Reliability and Validity of the Neo-Factor Inventory: A Systematic Review

Khushi B Raju[1] , Surej Unnikrishnan[2] , J Larissa Pinky[3] , Reneeta Lydia Winson[4]

[1,3,4] Integrated Bs. c. Ms. c. Psychology (Specialisation in Clinical Psychology)

[2]Assistant Professor, PG Department of Psychology, Jyoti Nivas College Autonomous

Abstract

Psychometric properties are the quantifiable aspects of a test that indicate its statistical strength or weakness. They are intrinsic components of a test and reveal information about a test's adequacy, relevance, and usefulness. Validity, reliability, and norming are fundamental psychometric properties of a test. The study investigates how the NEO Five-Factor Inventory (NEO-FFI) functions with respect to validity and reliability among various populations across different environments. The research reviews three key properties including internal consistency and retest reliability and construct validity from studies appearing since 2005 and focuses on these measures as they relate to emerging adults and stroke survivors and cross-cultural groups. Studies indicate robust psychometric measurement capabilities specifically for Neuroticism and Conscientiousness aspects yet Openness and Agreeableness domains present inconsistent results across various conditions. The evaluation discusses how assessment environments and cultural elements affect personality evaluation and argues for instrumentation that adjust to individual conditions. The evaluation process has two main weaknesses due to sample variability among participants and incomplete inclusion of English and grey literature sources.

Keywords: Cross-cultural, Emerging Adults, FFI, Five-Factor Model, Personality Assessment, Psychometric

Introduction

Personality is the consistent traits, behaviors, and emotional responses that define an individual's individualized adaptation to life. It consists of interests, values, self-concept, and skills. Personality psychology investigates the structure, formation, and diversity of personality with an emphasis on persistent individual differences. These theories include trait theory, psychoanalysis, and humanism, which provide varying explanations of personality development. Personality also affects behavior, motivation, and interpersonal relationships. Studies indicate that personality characteristics tend to be somewhat stable across time but can change as a result of experience and aging (APA, n.d.).

The Big Five personality factors—openness, conscientiousness, extraversion, agreeableness, and neuroticism—are generally accepted as basic dimensions of personality. They account for individual differences in behavior, thinking, and emotional reactions.

Openness is related to creativity and openness to new ideas, and conscientiousness is characterized by organization and goal-orientation. Extraversion refers to sociability and activity levels, and agreeableness is characterized by kindness and compassion. Neuroticism,

however, refers to emotional instability and stress sensitivity. Empirical findings have shown that the traits are somewhat stable across time and have genetic as well as environmental contributions. Evidence further points to the fact that as a person ages, he or she becomes more conscientious and agreeable but lower in extraversion and neuroticism. Cross-cultural universality of these traits lends credence to their value in understanding personality in humans. Psychologists tend to employ the Big Five model to measure personality in diverse arenas such as mental health, workplace efficiency, and interpersonal relationships.

Knowledge of one's personality profile can give great insights into areas of strength and areas for improvement. Though stable, personality traits can change based on life experience and altering conditions, affecting behavior and decision-making over the course of a lifetime (MSEd, K. C., 2025, January 29).

The NEO Five-Factor Inventory (NEO FFI) is a 60-question self-report test that measures the five-factor model (FFM) of personality: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. There are 12 items for each domain, which were chosen from the NEO Personality Inventory for having their highest correlations with corresponding factor scores. With satisfactory reliability, however, mixed results from factor analyses have indicated inconsistent latent structure. Exploratory structural equation

modeling (ESEM) indicates some of the items that are not functioning well but supports the five-factor structure. NEO FFI is still used extensively in clinical and research environments to investigate personality's correlation with psychopathology (Rosellini & Brown, 2010).

Personality dimensions, measured using the NEO Five-Factor Inventory (NEO FFI), are central to explaining individual differences among emerging adults. Emerging adults, during their transition to adulthood, display considerable variation in Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness that affect mental well-being and adjustment. The brief format of the NEO FFI renders it perfect for measuring personality in large samples, despite issues with its latent structure. Research applying exploratory structural equation modeling (ESEM) validates its effectiveness but identifies some poorly performing items, requiring revision for greater validity in emerging adult samples (Rosellini & Brown, 2010).

Review of Literature

McCrae's (1991) work was specifically aimed at presenting the Five-Factor Model (FFM) of personality and its assessment using the NEO Personality Inventory (NEO-PI) and determining the use of the NEO-PI in clinical contexts. The population of interest was adult populations and a sample of 297 adult volunteers. The aim was to examine the comprehensiveness, reliability, validity, and stability of the NEO-PI as a measure of the five personality factors: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. The approach involved cross-walks of NEO-PI scales with well-researched scales such as the Minnesota Multiphasic Personality Inventory (MMPI) and Millon Clinical Multiaxial Inventory (MCMI) in order to highlight differences and commonalities of normal and clinical testing. NEO-PI successfully measured these five dimensions, displaying acceptable reliability and validity, it was discovered by the study. Six facets that make up each

factor and total 30 subscales are Likert-rated by items. The NEO-PI's psychometric characteristics have been discovered to generalize across age, culture, and measurement method. The results indicate that the FFM is an extensive model of personality traits for both clinical and normal populations and can support its application in clinical evaluation (McCrae, 1991).

The first study, Reliability of the NEO Five-Factor Inventory Short Form for Assessing Personality After Stroke (Dwan et al., 2017), aimed to evaluate the reliability of the NEO-FFI in assessing personality changes after stroke. The study involved 40 stroke participants (52.5% male, M age = 65.55 years) and 40 matched controls. Participants completed the NEO-FFI alongside measures of depression and anxiety. The findings suggested that internal consistency was adequate (α = .57-.86), though agreeableness showed lower reliability. Concordance between self-reported and informant-rated NEO-FFI scores was fair to excellent (ICC = .58-.78). Neuroticism was positively correlated with mood disturbances, whereas extraversion and agreeableness were negatively linked to depression and anxiety. A notable finding was that extraversion levels decreased while agreeableness levels increased in stroke patients during the hospital-to-home transition. However, the study was limited by a modest sample size, leading to wide confidence intervals, particularly for agreeableness, conscientiousness, and openness. Additionally, the use of a convenience sample and participant attrition restricted the generalizability of the findings. Future research should focus on larger samples and a longer follow-up period to gain a deeper understanding of personality changes post-stroke.

The second study, Internal Consistency, Retest Reliability, and Their Implications for Personality Scale Validity (McCrae et al., 2011), examined the psychometric properties of personality assessments across a large sample of 34,108 individuals. The study assessed the reliability of personality facets in the NEO Inventories, comparing internal consistency and retest reliability as indicators of validity. Findings revealed that retest reliability was a better predictor of validity than internal consistency, which was useful for checking data quality but had limited utility in assessing the long-term stability of traits. The study highlighted that validity is better determined through measures such as cross-observer agreement and longitudinal stability rather than internal consistency alone. One of the key limitations of this research was its focus on the NEO Inventories, with limited exploration of other personality measures. Additionally, while the study emphasized the importance of retest reliability, further research is needed to understand its underlying determinants and how cultural and methodological factors influence reliability outcomes.

Aluja et al. (2005) carried out a study aimed at a comparison of the psychometric characteristics of three brief forms of the NEO Personality Inventory-Revised (NEO-PI-R): the initial NEO Five-Factor Inventory (NEO-FFI), the updated NEO-FFI (NEO-FFI-R), and a different 60-item form (NEO-60). The research included Swiss and Spanish samples, where 1,320 Spanish participants (mean age = 31.5 years) and 978 Swiss participants (mean age = 25.5 years) took part. The aim was to evaluate the internal consistency, factor structure, and cross-cultural equivalence of these measures. Methodologically, the researchers conducted confirmatory factor analyses and tested internal consistency with Cronbach's alpha coefficients. NEO-60 had the most optimal factor structure, and three items only failed to load

predominantly on their intended factors in both groups. Furthermore, NEO-60 had greater item-total correlations with the NEO-PI-R domain scores than the other forms. All three measures the five dimensions of personality: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. The research concluded that the NEO-60 is a valid and reliable tool for the cross-cultural assessment of the Five-Factor Model of personality (Aluja et al., 2005).

Anisi et al. (2012) carried out a study to test the reliability and validity of the NEO Five-Factor Inventory (NEO-FFI) in university students. The sample consisted of 630 students of a military corps college. Participants filled out the NEO-FFI in conjunction with the Adult Eysenck Personality Inventory to determine convergent validity. The NEO-FFI assesses five personality factors: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Internal consistency coefficients (Cronbach's alpha) for subscales were: Conscientiousness (0.83), Neuroticism (0.80), Agreeableness (0.60), Extraversion (0.58), and Openness to Experience (0.39). Convergent validity was supported by significant correlations between NEO-FFI and Eysenck Inventory subscales, i.e., Neuroticism (r = 0.68) and Extraversion (r = 0.47). The results indicate that, with the exception of the Openness to Experience subscale, the NEO-FFI has satisfactory reliability and validity for measuring personality traits in this group (Anisi et al., 2012).

Methodology

To find appropriate studies for the systematic review, the researchers employed scientific search engines such as Google Scholar, Research Gate, and Sci Hub. The following terms were used in different combinations in the keyword search: The Neo-5 personality test's validity and reliability. To make sure that just the most recent findings were included, a 15-year date limit was chosen.

The papers considered were, quantitative analysis, mixed method analyses, and an adult sample. studies published in English that only have abstracts available. The following criteria were used to exclude papers: studies of Indian/ Asian Population, basic correlation studies (such as Pearson product-moment) and meta-analyses correlation and Spearman rank order correlation), books or book chapters, theses. In addition, the authors also searched through reference lists of all identified studies to find further relevant articles. The titles and abstracts were reviewed to check whether the article was relevant or not.

The complete search method was carried out by three independent researchers (RLW, LP, and KBR). To locate related publications, the authors additionally looked through the reference lists of every study that was found. To determine if the article was appropriate, the abstracts and titles were examined.

Study Selection and Categorization

The abstract was examined to determine the studies' eligibility. The complete article was obtained and reexamined to confirm its eligibility if most of the inclusion requirements were satisfied. The article was eliminated without additional investigation if it was clear from the abstract that it was inappropriate.

Data Extraction

Year, Author, indexed in, Journal Name, Title of Study, Sample, Tools, Findings, Limitations, and References are the fields that were used to extract and implement the data into Microsoft Excel. Lastly, the process of choosing and rejecting manuscripts was illustrated using the PRISMA Flow diagram (refer to Figure 1).

Results

The extensive research was produced after an initial search of 20 studies, but not all the studies met the requirements of the study; 6 papers were excluded as it was in an Indian population study. 7 were irrelevant to the topic of the study, 2 papers were excluded as the language was a barrier to translation. The search ended on 30 March,2025. The participants were from a varied background in terms of age, sample, occupation, and location. The sample size of the studies ranged from 40 -34,000 individuals from the age range of young adults to old (17- 75 years). All included studies and their characteristics are listed in Table 1(Appendix C). All the studies are on the common aim of determining the reliability and validity of the Neo-5 Inventory

Figure1. PRISMA Flow Diagram
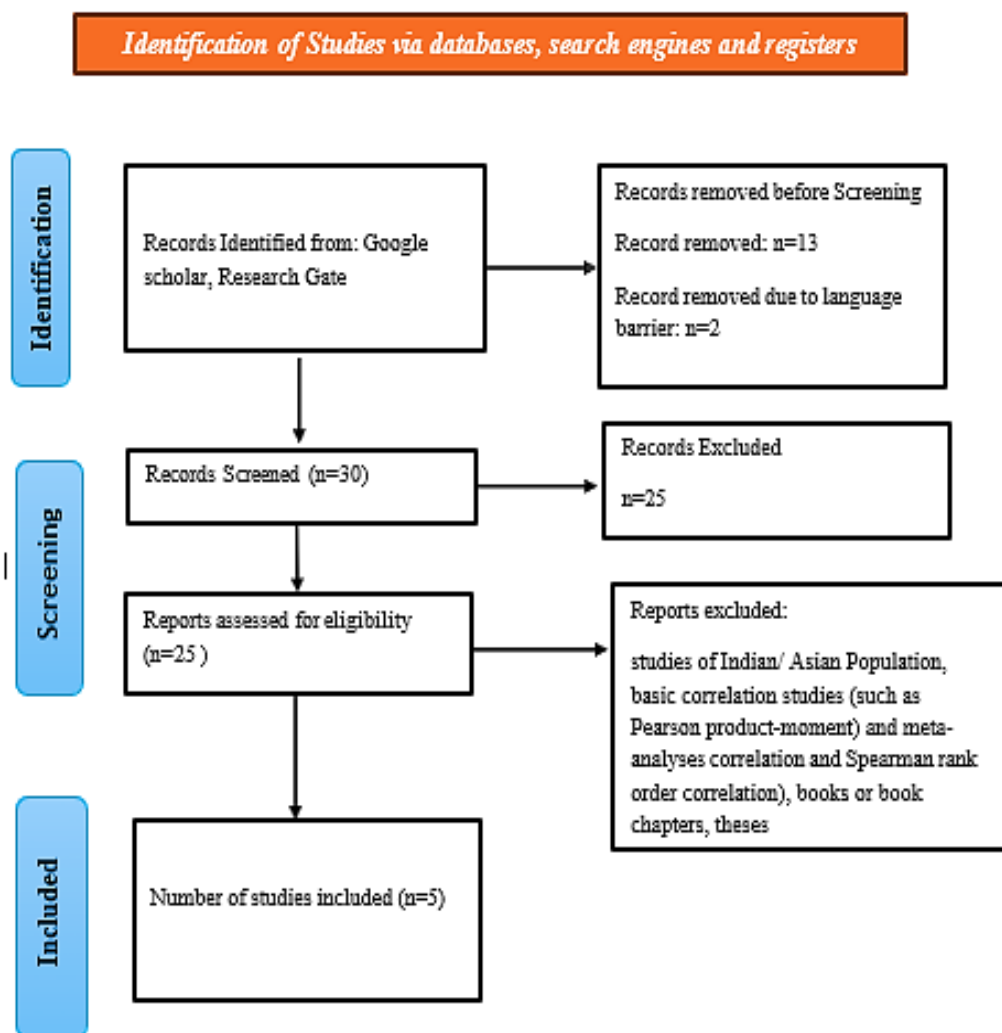
Discussion

The studies reviewed by the authors support the exceptional psychometric quality as well as multi-dimensional construct of NEO Personality Inventories. Multiple research studies conducted by McCrae et al. (2011), Aluja et al. (2005), and Anisi et al. (2012) found strong internal consistency measures together with retest reliability within Neuroticism and Conscientiousness domains of the NEO-PI-R, NEO-FFI, and NEO-60 instruments. The unreliable measure in Openness and Agreeableness domains when assessing across various populations warrants further development particularly during across-population and cross-context assessment.

The interpretation of personality scores requires analysis of contextual variables due to their impact on personality measurement results. The research conducted by Dwan et al. (2017) documented considerable changes in Extraversion and Agreeableness among stroke survivors who passed through hospital wards before settling into their homes. Data from this study disprove the theory that personality maintains a steady state and demonstrates that the NEO assessment tools are impacted by environmental elements. The findings present essential points for clinical and research work because they explain that enduring personality traits operate dynamically with environmental impacts

A major aspect Investigated by researchers involves how well the NEO instruments function across different cultures. The research by Aluja et al. (2005) proved that NEO-60 version achieved enhanced psychometric competence in Spanish and Swiss measurement contexts thus demonstrating its suitability for global testing. This discovery emphasizes the potential cultural prejudice that can exist during personality evaluations. None of the cited studies included Indian or Asian populations since the researchers wanted homogeneous samples but this shows the need for adapting NEO scales to become usable worldwide.

The conceptualization changes related to construct validity strengthen current discussions about the NEO inventories. According to McCrae et al. (2011) researchers should replace internal consistency assessments with an extensive validity model which involves secure validity (longitudinal reliability) and observer agreement (cross-observer validity) measures. The complete construct validity method produces detailed and scientifically solid personality assessment methods which better match personality's complex natural form.

The NEO Personality Inventories demonstrate strong psychometric reliability together with robust conceptual makeup according to literature review but still demand further refinement and contextual adjustments. The advancement in both theoretical framework and practical usage of widely used personality assessments depends heavily on addressing considerations regarding environmental factors together with cultural understanding and multidimensional assessment validity needs.

Inter-relatedness of themes

The examined research demonstrates a consistent relationship which exists between psychometric reliability features and personality structural elements and cross-cultural suitability while incorporating context-specific adaption techniques. These asset structures

remain highly connected within the NEO Personality Inventory assessment framework because they directly affect how the tools operate and perform. The foundation of reliability in establishing validity depends on two elements including internal consistency and retest reliability yet their practical value depends on factors like stroke events and demographic variations.

The necessity of adaptive measurement instruments arises from both cultural acceptance across all populations and diverse sample variations. Research conducted by Aluja et al. (2005) proves that Five-Factor Model global applicability requires psychometric stability across cultures yet Anisi et al. (2012) reported trait-specific instability when measuring Openness.

Environmental changes analyzed by Dwan et al. (2017) altered trait manifestation severely and provoked doubts about personality trait consistency. According to McCrae et al. (2011) the assessment of personality requires long-term evaluations supported by independent observer validation to establish deeper understanding. The NEO inventories allow researchers to assess personality using various influencing elements including methodological reliability and validity and cultural sensitivity alongside life context and trait malleability.

Limitations

All studies acknowledge limitations in generalizability due to sample size, sample type, cultural scope, or methodological focus. This study only considers published research; important unpublished information (grey literature), such as government reports, dissertations, and conference proceedings, is not included. Grey literature isn't not inherently any less significant or important than previously published works. The included studies differ in terms of demographic characteristics, methodology, and quality. Because the data is not uniform, this compromises the validity of the ensuing conclusions. Additionally, researchers may incorporate inherent bias into the selection, extraction, and analysis of studies. The authors acknowledge that although the results are logical and economical, they could not always be true in light of these restrictions.

Conclusion

Research indicates that the NEO Five-Factor Inventory with its derivatives presents excellent psychometric performance and valid measurement properties for different populations yet displays inconsistent results among Openness and Agreeableness traits.

Future expansion of these tools needs cultural flexibility and trait change comprehension for improved worldwide usefulness and precise diagnosis.

Conflicts of Interest

There is no conflict of interest among the authors.

References

Aluja, A., Garcıa, O., Rossier, J., & Garcıa, L. F. (2005). Comparison of the NEO-FFI, the NEO-FFI-R and an alternative short version of the NEO-PI-R (NEO-60) in Swiss and Spanish samples. *Personality and Individual Differences, 38*(3), 591–604.

American Psychological Association. (2025). *Personality*. American Psychological Association. https://www.apa.org/topics/personality

Anisi, J. (2012). Validity and reliability of NEO Five-Factor Inventory (NEO-FFI) on university students. *International Journal of Behavioral Sciences, 5*(4), 351–355.

Dwan, T., Ownsworth, T., Donovan, C., & Lo, A. H. Y. (2017). Reliability of the NEO Five Factor Inventory short form for assessing personality after stroke. *International*

https://www.verywellmind.com/the-big-five-personality-dimensions-2795422

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28–50.

MSEd, K. C. (2025, January 29). What are the big 5 personality traits? *Verywell Mind.*

Psychogeriatrics, 29(7), 1157–1168.

Rosellini, A. J., & Brown, T. A. (2010). The NEO Five-Factor Inventory: Latent structure and relationships with dimensions of anxiety and depressive disorders in a large clinical sample. *Assessment, 18*(1), 27–38. https://doi.org/10.1177/1073191110382848

Appendix

A. **Contributors**:

Khushi – Abstract, Introduction, Methods, PRISMA Model, Results, Discussion, Limitations, Conclusion, Formatting

Reneeta - Abstract, Introduction, Methods, PRISMA Model, Results, Discussion, Limitations, Conclusion, Formatting

Larissa - Abstract, Introduction, Methods, PRISMA Model, Results, Discussion, Limitations, Conclusion, Formatting

B. Proof

C. Studies Included in the Review

Sr. No	Year	Author	Title	Sample	Findings	Journal
1	2017	Dwan, T., Ownsworth, T., Donovan, T., & Lo, A.	Reliability of the NEO Five-Factor Inventory Short Form for Assessing Personality After Stroke	40 stroke participants and 40 matched controls (mean age = 65.55 years)	Internal consistency adequate (α = .57–.86); extraversion decreased and agreeableness increased after stroke; neuroticism correlated with mood disturbance.	*Archives of Physical Medicine and Rehabilitation*
2	2011	McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A.	Internal Consistency, Retest Reliability, and Their Implications for Personality Scale Validity	34,108 individuals	Retest reliability was a better predictor of validity than internal consistency; cross-observer agreement and longitudinal stability important for validity.	*Journal of Research in Personality*

3	2005	Aluja, A., García, Ó., & García, L. F.	Comparison of the NEO-FFI, the NEO-FFI-R and the NEO-60 in a Spanish and a Swiss Sample	1,320 Spanish participants (mean age = 31.5 years) and 978 Swiss participants (mean age = 25.5 years)	NEO-60 had optimal factor structure and higher item-total correlations; reliable for cross-cultural personality assessment.	*Personality and Individual Differences*
4	2012	Anisi, J., Majdian, M., & Piri, M	An Examination of the Validity and Reliability of the NEO Five-Factor Inventory (NEO-FFI)	630 military college students	Good convergent validity; internal consistency acceptable except for Openness (Cronbach's $\alpha = .39$).	*Iranian Journal of Psychiatry and Clinical Psychology*
5	2020	Dong Xie & Cory L. Cobb	Revised NEO Personality Inventory (NEO-PI-R)	Adults aged 17 years and older (general population)	Strong internal consistency and test-retest reliability; supports universality of the five-factor model; wide use across fields.	*The Wiley Encyclopaedia of Personality and Individual Differences: Measurement and Assessment*