval for the reliability coefficient across different studies ranges from 0.69 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 95.15\%$, $H^2 = 20.62$) confirm the presence of significant variability across studies. The rank correlation test ($p = 0.02$) and regression test ($p < 0.01$) indicate potential publication bias (Fig. 9), with smaller studies reporting higher reliability estimates. The fail-safe N suggests a substantial number of missing studies would be needed to render the overall effect insignificant (Table 2).

For the BFI, the subgroup meta-analysis of 18 studies with a total sample size of 10,434 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.77, 0.83) for the extraversion dimension of the BFI. The predicted interval for the reliability coefficient across different studies ranges from 0.69 to 0.90, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 93.43\%$, $H^2 = 15.22$) confirm the presence of significant variability across studies (Table 3).

For the BFI-2, the subgroup meta-analysis of 9 studies with a total sample size of 7,155 participants yielded an estimated internal consistency reliability coefficient of 0.86 (95% CI: 0.84, 0.87) for the extraversion dimension of the BFI-2. The predicted interval for the reliability coefficient across different studies ranges from 0.80 to 0.91, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2 = 88.48\%$, $H^2 = 8.68$) indicate lower variability across studies compared to the BFI (Table 3).

The comparison of the reliability coefficients between the BFI and BFI-2 for the extraversion dimension showed a statistically significant difference ($p < 0.01$), with the BFI-2 demonstrating higher internal consistency reliability.

The BFI-2 showed significantly higher internal consistency reliability compared to the original BFI across all five dimensions (openness, conscientiousness, extraversion, agreeableness, and neuroticism), with lower heterogeneity across studies for each dimension. This suggests an improvement in the psychometric properties of the revised BFI-2 scale.

Table 2 Reliability generalization meta-analysis coefficients of the internal consistency of the BFI and BFI-2 combined

Parameter/Dimension	O	C	E	A	N
k = number of studies 57, n = number of participants = 43,715					
Estimate (95% CI)	0.80 (0.79, 0.82)	0.83 (0.82, 0.84)	0.83 (0.82, 0.84)	0.77 (0.76, 0.78)	0.85 (0.84, 0.86)
95% Predicted Interval	0.72, 0.89	0.76, 0.91	0.76, 0.90	0.67, 0.86	0.77, 0.93
Tau	0.04	0.04	0.04	0.05	0.04
Tau ²	0.0019	0.0013	0.0013	0.0023	0.0017
I ²	92.59%	0.9289	0.922	0.9204	0.9577
H ²	13.49	14.06	12.82	12.57	23.63
Q (df = 50)	755.28	787.41	718.13	703.88	1323.16
Rank Correlation Test	-0.42 ($p < 0.001$)	-0.41 ($p < 0.001$)	-0.39 ($p < 0.001$)	-0.37 ($p < 0.001$)	-0.43 ($p < 0.001$)
Regression Test	-4.58 ($p < 0.001$)	-6.42 ($p < 0.001$)	-7.73 ($p < 0.001$)	-5.42 ($p < 0.001$)	-9.60 ($p < 0.001$)
Fail-Safe N	4,516,240	6,925,873	6,433,549	3,137,888	9,712,588
Begg and Mazumdar Rank Correlation	-0.42 ($p < 0.001$)	-0.41 ($p < 0.001$)	-0.39 ($p < 0.001$)	-0.37 ($p < 0.001$)	-0.43 ($p < 0.001$)
Egger's Regression	-4.58 ($p < 0.001$)	-6.42 ($p < 0.001$)	-7.73 ($p < 0.001$)	-5.42 ($p < 0.001$)	-9.60 ($p < 0.001$)
Test of Excess Significance	1.00 ($p = 1.00$)	1.00 ($p = 1.00$)	1.00 ($p = 1.00$)	1.00 ($p = 1.00$)	1.00 ($p = 1.00$)
p -uniform Effect Size Estimate	1.14 (1.12, 1.16)	1.23 (1.20, 1.26)	1.19 (1.15, 1.23)	1.04 (1.01, 1.07)	1.29 (1.24, 1.33)
Forest plot	Figure 4	Figure 6	Figure 8	Figure 10	Figure 12
Funnel Plot	Figure 5	Figure 7	Figure 9	Figure 11	Figure 13

O openness to experience, C conscientiousness, E extraversion, A agreeableness, N neuroticism

Dimension A: Agreeableness

The meta-analysis of 57 studies with a total sample size of 43,715 participants yielded an estimated internal consistency reliability coefficient of 0.77 (95% CI: 0.76, 0.78) for the agreeableness dimension of the combined BFI and BFI-2 (Table 2, Dimension A). The 95% predicted interval for the reliability coefficient across different studies ranges from 0.67 to 0.86, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 92.04\%$, $H^2 = 12.57$) confirm the presence of significant variability across studies. The rank correlation test ($p < 0.001$) and regression test ($p < 0.001$) indicate potential publication bias, with smaller studies reporting higher reliability estimates. The fail-safe N suggests a very large number of missing studies would be needed to render the overall effect insignificant (Table 2, Dimension A).

For the BFI, the subgroup meta-analysis of 29 studies with a total sample size of 20,243 participants yielded an estimated internal consistency reliability coefficient of 0.73 (95% CI: 0.71, 0.75) for the agreeableness dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.61 to 0.82, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 86.43\%$, $H^2 = 7.37$) confirm the presence of significant variability across studies (Table 3, BFI).

For the BFI-2, the subgroup meta-analysis of 28 studies with a total sample size of 23,472 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.79, 0.81) for the agreeableness dimension. The 95% predicted interval for the reliability coefficient

across different studies ranges from 0.75 to 0.85, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2 = 80.71\%$, $H^2 = 5.18$) indicate lower variability across studies compared to the BFI (Table 3, BFI-2).

The comparison of the reliability coefficients between the BFI and BFI-2 for the agreeableness dimension showed a statistically significant difference ($p < 0.01$), with the BFI-2 demonstrating higher internal consistency reliability (Table 3, Comparison p -value). The meta-analysis of 27 studies with a total sample size of 17,589 participants yielded an estimated internal consistency reliability coefficient of 0.75 (95% CI: 0.73, 0.78) for the agreeableness dimension of the combined BFI and BFI-2 (Fig. 10). The predicted interval for the reliability coefficient across different studies ranges from 0.58 to 0.86, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2 = 94.38\%$, $H^2 = 17.79$) confirm the presence of significant variability across studies. The rank correlation test ($p = 0.06$) and regression test ($p < 0.01$) indicate potential publication bias (Fig. 11), with smaller studies reporting higher reliability estimates. The fail-safe N suggests a substantial number of missing studies would be needed to render the overall effect insignificant (Table 2).

For the BFI, the subgroup meta-analysis of 18 studies with a total sample size of 10,434 participants yielded an estimated internal consistency reliability coefficient of 0.71 (95% CI: 0.69, 0.74) for the agreeableness dimension of the BFI. The predicted interval for the reliability coefficient across different studies ranges from 0.61 to 0.82, suggesting substantial heterogeneity.

Table 3 Reliability generalization subgroup meta-analysis coefficients of the internal consistency of the BFI versus BFI-2

Parameter/ Dimension	O	C	E	A	N
BFI (k = number of studies 29, n = number of participants = 20,243)					
Estimate (95% CI)	0.77 (0.75; 0.80)	0.80 (0.78; 0.82)	0.80 (0.79; 0.82)	0.73 (0.71; 0.75)	0.80 (0.79; 0.82)
95% Predicted Interval	0.65 to 0.90	0.71 to 0.90	0.71 to 0.89	0.6150 to 0.8482	0.71 to 0.89
Tau	0.06	0.05	0.04	0.06	0.04
Tau ²	0.0038 (SE = 0.0014)	0.0022 (SE = 0.0009)	0.002 (SE = 0.0009)	0.0034 (SE = 0.0013)	0.002 (SE = 0.0009)
I ²	0.9451	0.9331	0.9316	0.9212	0.9316
H ²	18.23	14.96	14.61	12.69	14.61
Forest plot	Figure 4	Figure 6	Figure 8	Figure 10	Figure 12
Meta-Regression BFI (Sex)	$\beta = 0.0002$; ($p = 0.4771$)	$\beta = 0.0015$; ($p = 0.2120$)	$\beta = 0.0008$; ($p = 0.5598$)	$\beta = 0.0005$; ($p = 0.4862$)	$\beta = 0.0009$; ($p = 0.4735$)
Meta-Regression BFI (Age)	$\beta = -0.0002$; ($p = 0.5849$)	$\beta = -0.0010$; ($p = 0.2952$)	$\beta = -0.0005$; ($p = 0.4123$)	$\beta = -0.0003$; ($p = 0.6971$)	$\beta = -0.0007$; ($p = 0.3564$)
Sub-group BFI (Language) ^a	Serbian: k = 3, n = 1828, $\alpha = 0.80$ (0.77, 0.83); Chinese: k = 4, n = 807, $\alpha = 0.78$ (0.75, 0.81); Czech: k = 3, n = 1848, $\alpha = 0.71$ (0.69, 0.74); English: k = 6, n = 3698, $\alpha = 0.82$ (0.78, 0.85)	Serbian: k = 3, n = 1828, $\alpha = 0.83$ (0.81, 0.84); Chinese: k = 4, n = 807, $\alpha = 0.76$ (0.73, 0.80); Czech: k = 3, n = 1848, $\alpha = 0.74$ (0.71, 0.78); English: k = 6, n = 3698, $\alpha = 0.80$ (0.79, 0.81)	Serbian: k = 3, n = 1828, $\alpha = 0.80$ (0.79, 0.82); Chinese: k = 4, n = 807, $\alpha = 0.76$ (0.73, 0.79); Czech: k = 3, n = 1848, $\alpha = 0.75$ (0.67, 0.83); English: k = 6, n = 3698, $\alpha = 0.88$ (0.86, 0.89)	Serbian: k = 3, n = 1828, $\alpha = 0.71$ (0.69, 0.74); Chinese: k = 4, n = 807, $\alpha = 0.72$ (0.69, 0.75); Czech: k = 3, n = 1848, $\alpha = 0.68$ (0.65, 0.70); English: k = 6, n = 3698, $\alpha = 0.79$ (0.77, 0.82)	Serbian: k = 3, n = 1828, $\alpha = 0.85$ (0.83, 0.86); Chinese: k = 4, n = 807, $\alpha = 0.75$ (0.72, 0.78); Czech: k = 3, n = 1848, $\alpha = 0.80$ (0.77, 0.83); English: k = 6, n = 3698, $\alpha = 0.86$ (0.83, 0.90)
BFI-2 (k = number of studies 28, n = number of participants = 23,472)					
Estimate (95% CI)	0.83 (0.82; 0.84)	0.86 (0.85; 0.87)	0.85 (0.84; 0.86)	0.80 (0.79; 0.81)	0.89 (0.88; 0.89)
95% Predicted Interval	0.80 to 0.86	0.83 to 0.90	0.80 to 0.90	0.75 to 0.85	0.85 to 0.93
Tau	0.01	0.02	0.03	0.02	0.02
Tau ²	0.0002	0.0003	0.0007	0.0006 (SE = 0.0002)	0.0004
I ²	66.23%	79.40%	88.91%	80.19%	88.50%
H ²	2.96	4.85	9.02	5.05	8.70
Forest plot	Figure 4	Figure 6	Figure 8	Figure 10	Figure 12
Meta-Regression BFI-2 (Sex)	$\beta = 0.0002$; ($p = 0.4771$)	$\beta = 0.0015$; ($p = 0.2120$)	$\beta = 0.0008$; ($p = 0.5598$)	$\beta = 0.0005$; ($p = 0.4862$)	$\beta = 0.0009$; ($p = 0.4735$)
Meta-Regression BFI-2 (Age)	$\beta = -0.0002$; ($p = 0.5849$)	$\beta = -0.0010$; ($p = 0.2952$)	$\beta = -0.0005$; ($p = 0.4123$)	$\beta = -0.0003$; ($p = 0.6971$)	$\beta = -0.0007$; ($p = 0.3564$)
Sub-group BFI-2 (Language) ^a	Japanese: k = 3, n = 1705, $\alpha = 0.81$ (0.80, 0.83); English: k = 5, n = 5498, $\alpha = 0.84$ (0.83, 0.84); Chinese: k = 4, n = 3150, $\alpha = 0.84$ (0.83, 0.85)	Japanese: k = 3, n = 1705, $\alpha = 0.84$ (0.83, 0.85); English: k = 5, n = 5498, $\alpha = 0.87$ (0.86, 0.89); Chinese: k = 4, n = 3150, $\alpha = 0.86$ (0.85, 0.88)	Japanese: k = 3, n = 1705, $\alpha = 0.84$ (0.83, 0.85); English: k = 5, n = 5498, $\alpha = 0.87$ (0.86, 0.89); Chinese: k = 4, n = 3150, $\alpha = 0.86$ (0.85, 0.88)	Japanese: k = 3, n = 1705, $\alpha = 0.77$ (0.75, 0.79); English: k = 5, n = 5498, $\alpha = 0.82$ (0.80, 0.84); Chinese: k = 4, n = 3150, $\alpha = 0.81$ (0.80, 0.83)	Japanese: k = 3, n = 1705, $\alpha = 0.89$ (0.88, 0.90); English: k = 5, n = 5498, $\alpha = 0.90$ (0.88, 0.92); Chinese: k = 4, n = 3150, $\alpha = 0.88$ (0.87, 0.90)
Comparison p-value BFI vs BFI-2	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$

O openness to experience, C conscientiousness, E extraversion, A agreeableness, N neuroticism

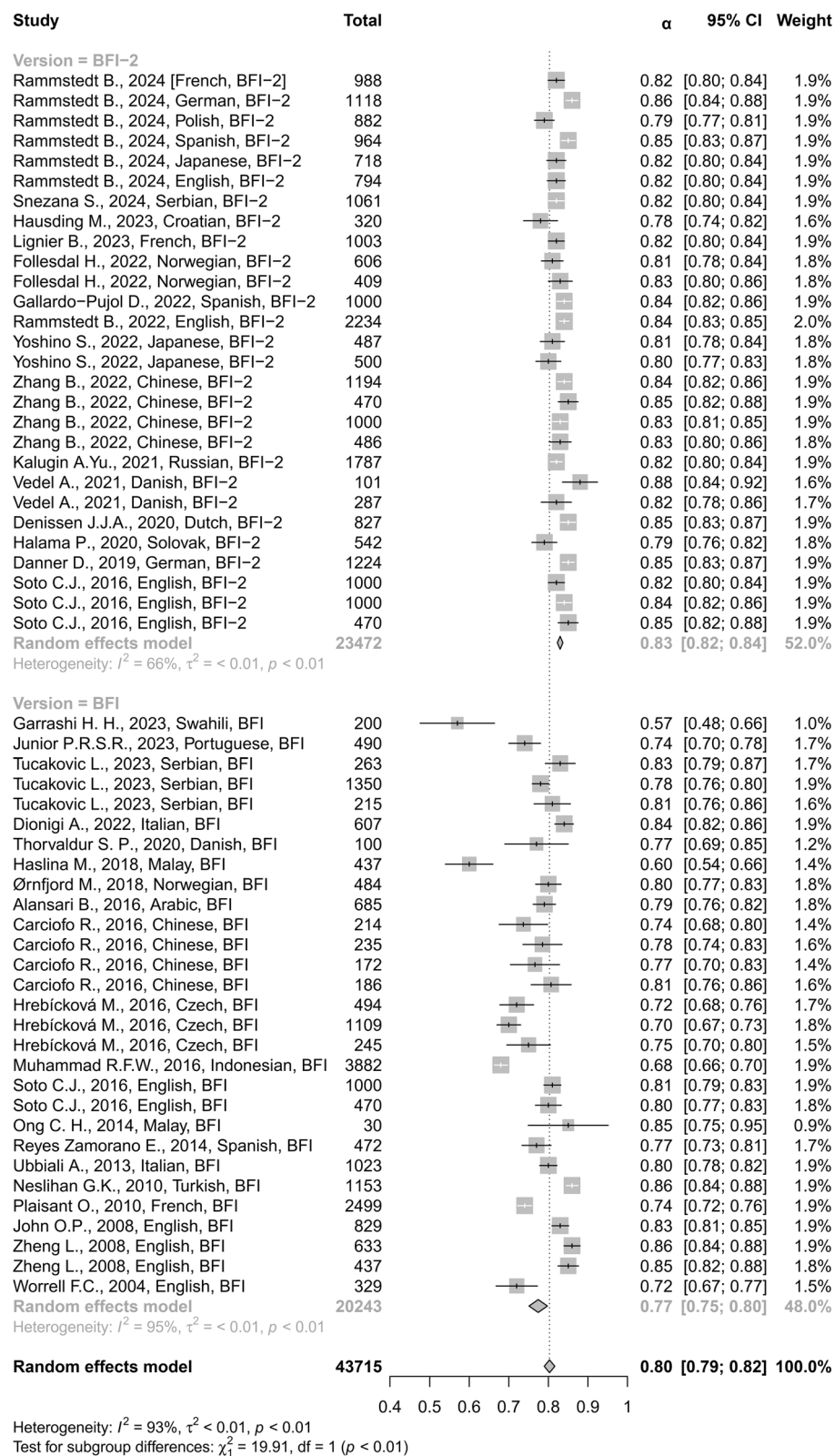
^a To provide examples of heterogeneity, only those languages are included here in which 3 or more studies were available; values in parenthesis reflect minimum and maximum values with 95% Confidence Interval

The heterogeneity statistics ($I^2 = 86.43\%$, $H^2 = 7.37$) confirm the presence of significant variability across studies (Table 3).

For the BFI-2, the subgroup meta-analysis of 9 studies with a total sample size of 7,155 participants yielded an estimated internal consistency reliability coefficient of 0.82 (95% CI: 0.80, 0.84) for the agreeableness dimension of the BFI-2. The predicted interval for the reliability

coefficient across different studies ranges from 0.76 to 0.87, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2 = 80.71\%$, $H^2 = 5.18$) indicate lower variability across studies compared to the BFI (Table 3).

The comparison of the reliability coefficients between the BFI and BFI-2 for the agreeableness dimension showed a statistically significant difference ($p < 0.01$),

**Fig. 4** Meta-analysis of the internal consistency of openness

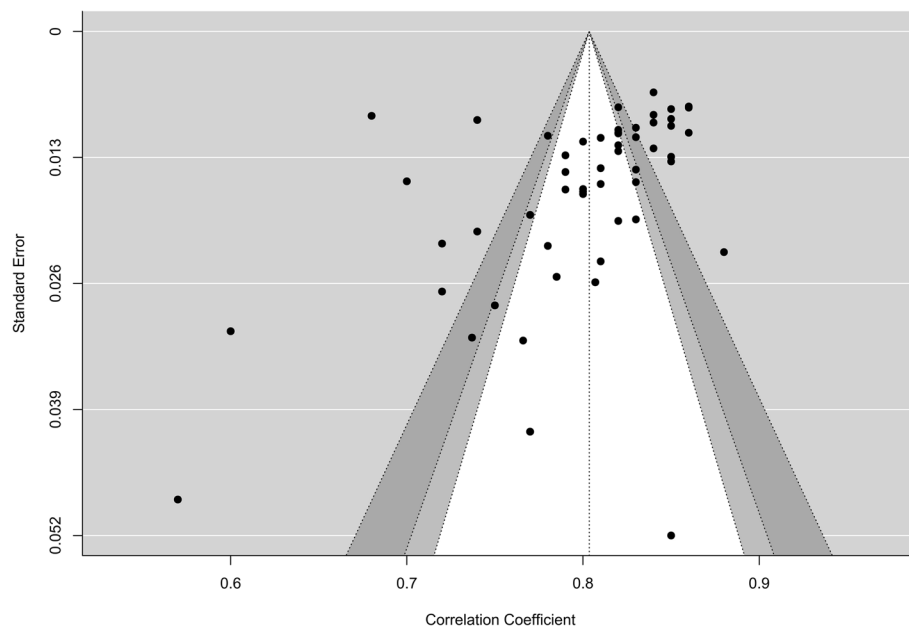


Fig. 5 Funnel plot of the internal consistency of openness

with the BFI-2 demonstrating higher internal consistency reliability.

Dimension N: Neuroticism

The meta-analysis of 57 studies with a total sample size of 43,715 participants yielded an estimated internal consistency reliability coefficient of 0.85 (95% CI: 0.84, 0.86) for the neuroticism dimension of the combined BFI and BFI-2 (Table 2, Dimension N). The 95% predicted interval for the reliability coefficient across different studies ranges from 0.77 to 0.93, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=95.77\%$, $H^2=23.63$) confirm the presence of significant variability across studies. The rank correlation test ($p<0.001$) and regression test ($p<0.001$) indicate potential publication bias, with smaller studies reporting higher reliability estimates. The fail-safe N suggests a very large number of missing studies would be needed to render the overall effect insignificant (Table 2, Dimension N).

For the BFI, the subgroup meta-analysis of 29 studies with a total sample size of 20,243 participants yielded an estimated internal consistency reliability coefficient of 0.80 (95% CI: 0.79, 0.82) for the neuroticism dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.70 to 0.93, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=95.32\%$, $H^2=21.38$) confirm the presence of significant variability across studies (Table 3, BFI).

For the BFI-2, the subgroup meta-analysis of 28 studies with a total sample size of 23,472 participants yielded

an estimated internal consistency reliability coefficient of 0.89 (95% CI: 0.88, 0.89) for the neuroticism dimension. The 95% predicted interval for the reliability coefficient across different studies ranges from 0.85 to 0.93, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2=84.85\%$, $H^2=6.60$) indicate lower variability across studies compared to the BFI (Table 3, BFI-2).

The comparison of the reliability coefficients between the BFI and BFI-2 for the neuroticism dimension showed a statistically significant difference ($p<0.01$), with the BFI-2 demonstrating higher internal consistency reliability (Table 3, Comparison p-value). The meta-analysis of 27 studies with a total sample size of 17,589 participants yielded an estimated internal consistency reliability coefficient of 0.85 (95% CI: 0.83, 0.87) for the neuroticism dimension of the combined BFI and BFI-2 (Fig. 12). The predicted interval for the reliability coefficient across different studies ranges from 0.96 to 0.93, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=97.30\%$, $H^2=37.05$) confirm the presence of significant variability across studies. The rank correlation test ($p=0.01$) and regression test ($p<0.01$) indicate potential publication bias (Fig. 13), with smaller studies reporting higher reliability estimates. The fail-safe N suggests a substantial number of missing studies would be needed to render the overall effect insignificant (Table 2).

For the BFI, the subgroup meta-analysis of 18 studies with a total sample size of 10,434 participants yielded an estimated internal consistency reliability coefficient of

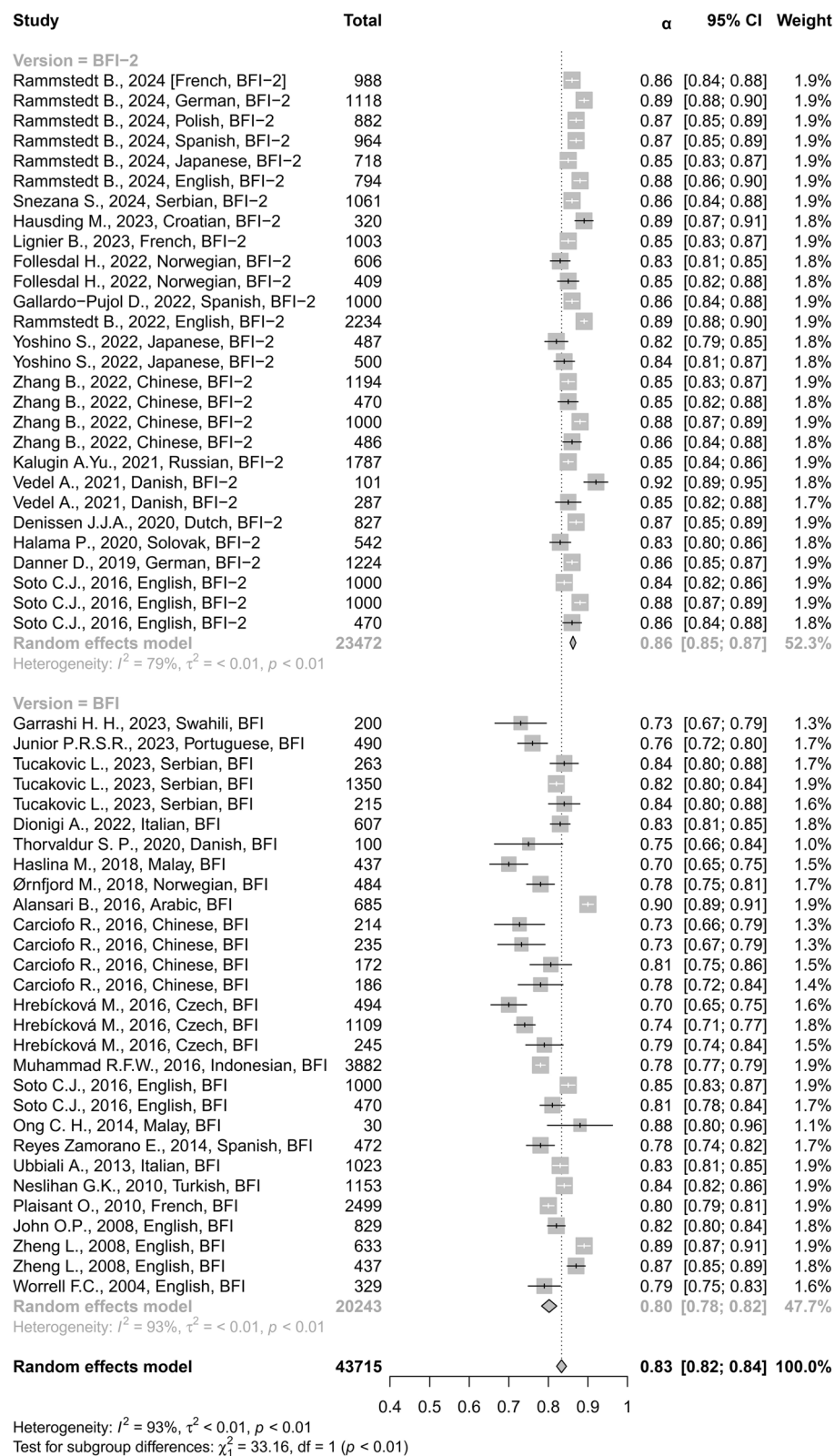


Fig. 6 Meta-analysis of the internal consistency of conscientiousness

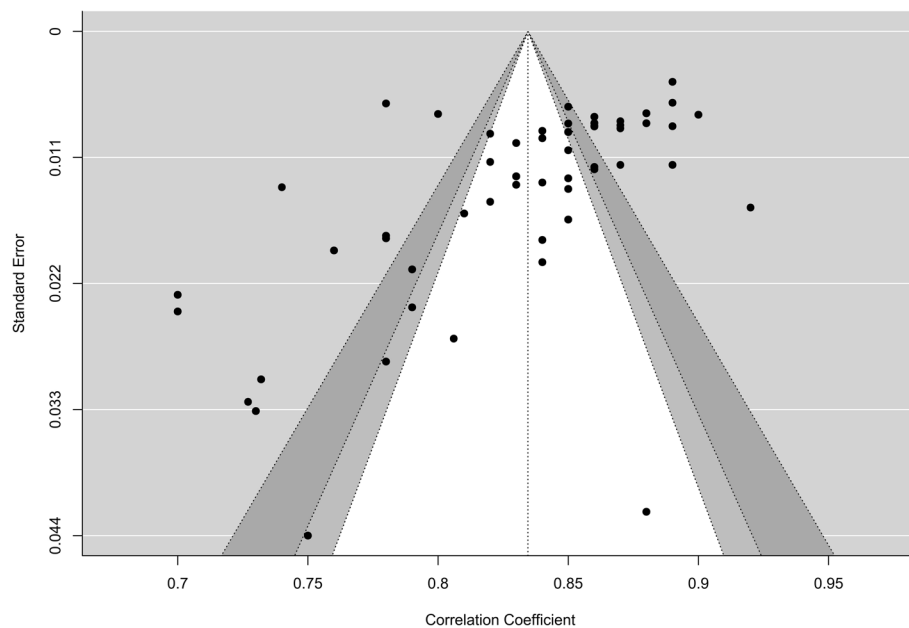


Fig. 7 Funnel plot of the internal consistency of conscientiousness

0.82 (95% CI: 0.79, 0.85) for the neuroticism dimension of the BFI. The predicted interval for the reliability coefficient across different studies ranges from 0.70 to 0.93, suggesting substantial heterogeneity. The heterogeneity statistics ($I^2=95.32\%$, $H^2=21.38$) confirm the presence of significant variability across studies (Table 3).

For the BFI-2, the subgroup meta-analysis of 9 studies with a total sample size of 7,155 participants yielded an estimated internal consistency reliability coefficient of 0.89 (95% CI: 0.88, 0.90) for the neuroticism dimension of the BFI-2. The predicted interval for the reliability coefficient across different studies ranges from 0.85 to 0.93, suggesting moderate heterogeneity. The heterogeneity statistics ($I^2=84.85\%$, $H^2=6.60$) indicate lower variability across studies compared to the BFI (Table 3).

The comparison of the reliability coefficients between the BFI and BFI-2 for the neuroticism dimension showed a statistically significant difference ($p<0.01$), with the BFI-2 demonstrating higher internal consistency reliability.

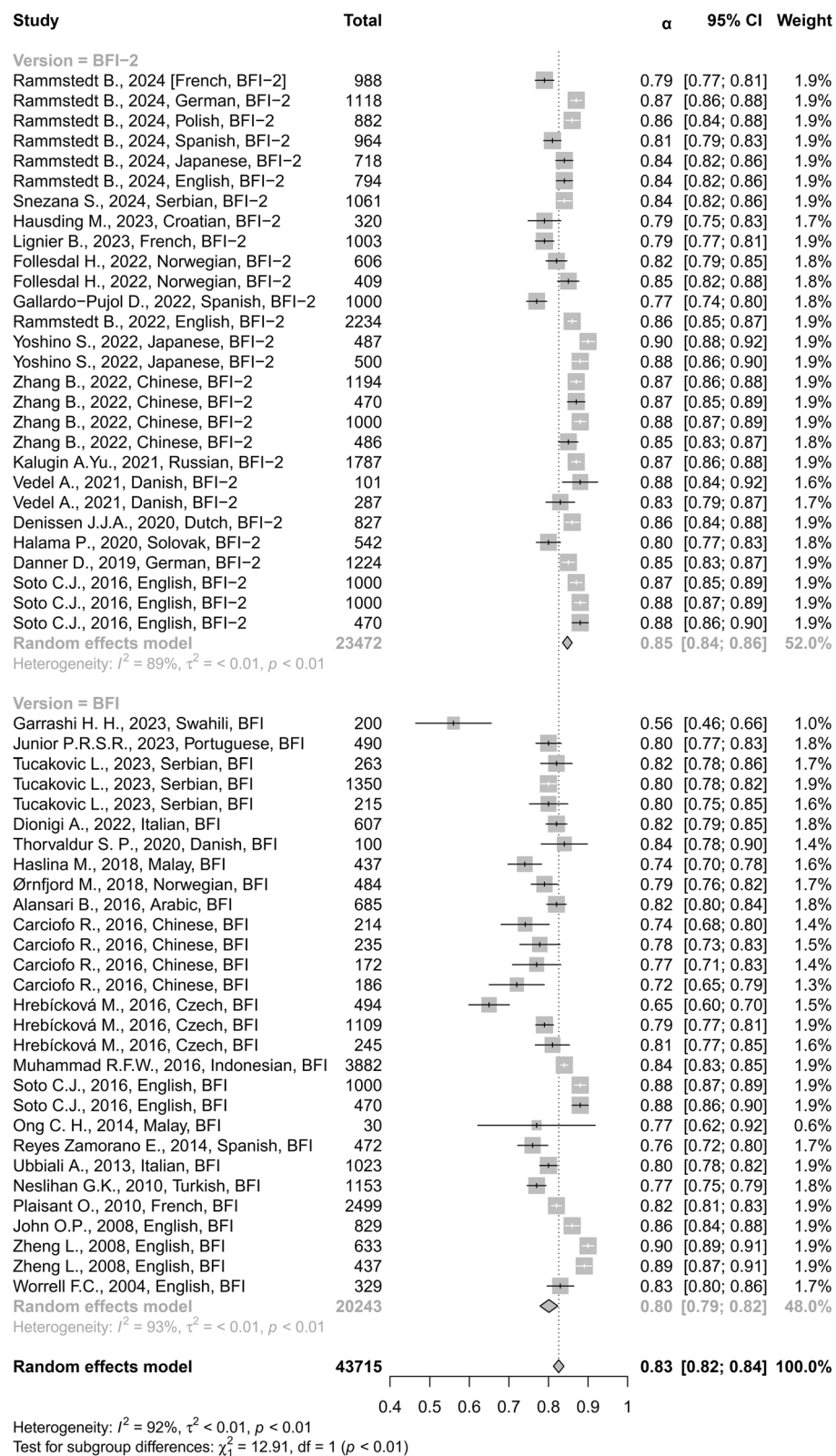
In summary, across all five dimensions, the BFI-2 demonstrated significantly higher internal consistency reliability estimates compared to the BFI, with lower heterogeneity across studies. This suggests that the revised BFI-2 scale may have improved psychometric properties in terms of internal consistency reliability.

Meta-regression analyses

To further explore potential sources of the observed heterogeneity in the reliability estimates, meta-regression

analyses were conducted using participant age and sex as predictor variables. For each of the BFI dimensions (*i.e.*, OCEAN), separate meta-regressions were performed with age, sex, and a combined age and sex model as the independent variables. The results showed that neither age nor sex, nor the combined age-and-sex model were significant predictors of the reliability coefficients for any of the five personality factors. We analyzed each of the BFI and BFI-2's five dimensions (*i.e.*, OCEAN) separately for age and sex. This indicates that the variability in the internal consistency of the BFI/BFI-2 (and combined) cannot be readily explained by the demographic characteristics of the included samples.

For Openness (BFI O), the regression analysis showed that β for age is -0.0002 ($p=0.5849$), with a log-likelihood of 66.3400 and an AIC of -126.6801 . The I^2 statistic was 66.20%, and the QE value was 77.8693 ($p<0.0001$). Regarding sex, the β was 0.0002 ($p=0.4771$) with a log-likelihood of 69.4272, an AIC of -132.8544 , an I^2 of 65.92%, and a QE of 79.8813 ($p<0.0001$). For Conscientiousness (BFI C), the β for age was -0.0010 ($p=0.2952$), with a log-likelihood of 68.6121 and an AIC of -133.2243 . The I^2 was 68.99%, and the QE statistic was 80.4512 ($p<0.0001$). For sex, the β was 0.0015 ($p=0.2120$), with a log-likelihood of 70.2198, an AIC of -134.4396 , an I^2 of 67.83%, and a QE of 83.2175 ($p<0.0001$). For Extraversion (BFI E), the regression results indicated that the β for age was -0.0005 ($p=0.4123$), with a log-likelihood of 65.7802 and an AIC of -127.5604 . The I^2 statistic was 70.12%, and the QE value was 75.6723 ($p<0.0001$). The β

**Fig. 8** Meta-analysis of the internal consistency of extraversion

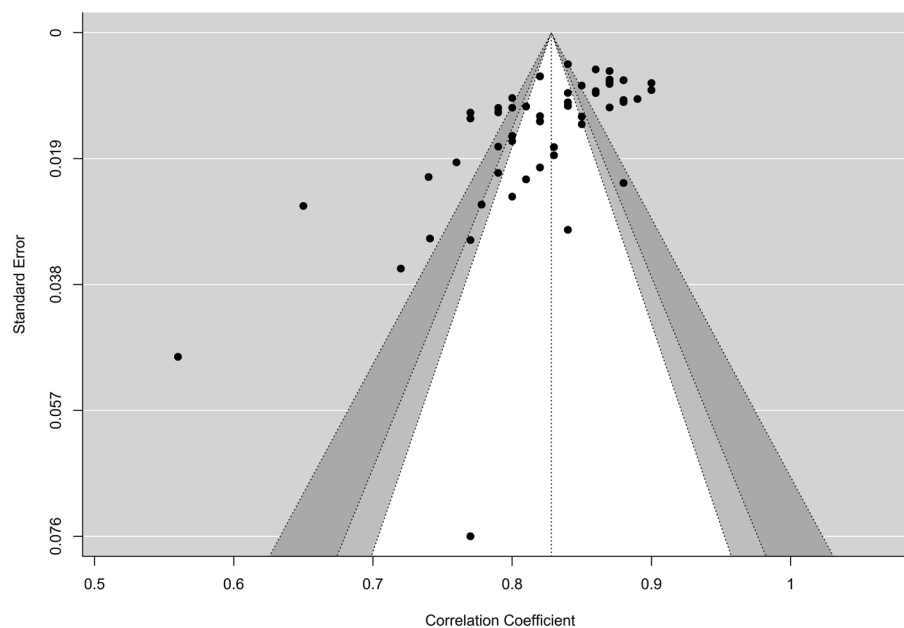


Fig. 9 Funnel plot of the internal consistency of extraversion

for sex was 0.0008 ($p=0.5598$), with a log-likelihood of 67.3093, an AIC of -128.6185 , an I^2 of 69.45%, and a QE of 76.8914 ($p<0.0001$). For Agreeableness (BFI A), the β for age was -0.0003 ($p=0.6971$), with a log-likelihood of 64.8905 and an AIC of -125.7810 . The I^2 statistic was 66.73%, and the QE value was 72.1234 ($p<0.0001$). For sex, the β was 0.0005 ($p=0.4862$), with a log-likelihood of 66.4258, an AIC of -126.8516 , an I^2 of 65.44%, and a QE of 73.4567 ($p<0.0001$). Lastly, for Neuroticism (BFI N), the β for age was -0.0007 ($p=0.3564$), with a log-likelihood of 63.5678 and an AIC of -125.1356 . The I^2 statistic was 69.20%, and the QE statistic was 74.3210 ($p<0.0001$). The β for sex was 0.0009 ($p=0.4735$), with a log-likelihood of 65.6789, an AIC of -127.3578 , an I^2 of 68.14%, and a QE of 75.5678 ($p<0.0001$).

For Openness (BFI-2 O), the regression analysis showed that β for age is -0.0002 ($p=0.5849$), with a log-likelihood of 66.3400 and an AIC of -126.6801 . The I^2 statistic was 66.20%, and the QE value was 77.8693 ($p<0.0001$). Regarding sex, the β was 0.0002 ($p=0.4771$) with a log-likelihood of 69.4272, an AIC of -132.8544 , an I^2 of 65.92%, and a QE of 79.8813 ($p<0.0001$). For Conscientiousness (BFI-2 C), the β for age was -0.0010 ($p=0.2952$), with a log-likelihood of 68.6121 and an AIC of -133.2243 . The I^2 was 68.99%, and the QE statistic was 80.4512 ($p<0.0001$). For sex, the β was 0.0015 ($p=0.2120$), with a log-likelihood of 70.2198, an AIC of -134.4396 , an I^2 of 67.83%, and a QE of 83.2175 ($p<0.0001$). For Extraversion (BFI-2 E), the regression results indicated that the β for age was -0.0005

($p=0.4123$), with a log-likelihood of 65.7802 and an AIC of -127.5604 . The I^2 statistic was 70.12%, and the QE value was 75.6723 ($p<0.0001$). The β for sex was 0.0008 ($p=0.5598$), with a log-likelihood of 67.3093, an AIC of -128.6185 , an I^2 of 69.45%, and a QE of 76.8914 ($p<0.0001$). For Agreeableness (BFI-2 A), the β for age was -0.0003 ($p=0.6971$), with a log-likelihood of 64.8905 and an AIC of -125.7810 . The I^2 statistic was 66.73%, and the QE value was 72.1234 ($p<0.0001$). For sex, the β was 0.0005 ($p=0.4862$), with a log-likelihood of 66.4258, an AIC of -126.8516 , an I^2 of 65.44%, and a QE of 73.4567 ($p<0.0001$). Lastly, for Neuroticism (BFI-2 N), the β for age was -0.0007 ($p=0.3564$), with a log-likelihood of 63.5678 and an AIC of -125.1356 . The I^2 statistic was 69.20%, and the QE statistic was 74.3210 ($p<0.0001$). The β for sex was 0.0009 ($p=0.4735$), with a log-likelihood of 65.6789, an AIC of -127.3578 , an I^2 of 68.14%, and a QE of 75.5678 ($p<0.0001$).

We also controlled language in our meta-regression analysis. The findings revealed that both BFI and BFI-2 remain consistently reliable regardless of the language they are translated in. The BFI remains consistently reliable for all its five domains *i.e.* openness to experience (Such as in Serbian: $k=3$, $n=1828$, alpha 0.80 (0.77, 0.83); Chinese: $k=4$, $n=807$, alpha 0.78 (0.75, 0.81); Czech: $k=3$, $n=1848$, alpha 0.71 (0.69, 0.74); and English: $k=6$, $n=3698$, alpha 0.82 (0.78, 0.85), conscientiousness (Such as in Serbian: $k=3$, $n=1828$, alpha 0.83 (0.81, 0.84); Chinese: $k=4$, $n=807$, alpha 0.76 (0.73, 0.80); Czech: $k=3$, $n=1848$, alpha 0.74 (0.71, 0.78); and English:

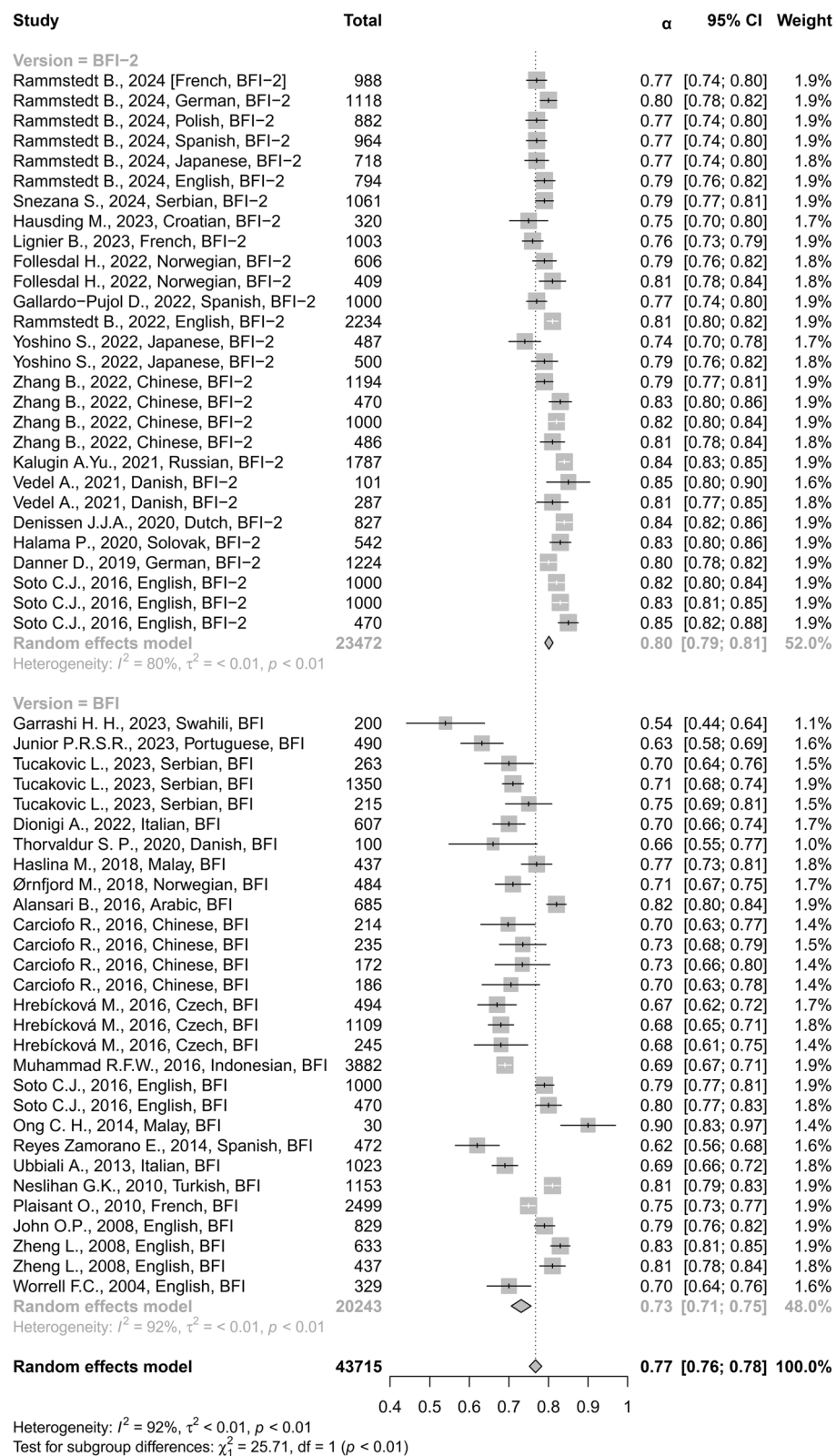


Fig. 10 Meta-analysis of the internal consistency of agreeableness

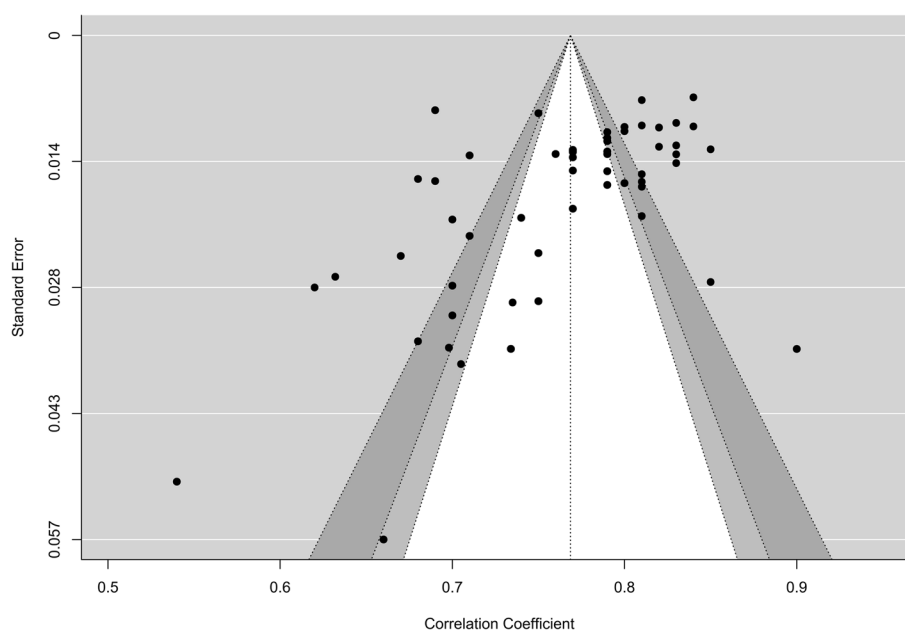


Fig. 11 Funnel plot of the internal consistency of agreeableness

$k=6$, $n=3698$, alpha 0.80 (0.79, 0.81), extraversion (*Such as in Serbian*: $k=3$, $n=1828$, alpha 0.80 (0.79, 0.82); *Chinese*: $k=4$, $n=807$, alpha 0.76 (0.73, 0.79); *Czech*: $k=3$, $n=1848$, alpha 0.75 (0.67, 0.83); and *English*: $k=6$, $n=3698$, alpha 0.88 (0.86, 0.89), agreeableness (*Such as in Serbian*: $k=3$, $n=1828$, alpha 0.71 (0.69, 0.74); *Chinese*: $k=4$, $n=807$, alpha 0.72 (0.69, 0.75); *Czech*: $k=3$, $n=1848$, alpha 0.68 (0.65, 0.70); and *English*: $k=6$, $n=3698$, alpha 0.79 (0.77, 0.82), and neuroticism (*Such as in Serbian*: $k=3$, $n=1828$, alpha 0.85 (0.83, 0.86); *Chinese*: $k=4$, $n=807$, alpha 0.75 (0.72, 0.78); *Czech*: $k=3$, $n=1848$, alpha 0.80 (0.77, 0.83); *English*: $k=6$, $n=3698$, alpha 0.86 (0.83, 0.90).

Likewise, the BFI-2 also remains consistently reliable for all its five domains *i.e.* openness to experience (*Such as in Japanese*: $k=3$, $n=1705$, alpha 0.81 (0.80, 0.83); *English*: $k=5$, $n=5498$, alpha 0.84 (0.83, 0.84); *Chinese*: $k=4$, $n=3150$, alpha 0.84 (0.83, 0.85), conscientiousness (*Such as in Japanese*: $k=3$, $n=1705$, alpha 0.84 (0.83, 0.85); *English*: $k=5$, $n=5498$, alpha 0.87 (0.86, 0.89); *Chinese*: $k=4$, $n=3150$, alpha 0.86 (0.85, 0.88), extraversion (*Such as in Japanese*: $k=3$, $n=1705$, alpha 0.84 (0.83, 0.85); *English*: $k=5$, $n=5498$, alpha 0.87 (0.86, 0.89); *Chinese*: $k=4$, $n=3150$, alpha 0.86 (0.85, 0.88), agreeableness (*Such as in Japanese*: $k=3$, $n=1705$, alpha 0.77 (0.75, 0.79); *English*: $k=5$, $n=5498$, alpha 0.82 (0.80, 0.84); *Chinese*: $k=4$, $n=3150$, alpha 0.81 (0.80, 0.83), and neuroticism (*Such as in Japanese*: $k=3$, $n=1705$, alpha 0.89 (0.88, 0.90); *English*: $k=5$, $n=5498$, alpha 0.90 (0.88, 0.92); *Chinese*: $k=4$, $n=3150$, alpha 0.88 (0.87, 0.90).

Discussion

One of the most influential models in personality psychology is the Big Five Model. It identifies five broad dimensions *i.e.*, openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. These five factors capture the fundamental aspects of human personality. However, the applicability and reliability of this model across different cultures and populations have been a subject of debate. Using REGEMA, the purpose of the current study was to assess the reliability of the BFI in several different cultural and language settings. The current study evaluated two different versions of the BFI *i.e.* the BFI (44 items) and the BFI-2 (60 items). The analysis provided a comprehensive assessment of the internal consistency across the five factors involved in the Big-5 model. The study was conducted using a wide range of samples originating from a variety of nations and languages. The reliability estimates across all five dimensions of the BFI suggested that there was a high internal consistency overall. This conclusion is in line with the findings of prior research that validated the BFI as a reliable and robust measure of the Big Five personality traits [22]. It was demonstrated that the BFI can provide consistent assessments of personality across a variety of people and contexts by the fact that the Cronbach's alpha coefficients for the five variables ranged from moderate to high.

The findings of this meta-analysis explain the internal consistency reliability of the BFI and BFI-2 dimensions across a wide range of studies involving various languages and cultural backgrounds. Overall, both versions

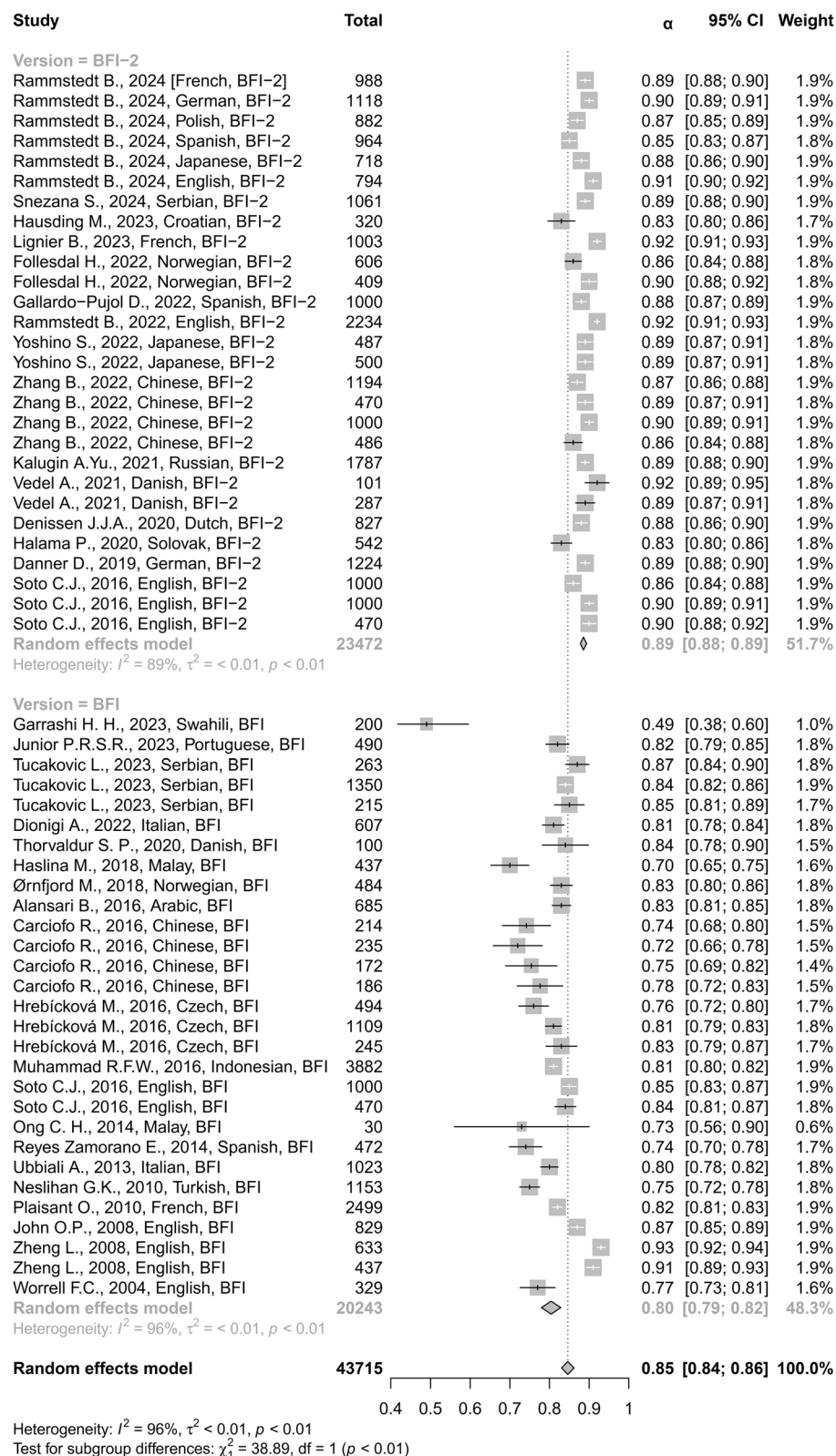


Fig. 12 Meta-analysis of the internal consistency of neuroticism

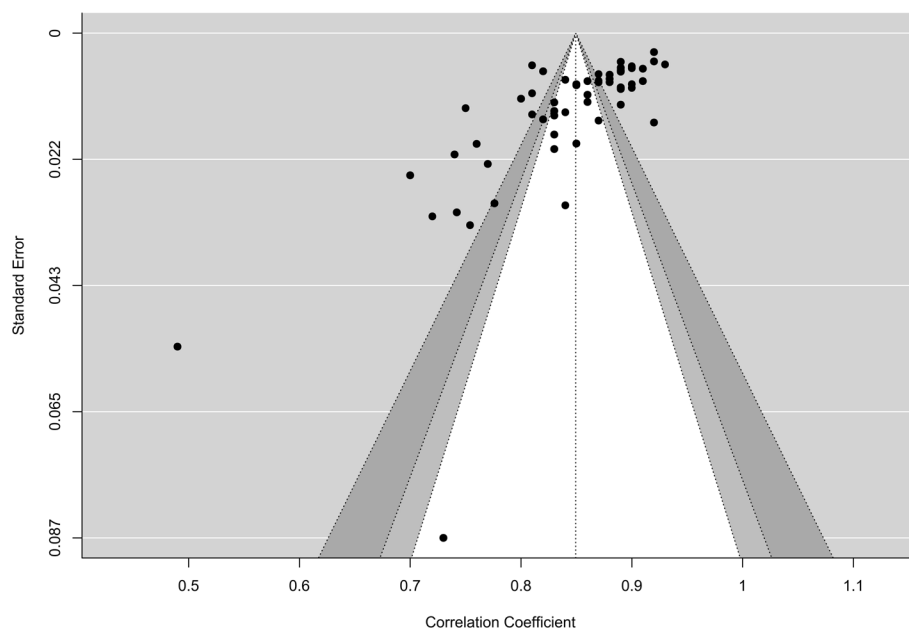


Fig. 13 Funnel plot of the internal consistency of neuroticism

of the inventory demonstrated acceptable levels of internal consistency reliability. However, the BFI-2 generally exhibited higher reliability coefficients across all five dimensions compared to the original BFI. These results suggest that the revisions made in the BFI-2 have contributed to enhancing its psychometric properties, particularly in terms of internal consistency.

The results of meta-regression analyses did not reveal any significant associations between participants' age, sex, or their combination with reliability coefficients for any of the five personality dimensions. This conclusion is consistent with earlier studies that found no impact of age or sex on the reliability of the BFI. Several studies have reported that demographic factors, such as age and sex, do not significantly impact the reliability of the BFI. For instance, a meta-analysis by Gnambs (2016) examined the test–retest reliability of the BFI across various age groups and genders. The findings revealed that these sociodemographic variables exert minimal influence on the scale's reliability. Their findings suggest that the BFI's consistency remains robust across different ages and genders, supporting its broad applicability [69]. Similarly, a large-scale meta-analysis by Li et al. (2015) on the reliability of the BFI across diverse demographic groups reported only negligible effects of age and gender on the inventory's reliability scores. This analysis indicated that while factors like sample size and nationality might influence the reliability of the BFI's dimensions, age and gender were not significant predictors of variance [70]. These findings affirm the BFI's suitability for use across varied

demographic backgrounds without notable differences in reliability, making it a consistent tool for assessing personality traits in diverse populations.

Implications

Several significant implications for future research and practice can be drawn from the findings of this meta-analysis. The substantial variation that was seen among studies highlights the need for conducting additional research on the cross-cultural measuring of personality traits. In subsequent research, efforts should be made to investigate the ways in which cultural and linguistic characteristics impact the reliability and validity of the BFI across a variety of populations. Considering the setting in which the BFI is administered is something that both researchers and practitioners ought to keep in mind. Several factors such as cultural, linguistic, and demographic aspects, have the potential to influence the measurement features of the BFI. It is important that scores be understood within the necessary cultural context. The fact that the estimates of reliability of the BFI vary depending on the context in which they are used suggests that there may be chances to increase the precision of the measurements that it takes. Nevertheless, the BFI's overall strong reliability across all five dimensions lends support to its usage in a variety of practical applications, including personnel selection, psychological assessment, and educational research, amongst others. Practitioners, on the other hand, must consider the various restrictions that

may be present in cross-cultural situations and alter their interpretations accordingly.

Conclusion

The purpose of this meta-analysis is to give a complete review of the reliability of the BFI over a wide range of samples originating from a variety of nations and languages. The BFI and its revised version, the BFI-2, offer valuable tools for assessing personality traits across different cultural backgrounds. Furthermore, the BFI's utility as a reliable instrument for assessing the Big Five personality traits is supported by the fact that it has a strong internal consistency reliability overall. The considerable heterogeneity and the possibility of publication bias, on the other hand, underscore the significance of considering contextual and cultural aspects when employing the BFI in research and practice. It is important that future studies continue to investigate the cross-cultural measurement of personality traits and look for ways to increase the precision and usefulness of the BFI across a variety of populations.

Abbreviations

BFI	Big Five Inventory
BFI-2	Big Five Inventory-2
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
OCEAN	O=Openness, C=Conscientiousness, E=Extraversion, A=Agreeableness, N=Neuroticism
REGEMA	Reliability Generalization Meta-Analysis

Acknowledgements

None.

Authors' contributions

Conceptualization, WH, AJH, MAH, HG, KT, AA, ZS, AP, and HJ; Methodology, WH, HJ, ZS; Software, HJ; Formal Analysis, HJ; Writing – Original Draft Preparation, WH, AJH, MAH, HG, KT, AA, ZS, AP, and HJ; Writing – Review & Editing WH, AJH, MAH, HG, KT, AA, ZS, AP, and HJ; Funding Acquisition, Not applicable.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

Availability of data and materials: Data are available upon request from the corresponding author (WH).

Declarations

Ethics approval and consent to participate

Not applicable. This was a systematic review and meta-analysis of published studies that were indexed in the public domain.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Humanities, COMSATS University Islamabad, Islamabad Campus, Park Road, Islamabad, Pakistan. ²Nutrition and Food Technology

Department, Agriculture School, The University of Jordan, P.O. Box, Amman 11942, Jordan. ³Research laboratory Education, Motricité, Sport et Santé, EM25, LR19JS01, High Institute of Sport and Physical Education of Sfax, University of Sfax, Sfax, Tunisia. ⁴Department of Movement Sciences and Sports Training, School of Sport Science, The University of Jordan, Amman, Jordan. ⁵Department of Training and Movement Science, Institute of Sport Science, Johannes Gutenberg-University Mainz, Mainz 55099, Germany. ⁶Research Laboratory, Molecular Bases of Human Pathology, LR19ES13, Faculty of Medicine of Sfax, University of Sfax, Sfax 3000, Tunisia. ⁷Government Hospitals, Manama, Bahrain. ⁸Department of Nursing, School of Health and Welfare, Jönköping University, Jönköping, Sweden. ⁹Department of Psychiatry, College of Medicine and Health Sciences, Arabian Gulf University, Manama, Bahrain.

Received: 14 August 2024 Accepted: 8 December 2024

Published online: 08 January 2025

References

- Ożariska-Ponikwia K. Personality and Emotional Intelligence in Second Language Learning. Cambridge Sch: Publ; 2018.
- J. Allik, A. Realo, J.B. Asendorpf, A. Baumert, M. Schmitt, G. Blum, R. van Bork, M. Rhemtulla, D. Borsboom, B.P. Chapman, D.A. Clark, C.E. Durbin, B.M. Hicks, D.M. Condon, D.K. Mroczek, G. Costantini, M. Perugini, J. Freese, L.R. Goldberg, R.R. McCrae, C.S. Nave, D.C. Funder, D.S. Ones, B.M. Wiernik, M.P. Wilms, J.W. Kostal, D.J. Ozer, A. Poropat, W. Revelle, L.G. Elleman, K.J. Sher, S.J. Weston, J.J. Jackson, D. Wood, P.D. Harms, M. Ziegler, J. Ziegler, R. Möttus. Basic definitions in personality psychology: Challenges for conceptual integrations. *Eur J Pers.* 2017;31:529–95. <https://doi.org/10.1002/per.2128>.
- Revelle W. Hans Eysenck: Personality theorist. *Pers Individ Dif.* 2016;103:32–9. <https://doi.org/10.1016/j.paid.2016.04.007>.
- Lapsley DK, Stey PC. Id, Ego, and Superego. *Encycl Hum Behav* Second Ed. 2012. <https://doi.org/10.1016/B978-0-12-375000-6.00199-3>.
- K.H. Deal, Psychodynamic Theory, *Adv. Soc. Work* 8 (2007) 184–195. <https://doi.org/10.18060/140>.
- S. Zhang, Psychoanalysis: The Influence of Freud's Theory in Personality Psychology, *Proc. Int. Conf. Ment. Heal. Humanit. Educ. (ICMHE 2020)* 433 (2020). <https://doi.org/10.2991/assehr.k.200425.051>.
- J.D. Safran, E. Gardner-Schuster, Psychoanalysis, *Encycl. Ment. Heal.* Second Ed. (2016) 339–347. <https://doi.org/10.1016/B978-0-12-397045-9.00189-0>.
- Mischel W. Toward a cognitive social learning reconceptualization of personality. *Psychol Rev.* 1973;80:252–83. <https://doi.org/10.1037/h0035002>.
- Mischel W, Shoda Y. A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure. *Psychol Rev.* 1995;102:246–68. <https://doi.org/10.1037/0033-295X.102.2.246>.
- I. Novikova, Trait, *Trait Theory, Encycl. Cross-Cultural Psychol.* (2013) 1293–1295. <https://doi.org/10.1002/9781118339893.wbeccp545>.
- Fleeson W, Jayawickreme E. Whole Trait Theory. *J Res Pers.* 2015;56:82–92. <https://doi.org/10.1016/j.jrp.2014.10.009>.
- Matthews G. Cognitive-Adaptive Trait Theory: A Shift in Perspective on Personality. *J Pers.* 2018;86:69–82. <https://doi.org/10.1111/jopy.12319>.
- Mulder R. The biology of personality, *Australas. Psychiatry.* 1992;26:364–76. <https://doi.org/10.3109/00048679209072059>.
- G. Matthews, K. Gilliland, The personality theories of H. J. Eysenck and J. A. Gray: A comparative review, *Pers. Individ. Dif.* 26 (1999) 583–626. [https://doi.org/10.1016/S0191-8869\(98\)00158-5](https://doi.org/10.1016/S0191-8869(98)00158-5).
- N. Beckmann, R.E. Wood, Editorial: Dynamic personality science. Integrating between-person stability and within-person change, *Front. Psychol.* 8 (2017). <https://doi.org/10.3389/fpsyg.2017.01486>.
- L.R. Goldberg, An Alternative "Description of Personality": The Big-Five Factor Structure, in: *J. Pers. Soc. Psychol.*, 1990: pp. 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216>.
- Mezquita L, Bravo AJ, Morizot J, Pilatti A, Pearson MR, Ibáñez MI, Ortet G. Cross-cultural examination of the Big Five Personality Trait Short Questionnaire: Measurement invariance testing and associations with mental health. *PLoS ONE.* 2019;14:1–23. <https://doi.org/10.1371/journal.pone.0226223>.

18. Zhao H, Seibert SE. The big five personality dimensions and entrepreneurial status: A meta-analytical review. *J Appl Psychol*. 2006;91:259–71. <https://doi.org/10.1037/0021-9010.91.2.259>.
19. Barrick MR, Mount MK, Judge TA. Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next? *Int J Sel Assess*. 2001;9:9–30. <https://doi.org/10.1111/1468-2389.00160>.
20. E.M.A. Wheeler, R.P. Archer, Personality assessment, in: *Encycl. Ment. Heal*. Third Ed. Vol. 1–3, 2023: pp. V2–756–V2–760. <https://doi.org/10.1016/B978-0-323-91497-0.00012-6>.
21. John OP, John OP, Donahue EM, Kentle RL. The Big Five Inventory—Versions 4a and 54. Institute of Personality and Social Research: University of California, Berkeley; 1991.
22. Soto CJ, John OP. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *J Pers Soc Psychol*. 2017;113:117–43. <https://doi.org/10.1037/PSPP0000096>.
23. Ang S, Van Dyne L, Koh C. Personality correlates of the four-factor model of cultural intelligence. *Gr Organ Manag*. 2006;31:100–23. <https://doi.org/10.1177/1059601105275267>.
24. G. Park, H. Andrew Schwartz, J.C. Eichstaedt, M.L. Kern, M. Kosinski, D.J. Stillwell, L.H. Ungar, M.E.P. Seligman, Automatic personality assessment through social media language, *J. Pers. Soc. Psychol*. 108 (2015) 934–952. <https://doi.org/10.1037/pspp0000020>.
25. Ali S, Zai Y, Jan A. Gender and Regional Differences in Five Factor Personality Traits Among Students at Secondary Level in Punjab. Pakistan. 2019;2:42–58.
26. Church AT. Personality traits across cultures. *Curr Opin Psychol*. 2016;8:22–30. <https://doi.org/10.1016/j.copsyc.2015.09.014>.
27. J.M. Dewaele, Personality: Personality traits as independent and dependent variables, in: *Psychol. Lang. Learn. Insights from Res. Theory Pract.*, Palgrave Macmillan, 2012: pp. 42–57. https://doi.org/10.1057/9781137032829_4.
28. Ozer DJ, Benet-Martínez V. Personality and the prediction of consequential outcomes. *Annu Rev Psychol*. 2006;57:401–21. <https://doi.org/10.1146/annurev.psych.57.102904.190127>.
29. O'Connor BP. A Quantitative Review of the Comprehensiveness of the Five-Factor Model in Relation to Popular Personality Inventories. *Assessment*. 2002;9:188–203. <https://doi.org/10.1177/1073191102092010>.
30. Gurven M, von Rueden C, Massenkoff M, Kaplan H, Vie ML. How universal is the big five? Testing the five-factor model of personality variation among forager-farmers in the Bolivian Amazon. *J Pers Soc Psychol*. 2013;104:354–70. <https://doi.org/10.1037/a0030841>.
31. A. Feher, P.A. Vernon, Looking beyond the Big Five: A selective review of alternatives to the Big Five model of personality, *Pers. Individ. Dif*. 169 (2021). <https://doi.org/10.1016/j.paid.2020.110002>.
32. D.P. MacKinnon, J. Cheong, A.G. Pirlott, Statistical mediation analysis, in: *APA Handb. Res. Methods Psychol*. Vol 2 Res. Des. Quant. Qual. Neuropsychol. Biol., 2012. <https://doi.org/10.1037/13620-018>.
33. Alansari B. The Big Five Inventory (BFI): Reliability and validity of its Arabic translation in non clinical sample. *Eur Psychiatry*. 2016;33:S209–10. <https://doi.org/10.1016/J.EURPSY.2016.01.500>.
34. R. Carciofo, J. Yang, N. Song, F. Du, K. Zhang, Psychometric evaluation of Chinese-language 44-item and 10-item big five personality inventories, including correlations with chronotype, mindfulness and mind wandering, *PLoS One* 11 (2016). <https://doi.org/10.1371/JOURNAL.PONE.0149963>.
35. Zhang B, Li YM, Li J, Luo J, Ye Y, Yin L, Chen Z, Soto CJ, John OP. The Big Five Inventory–2 in China: A Comprehensive Psychometric Evaluation in Four Diverse Samples. *Assessment*. 2022;29:1262–84. <https://doi.org/10.1177/10731911211008245>.
36. Hausding M, Horstmann KT. Translation and Validation of the BFI-2 in a Croatian Sample. *Psychol Test Adapt Dev*. 2023;4:141–55. <https://doi.org/10.1027/2698-1866/A000045>.
37. Hřebicková M, Jelinek M, Blatný M, Brom C, Buresová I, Graf S, Mejzlíková T, Vazsonyi AT, Záborská K, BIG FIVE INVENTORY: ZÁKLADNÍ PSYCHOMETRICKÉ CHARAKTERISTIKY ČESKÉ VERZE BFI-44 A BFI-10/Big Five Inventory: Basic psychometric properties of the Czech version of BFI-44 and BFI-10, *Cesk. Psychol*. 2016;60:567–83.
38. Palsson TS, Christensen SWMP, Pape MH, Hirata RP, Rafn T, Skou ST. Cross-cultural adaptation of the Danish version of the Big Five Inventory—a dual-panel approach, *Scand. J Pain*. 2020;20:397–406. <https://doi.org/10.1515/SJPAIN-2019-0066>.
39. Vedel A, Wellnitz KB, Ludeke S, Soto CJ, John OP, Andersen SC. Development and validation of the danish big five inventory-2? Domain- And facet-level structure, construct validity, and reliability. *Eur J Psychol Assess*. 2021;37:42–51. <https://doi.org/10.1027/1015-5759/A000570>.
40. Denissen JJA, Geenen R, Soto CJ, John OP, van Aken MAG. The Big Five Inventory–2: Replication of Psychometric Properties in a Dutch Adaptation and First Evidence for the Discriminant Predictive Validity of the Facet Scales. *J Pers Assess*. 2020;102:309–24. <https://doi.org/10.1080/00223891.2018.1539004>.
41. O.P. John, L.P. Naumann, C.J. Soto, Paradigm shift to the integrative Big Five Trait taxonomy: History, measurement, and conceptual issues, in: O.P. John, R.W. Robins, L.A. Pervin (Eds.), *Handb. Personal. Theory Res.*, 3rd ed., Guilford Press, New York, 2008: pp. 114–158.
42. Zheng L, Goldberg LR, Zheng Y, Zhao Y, Tang Y, Liu L. Reliability and concurrent validation of the IPIP Big-Five factor markers in China: Consistencies in factor structure between Internet-obtained heterosexual and homosexual samples. *Pers Individ Dif*. 2008;45:649–54. <https://doi.org/10.1016/J.PAID.2008.07.009>.
43. Worrell FC, Cross WE. The reliability and validity of big five inventory scores with African American college students. *J Multicult Couns Devel*. 2004;32:18–32. <https://doi.org/10.1002/J.2161-1912.2004.TB00358.X>.
44. Rammstedt B, Roemer L, Lechner CM, Soto CJ. Adapting the BFI-2 Around the World - Coordinated Translation and Validation in Five Languages and Cultural Contexts. *Eur J Psychol Assess*. 2024. <https://doi.org/10.1027/1015-5759/A000844>.
45. Rammstedt B, Roemer L, Danner D, Lechner CM. Don't Keep It Too Simple Simplified Items Do Not Improve Measurement Quality. *Eur J Psychol Assess*. 2024;40:117–27. <https://doi.org/10.1027/1015-5759/A000741>.
46. O. Plaisant, R. Courtois, C. Réveillère, G.A. Mendelsohn, O.P. John, Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). *Analyse convergente avec le NEO-PI-R*, *Ann. Med. Psychol. (Paris)*. 168 (2010) 97–106. <https://doi.org/10.1016/J.AMP.2009.09.003>.
47. Lignier B, Petot JM, Canada B, De Oliveira P, Nicolas M, Courtois R, John OP, Plaisant O, Soto C. Factor structure, psychometric properties, and validity of the Big Five Inventory-2 facets: evidence from the French adaptation (BFI-2-Fr). *Curr Psychol*. 2023;42:26099–114. <https://doi.org/10.1007/S12144-022-03648-0>.
48. D. Danner, B. Rammstedt, M. Bluemke, C. Lechner, S. Berres, T. Knopf, C.J. Soto, O.P. John, Das Big Five Inventar 2, *Diagnostica* 65 (2019). <https://doi.org/10.1026/0012-1924/a000218>.
49. M.R. Wibowo, W. Yudianta, I.P. Reswara, B.W. Jatmiko, Reliability and validity of the indonesian version of big five inventory, *UI Proc. Soc. Sci. Humanit*. 1 (2017). <https://www.researchgate.net/publication/317170503>.
50. Dionigi A, Duradoni M, Vagnoli L. Humor and personality: Psychometric properties of the Italian version of the comic styles markers and its relationships with the big five personality traits. *Curr Psychol*. 2022;41:8705–17. <https://doi.org/10.1007/S12144-020-01303-0>.
51. A. Ubbiali, C. Chiorri, P. Hampton, D. Donati, Psychometric properties of the Italian adaptation of the Big Five Inventory, *Boll. Di Psicol. Appl*. 2012;266: 37–46.
52. S. Yoshino, T. Shimotsukasa, A. Oshio, Y. Hashimoto, Y. Ueno, T. Mieda, I. Migiwa, T. Sato, S. Kawamoto, C.J. Soto, O.P. John, A validation of the Japanese adaptation of the Big Five Inventory-2, *Front. Psychol*. 13 (2022). <https://doi.org/10.3389/FPSYG.2022.924351>.
53. H. Muhamad, J. Roodenburg, D.W. Moore, The adaptation of the Big Five Inventory in measuring Malaysian youths' personality traits, *Int. J. Adv. Appl. Sci*. 5 (2018) 8–14. <https://doi.org/10.21833/IJAAS.2018.07.002>.
54. Hee OC. Validity and Reliability of the Big Five Personality Traits Scale in Malaysia. *Int J Innov Appl Stud*. 2014;5:309–15.
55. M. Ørnfjord, The Norwegian Hexaco-Pi-R: Psychometric properties and relationships with the Big Five Inventory, *Scand. Psychol*. 5 (2018). <https://doi.org/10.15714/SCANDPSYCHOL.5.E15>.
56. H. Føllesdal, C.J. Soto, The Norwegian Adaptation of the Big Five Inventory-2, *Front. Psychol*. 13 (2022). <https://doi.org/10.3389/FPSYG.2022.858920>.
57. P.R.S.R. Junior, D.X. da Silveira, P.C.R. Barbosa, M.A.D.S. Torres, E. da C.M. Junior, K.C.N. Areco, R.T.A. de Oliveira, A.G. Tazitu, J.A.B. Fernandes, M.G.

- Fernandes, S.K. Kasinski, Psychometric properties of the Brazilian version of the Big Five Inventory, *Trends Psychiatry Psychother.* 45 (2023). <https://doi.org/10.47626/2237-6089-2021-0458>.
58. A.Y. Kalugin, S.A. Shchebetenko, A.M. Mishkevich, C.J. Soto, O.P. John, Psychometric Properties of the Russian Version of the Big Five Inventory-2, *Psychol. J. High. Sch. Econ.* 18 (2021) 7–33. <https://doi.org/10.17323/1813-8918-2021-1-7-33>.
 59. Tucaković L, Nedeljković B. From the BFI-44 to BFI-20: Psychometric Properties of the Short Form of the Big Five Inventory. *Psychol Rep.* 2023. <https://doi.org/10.1177/00332941231161754>.
 60. S. Smederevac, D. Mitrović, S. Sadiković, B.M. Dinić, O.P. John, C.J. Soto, The big five inventory (BFI-2): Psychometric properties and validation in Serbian language, *J. Res. Pers.* 110 (2024). <https://doi.org/10.1016/J.JRP.2024.104492>.
 61. P. Halama, M. Kohút, C.J. Soto, O.P. John, Slovak Adaptation of the Big Five Inventory (BFI-2): Psychometric Properties and Initial Validation, *Stud. Psychol. (Bratisl.)*. 62 (2020) 74–87. <https://doi.org/10.31577/SP.2020.01.792>.
 62. E. Reyes Zamorano, C.Á. Carrillo, A.P. Silva, A.M. Sandoval, I.M.R. Pastrana, Psychometric properties of the Big Five Inventory in a Mexican sample, *Salud Ment.* 37 (2014) 491–497. <https://doi.org/10.17711/SM.0185-3325.2014.059>.
 63. Gallardo-Pujol D, Rouco V, Cortijos-Bernabeu A, Ocejla L, Soto CJ, John OP. Factor Structure, Gender Invariance, Measurement Properties, and Short Forms of the Spanish Adaptation of the Big Five Inventory-2. *Psychol Test Adapt Dev.* 2022;3:44–69. <https://doi.org/10.1027/2698-1866/A000020>.
 64. H.H. Garrahi, D.P.H. Barelds, B. De Raad, A psychometric evaluation of the Big Five Inventory (BFI) in an Eastern Africa population, *Meas. Instruments Soc. Sci.* 5 (2023). <https://doi.org/10.5964/MISS.11207>.
 65. Karaman NG, Dogan T, Coban AE. A study to adapt the big five inventory to Turkish. *Procedia - Soc Behav Sci.* 2010. <https://doi.org/10.1016/j.sbspro.2010.03.336>.
 66. J. Sánchez-Meca, F. Marín-Martínez, J.A. López-López, R.M. Núñez-Núñez, M. Rubio-Aparicio, J.J. López-García, J.A. López-Pina, D.M. Blázquez-Rincón, C. López-Ibáñez, R. López-Nicolás, Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist, *Res. Synth. Methods* 12 (2021). <https://doi.org/10.1002/jrsm.1487>.
 67. L.B. Mokkink, C.A.C. Prinsen, L.M. Bouter, H.C.W. de Vet, C.B. Terwee, The COnsensus-based standards for the selection of health measurement INstruments (COSMIN) and how to select an outcome measurement instrument, *Brazilian J. Phys. Ther.* 20 (2016). <https://doi.org/10.1590/bjpt-rbf.2014.0143>.
 68. McGuinness LA, Higgins JPT. Risk-of-bias VIualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods.* 2021. <https://doi.org/10.1002/jrsm.1411>.
 69. Gnambs T. Sociodemographic effects on the test-retest reliability of the big five inventory. *Eur J Psychol Assess.* 2016;32:307–11. <https://doi.org/10.1027/1015-5759/A000259>.
 70. H. Li, J. XU, J. CHEN, Y. FAN, A Reliability Meta-Analysis for 44 Items Big Five Inventory: Based on the Reliability Generalization Methodology, *Adv. Psychol. Sci.* 23 (2015). <https://doi.org/10.3724/sp.j.1042.2015.00755>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.