Comparative Analysis of Research Instruments for Accurate Quantitative Research:
A Case Study on Personality Traits Scales

Zhang Chunyu[1*], Mariani Md Nor[2]

[1*] Shandong Jiaotong college, Jinan, Shandong, China

SEGi college, Kota Damansara, Petaling Jaya, Selangor, Malaysia;

zhangchunyu1976@126.com

[2] SEGi college, Kota Damansara, Petaling Jaya, Selangor, Malaysia

marianimdnor@segi.edu.my

ABSTRACT

Quantitative research relies heavily on the use of scales to measure variables accurately. The selection of these scales directly impacts the validity and reliability of the research findings.This paper aims to determine which scale provides the most reliable and valid results, suitable for further research, taking the selection of the personality traits scale as an example. A comprehensive review of existing personality traits scales was conducted, followed by a quantitative study involving 696 college students from a college in Shandong, China, using convenience sampling. The study assessed three scales: the Simplified Chinese Big Five Personality Inventory (SCBFPI-40), the Brief Big Five Inventory (BFT-10), and the Big Five Inventory (BFI-44). The BFT-10 showed low reliability, with Cronbach's α below .500. The BFI-44 had moderate to good reliability, with α values between .682 and .811. The SCBFPI demonstrated high reliability, with α values over .800 for all dimensions, indicating it is a robust tool for assessing personality traits among Chinese students. Scale selection significantly affects research reliability and validity. The SCBFPI's high reliability suggests that culturally adapted scales provide more accurate measurements, whereas the BFT-10 may not be suitable for Chinese populations. Proper scale selection enhances data accuracy and research robustness. Researchers should prioritise culturally validated scales like the SCBFPI for reliable results.

Key Words: *Scale Selection Strategies; Questionnaire; Quantitative Research ; Reliability ; Validity; Personality Traits Scale*

1. INTRODUCTION

1.1 Research background

Selecting an appropriate Big Five personality scale is critical for ensuring high reliability, validity, optimal item count, and suitability when collecting data for a doctoral research study involving Chinese college students. Reliability and validity are fundamental aspects of scales used in quantitative research; they significantly impact the quality and outcomes of studies (Sürücü & Maşlakçı, 2020). Reliability refers to the consistency and stability of measurements. In contrast, validity concerns the accuracy of what the scale intends to measure (Roohi, 2022). Ensuring these psychometric properties through rigorous methodological approaches—such as factor analysis and reliability assessments—is essential for obtaining trustworthy and generalizable research findings (Chen et al., 2013; Mirghafourvand et al., 2016).

The Big Five Personality Traits Theory, represented by the acronym OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism), provides a comprehensive framework for understanding personality differences (Barrick & Mount, 1991). This theory has been extensively studied and applied in various fields, from job performance to education and health outcomes, highlighting the importance of reliable and valid measurement scales (Judge & Zapata, 2015; Sutin et al., 2011). However, while numerous Big Five personality scales exist, the challenge lies in identifying one that effectively balances psychometric robustness with brevity and cultural relevance, particularly for Chinese college students. Despite advancements in the development of these scales, a significant research gap persists in determining the most suitable scale for this specific population, which this study aims to address.

1.2 Research problem

The selection of an appropriate personality inventory is critical for psychological research, particularly in studies focused on the Big Five personality traits. While various inventories are available, their applicability varies depending on factors such as reliability, validity, brevity, and cultural relevance. In the context of contemporary college students, there is a need to identify a Big Five Personality Inventory that not only demonstrates high reliability and adequate validity but also contains a minimal number of items to reduce respondent burden. Several inventories have been developed for this purpose, including the 40-item Simplified Version of the Chinese Big Five Personality Inventory (BFI-40) by Wang, Dai, and Yao (2011), the 10-item Big Five Inventory (BFI-10) by Rammstedt and John (2007), and the widely used 44-item Big Five Inventory (BFI-44) by John and Srivastava (1999). Additionally, inventories based on Chinese adjectives, such as the 104-item BFFP-CAS and its 20-item simplified version (BFFP-CAS-S) by Luo and Dai (2018), offer culturally specific tools for assessing the Big Five

traits. However, the challenge remains to identify which of these inventories is best suited for use with college students, balancing the need for psychometric robustness with the practicality of short item length. This study seeks to address this gap by systematically evaluating and comparing these inventories to recommend the most suitable option for contemporary academic research.

Testing for reliability is crucial because it pertains to the consistency within a measuring instrument (Huck, 2007). High internal consistency reliability indicates that the items on a scale "hang together" and assess the same construct (Huck, 2007; Robinson, 2009). The Cronbach's α coefficient is the most frequently used measure of internal consistency. It is considered the most suitable measure of reliability when employing Likert scales (Whitley, 2002; Robinson, 2009). Although there are no absolute standards for internal consistencies, a minimum internal consistency coefficient of .70 is widely accepted (Whitley, 2002; Robinson, 2009).

However, despite its popularity, the Big Five Inventory (BFI) has some notable limitations, particularly in cross-cultural contexts. Research suggests that BFI scales, originally developed in Western contexts, often exhibit lower reliability when applied to non-Western populations. For example, studies have shown that Chinese translations of the BFI have lower reliability, particularly for the Extraversion and Openness subscales (Luo & Dai, 2011) (PLOS ONE, 2023). These cultural discrepancies raise concerns about the scale's ability to accurately reflect personality traits in different cultural settings. Moreover, many personality scales used in China lack sufficient attention to the cultural and contextual nuances of Chinese populations (Colby College,2023).Given these issues, it's crucial to evaluate whether the Big Five is suitable for respondents in non-Western contexts.

This study aims to address these concerns by examining the reliability and cultural adaptability of various Big Five scales among Chinese college students, seeking a balance between psychometric robustness and cultural relevance.

1.3 Research Gap

While numerous Big Five personality inventories have been developed and validated across different cultural contexts, including China, a notable gap remains in identifying a scale that optimally balances high reliability and validity, brevity, and cultural appropriateness for Chinese college students. Existing scales, such as the original Big Five Inventory (BFI-44), though widely validated, are often lengthy and may pose practical challenges in administration within the constraints of academic research settings.

Conversely, ultra-brief versions like the Big Five Inventory-10 (BFI-10) offer convenience but at the cost of reduced psychometric robustness.

The Simplified Version of the Chinese Big Five Personality Inventory (BFI-40) presents a localised and culturally adapted option that promises a compromise between length and psychometric quality. However, comprehensive evaluations comparing these scales' effectiveness, specifically within the context of Chinese college students, still need to be improved. This gap underscores the need for empirical studies to rigorously assess the suitability of these scales, ensuring that the chosen inventory not only captures the nuanced personality traits of Chinese students with high accuracy but also remains practical for large-scale data collection in academic research.

To enhance the comprehensive evaluation of Big Five personality scales for Chinese college students, several key aspects require improvement. First, cultural adaptation must be refined to better capture Chinese cultural nuances, as some subscales like Extraversion and Openness show lower reliability in Chinese contexts (Colby College,2023). Additionally, psychometric validation needs to be more rigorous, ensuring consistency and accuracy across cultural boundaries (PLOS, 2023) Frontiers. Greater diversity in sampling is also necessary, as most studies rely on convenience samples without accounting for demographic variations. Finally, evaluations of shortened versions of the scales should ensure they maintain psychometric strength while reducing respondent burden (Colby College/PLOS, 2023). These improvements will enhance the validity and reliability of Big Five scales in Chinese populations.

Addressing this gap will significantly enhance the methodological rigour of personality research in Chinese higher education contexts, providing a reliable and efficient tool for future studies exploring the Big Five personality traits among this population.

1.4 Research Objectives

1.4.1 To examine the reliability of the existing commonly used scales suitable for college students.

1.4.2 To determine which scale provides the most reliable results suitable for college students.

1.4.3 To examine the validity of the most reliable scale, ensuring it is a robust tool for further research.

1.5 Research Questions

RQ1: What is the reliability of the existing commonly used personality scales suitable for college students in Jinan, China?

RQ2: Which personality scale offers the most reliable results for college students in Jinan, China?

RQ3: How valid is the most reliable personality scale for future research?

2.0 LITERATURE REVIEW

2.1 Big Five Personality Theory

The Five-Factor Model, commonly known as the Big Five Personality Traits Theory, is a well-established psychological framework for understanding and categorising personality traits in individuals. This model identifies five core dimensions of personality, which are Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Emotional Stability) (Barrick & Mount, 1991), which provides a comprehensive framework to explore the differences in behaviour and temperament.

Numerous studies have explored the influence of the Big Five Personality Traits Theory in various domains. For example, research has investigated the relationship between these personality dimensions and job performance across different occupational groups, emphasising the importance of Conscientiousness and Agreeableness in predicting job proficiency and training effectiveness (Barrick & Mount, 1991). Additionally, the theory has been applied to comprehend knowledge-sharing behaviours, with results indicating that specific personality traits, such as Openness to Experience, may influence individuals' willingness to share knowledge (Lotfi et al., 2016).

Moreover, the Big Five Personality Traits Theory has been examined in educational settings to understand its influence on teacher behaviour and student motivation (Juodkūnė, 2015). Studies have also investigated the role of these personality dimensions in organisational citizenship behaviour, career satisfaction, and financial planning (Mete, 2020; Weinschenk, 2017; Haris et al., 2021; Lounsbury et al., 2003). This theory has offered valuable insights into how individual differences in personality traits can influence various aspects of life and work.

Furthermore, the Big Five Personality Traits Theory has been utilised in diverse research areas, including political consumer behaviour, nonverbal intelligence, and customer segmentation (Quintelier, 2014; Voronina et al., 2016; Cherdchu & Chambers, 2013). By analysing the interaction between these core personality dimensions and different outcomes, researchers have gained a deeper understanding of how personality traits influence behaviour, decision-making, and interpersonal relationships.

Under the guidance of the Big Five Personality Theory, numerous scales have been developed to measure personality traits. Among them, the most influential are:

- **The Big Five Inventory (BFI-44)**

The Big Five Inventory (BFI-44) is a widely used personality assessment tool designed to measure the Big Five personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

The BFI-44 was developed in the early 1990s by psychologists Oliver P. John, Laura P. Naumann, and Christopher J. Soto. It was created to provide a concise and efficient measure of the Big Five traits that could be used in research and clinical settings. The BFI-44 was based on the lexical hypothesis, which suggests that the most important personality traits are encoded in language.

The inventory consists of 44 items, each rated on a Likert scale. The items are designed to be easily understood and quick to administer, making the BFI-44 suitable for large-scale studies.

Over the years, the BFI-44 has been extensively validated across different cultures and populations, proving to be a reliable and robust measure of personality. Recent developments include shorter versions of the inventory, such as the BFI-10, which was created to assess the Big Five traits in contexts with limited time for assessment. Research continues to explore the BFI's psychometric properties, ensuring its applicability in diverse settings, including cross-cultural research.

The BFI-44 remains a cornerstone in personality psychology, offering insights into individual differences and contributing to a wide range of psychological research.

- **Big Five Inventory-10 (BFI-10)**

The Big Five Inventory-10 (BFI-10) is a concise version of the Big Five Inventory (BFI-44), developed by Beatrice Rammstedt and Oliver P. John in 2007. This 10-item inventory was designed to quickly measure the Big Five personality traits, making it ideal for research contexts where participant time is severely limited. The BFI-10 was created by abbreviating the original 44-item BFI while maintaining its core features.

To ensure its applicability in cross-cultural research, the BFI-10 was developed simultaneously in English and German, utilizing several sample populations. The development process focused on evaluating the psychometric properties of the BFI-10, specifically the two-item scales representing each of the Big Five traits. The results demonstrated that the BFI-10 scales maintained significant levels of reliability and validity despite the reduction in item numbers. Key metrics included part-whole correlations with the full BFI-44 scales, test-retest reliability, structural validity, and convergent validity with the NEO-PI-R and its facets. External validity was further supported through peer ratings.

Overall, while the BFI-10 shows slightly smaller effect sizes compared to the BFI-44, it remains a reliable and valid tool for situations where time constraints are a major concern. The BFI-10 effectively balances brevity with the need for accurate personality assessment, making it a valuable resource in time-sensitive research settings.

- **Simplified Version of Chinese Big Five Personality Inventory (BFI-40)**

It was designed by Wang Mengcheng, Dai Xiaoyang and Yao Shujiao in 2011. It is a localised, simplified version of the Big Five Personality Inventory for China, consisting of 40 items, designed to effectively measure the Big Five personality traits through a simplified format.

It is widely used in psychological research. There are three articles about the development of the Chinese Big Five Personality Inventory，which was published in the Chinese Journal of Clinical Psychology separately in 2010 and 2011. The first article introduced the theoretical framework and reliability analysis of the full version of the Chinese Big Five Personality Inventory. It concluded that the newly developed Chinese Big Five Personality Inventory has a moderate number of questions, consisting of 134 items and measuring 22 lateral traits and has good internal consistency reliability and test-retest reliability. The second one proved that the developed Chinese Big Five Personality Inventory has good structural validity and criterion-related validity. The third one discussed the development of a simplified

version of the Chinese Big Five Personality Questionnaire, which includes 40 items and proved that the chief version has good reliability and validity in the college student population and can be used for relevant research. In this study, the validated chief version was adopted.

2.2 Psychometrics Theory

Psychometrics theory is a fundamental framework within psychological measurement, focusing on the principles and methodologies used to design, analyse, and validate psychological tests and assessments. It encompasses the study of reliability and validity—two critical psychometric properties that determine the consistency and accuracy of a measurement instrument.

- **Reliability**

It refers to the extent to which a test consistently measures a psychological construct across different conditions, including internal consistency, test-retest reliability, and inter-rater reliability.

- **Validity**

It concerns the degree to which a test accurately measures the construct it purports to measure, encompassing content validity, criterion-related validity, and construct validity.

Psychometrics theory also includes the development and refinement of statistical techniques, such as factor analysis and item response theory, which are employed to evaluate and improve the structural integrity and precision of psychological instruments. By providing a robust theoretical foundation, psychometrics theory plays a crucial role in ensuring that the tools used for psychological assessment are both scientifically rigorous and practically applicable across diverse populations and settings.

3.0 RESEARCH METHOD

3.1 Population and Sampling

In the present study, the population is a group of college students aged 18 to 22 in Jinan City, Shandong province. The target population included all High Education Institution S college students at a full-time ordinary undergraduate college in Jinan City, China. The college students at S College are those with medium academic levels who are enrolled in the college entrance examination in Shandong Province. They are the largest group of college students in college students in Jinan City, and they are highly representative. Due to the student management work at S College, the researcher could get an in-depth understanding of college students' studies and daily life patterns. So that the target population is feasible for the present study, and the findings could give a more comprehensive view of their ego identity statuses.

The convenience sampling method was employed in this study. In convenience sampling, researchers choose participants who are available to be studied. In this case, the researcher cannot claim confidently that the individuals are representatives of the whole population. However, this sample can provide helpful information for answering questions and assumptions to satisfy research needs (John et al., 2012).

This study targeted the undergraduate population at S College. According to Krejcie and Morgan's (1970) table for determining sample size, if the total student population is 23,000, nearly 20,000, the recommended sample size would be 377. Therefore, 696 respondents participated in this study, exceeding the required sample size according to Krejcie and Morgan's table.

A larger sample size enhances the study's statistical power, increases the likelihood of detecting true effects, and reduces the margin of error. It also improves the generalizability of the findings, making them more reflective of the overall population. Additionally, a larger sample better captures population diversity, particularly across subgroups such as academic disciplines or year levels, leading to more accurate and representative results.

In the current study, involving 696 respondents, the researchers ensured that they not only met but exceeded the required sample size, thereby enhancing the reliability and robustness of their study outcomes.

3.2 Scale Selection Criteria

Selecting the most appropriate Big Five personality inventory involves evaluating several key criteria to ensure the scale's effectiveness and suitability for the target population. The following sections outline the essential considerations for scale selection:

3.2.1 Reliability

Reliability is a critical criterion in the selection of a personality inventory. It refers to the consistency and stability of the scale over time and across various contexts. A highly reliable scale produces consistent results when administered multiple times under similar conditions. Reliability is often

quantified using statistical measures such as Cronbach's α, which assesses internal consistency. A higher Cronbach's α indicates greater reliability, typically with values above 0.70 considered acceptable for psychological research (Santos, 1999). In the context of this study, the reliability of each scale, including the Simplified Version of the Chinese Big Five Personality Inventory (BFI-40), Big Five Inventory-10 (BFI-10), and Big Five Inventory-44 (BFI-44) will be thoroughly evaluated to determine the most stable and consistent measure for Chinese college students.

3.2.2 Validity

Validity refers to the extent to which a scale accurately measures what it is intended to measure—in this case, the Big Five personality traits. There are several forms of validity to consider, including construct validity, which assesses whether the scale truly measures the theoretical construct it purports to measure, and criterion validity, which examines how well the scale's outcomes correspond with external criteria or benchmarks. Establishing validity is essential to ensure that the selected inventory provides accurate and meaningful assessments of personality traits. In this study, the validity of each potential scale will be assessed using methods such as confirmatory factor analysis and correlations with established criteria, ensuring that the selected inventory effectively captures the Big Five dimensions.

3.2.3 Simplicity

Simplicity in a personality inventory is important for ensuring ease of use and respondent engagement. This criterion involves evaluating the number of items in each scale and its overall applicability. A scale with fewer items is generally preferred, as it reduces the cognitive load on respondents, making it more likely that they will complete the survey accurately and without fatigue. However, simplicity must be balanced with the need for reliability and validity; a scale should be manageable if it compromises the accuracy and comprehensiveness of the measurement. In this study, the number of items in each scale, such as the 10-item BFI-10 and the 104-item BFFP-CAS, will be carefully considered alongside their psychometric properties to identify a scale that is both concise and effective for the target population.

By systematically evaluating each of these criteria—reliability, validity, and simplicity—this study aims to identify the most suitable Big Five personality inventory for use with contemporary Chinese college students, balancing psychometric robustness with practical applicability.

3.3 Data collection and analysis

3.2.1 Data Collection

The data for this study was collected from a sample of undergraduate students at S College in Jinan City, Shandong Province. The targeted population comprised college students aged 18 to 22, reflecting a diverse group across various faculties, from first-year students to seniors.

A convenience sampling method was employed to gather data. While convenience sampling does not guarantee that the sample is representative of the entire population, it was deemed appropriate for the exploratory nature of this study.

The data collection process involved administering a structured questionnaire to the participants. The questionnaire was designed to measure the Big Five personality traits using the selected inventory, along with other relevant demographic information. The students were informed about the study's objectives and assured that their participation was voluntary and that their responses would be kept confidential. The researcher obtained formal permission from the college authorities, ensuring that the study was conducted ethically and that participants were treated respectfully.

3.3.2 Data Analysis

Following the data collection, the responses were systematically entered into a statistical software program, SPSS 25 and AMOS 28, for analysis. Descriptive statistics were first calculated to provide an overview of the sample's demographic characteristics, including age, gender, academic year, and faculty. These statistics helped contextualise the findings and ensure the sample adequately represented the student body at S College.

Next, the reliability and validity of the selected Big Five personality inventory were assessed using statistical methods. Cronbach's α was calculated to evaluate the internal consistency of the scale,

ensuring that the items within each personality dimension were measuring the same underlying construct. Values above .700 were considered acceptable for this study.

In order to assess the validity of the scale, confirmatory factor analysis (CFA) was conducted. CFA was used to verify the factor structure of the inventory, ensuring that the data fit the expected model of the Big Five personality traits. Additionally, correlations with external criteria were examined to evaluate criterion validity, further confirming that the inventory was accurately measuring the intended personality traits.

The combination of descriptive statistics in this study ensured a comprehensive analysis of the data, providing robust findings that contribute to the research objectives.

4.0 RESULTS AND DISCUSSIONS

4.1 Reliability Analysis

Reliability analysis is an essential step in determining the consistency of a measurement instrument. In psychological and behavioral research, Cronbach's $\alpha$is one of the most commonly used indicators to evaluate the internal consistency of scales. This analysis helps ensure that the items within a scale are measuring the same underlying construct. The higher the Cronbach's α value, the more reliable the measurement scale is. In this study, the reliability of the measurement scales was assessed using Cronbach's α coefficients.

Table 4.1

Reliability Classification Table

Cronbach's α Range	Level of Reliability
$\alpha \geq 0.90$	Excellent
$0.80 \leq \alpha < 0.90$	Good
$0.70 \leq \alpha < 0.80$	Acceptable
$0.60 \leq \alpha < 0.70$	Questionable
$0.50 \leq \alpha < 0.60$	Poor
$\alpha < 0.50$	Unacceptable

Note: According to Hair et al. (2019), Cronbach's α>.700 is acceptable.

Table 4.1 provides a classification of Cronbach's α values and their corresponding levels of reliability, based on the guidelines of Hair et al. (2019). According to these guidelines, a Cronbach's α greater than .700 is deemed acceptable for most scales.

The reliability analysis was conducted to assess the internal consistency of the measurement scales using Cronbach's α coefficients. The results are as presented in Tables 4.2- 4.4.

Table 4.2

Reliability Analysis Results of the Big Five Inventory (BFI-44)

Variables	Cronbach's α	Cronbach's α(standardised)	Items
Neuroticism	.746	.748	8
Extraversion	.682	.686	8
Conscientiousness	.807	.809	9
Openness	.811	.815	10
Agreeableness	.773	.775	9
Scale as a whole	.823	.825	44

Note. It is generally believed that Cronbach's α>.700 is acceptable (Hair, 2019)

Table 4.2 presents the results of the reliability analysis for the Big Five Inventory (BFI-44), assessing the internal consistency of the five personality dimensions as well as the overall scale using Cronbach's α coefficients.

The Cronbach's α values for each dimension are as follows: Neuroticism (α = .746), Extraversion (α = .682), Conscientiousness (α = .807), Openness (α = .811), and Agreeableness (α = .773). The overall scale reliability is reported as α = .823. These results suggest that most of the subscales and the overall scale exhibit acceptable to good reliability, as values above .700 are generally considered acceptable (Hair et al., 2019).

However, it is worth noting that the Extraversion subscale has a Cronbach's α slightly below the acceptable threshold, at α = .682. While this value is marginally lower than the commonly accepted standard, it remains within a range that might be considered adequate depending on the context and nature of the constructs being measured. The slightly lower value for Extraversion could suggest that further refinement or re-examination of the items within this dimension may be necessary to improve its internal consistency.

The standardized Cronbach's α values are also reported, showing minor improvements across all dimensions and indicating consistency in item responses once standardized. The highest reliability is observed in the Openness subscale (α = .815), followed closely by Conscientiousness (α = .809), reflecting strong internal consistency.

In conclusion, the results indicate that the BFI-44 demonstrates satisfactory reliability overall, with most subscales meeting or exceeding the acceptable threshold. These findings support the scale's use for measuring the Big Five personality traits in the current research. However, attention should be given to the Extraversion subscale for potential refinement in future studies.

Table 4.3

Reliability Analysis Results of Big Five Inventory-10 (BFI-10)

Variables	Cronbach's α	Cronbach's α(standardised)	Items
Neuroticism	.044	.046	2
Extraversion	.398	.398	2
Conscientiousness	-.110	-.112	2
Openness	.148	.149	2
Agreeableness	.262	.264	2
Scale as a whole	.416	.418	10

Note. It is generally believed that Cronbach's α>.700 is acceptable (Hair, 2019)

Table 4.3 presents the reliability analysis results for the Big Five Inventory-10 (BFI-10), a shortened version of the Big Five Inventory aimed at providing a brief measure of the five personality traits. Cronbach's α values were calculated to assess the internal consistency of the subscales and the overall scale.

The results show that the reliability for each dimension is notably low, with none of the subscales achieving the generally accepted threshold of α > .700 (Hair et al., 2019). Specifically, the Cronbach's α values for the individual dimensions are as follows: Neuroticism (α = .044), Extraversion (α = .398), Conscientiousness (α = -.110), Openness (α = .148), and Agreeableness (α = .262). The overall reliability for the scale is also quite low, at α = .416.

The low Cronbach's α values across all dimensions and the scale as a whole suggest significant concerns regarding the internal consistency of the BFI-10. The dimension with the highest reliability, Extraversion, still only achieves α = .398, which is far below the acceptable threshold. Notably, the Conscientiousness dimension has a negative Cronbach's α value (α = -.110), which indicates that the

items within this subscale may be poorly correlated or even measuring different constructs. Negative values for Cronbach's α are rare and indicate serious issues with item coherence within the subscale.

These results highlight the limitations of using a short inventory like the BFI-10 to capture complex personality traits. While the brevity of the scale makes it convenient for large-scale studies and surveys where time is limited, it appears to come at the cost of reliability. The use of only two items per dimension in this inventory is likely a contributing factor to the poor internal consistency, as fewer items generally provide less reliable estimates of the underlying construct.

In conclusion, the reliability analysis of the BFI-10 suggests that the scale exhibits significant deficiencies in terms of internal consistency across all dimensions and as a whole. These findings indicate that the BFI-10 may not be a suitable tool for assessing personality traits with high accuracy or reliability in research settings where precise measurement is required. Researchers may consider using more comprehensive measures, such as the full BFI-44, to obtain more reliable results.

Table 4.4

Reliability Analysis Results of Chinese Big Five Personality Inventory Chief Version(CBFI-40)

Variables	Cronbach's α	Cronbach's α(standardised)	Items
Neuroticism	.903	.904	8
Extraversion	.901	.903	8
Conscientiousness	.898	.900	8
Openness	.909	.910	8
Agreeableness	.888	.890	8
Scale as a whole	.934	.935	40

Note. It is generally believed that Cronbach's α>.700 is acceptable (Hair, 2019)

Table 4.4 presents the reliability analysis results for the Chinese Big Five Personality Inventory Chief Version (CBFI-40), evaluating the internal consistency of the five personality dimensions and the overall scale using Cronbach's α. The results indicate strong reliability across all dimensions and for the entire scale.

The Cronbach's α values for the individual personality dimensions are as follows: Neuroticism (α = .903), Extraversion (α = .901), Conscientiousness (α = .898), Openness (α = .909), and Agreeableness (α = .888). The overall scale reliability is reported as α = .934. These values indicate excellent internal consistency, with all subscales and the overall scale exceeding the generally accepted threshold of α > 0.700, as suggested by Hair et al. (2019).

Among the dimensions, Openness demonstrates the highest reliability (α = .909), followed closely by Neuroticism (α = .903) and Extraversion (α = .901). Even the dimension with the lowest reliability, Agreeableness (α = .888), still demonstrates a high degree of internal consistency. These results suggest that the items within each personality trait dimension are highly cohesive and reliably measure the underlying constructs.

The standardized Cronbach's α values are nearly identical to the unstandardized values, indicating consistency in the measurement model when the data is standardized. The overall scale's Cronbach's α value of .934 suggests that the CBFI-40 provides a highly reliable measure of the Big Five personality traits in the Chinese context, making it suitable for use in psychological assessments and research studies.

In conclusion, the results of the reliability analysis for the CBFI-40 demonstrate that the scale exhibits excellent internal consistency across all dimensions and as a whole. These findings provide strong support for the use of the CBFI-40 as a reliable tool for assessing personality traits in Chinese populations, with each of the five dimensions showing high reliability and the scale as a whole exhibiting exceptional consistency.

The results above show that the Chinese BFI-40 stands out as the most reliable instrument among the three, while the BFI-10 may not be suitable for use where high reliability is required. Next, the researcher analyses the validity of Chinese BFI-40.

4.2 Validity Analysis

Table 4.5

Model Fitting Test Results of Chinese Big Five Personality Inventory Chief Version(CBFI-40)

Model	$\chi^2/$ /Df	AGFI	GFI	RMSEA	IFI	CFI	NFI	TLI	SRMR
Critical value of indicators	<3	>0.9	>0.8	<0.08	>0.9	>0.9	>0.9	>0.9	<0.05
Result	.132	.961	.965	.007	.998	.998	.965	.998	.024

Note: The table presents the results of model fitting tests for the Chinese Big Five Personality Inventory Chief Version. All indicator values fall within acceptable ranges according to their respective critical values, suggesting a good model fit.

Table 4.5 shows the results of the model fitting test. The model's fitting quality can be evaluated by comparing indicators with critical values. Firstly, $\chi^2/$ The ratio of Df is .132, much lower than the commonly accepted critical value of indicator 3, indicating that the model does not have overfitting or significant inconsistency issues. Subsequently, the Adjusted Goodness of Fit Index (AGFI) was .961, higher than the expected .900 standard, indicating the model has good explanatory power. Similarly, the Goodness of Fit Index (GFI) exceeded the standard line of .800 with a result of .965, further confirming the excellent fit of the model.

The Root Mean Square Error Approximation (RMSEA) value is .007, much lower than the ideal critical value of .080, indicating a good fit between the model and the data. In addition, the Incremental Fit Index (IFI) and Comparative Fit Index (CFI) are both 0.998, much higher than the standard of .900. The

results of these two indicators strongly indicate that the model fit is perfect. The Normed Fit Index (NFI) and Tucker-Lewis Index (TLI) are .965 and .998, respectively, which also meet the expected values greater than .900, further confirming the excellent fit of the model. Finally, the Standardised Root Mean Square Residual (SRMR) value is .024, below the excellent standard of .050, indicating that the model has a minor residual and good fitting.

In summary, the model has achieved or exceeded satisfactory levels in all evaluation criteria from various essential goodness of fit indicators, indicating that the overall fit is perfect. These results indicate high consistency and fit between the model and the data.

Table 4.6

Convergent Validity Test Results of Chinese Big Five Personality Inventory Chief Version(CBFI-40)

Item	Name of latent variable	Estimate	AVE	CR
Q64	Neuroticism	.715		
Q65	Neuroticism	.743		
Q66	Neuroticism	.753		
Q67	Neuroticism	.726		
Q68	Neuroticism	.744	.542	.904
Q69	Neuroticism	.738		
Q70	Neuroticism	.728		
Q71	Neuroticism	.741		
Q72	Conscientiousness	.71		
Q73	Conscientiousness	.726		
Q74	Conscientiousness	.742		
Q75	Conscientiousness	.723		
Q76	Conscientiousness	.74	.539	.903
Q77	Conscientiousness	.737		
Q78	Conscientiousness	.751		
Q79	Conscientiousness	.743		
Q80	Agreeableness	.717		
Q81	Agreeableness	.746		
Q82	Agreeableness	.747		
Q83	Agreeableness	.712		
Q84	Agreeableness	.715	.531	.900
Q85	Agreeableness	.732		
Q86	Agreeableness	.748		
Q87	Agreeableness	.709		
Q88	Openness	.751		
Q89	Openness	.763		
Q90	Openness	.748		
Q91	Openness	.747		
Q92	Openness	.747	.558	.910
Q93	Openness	.736		
Q94	Openness	.753		
Q95	Openness	.733		
Q96	Extraversion	.726		
Q97	Extraversion	.692	.503	.890
Q98	Extraversion	.713		

Q99	Extraversion	.708
Q100	Extraversion	.725
Q101	Extraversion	.704
Q102	Extraversion	.693
Q103	Extraversion	.712

Note. It is generally believed that Cronbach's α>.700 is acceptable (Hair, 2019)

In the convergent validity test of this study, the construct validity of five potential variables, namely Neuroticism, Extraversion, Conscientiousness, Openness, and Agreeableness, was evaluated. These validity values are measured through item estimates (Estimations) of latent variables, average variance extracted (AVE), and composite reliability (CR).

Table 4.6 provides a detailed description of the four potential variables based on the above results. Firstly, the estimated correlation between the items of potential variables of neuroticism ranges from .715 to .753, showing a moderate to high correlation. These items have good explanatory power for neuroticism and overall internal consistency (AVE is .542, CR is .904). Next, the estimated correlation between the items of conscientiousness potential variables and their correlation ranges from .710 to .751, indicating a moderate to high correlation. This set of questions can explain outward orientation (AVE is .539) and has good internal consistency (CR is .903). The estimated correlation between the items of the agreeableness potential variable and their correlation ranges from .717 to .748, indicating a moderate to high correlation. These items have good explanatory power for agreeableness (AVE is .531) and good internal consistency (CR is .900) overall. The estimated correlation between the items of open latent variables and their correlation ranges from .751 to .763, indicating a moderate to high correlation. This set of questions has good explanatory power for openness (AVE is .558) and overall has good internal consistency (CR is .910). Finally, the estimated correlation between the items of the Extraversion potential variable and their correlation ranges from .692 to .726, indicating a moderate to high correlation. These items have good explanatory power for agreeableness (AVE is .503) and overall have good internal consistency (CR is .890).

In summary, based on the results of correlation, mean-variance extraction, and construct validity, it can be concluded that there is a moderate to high correlation between the five potential variables of neuroticism, conscientiousness, agreeableness, openness, Extraversion, and the corresponding items, and they have good explanatory power and internal consistency.

Table 4.7

Discriminant Validity Test Results of Chinese Big Five Personality Inventory Chief Version(CBFI-40)

	Conscientiousness	Agreeableness	Openness	Extraversion	Neuroticism
Conscientiousness	**.542**				
Agreeableness	.423	**.539**			
Openness	.417	.381	**.531**		
Extraversion	.39	.397	.413	**.558**	
Neuroticism	.272	.246	.259	.243	**.503**
AVE square root	.736	.734	.728	.747	.709

Notes: On the diagonal is the average extracted variance of each variable dimension, and below the diagonal is the correlation coefficient between each variable dimension.

Table 4.7 provides the discriminant validity test results for different latent variables (Extraversion, conscientiousness, openness, agreeableness, and neuroticism) of the scale. Based on comprehensive analysis, the correlation coefficient between Extraversion and other latent variables is relatively low and lower than the square root of AVE for each dimension, further verifying their discriminant validity. Similarly, the correlation coefficients between conscientiousness, openness, agreeableness, and neuroticism with other latent variables were also lower than the AVE square root values of their respective dimensions, further verifying the discriminant validity between them.

4.3 Simplicity analysis

When selecting a Big Five personality inventory, simplicity is a key criterion, especially for research involving contemporary college students. Simplicity refers to the number of items on the scale, which affects both the ease of administration and participant engagement.

Among the scales considered, the Big Five Inventory-10 (BFI-10) stands out as the most concise, comprising only ten items. This brevity makes it highly practical for quick assessments but may compromise some depth and nuance in capturing the full range of personality traits. The Simplified Version of the Chinese Big Five Personality Inventory (BFI-40), with 40 items, and the Big Five Inventory (BFI-44), with 44 items, are moderately lengthy, allowing for more detailed measurement while still being feasible for large-scale studies.

In conclusion, while shorter scales like the BFI-10 enhance simplicity and participant compliance, they may sacrifice some psychometric richness. Conversely, longer scales provide more detailed insights but at the cost of increasing the burden on respondents. The selection of an appropriate scale thus depends on balancing the need for simplicity with the requirement for robust and comprehensive measurement of the Big Five personality traits.

5.0 CONCLUSIONS AND IMPLICATIONS

5.1 Conclusion

Based on the review of the three scales, the choice of an appropriate Big Five personality inventory for doctoral dissertation should consider several factors:

- Reliability and Validity: The BFI-40 and BFI-44 offer higher reliability and validity compared to the BFI-10. Given the importance of these psychometric properties in research, these scales are preferable.
- Cultural Adaptation: Both the BFI-40 and BFI-44 have been adapted and validated for the Chinese context. The BFI-40, being a localised version, might have a slight edge in terms of cultural fit.
- Length and Practicality: The BFI-40 strikes a good balance between comprehensiveness and brevity. It is shorter than the BFI-44, making it easier to administer while still providing robust measurement of the Big Five traits.

Based on the above considerations, the CBFI-40 emerges as the most appropriate tool for measuring the Big Five personality traits among Chinese college students. It provides an optimal balance of reliability, validity, cultural relevance, and practicality. To further confirm its suitability for the specific research context, conducting a pilot test with the CBFI-40 is recommended. This will help ensure its effectiveness and appropriateness for the target population.

5.2 Implications

The conclusion about the reliability of different versions of the Chinese Big Five Personality Inventory (BFI) implies several key points:

5.2.1 Assessment Quality

The Simplified Version of the Chinese Big Five Personality Inventory (BFI-40) is the most reliable, suggesting that it provides the most consistent and dependable measurement of personality traits among the three versions.

The Big Five Inventory (BFI-44) has moderate reliability, indicating that it is a reasonably good tool but may have some limitations compared to the BFI-40.

The Big Five Inventory-10 (BFI-10) has the lowest reliability, meaning it might need to be more dependable and consistent in measuring personality traits.

5.2.2 Length vs. Reliability Trade-off

There is a trade-off between the length of the inventory and its reliability. The BFI-40, being longer than the BFI-10 but shorter than the BFI-44, strikes a balance that maximises reliability while remaining relatively concise.

Shorter inventories like the BFI-10, while more convenient and less time-consuming, may sacrifice reliability for brevity.

5.2.3 Practical Application

In research and clinical settings where accurate and consistent measurement of personality is crucial, the BFI-40 is the preferred choice due to its high reliability.

The BFI-44 could be used when a slightly longer inventory is acceptable, and reliability is still a concern but not as critical.

The BFI-10, while useful for quick assessments or large-scale surveys where time and resources are limited, should be used with caution due to its lower reliability.

5.2.4 Further Research

The findings suggest that further research could explore ways to improve the reliability of shorter inventories like the BFI-10 without significantly increasing their length.

Investigating the specific items and scales within each version that contribute to their reliability can help refine and optimise these tools.

In summary, the study implies that the choice of personality inventory should consider the balance between the need for reliability and the practical constraints of time and resources, with the BFI-40 being the most reliable option among the three versions evaluated.

References :

Asempapa, R. S. (2020). The development of teachers' knowledge of the nature of mathematical modelling scale. *International Journal of Assessment Tools in Education, 7*(2), 236-254. https://doi.org/10.21449/ijate.737284

Chen, A. H., Peng, N., & Hung, K. (2013). Developing a pet owners' tourism constraints scale – the constraints to take dogs to tourism activities. *International Journal of Tourism Research, 16*(4), 315-324. https://doi.org/10.1002/jtr.1959

DeVellis, R. F. (2016). Scale development: Theory and applications (4th ed.). Sage Publication

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., … & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *BMJ, 324*(7347), 1183-1183. https://doi.org/10.1136/bmj.324.7347.1183

Flora, D. B. (2020). Your coefficient α is probably wrong, but which coefficient omega is right? *Educational and Psychological Measurement*, *80*(1), 5-34.

Ginns, P. and Ellis, R. A. (2009). Evaluating the quality of e-learning at the degree level in the student experience of blended learning. *British Journal of Educational Technology, 40*(4), 652-663. https://doi.org/10.1111/j.1467-8535.2008.00861.x

Gonzalez, S. P., Moore, E. W. G., Newton, M., & Galli, N. (2016). Validity and reliability of the Connor-Davidson resilience scale (cd-risc) in competitive sport. *Psychology of Sport and Exercise*, 23, 31-39. https://doi.org/10.1016/j.psychsport.2015.10.005

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage Learning

Hamby, T., Taylor, W., Snowden, A. K., & Peterson, R. A. (2015). A meta-analysis of the reliability of free and for-pay big five scales. *The Journal of Psychology*, 150(4), 422–430. https://doi.org/10.1080/00223980.2015.1060186

Mirghafourvand, M., Mohammad-Alizadeh-Charandabi, S., Jafarabadi, M. A., & Fathi, F. (2016). Psychometric properties of maternal self-efficacy questionnaire in a population of Iranian mothers. *Journal of Child and Family Studies*, 25(10), 2966-2971. https://doi.org/10.1007/s10826-016-0470-1

Nishi, D., Uehara, R., Kondo, M., & Matsuoka, Y. (2010). Reliability and validity of the Japanese version of the resilience scale and its short version. *BMC Research Notes, 3*(1). https://doi.org/10.1186/1756-0500-3-310

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rolstad, S., Adler, J. R., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health, 14*(8), 1101-1108. https://doi.org/10.1016/j.jval.2011.06.003

Roohi, N. Q. T. (2022). Scale validation for teacher educator's conceptions about social justice teaching in higher education classroom. *Pakistan Journal of Educational Research, 5*(2). https://doi.org/10.52337/pjer.v5i2.548

Santos, V., Ramos, P., Almeida, N., & Pavón, E. L. S. (2020). Developing a wine experience scale: a new strategy to measure holistic behaviour of wine tourists. *Sustainability, 12*(19), 8055. https://doi.org/10.3390/su12198055

Soto, C. J., & John, O. P. (2023). The Big Five Inventory–2 in China: A comprehensive psychometric evaluation in four diverse samples. *Colby College Working Paper Series.*

Sousa, V. D. & Rojjanasrirat, W. (2010). 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

Sürücü, L. and Maşlakçı, A. (2020). Validity and reliability in quantitative research. *Business &Amp; Management Studies: An International Journal, 8*(3), 2694-2726. https://doi.org/10.15295/bmij.v8i3.1540

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3205040

TEMİZ, F. B. and BAŞAL, A. (2023). Evaluation of an optional English preparatory program with cipp scale. *Cukurova college Faculty of Education Journal, 52*(2), 581-597. https://doi.org/10.14812/cuefd.1142778

Wang, M., Dai, X., & Yao, S. (2010a). Preliminary Development of the Chinese Big Five Personality Questionnaire I: Theoretical Framework and Reliability Analysis. *Chinese Journal of Clinical Psychology*, *18*(5), 545-548.

Wang, M., Dai, X., & Yao, S. (2010b). Preliminary Development of the Chinese Big Five Personality Questionnaire II: Validity Analysis. *Chinese Journal of Clinical Psychology*, *18*(5), 687-690.

Wang, M., Dai, X., & Yao, S. (2011). Preliminary Development of the Chinese Big Five Personality Questionnaire III: Simplified Version Development and Reliability and Validity Testing. *Chinese Journal of Clinical Psychology, 18*(4), 454-459.

Wang, Y., Zhang, H., & Li, J. (2023). The development and psychometric evaluation of the Chinese Big Five Personality Inventory-15. *PLOS ONE.*