Edu Sportivo
Indonesian Journal of Physical Education
e-ISSN 2745-942X
Journal Homepage: https://journal.uir.ac.id/index.php/SPORTIVO

RESEARCH ARTICLE

Cross-cultural adaptation and psychometric validation of the Indonesian IPIP–NEO–60 among junior and senior high school students

[1acde]Burhan Hambali*, [1abcd]Akhmad Mar'i Muyassar, [2ade]Yusuf Hidayat, [2acd]Carsiwan, & [3ace]Wahidah Tumijan

Department of Sports Education, Faculty of Postgraduate School, Universitas Pendidikan Indonesia, Bandung, Indonesia[1]
Department of Physical Education, Health and Recreation, Faculty of Sports and Health Education, Universitas Pendidikan Indonesia, Bandung, Indonesia[2]
Faculty of Sports Science and Recreation, Universiti Teknologi MARA, Negeri Sembilan, Malaysia[3]

Received 26 June 2026; Accepted 20 August 2026; Published 29 August 2026

OPEN ACCESS

ABSTRACT

Background: The International Personality Item Pool–NEO–60 (IPIP–NEO–60) is an open-access personality inventory based on the Five-Factor Model (FFM). However, evidence on a culturally adapted and psychometrically evaluated Indonesian version remains limited. **Objectives**: This study aimed to cross-culturally adapt and psychometrically validate the Indonesian IPIP–NEO–60 among junior and senior high school students. **Methods:** A cross-sectional psychometric validation study involved 720 students aged 12–18 years, comprising 504 males (70.0%) and 216 females (30.0%). The instrument underwent forward and backward translation, multidisciplinary expert review, and cognitive interviews. Psychometric evaluation examined the hypothesised five-factor measurement model, internal consistency, convergent validity, and discriminant validity. **Results:** The measurement model showed an acceptable fit across several indices (RMSEA = 0.060, SRMR = 0.062, TLI = 0.952, CFI = 0.971). Standardised factor loadings ranged from 0.706 to 0.940. Cronbach's alpha and composite reliability ranged from 0.950 to 0.969, indicating high internal consistency. Average variance extracted ranged from 0.615 to 0.731, supporting convergent validity. Discriminant validity was supported by the Fornell–Larcker criterion and Heterotrait–Monotrait ratios, with the highest HTMT value of 0.795. **Conclusion:** The findings provide preliminary psychometric evidence supporting the Indonesian adaptation of the IPIP–NEO–60 among junior and senior high school students in West Java and Banten. Further research should examine measurement invariance, temporal stability, and external criterion-related validity.

Keywords: IPIP–NEO–60; five-factor model; personality assessment; psychometric validation; cross-cultural adaptation; adolescents

***Corresponding Author**
burhanhambali@upi.edu

10.25299/es:ijope.2026.vol7(2).29085

How to Cite: Hambali, B., Muyassar, A. M., Hidayat, Y., Carsiwan, & Tumijan, W. (2026). Cross-cultural adaptation and psychometric validation of the Indonesian IPIP–NEO–60 among junior and senior high school students. *Edu Sportivo: Indonesian Journal of Physical Education, 7*(2), 152-169. https://doi.org/10.25299/es:ijope.2026.vol7(2).29085

Authors' Contribution: a. Study Design; b. Data Collection; c. Statistical Analysis; d. Manuscript Preparation; e. Funds Collection

INTRODUCTION

Personality traits represent relatively stable individual differences in patterns of cognition, emotion, motivation, and behaviour and are associated with important psychological, educational, and social outcomes. In educational contexts, personality has been linked to academic achievement, engagement, adjustment, and other indicators of student development (Bardach et al., 2023; Mammadov, 2022; Ozer & Benet-Martinez, 2006; Poropat, 2009). Recent meta-analytic evidence further indicates that the associations between Big Five personality traits and student achievement are robust but vary according to academic domain and the type of achievement measure used (Meyer et al., 2023). These associations are particularly relevant during adolescence, a developmental period in which personality traits undergo meaningful changes in both mean levels and stability (Bleidorn et al., 2013; Klimstra et al., 2009). Longitudinal evidence indicates that Big Five traits continue to develop across adolescence and young adulthood, with age-related changes differing across personality domains (Jones et al., 2022). Recent longitudinal evidence also indicates that personality and academic development may interact dynamically during secondary education (Bardach et al., 2023). Consequently, reliable and valid personality assessment is important for research seeking to understand individual differences among adolescents.

Among contemporary models of personality, the five-factor model (FFM) has received extensive empirical support and remains one of the most influential frameworks for organising individual differences in personality (John et al., 2008; McCrae & Costa Jr, 2008). The FFM describes personality through five broad domains neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness which capture broad but distinct patterns of emotional functioning, interpersonal behaviour, self-regulation, and cognitive orientation (Costa Jr & McCrae, 1992; John et al., 2008). Recent cross-cultural research continues to support the usefulness of the Big Five framework while demonstrating that the expression and relationships of personality traits may vary across cultural contexts (Cupani et al., 2025). Evidence from cross-national research indicates that the broad five-factor structure can be identified across diverse cultural settings, although the degree of structural and metric equivalence may vary across languages and populations (De Fruyt et al., 2009; Schmitt et al., 2007; Soto & Jackson, 2020). This issue remains relevant in recent research, which shows that cross-cultural patterns of personality should be interpreted with attention to cultural context rather than assuming complete equivalence across populations (Cupani et al., 2025). Importantly, the apparent cross-cultural stability of the FFM does not eliminate the need to establish the validity and equivalence of specific instruments within each target population.

Several instruments have been developed to operationalise the FFM, differing in breadth, facet coverage, administration time, and accessibility. The NEO PI-R provides comprehensive assessment of the five domains and their underlying facets, but its proprietary nature, cost, and administration burden may constrain its use in large-scale research (Costa Jr & McCrae, 2008). Shorter instruments such as the Big Five Inventory and the Ten-Item Personality Inventory reduce respondent burden but provide less comprehensive coverage of the personality construct (Soto & John, 2017). The International Personality Item Pool (IPIP) provides an open-access framework for personality assessment and has been used extensively as an alternative to proprietary inventories (Goldberg et al., 2006; Johnson, 2014). Recent evidence from a reliability generalisation meta-analysis covering 57 datasets and more than 43,000 participants across multiple languages and cultures further demonstrates the importance of

evaluating the reliability of Big Five instruments within specific linguistic and cultural contexts rather than assuming that reliability coefficients generalise automatically across populations (Husain et al., 2025). This combination of accessibility, theoretical comparability, and flexibility makes IPIP-based measures particularly attractive for research conducted in educational and resource-constrained settings.

The IPIP–NEO–60 was subsequently developed as a concise representation of the NEO PI-R using the International Personality Item Pool. Unlike simple item reduction, its development employed Item Response Theory to select items with strong psychometric characteristics while maintaining representation of the 30 facets corresponding to the Five-Factor Model (Maples-Keller et al., 2019). In the original development study, the IPIP–NEO–60 demonstrated good reliability and convergent and criterion-related validity, with performance comparable to longer personality measures while substantially reducing respondent burden. More recent applications have continued to demonstrate the utility of the IPIP–NEO–60 in empirical research, including studies examining associations between personality traits and language performance in educational contexts (Kim & Hwang, 2024). The instrument therefore offers a potentially efficient alternative for studies requiring broad FFM coverage within a relatively short assessment format.

Despite the established psychometric evidence for the original instrument, measurement properties cannot automatically be assumed to remain equivalent following translation into another language and cultural context. Personality items may be interpreted differently because linguistic expressions, social norms, cultural values, and behavioral expectations vary across populations (Lu et al., 2023; McCrae, 2008; Schmitt et al., 2007). Recent methodological guidance emphasizes that cross-cultural validation involves more than literal translation and requires systematic consideration of translation, cultural adaptation, pre-testing, and psychometric evaluation to minimise sources of measurement bias (Cruchinho et al., 2024). Accordingly, cross-cultural adaptation should extend beyond literal translation and address semantic, conceptual, experiential, and cultural equivalence before the adapted instrument is evaluated psychometrically (Beaton et al., 2000; Epstein et al., 2015; Sousa & Rojjanasrirat, 2011). Contemporary guidance further recommends treating translation, cultural adaptation, and psychometric validation as interconnected but analytically distinguishable stages of instrument development and evaluation (Cruchinho et al., 2024).

Cross-cultural research further demonstrates that the structural stability of the Five-Factor Model does not necessarily imply complete measurement equivalence across populations. Studies involving adolescents have shown that the Big Five can be identified across cultures while also highlighting challenges in establishing equivalent item functioning and factor structures across groups (De Fruyt et al., 2009; Morizot, 2014). Similarly, large multinational studies have reported substantial cross-cultural commonality in Big Five structure but have cautioned that metric comparability may not always be achieved for individual items (Minkov et al., 2019). Recent cross-cultural measurement research likewise demonstrates that the validity of comparisons across cultural groups depends on the degree to which measurement models exhibit invariance across languages and populations (Sorrel et al., 2021). More recent cross-cultural validation research continues to emphasize the importance of testing measurement invariance before making comparisons across languages, countries, or demographic groups (Merino-Soto et al., 2025). Thus, evidence from a translated instrument should be interpreted as population-specific unless structural and measurement equivalence are explicitly evaluated.

In Indonesia, several Big Five-based instruments have already been adapted and evaluated, indicating growing interest in culturally appropriate personality assessment.

For example, Ahya and Siaputra (2022) evaluated the Indonesian version of the Big Five Inventory–2 using principal component analysis, confirmatory factor analysis, and convergent validity evidence among 853 university students. Their findings demonstrated that the Indonesian BFI-2 showed evidence of internal structural validity and convergent validity, while also illustrating the importance of evaluating the factor structure of translated personality instruments within the Indonesian context (Ahya & Siaputra, 2022). However, recent international research indicates that psychometric properties such as reliability and structural validity require careful evaluation across languages and cultural settings, reinforcing the need for population-specific validation (Rammstedt et al., 2026). However, evidence specifically concerning the Indonesian adaptation of the IPIP–NEO–60 remains limited. This distinction is important because different personality inventories are not psychometrically interchangeable, even when they are based on the same Five-Factor Model. The availability of a validated Indonesian IPIP–NEO–60 would therefore complement, rather than duplicate, existing Indonesian Big Five measures by providing an open-access instrument with broader facet representation in a relatively concise format.

Adolescence represents an important population for personality research because personality traits continue to mature during this developmental period, with changes observed in both trait levels and stability across early, middle, and late adolescence (Bleidorn et al., 2013; Klimstra et al., 2009). Recent longitudinal research provides further evidence that personality development continues throughout adolescence and young adulthood, with distinct developmental trajectories across the Big Five domains (Jones et al., 2022). Valid personality assessment among junior and senior high school students can therefore support research examining educational, psychological, social, and behavioural outcomes, including academic performance, engagement, self-regulation, and well-being (Bardach et al., 2023; Mammadov, 2022). Recent meta-analytic evidence also indicates that personality is meaningfully associated with academic achievement among school-aged students, although the magnitude and direction of these associations depend on the personality domain, academic subject, and achievement indicator (Meyer et al., 2023). Against this background, the present study aimed to cross-culturally adapt and psychometrically validate the Indonesian IPIP–NEO–60 among junior and senior high school students in West Java and Banten. Specifically, the study examined the hypothesised five-factor structure and evaluated internal consistency, convergent validity, and discriminant validity. By addressing this measurement gap, the study provides psychometric evidence for an Indonesian adaptation of an open-access FFM-based inventory and establishes a basis for its further evaluation across broader demographic and regional populations.

METHODS

Research Design

This study employed a cross-sectional psychometric validation design to cross-culturally adapt and validate the Indonesian version of the International Personality Item Pool–NEO–60 (IPIP–NEO–60). The adaptation procedure was guided by established recommendations for cross-cultural instrument adaptation and psychometric evaluation (Beaton et al., 2000; Boateng et al., 2018; Sousa & Rojjanasrirat, 2011). As the IPIP–NEO–60 is an established instrument, no new items or measurement framework were developed. The original 60 items and five-factor structure were retained, while the adaptation process involved forward and backward translation, expert review, and cognitive interviewing to ensure semantic, conceptual, and cultural appropriateness.

The preliminary Indonesian version was pre-tested among members of the target population to assess item clarity and comprehensibility, and necessary linguistic revisions were subsequently made. Psychometric evaluation examined the hypothesised five-factor structure using confirmatory factor analysis (CFA). Model fit was assessed using χ^2/df, RMSEA, SRMR, CFI, TLI, and NFI. Internal consistency was evaluated using Cronbach's alpha and Composite Reliability (CR), convergent validity using Average Variance Extracted (AVE), and discriminant validity using the Fornell–Larcker criterion and Heterotrait–Monotrait (HTMT) ratio.

Instrument

The International Personality Item Pool–NEO–60 (IPIP–NEO–60) is a 60-item personality inventory developed as a concise representation of the Five-Factor Model (FFM) based on the International Personality Item Pool (IPIP) and the NEO Personality Inventory framework (Maples-Keller et al., 2019). The instrument assesses five broad personality domains: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness, represented by 30 facets, with six facets nested within each domain. The IPIP–NEO–60 was developed using an Item Response Theory-based item selection procedure to provide a shorter alternative to longer NEO-based measures while retaining broad facet coverage (Maples-Keller et al., 2019).

Each item is rated using a [five-point] Likert-type response scale, ranging from [response option 1] to [response option 5]. The original English-language items were obtained from the publicly available IPIP–NEO–60 source and served as the basis for the Indonesian cross-cultural adaptation. The original item wording, item order, and five-domain conceptual structure were retained during the adaptation process, with linguistic modifications made only where necessary to achieve semantic and cultural appropriateness.

Procedures

The cross-cultural adaptation and psychometric evaluation of the IPIP–NEO–60 were conducted in four sequential stages: instrument selection and translation, expert review and content evaluation, cognitive interviewing, and psychometric evaluation.

Instrument Selection and Translation

The original IPIP–NEO–60 was selected as the target instrument because it provides a concise representation of the Five-Factor Model (FFM) while retaining the original 60-item structure (Maples-Keller et al., 2019). The conceptual structure and all 60 items were retained because the purpose of the study was to adapt an established instrument rather than develop a new measure. Two bilingual psychologists independently translated the original English items into Indonesian. The two translations were compared and reconciled through discussion to produce a preliminary Indonesian version. The reconciled version was subsequently back-translated into English by an independent bilingual translator who was not involved in the forward translation. The back-translated version was compared with the original items to identify discrepancies in semantic and conceptual meaning, which were resolved through consensus by the research team.

Expert Review and Content Evaluation

The preliminary Indonesian version was reviewed by a multidisciplinary panel of four experts with expertise in personality psychology, psychometrics, educational psychology, linguistics, and educational sociology. The experts evaluated the items for semantic equivalence, conceptual equivalence, clarity, and cultural appropriateness. Their

feedback was discussed by the research team and used to revise ambiguous wording, unfamiliar terminology, and sentence structure while preserving the intended meaning of the original items. The final Indonesian version was prepared for cognitive interviewing following the expert review.

Cognitive Interviewing

Cognitive interviews were conducted with 10 adolescents aged 12-18 years who were not included in the main validation sample. A verbal probing technique was used to examine participants' understanding, interpretation, and perceived clarity of the translated items. Two rounds of cognitive interviews were conducted. Feedback primarily identified issues related to wording and sentence structure, resulting in minor revisions to five items. Revisions were made to improve comprehensibility and interpretability while maintaining the original conceptual meaning of each item.

Psychometric Evaluation and Data Collection

Following the adaptation and pre-testing stages, the finalized Indonesian IPIP–NEO–60 was administered to the main sample of 720 junior and senior high school students aged 12-18 years from West Java and Banten, Indonesia. Data were collected electronically using the Open Data Kit (ODK) platform. The questionnaire was configured to require responses to all items before submission, resulting in no missing item-level responses. Participants were informed about the study objectives, voluntary nature of participation, confidentiality, and their right to withdraw. The collected data were subsequently subjected to preliminary item analysis and psychometric evaluation, including assessment of the hypothesised five-factor measurement model, internal consistency, convergent validity, and discriminant validity.

Participants and Data Collection

A total of 720 junior and senior high school students aged 12-18 years from West Java and Banten, Indonesia, participated in the psychometric evaluation of the Indonesian IPIP–NEO–60. The sample comprised 504 males (70.0%) and 216 females (30.0%). Participants were recruited from junior and senior high schools in the study area based on the predefined eligibility criteria. The sample size exceeded the minimum target of 600 participants based on the commonly applied guideline of approximately 10 participants per item for a 60-item instrument.

Data were collected electronically using the Open Data Kit (ODK) platform. Participants completed all 60 IPIP–NEO–60 items independently. The questionnaire was configured to require a response to each item before submission; consequently, no item-level missing data were recorded. Before completing the questionnaire, participants received information about the study objectives, voluntary participation, confidentiality, and their right to withdraw from the study. The collected responses were subsequently screened for data quality before psychometric analysis.

Data Analysis

Data were analysed using a series of psychometric procedures to evaluate the measurement properties of the Indonesian IPIP–NEO–60. The analysis comprised preliminary item analysis, confirmatory factor analysis, reliability assessment, and evaluation of convergent and discriminant validity.

Preliminary Item Analysis

Item performance was examined using corrected item–total correlations and inter-item correlations. Corrected item–total correlation values of ≥0.30 were considered acceptable. Item-level results were also examined for conceptual consistency with their hypothesised personality domains. Items were retained when they demonstrated satisfactory statistical and theoretical performance.

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was conducted to evaluate the hypothesised five-factor measurement model, comprising Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each item was specified to load on its theoretically corresponding factor, and the five latent factors were allowed to correlate. Model fit was evaluated using χ^2/df, RMSEA, SRMR, CFI, TLI, and NFI. Standardized factor loadings were examined to assess the relationship between each item and its respective latent construct, with values ≥ 0.50 considered acceptable and ≥ 0.70 indicating strong loadings.

Reliability

Internal consistency was evaluated using Cronbach's alpha (α) and Composite Reliability (CR). Coefficients of ≥ 0.70 were considered indicative of acceptable internal consistency.

Convergent Validity

Convergent validity was assessed using Average Variance Extracted (AVE). An AVE value of ≥ 0.50 was considered evidence that a construct explained a sufficient proportion of the variance in its indicators.

Discriminant Validity

Discriminant validity was evaluated using the Fornell–Larcker criterion and the Heterotrait–Monotrait (HTMT) ratio. Adequate discriminant validity was indicated when each construct shared greater variance with its own indicators than with other constructs and when HTMT values remained below the predefined threshold.

Ethical Considerations

The study was conducted in accordance with established ethical principles for research involving human participants. Ethical approval was obtained from [Universitas Pendidikan Indonesia] under approval number [347/UN40.F7/AM.02.01/2026]. Before data collection, participants were informed about the study objectives, procedures, voluntary nature of participation, confidentiality, and their right to withdraw at any time without consequences. For participants under 18 years of age, parental or legal guardian consent and participant assent were obtained before participation. Data were collected anonymously and used solely for research purposes, with access restricted to the research team.

RESULTS

Preliminary Item Analysis

Preliminary item analysis was conducted to examine the performance of the 60 translated items. All items met the predefined criteria for retention and were therefore retained for subsequent confirmatory factor analysis. No items were removed during the preliminary analysis.

Confirmatory Factor Analysis

A five-factor confirmatory factor analysis (CFA) was conducted to evaluate the hypothesised measurement structure of the Indonesian IPIP–NEO–60. The model consisted of five correlated personality domains: Extraversion, Agreeableness, Conscientiousness, Openness, and Neuroticism. As presented in **Table 1**, the model demonstrated acceptable fit across several complementary fit indices. RMSEA = 0.060 and SRMR = 0.062 met the recommended criteria, while TLI = 0.952 and CFI = 0.971 indicated good incremental fit. However, NFI = 0.858 was below the conventional criterion of 0.90, and the χ^2/df ratio = 19.352 exceeded the commonly used threshold. Given the sensitivity of the chi-square statistic to sample size, the fit indices were interpreted collectively rather than relying on a single index. Overall, the pattern of results provided sufficient evidence to retain the hypothesised five-factor measurement model. No post-hoc modifications based on modification indices were applied.

Figure 1 presents the standardised five-factor measurement model. The standardised factor loadings of the 60 items ranged from 0.706 to 0.940, with all items exceeding the predefined criterion of 0.70. The highest loading was observed for P9 (0.940), whereas the lowest was observed for P17 (0.706). These results indicate that all items demonstrated substantial relationships with their respective latent personality domains.

Table 1. Goodness-of-Fit Indices for the Five-Factor Measurement Model

Fit index	Criterion	Obtained value
χ^2/df	≤ 5.00	19.352
RMSEA	≤ 0.08	0.060
SRMR	≤ 0.08	0.062
NFI	≥ 0.90	0.858
TLI	≥ 0.90	0.952
CFI	≥ 0.90	0.971

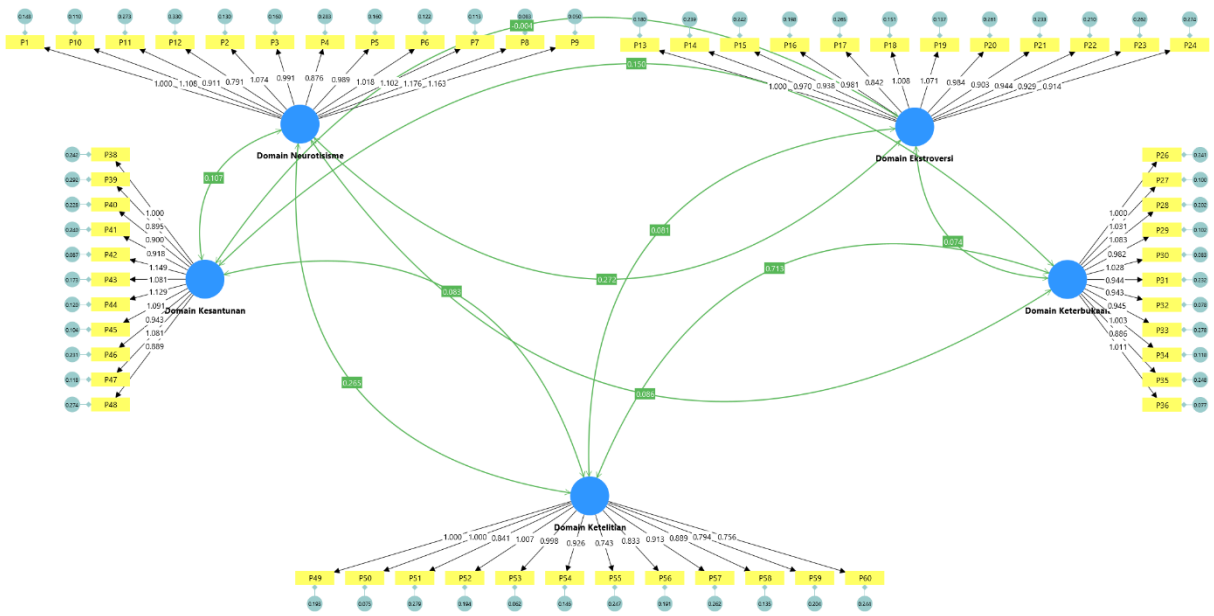

Figure 1. Reflective Measurement Model of the Indonesian Version of the IPIP–NEO–60

Reliability

The internal consistency results are presented in **Table 2**. Cronbach's alpha coefficients ranged from 0.950 to 0.969, while Composite Reliability (CR) values ranged from 0.950 to 0.969 across the five personality domains. All coefficients exceeded the

predefined criterion of 0.70, indicating high internal consistency. Openness demonstrated the highest reliability (α = 0.969; CR = 0.966), whereas Extraversion showed the lowest (α = 0.950; CR = 0.950). The very high reliability coefficients may warrant further examination of potential item redundancy in future studies.

Table 2. Internal Consistency Reliability of the Five Personality Domains

Domain	Cronbach's α	CR
Extraversion	0.950	0.950
Agreeableness	0.954	0.952
Conscientiousness	0.960	0.959
Openness	0.969	0.966
Neuroticism	0.968	0.969

Convergent Validity

Convergent validity was further assessed using Average Variance Extracted (AVE). As shown in **Table 3**, AVE values ranged from 0.615 to 0.731, exceeding the recommended minimum of 0.50 for all five personality domains. Openness demonstrated the highest AVE (0.731), whereas Extraversion showed the lowest (0.615). These findings provide evidence of adequate convergent validity across the five personality domains.

Table 3. Convergent Validity of the Five Personality Domains

Domain	AVE
Extraversion	0.615
Agreeableness	0.648
Conscientiousness	0.665
Openness	0.731
Neuroticism	0.721

Discriminant Validity

Discriminant validity was evaluated using the Fornell–Larcker criterion and Heterotrait–Monotrait (HTMT) ratio. As presented in **Table 4**, the square roots of AVE ranged from 0.784 to 0.855 and exceeded the corresponding inter-construct correlations, satisfying the Fornell–Larcker criterion. The strongest inter-construct correlation was observed between Conscientiousness and Openness (r = 0.713), which remained lower than the square roots of AVE for both constructs (0.815 and 0.855, respectively).

As shown in **Table 5**, the HTMT values ranged from 0.051 to 0.795, with all values below the recommended threshold of 0.85. The highest HTMT value was observed between Conscientiousness and Openness (HTMT = 0.795), whereas the lowest was observed between Extraversion and Agreeableness (HTMT = 0.051). These findings provide additional evidence supporting the discriminant validity of the five-factor measurement model.

Table 4. Fornell–Larcker Criterion

Domain	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Extraversion	**0.784**				
Agreeableness	-0.004	**0.805**			
Conscientiousness	0.251	0.276	**0.815**		
Openness	0.290	0.316	0.713	**0.855**	
Neuroticism	0.182	0.147	0.230	0.268	**0.849**

Note: Bold diagonal values represent the square root of AVE.

Table 5. Heterotrait–Monotrait (HTMT) Ratio

Domain	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Extraversion	—	0.051	0.269	0.306	0.193
Agreeableness	0.051	—	0.291	0.337	0.154
Conscientiousness	0.269	0.291	—	**0.795**	0.242
Openness	0.306	0.337	**0.795**	—	0.277
Neuroticism	0.193	0.154	0.242	0.277	—

Note: HTMT values below 0.85 indicate adequate discriminant validity.

DISCUSSION

Main Findings and Five-Factor Structure

This study aimed to cross-culturally adapt and evaluate the psychometric properties of the Indonesian IPIP–NEO–60 among junior and senior high school students. Overall, the findings provide supporting psychometric evidence for the hypothesised five-factor structure of the instrument. The CFA results showed acceptable values for RMSEA, SRMR, CFI, and TLI, although the χ^2/df ratio and NFI did not meet the conventional criteria. Given that chi-square-based statistics are sensitive to sample size, model fit should be interpreted using multiple complementary indices rather than a single criterion (Hu & Bentler, 1999; Kline, 2023). In addition, all 60 items demonstrated substantial standardised factor loadings ranging from 0.706 to 0.940, indicating strong relationships between the observed items and their respective latent personality domains. These findings support the representation of extraversion, agreeableness, conscientiousness, openness, and neuroticism as five correlated personality dimensions in the present Indonesian adolescent sample.

The findings are broadly consistent with the theoretical foundation of the five-factor model (FFM), which conceptualises personality as a multidimensional structure comprising relatively distinct but correlated domains (McCrae & Costa, 2008; Soto & John, 2017). Previous research has also demonstrated that the FFM can be identified across different cultural and linguistic contexts, although the strength and organisation of individual traits may vary across populations (Church, 2016; Schmitt et al., 2007). Evidence from studies using IPIP-based instruments further indicates that abbreviated personality measures can retain meaningful representations of the broader FFM when item selection and construct coverage are theoretically grounded (Johnson, 2014; Maples-Keller et al., 2019). More recent cross-cultural psychometric research similarly emphasises the importance of evaluating both structural fit and cultural appropriateness when adapting personality instruments across languages and populations (Cruchinho et al., 2024; Le Corff et al., 2025).

The present findings should nevertheless be interpreted as evidence supporting the five-factor structure within the current Indonesian adolescent sample, rather than as definitive proof that the structure is universally invariant across cultures. Cross-cultural personality research has shown that broad personality dimensions are generally reproducible, but cultural, linguistic, and sampling differences can influence item functioning and factor structure (Church, 2016; Schmitt et al., 2007; van de Vijver & Leung, 2021). Therefore, replication with samples from other Indonesian regions and demographic groups remains necessary.

Cross-Cultural Adaptation and Preservation of Construct Meaning

An important contribution of this study is the systematic adaptation procedure used to preserve the intended meaning of the original IPIP–NEO–60. Forward translation, reconciliation, backward translation, multidisciplinary expert review, and cognitive interviewing were used to identify linguistic ambiguity and potential differences in

interpretation before large-scale administration. Such procedures are widely recommended in cross-cultural instrument adaptation because literal translation alone may not guarantee semantic, conceptual, or cultural equivalence between the source and target versions (Beaton et al., 2000; Epstein et al., 2015; Sousa & Rojjanasrirat, 2011).

The cognitive interviewing stage provided an additional opportunity to examine how adolescents understood and interpreted the translated items. Cognitive interviewing is particularly useful when adapting self-report instruments because apparently equivalent words or expressions may be interpreted differently across linguistic and cultural groups (Daouk-Öyry & McDowal, 2013; Willis & Miller, 2011). In the present study, feedback from the cognitive interviews resulted in minor revisions to five items, primarily involving wording and sentence structure. Importantly, all 60 items were retained, suggesting that the adaptation process addressed linguistic and interpretive issues without requiring substantial alteration of the original item content.

The strong standardised factor loadings observed in the subsequent CFA provide additional evidence that the translated items maintained meaningful relationships with their hypothesised personality domains. However, these findings should not be interpreted as direct evidence of measurement equivalence across cultures. Measurement equivalence requires explicit testing of measurement invariance, such as configural, metric, and scalar invariance across relevant groups (Putnick & Bornstein, 2016). Thus, the present findings provide evidence of successful adaptation and structural validity in the Indonesian sample, while formal cross-cultural measurement invariance remains an important direction for future research.

Reliability and Convergent Validity

The reliability findings provide further support for the psychometric quality of the adapted instrument. Cronbach's alpha coefficients ranged from 0.950 to 0.969, while Composite Reliability values ranged from 0.950 to 0.969, substantially exceeding the commonly accepted minimum criterion of 0.70 (Hair et al., 2019). These findings indicate strong internal consistency across the five personality domains and are broadly consistent with previous research showing that IPIP-based personality measures can demonstrate adequate to high reliability despite their relatively concise format (Johnson, 2014; Maples-Keller et al., 2019).

However, the very high reliability coefficients should also be interpreted cautiously. Extremely high internal consistency coefficients may indicate that items within a construct are highly homogeneous and potentially redundant rather than simply demonstrating superior measurement quality (Streiner, 2003; Taber, 2018). This issue is particularly relevant for the present instrument because all 60 items demonstrated relatively high standardised factor loadings. Future research could therefore examine item redundancy and whether a more parsimonious item set could provide comparable psychometric performance without unnecessarily increasing respondent burden.

Convergent validity was also supported by the AVE values, which ranged from 0.615 to 0.731, exceeding the commonly recommended threshold of 0.50 (Hair et al., 2019). These results indicate that each latent personality domain explained a substantial proportion of the variance in its corresponding indicators. The findings are consistent with the broader psychometric literature suggesting that adequate factor loadings combined with sufficient AVE provide evidence that indicators converge on their intended latent constructs (Hair et al., 2019; Sarstedt et al., 2022). Thus, the Indonesian adaptation demonstrated not only high internal consistency but also satisfactory convergence among items representing each personality domain.

Discriminant Validity

The discriminant validity findings provide additional evidence that the five personality domains remained empirically distinguishable. The Fornell–Larcker criterion was satisfied because the square root of AVE for each construct exceeded its correlations with the other domains. In addition, all HTMT values were below the predefined threshold of 0.85, with the highest value observed between Conscientiousness and Openness (HTMT = 0.795). The combination of these findings indicates that the five domains demonstrated adequate empirical distinctiveness despite their expected theoretical associations (Henseler et al., 2015).

The relatively stronger association between conscientiousness and openness is theoretically plausible because personality domains are not expected to be completely independent. The FFM represents related dimensions of personality rather than five statistically orthogonal constructs (DeYoung, 2015; McCrae & Costa, 2008). Therefore, moderate associations should not automatically be interpreted as evidence of construct redundancy. Rather, the important consideration is whether the degree of overlap compromises the empirical distinction between constructs. In the present study, the conscientiousness–openness relationship remained below the square roots of AVE for both constructs, while the HTMT value remained below 0.85, providing evidence that the two domains remained distinguishable. These findings are also consistent with recommendations that discriminant validity should preferably be evaluated using multiple sources of evidence rather than relying on a single criterion (Henseler et al., 2015; Voorhees et al., 2016). The convergence of the Fornell–Larcker and HTMT results therefore strengthens the conclusion that the Indonesian IPIP–NEO–60 retains distinguishable personality domains within the present sample.

Implications for Educational and Physical Education Research

The availability of a psychometrically supported Indonesian IPIP–NEO–60 has potential implications for research in psychology, education, and sport science. A brief personality measure may provide researchers with a practical instrument for examining relationships between personality and educational or behavioural outcomes without imposing the burden associated with longer personality inventories (Johnson, 2014; Maples-Keller et al., 2019). The implications may be particularly relevant to physical education (PE) research. Personality characteristics have been discussed in relation to students' motivation, engagement, interpersonal behaviour, persistence, and participation in physical activity (Bailey et al., 2009; Beni et al., 2017; Whitehead, 2019). A culturally appropriate personality measure could therefore support future investigations into whether individual differences in traits such as conscientiousness, extraversion, or agreeableness are associated with learning engagement, cooperation, self-regulation, or participation in PE activities (Beni et al., 2017; Ntoumanis, 2001).

However, these applications should be interpreted as potential implications rather than direct findings of the present study. The current validation was conducted among junior and senior high school students and did not specifically test whether the instrument performs differently within PE classes or under PE-specific conditions. Therefore, future research should investigate the applicability of the Indonesian IPIP–NEO–60 across different educational and sport contexts before drawing context-specific conclusions.

Limitations and Future Research

Several limitations should be considered when interpreting the findings. First, participants were recruited from West Java and Banten, which may limit the generalisability of the findings to adolescents from other regions of Indonesia. Indonesia is characterised by substantial linguistic, cultural, and socio-demographic diversity; therefore, replication across additional regions is necessary to establish broader applicability of the adapted instrument (van de Vijver & Leung, 2021). Second, the sample was restricted to junior and senior high school students aged 12-18 years. Consequently, the psychometric properties reported in this study should not automatically be generalised to younger children, university students, adults, or other populations. Personality measurement can be influenced by developmental stage, and evidence obtained from adolescents may not necessarily transfer directly to other age groups (Soto & Tackett, 2015).

Third, the present study focused on structural validity, internal consistency, convergent validity, and discriminant validity. Other forms of psychometric evidence, including test–retest reliability, criterion-related validity, predictive validity, and measurement invariance across gender, age, educational level, and geographic region, were not examined. Future studies should therefore extend the validation programme by evaluating the temporal stability of scores and testing whether the measurement model operates equivalently across relevant demographic group (Putnick & Bornstein, 2016). Finally, although the present findings provide encouraging evidence for the Indonesian adaptation, further research should examine the instrument using independent samples and, where possible, more diverse sampling strategies. Replication studies would help determine whether the observed factor structure and psychometric properties are stable beyond the current sample and geographic context.

CONCLUSION

This study conducted the cross-cultural adaptation and psychometric validation of the Indonesian version of the IPIP–NEO–60 among junior and senior high school students. The findings provided evidence supporting the hypothesised five-factor structure, comprising extraversion, agreeableness, conscientiousness, openness to experience, and neuroticism. The 60 items demonstrated substantial standardised factor loadings, while the instrument showed high internal consistency and satisfactory convergent and discriminant validity. Overall, the findings indicate that the Indonesian IPIP–NEO–60 provides promising psychometric evidence for assessing the five broad personality domains among Indonesian adolescents in the sampled population.

Several limitations should be considered. The validation sample was limited to junior and senior high school students from West Java and Banten, which may restrict generalisability to other age groups, educational settings, and regions of Indonesia. In addition, the study did not examine measurement invariance, test–retest reliability, or criterion-related validity. Therefore, the findings should be interpreted as initial psychometric evidence rather than comprehensive validation across all relevant populations and measurement conditions.

Future research should extend the validation of the Indonesian IPIP–NEO–60 by examining measurement invariance across gender, age, educational level, and geographic regions, as well as test–retest reliability to establish temporal stability. Further studies could also apply Item Response Theory (IRT) to investigate item functioning and differential item functioning across relevant groups. In addition, research using independent and more geographically diverse samples should examine the relationships between the five personality domains and relevant psychological, educational, and behavioural outcomes. Such work would strengthen the evidence base and clarify the

applicability of the Indonesian IPIP–NEO–60 across diverse Indonesian populations and research contexts.

ACKNOWLEDGEMENTS

The authors would like to express their deepest gratitude to all parties who have supported the implementation of this programme. This community service program is funded by the Universitas Pendidikan Indonesia through the community service program contract No. 347/UN40.F7/AM.02.01/2026.

CONFLICT OF INTEREST

Regarding the publishing of this work, the authors affirm that they have no conflicts of interest.

AI DISCLOSURE STATEMENT

The authors state that ChatGPT was utilised to enhance the manuscript's readability by helping to correct grammar and sharpen the discussion part. The originality, scientific content, and interpretation of the research findings were unaffected by the usage of this technology. The content of this article is entirely the writers' responsibility.

REFERENCES

Ahya, A., & Siaputra, I. B. (2022). Validation Of Big Five Inventory-2 (BFI-2) for Indonesia: not Perfect but Still Valid and Reliabel to Measure Personality. *Jurnal Psikologi Ulayat*, *9*(1), 179–203. https://doi.org/10.24854/jpu458

Bailey, R., Armour, K., Kirk, D., Jess, M., Pickup, I., Sandford, R., & Education, B. P. (2009). The Educational Benefits Claimed for Physical Education and School Sport: An Academic Review. *Research Papers in Education*, *24*(1), 1–27. https://doi.org/10.1080/02671520701809817

Bardach, L., Hübner, N., Nagengast, B., Trautwein, U., & Von Stumm, S. (2023). Personality, Intelligence, and Academic Achievement: Charting Their Developmental Interplay. *Journal of Personality*, *91*(6), 1326–1343. https://doi.org/10.1111/jopy.12810

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Beni, S., Fletcher, T., & N'i Chróin'in, D. (2017). Meaningful Experiences in Physical Education and Youth Sport: A Review of the Literature. *Quest*, *69*(3), 291–312. https://doi.org/10.1080/00336297.2016.1224192

Bleidorn, W., Klimstra, T. A., Denissen, J. J. A., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2013). Personality Maturation around the World: A Cross-Cultural Examination of Social-Investment Theory. *Psychological Science*, *24*(12), 2530–2540. https://doi.org/10.1177/0956797613498396

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Church, A. T. (2016). Personality Traits Across Cultures. *Current Opinion in Psychology*, *8*, 22–30. https://doi.org/10.1016/j.copsyc.2015.09.014

Costa Jr, P. T., & McCrae, R. R. (1992). The Five-Factor Model of Personality and its Relevance to Personality Disorders. *Journal of Personality Disorders*, *6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

Costa Jr, P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment, 2*. 179–198. Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Cruchinho, P., Dolores, M. D. L.-F., Capelas, M. L., Almeida, S., Bennett, P. M., da Silva, M., Teixeira, G., Nunes, E., Lucas, P., & Gaspar, F. (2024). Translation, Cross-Cultural Adaptation, and Validation of Measurement Instruments: A Practical Guideline for Novice Researchers. *Journal of Multidisciplinary Healthcare*, *17*, 2701–2728. https://doi.org/10.2147/JMDH.S419714

Cupani, M., Lorenzo-Seva, U., & Morán, V. (2025). Personality Traits in Latin America: A Cross-Cultural Study of the Big Five Factor Structure and its Relationship with Self-Reported Daily Behaviors. *Journal of Personality Assessment*, *107*(1), 89–99. https://doi.org/10.1080/00223891.2024.2353139

Daouk-Öyry, L., & McDowal, A. (2013). Using Cognitive Interviewing for the Semantic Enhancement of Multilingual Versions of Personality Questionnaires. *Journal of Personality Assessment*, *95*(4), 407–416. https://doi.org/10.1080/00223891.2012.735300

De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., Costa Jr, P. T., & & Collaborators of the Adolescent Personality Profiles of Cultures Project (2009). Assessing the Universal Structure of Personality in Early Adolescence: The NEO-PI-R and NEO-PI-3 In 24 Cultures. *Assessment*, *16*(3), 301–311. https://doi.org/10.1177/1073191109333760

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Epstein, J., Santo, R. M., & Guillemin, F. (2015). A Review of Guidelines for Cross-Cultural Adaptation of Questionnaires Could not Bring Out a Consensus. *Journal of Clinical Epidemiology*, *68*(4), 435–441. https://doi.org/10.1016/j.jclinepi.2014.11.021

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the Future of Public-Domain Personality Measures. *Journal of Research in Personality*, *40*(1), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis (8th Ed.)*. Cengage Learning

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Husain, W., Haddad, A. J., Husain, M. A., Ghazzawi, H., Trabelsi, K., Ammar, A., Saif, Z., Pakpour, A., & Jahrami, H. (2025). 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. *BMC Psychology*, *13*(1), 20. https://doi.org/10.1186/s40359-024-02271-x

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm Shift to the Integrative Big Five Trait Taxonomy. In book: *Handbook of personality: Theory and research (3rd edition)*. Guilford Press

Johnson, J. A. (2014). Measuring Thirty Facets of the Five Factor Model with a 120-Item Public Domain Inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, *51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Jones, S. A., Van Fossen, R. P., Thompson, W. K., Baker, F. C., Clark, D. B., & Nagel, B. J. (2022). Developmental Trajectories of Big Five Personality Traits among Adolescents and Young Adults: Differences by Sex, Alcohol Use, and Marijuana Use. *Journal of Personality*, *90*(5), 748–761. https://doi.org/10.1111/jopy.12694

Kim, J., & Hwang, Y. (2024). How are Personality Factors Connected to EFL Learners' Oral Performance? A Psychological Network Analysis. *System*, *127*, 103516. https://doi.org/10.1016/j.system.2024.103516

Klimstra, T. A., Hale III, W. W., Raaijmakers, Q. A. W., Branje, S. J. T., & Meeus, W. H. J. (2009). Maturation of Personality in Adolescence. *Journal of Personality and Social Psychology*, *96*(4), 898-912. https://doi.org/10.1037/a0014746

Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling (5th ed.)*. The Guilford Press

Le Corff, Y., Aluja, A., Atitsogbe, K. A., Courtois, R., Dahourou, D., Forget, K., Hansenne, M., Kavitha, D., Kounou, K. B., Lapalme, M., & others. (2025). Cross-Cultural Validity of the Five-Factor Personality Inventory for ICD-11 Across Nine Countries and Validation of a French Translation. *Personality Disorders, 16*(5), 476–490. https://doi.org/10.1037/per0000711

Lu, J. G., Benet-Martínez, V., & Wang, L. C. (2023). A Socioecological-Genetic Framework of Culture and Personality: Their Roots, Trends, and Interplay. *Annual Review of Psychology*, *74*(1), 363–390. https://doi.org/10.1146/annurev-psych-032420-032631

Mammadov, S. (2022). Big Five Personality Traits and Academic Performance: A Meta-Analysis. *Journal of Personality*, *90*(2), 222–255. https://doi.org/10.1111/jopy.12663

Maples-Keller, J. L., Williamson, R. L., Sleep, C. E., Carter, N. T., Campbell, W. K., & Miller, J. D. (2019). Using Item Response Theory to Develop A 60-Item Representation of the NEO PI--R using the International Personality Item Pool: Development of the IPIP--NEO--60. *Journal of Personality Assessment*, *101*(1), 4–15. https://doi.org/10.1080/00223891.2017.1381968

McCrae, R. R. (2008). A Note on Some Measures of Profile Agreement. *Journal of Personality Assessment*, *90*(2), 105–109. https://doi.org/10.1080/00223890701845104

McCrae, R. R., & Costa, P. T. (2008). Empirical and Theoretical Status of the Five-Factor Model of Personality Traits. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment, Vol. 1. Personality theories and models (pp. 273–294)*. Sage Publications, Inc. https://doi.org/10.4135/9781849200462.n13

Merino-Soto, C., Dominguez-Lara, S., Rodriguez, L. M., Chans, G. M., & Marti-Vilar, M. (2025). Cross-Cultural Measurement Invariance of the BFI-15p in University Students from Argentina, Spain, and Peru. *Acta Psychologica*, *256*, 104993. https://doi.org/10.1016/j.actpsy.2025.104993

Meyer, J., Jansen, T., Hübner, N., & Lüdtke, O. (2023). Disentangling the Association between the Big Five Personality Traits and Student Achievement: Meta-Analytic Evidence on the Role of Domain Specificity and Achievement Measures. *Educational Psychology Review*, *35*(1), 12. https://doi.org/10.1007/s10648-023-09736-2

Minkov, M., van de Vijver, F. J. R., & Schachner, M. (2019). A Test of a New Short Big-Five Tool in Large Probabilistic Samples from 19 Countries. *Personality and Individual Differences*, *151*, 109519. https://doi.org/10.1016/j.paid.2019.109519

Morizot, J. (2014). Construct Validity of Adolescents' Self-Reported Big Five Personality Traits: Importance of Conceptual Breadth and Initial Validation of a Short Measure. *Assessment*, *21*(5), 580–606. https://doi.org/10.1177/1073191114524015

Ntoumanis, N. (2001). A Self-Determination Approach to the Understanding of Motivation in Physical Education. *The British journal of educational psychology, 71*(Pt 2), 225–242. https://doi.org/10.1348/000709901158497

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the Prediction of Consequential Outcomes. *Annu. Rev. Psychol.*, *57*(1), 401–421. https://doi.org/10.1146/annurev.psych.57.102904.190127

Poropat, A. E. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin*, *135*(2), 322-338. https://doi.org/10.1037/a0014996

Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review: DR, 41,* 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rammstedt, B., Roemer, L., Lechner, C. M., & Soto, C. J. (2026). Adapting the BFI-2 Around the World—Coordinated Translation and Validation in Five Languages and Cultural Contexts. *European Journal of Psychological Assessment*, *42*(2), 99-111. https://doi.org/10.1027/1015-5759/a000844

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2022). Partial Least Squares Structural Equation Modeling. In: Homburg, C., Klarmann, M., Vomberg, A. *(eds) Handbook of Market Research*. Springer, Cham. https://doi.org/10.1007/978-3-319-57413-4_15

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Mart'inez, V. (2007). The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations. *Journal of Cross-Cultural Psychology*, *38*(2), 173–212. https://doi.org/10.1177/0022022106297299

Sorrel, M. A., Garc'ia, L. F., Aluja, A., Rolland, J.-P., Rossier, J., Roskam, I., & Abad, F. J. (2021). Cross-cultural measurement invariance in the Personality Inventory for DSM-5☆. *Psychiatry Research*, *304*, 114134. https://doi.org/0.1016/j.psychres.2021.114134

Soto, C. J., & Tackett, J. L. (2015). Personality Traits in Childhood and Adolescence: Structure, Development, and Outcomes. *Current Directions in Psychological Science*, *24*(5), 358–362. https://doi.org/10.1177/0963721415589345

Soto, C. J., & John, O. P. (2017). The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model with 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power. *Journal of Personality and Social Psychology*, *113*(1), 117-143. https://doi.org/10.1037/pspp0000096

Soto, C., & Jackson, J. (2020). Five-Factor Model of Personality. In D. S. Dunn (Ed.), *Five-Factor Model of Personality*. Oxford University Press. https://doi.org/10.1093/obo/9780199828340-0120

Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, Adaptation and Validation of Instruments or Scales for Use in Cross-Cultural Health Care Research: A Clear and User-Friendly Guideline. *Journal of Evaluation in Clinical Practice*, *17*(2), 268–274. https://doi.org/10.1111/j.1365-2753.2010.01434.x

Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Taber, K. S. (2018). The Use of Cronbach's Alpha when Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

van de Vijver, F. J. R., & Leung, K. (2021). *Methods and Data Analysis for Cross-Cultural Research.* (V. H. Fetvadjiev, J. He, & J. R. J. Fontaine, Eds.) (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107415188

Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant Validity Testing in Marketing: An Analysis, Causes for Concern, and Proposed Remedies. *Journal of the Academy of Marketing Science*, *44*(1), 119–134. https://doi.org/10.1007/s11747-015-0455-4

Whitehead, M. (2019). *Physical Literacy across the World (1st ed.).* Routledge. https://doi.org/10.4324/9780203702697

Willis, G. B., & Miller, K. (2011). Cross-Cultural Cognitive Interviewing: Seeking Comparability and Enhancing Understanding. *Field Methods*, *23*(4), 331–341. https://doi.org/10.1177/1525822X11416092