JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)
DOI: http://dx.doi.org/10.15408/jp3i.v14i2.31689
http://journal.uinjkt.ac.id/index.php/jp3i

Investigating the Factor Structure of the Myers-Briggs Type Indicator (MBTI): Case of Indonesian Students

Dwi Yan Nugraha[1,2,3], Widyastuti[1], Ahmad Ridfah[1]

Faculty of Psychology, Universitas Negeri Makassar, Indonesia[1]
Faculty of Psychology, Universitas Gadjah Mada, Indonesia[2]
Faculty of Psychology, Universitas Islam Indonesia, Indonesia[3]

dwiyannugraha20@gmail.com

Abstract

Psychological tests require continuous refinement and evaluation to ensure their effectiveness. This study aimed to evaluate the factor structure, invariance, item quality, and differential item functioning (DIF) of the 60-item Myers–Briggs Type Indicator (MBTI) among Indonesian students using modern psychometric methods. Involving 7,526 participants, the results of the item factor analysis (IFA) indicated that a single-factor model for each MBTI dimension adequately fit the data, supporting satisfactory construct validity. The Infit and Outfit MNSQ values ranged between 0.5 and 1.5, demonstrating good item quality. Moreover, no gender bias was detected based on the DIF Contrast effect size, indicating that MBTI items function equivalently for male and female students. These findings provide strong empirical evidence for the psychometric validity and reliability of the MBTI in the Indonesian context and represent the first large-scale study contributing to the refinement and modernization of the instrument in alignment with national legislative standards for psychological test use.

Keywords: item factor analysis, rasch measurement model, myers-briggs type indicator test, students.

Abstrak

Tes psikologi memerlukan penyempurnaan dan evaluasi berkelanjutan untuk memastikan efektivitas penggunaannya. Penelitian ini bertujuan mengevaluasi struktur faktor, invariansi, kualitas item, dan differential item functioning (DIF) dari Myers-Briggs Type Indicator (MBTI) versi 60 item pada siswa di Indonesia menggunakan pendekatan psikometrik modern. Dengan melibatkan 7.526 siswa, hasil item factor analysis (IFA) menunjukkan bahwa model satu faktor untuk setiap dimensi MBTI sesuai dengan data, menegaskan validitas konstruk yang memadai. Nilai infit dan outfit MNSQ berada dalam rentang 0,5–1,5, menandakan kualitas item yang baik. Selain itu, tidak ditemukan bias antarjenis kelamin berdasarkan DIF Contrast, sehingga item MBTI bersifat setara bagi siswa laki-laki dan perempuan. Temuan ini memberikan bukti empiris kuat atas validitas dan reliabilitas psikometrik MBTI dalam konteks Indonesia serta menjadi studi berskala besar pertama yang mendukung pembaruan instrumen sesuai standar penggunaan tes psikologis nasional.

Kata kunci: *item factor analysis, rasch measurement model, myers-briggs type indicator test, siswa.*

Introduction

Education in Indonesia continues to advance, and so does the use of psychological tests to understand individual personality. Kaplan and Saccuzzo (2018) divide psychological tests into ability and personality tests, with the latter designed to reveal personality structures through projective or non-projective approaches. One of the most widely used instruments is the Myers-Briggs Type Indicator (MBTI) (van Zyl & Tylor, 2012; Stein & Swan, 2019), which has also gained popularity in Indonesia (Naisaban, 2003).

However, despite extensive international research on MBTI's psychometric functions (Harvey et al., 1995; Myers et al., 1998), studies in Indonesia remain limited (Periantalo & Azwar, 2017; Susanto & Mudaim, 2017). Previous studies in the Indonesian context tended to utilize the older MBTI versions and relied on classical methods. Therefore, there is an urgency to evaluate the MBTI factor structure using modern psychometric approaches to address this issue and strengthen its applicability in Indonesia's educational and cultural contexts.

Modern psychometrics offers three main measurement traditions: Classical Test Theory (CTT), Rasch Measurement Theory (RMT), and Structural Equation Modeling (SEM) (Engelhard & Wind, 2021). Although CTT remains common, it suffers from sample dependence (Kaplan & Saccuzzo, 2009). In contrast, RMT ensures invariance, thereby allowing unbiased person–item comparisons (Andrich & Marais, 2019). Within the structural approach, confirmatory factor analysis has increasingly been replaced by item factor analysis (IFA), equivalent to IRT 2-PL (Wirth & Edwards, 2007)—because it better accommodates ordinal and dichotomous data (Rhemtulla et al., 2012). Meanwhile, the Rasch model (Rasch, 1960) has gained prominence for its robustness in construct validation, item calibration, and bias detection (Hayat et al., 2021; DiStefano et al., 2019). This model, rooted in latent trait theory, is flexible for diverse response formats (Suryadi et al., 2020; Padgett & Morgan, 2020) and widely recognized as an influential psychometric approach (Aryadoust et al., 2019; Edelsbrunner & Dablander, 2019).

Moreover, the Rasch model enables rigorous examination of item functioning across groups and cultures (Eid & Rauber, 2000; De Jong et al., 2008) and provides a foundation for differential item functioning (DIF) analysis (Osterlind, 1983; Wu et al., 2016; Temel et al., 2022). In multicultural contexts such as Indonesia, DIF analysis ensures fairness by detecting potential gender or cultural bias (Baylor et al., 2014; Saggino et al., 2020). The unique psychometric advantages of the Rasch model—specific objectivity, additivity, and parameter separation—make it an ideal method to confirm MBTI's fairness and validity (Mesbah & Kreiner, 2012; Wu et al., 2016).

To our knowledge, this study is the first to evaluate the MBTI's psychometric properties in the Indonesian context using both Item Factor Analysis and Rasch Measurement Theory. By integrating these modern psychometric frameworks, this research advances the methodological rigor in personality assessment and provides cross-cultural evidence for the MBTI's validity and fairness. Therefore, this study aims to (1) evaluate the factor structure of the MBTI using IFA, (2) assess item quality through the Rasch model, and (3) detect gender-based DIF. Collectively, these analyses provide comprehensive evidence of MBTI's psychometric validity and fairness within the Indonesian context.

Review of the Literature

Development of the MBTI Test

The development of the Myers-Briggs Type Indicator (MBTI) test began with the theory of psychological types developed by Carl Jung, which states that seemingly random variations in behaviour are actually orderly and consistent, caused by basic differences in how individuals prefer to use their perceptions and judgments (Myers et al., 1998). Jung divided personality types into three main categories: attitudes (Extraversion-Introversion), perceptual functions (Sensation-Intuition), and judgment functions (Thinking-Feeling), which are mutually exclusive and opposite (Read et al., 1974).

Based on Jung's theory, Katherine Cook Briggs and Isabel Briggs Myers developed the MBTI test with the aim of applying the theory in a practical and easy-to-understand way (Myers et al., 1998). The MBTI aims to classify individuals based on their preferences on four dichotomous dimensions: Extrovert-Introvert (EI), Sensing-Intuition (SI), Thinking-Feeling (TF), and Judging-Perceiving (JP). This test focuses more on grouping individuals according to their type preferences than measuring them. The MBTI has undergone several developments and revisions since it was first published in 1962, starting with the 166-item MBTI Form F, followed by the 126-item MBTI Form G in 1978, and the 93-item MBTI Form M in 1998 (Quenk, 2009). In 2001, the 144-item MBTI Form Q was introduced, and in 2009, the latest versions of the MBTI Step I, Step II, and Step III were released for use in training, career development, and research in education and organisations (Quenk, 2009). The MBTI test serves as a tool for understanding individual differences in how they behave and interact, with many benefits both personally and in organisational contexts.

Psychometric Research

Research on the reliability and validity of the MBTI instrument shows that this measuring instrument has a good level of reliability. Several studies conducted in various social and cultural contexts, including research by Myers et al. (1998), reported Cronbach's alpha coefficient results ranging from 0.89 to 0.92, indicating high internal consistency across the four dimensions of the MBTI. More recent research, such as that conducted by Schaubhut et al. (2009), also supports these results by reporting alpha coefficients varying from 0.83 to 0.92 across different ethnic groups, ages, and employment statuses. The validity and reliability of the MBTI instrument are acceptable across samples, including groups diverse in terms of gender, age, and ethnicity, indicating that this instrument can be used reliably across various population groups. In order to test the structural validity of the MBTI, many factor analytic studies have been conducted to ensure whether the resulting structure is in accordance with the theory developed by Isabel Briggs Myers. Several studies using exploratory factor analysis (EFA) have found that four major factors emerge consistent with the theoretical model of the MBTI (e.g., Harvey et al., 1995; Thompson & Borrello, 1986). However, there have also been studies reporting results inconsistent with the predicted structure, such as those found by Comrey (1983) and Sipps et al. (1985), suggesting that other factors influence the MBTI factor structure. However, further research using confirmatory factor analysis (CFA) has shown strong support for the theoretical and widely used four-factor model (Schaubhut et al., 2009).

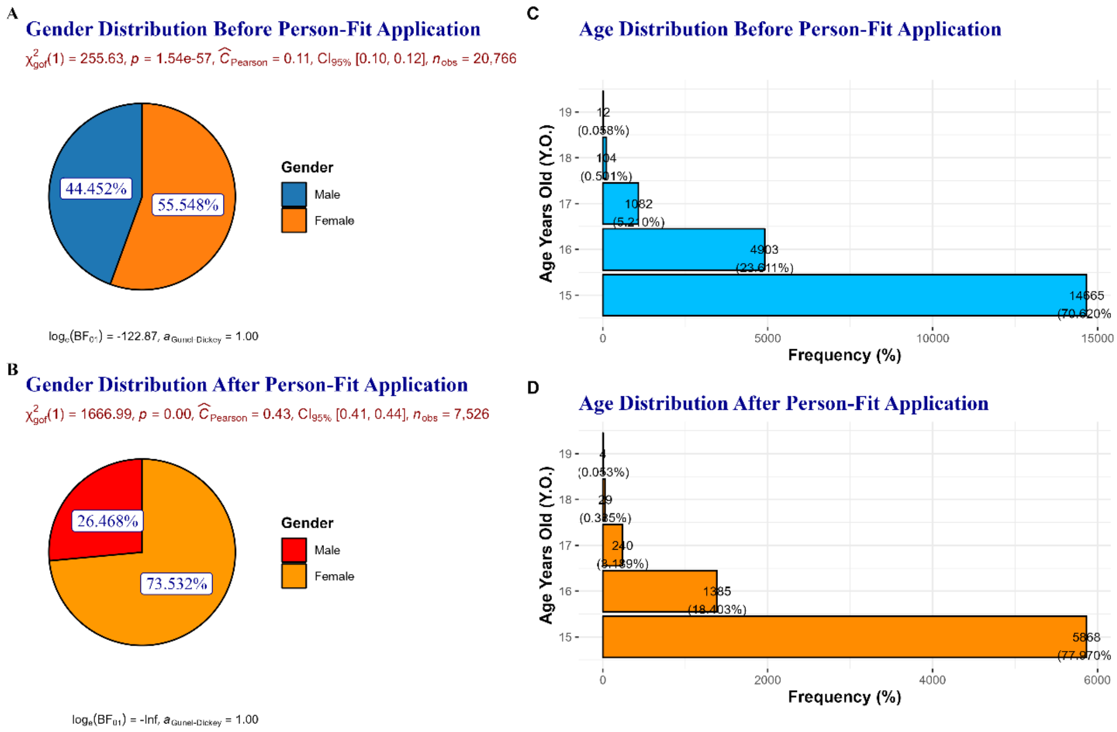
In addition, research in Indonesia conducted by Susanto and Mudaim (2017) showed that the revised MBTI instrument met strict validity and reliability criteria, with Rasch analysis results showing a good match between items and respondents. The Outfit Mean Square (MNSQ) and Outfit Z-Standard (ZSTD) values obtained were within the acceptable range, indicating that this instrument is valid and reliable in the Indonesian context. Testing of differential item functioning (DIF) also showed no significant item bias based on differences in respondent characteristics, such as age, gender, and race, which further strengthens the conclusion that this MBTI instrument is suitable for use in various population groups. Finally, research conducted by Periantalo and Azwar (2017) showed that the personality scale developed based on Jung's personality types and MBTI has good construct validity, with no significant correlation between the dimensions of the personality types tested. By using a trial method with different time intervals, the reliability of the measurement in this scale is also quite high, with a reliability coefficient of 0.91 for a one-day interval and 0.81 for a one-week interval. This finding confirms that this instrument can be used in various educational and counselling psychology contexts to help understand and utilise personality types in learning strategies and individual potential development.

Methods

This study follows the quantitative paradigm with an emphasis on the field of psychometrics using a cross-sectional survey design. This quantitative psychometric study aims to evaluate the attributes of psychological tests, such as the type of information generated, reliability, and validity of the data (Furr, 2022; Borsboom & Molenaar, 2015).

Participants

Participants in this study were participants in the aptitude and interest test conducted by the Applied Psychology Institute X, totalling 20,766 students, consisting of 9,231 males (44.452%) and 11,535 females (55.548%) with an age range of 15-19 years (*M* = 15.357, *SD* = 0.609). The data were then tabulated for further analysis. The results of data screening, conducted through twenty-two trials using the person-fit order method as part of the Rasch Model measurement, yielded an average logit value and standard deviation of 1.11. Based on this criterion, 13,240 students were identified as misfits because their logit values exceeded 1.11. After excluding the misfit cases, 7,526 students remained and were included in the main data analysis (Figure 1). Of these 7,526 students, 1,992 were male (26.468%) and 5,534 were female (73.532%), with age range of 15 to 19 years (*M* = 15.26; *SD* = 0.54).



Sources: Personal data (2019).

Figure 1. Demographic characteristics of participants

Intruments

The measuring instrument used in this study was the MBTI test (Myers, 1998), which was obtained from the Institute of Applied Psychology X and has undergone a development stage. This instrument is a well-known and widely used personality assessment, especially in the context of measuring student talents and interests. MBTI is designed to measure and classify individuals into psychological types based on Carl Gustav Jung's theory (Read et al., 1974). Individual preferences are measured through 60 items covering four dichotomies, namely: (1) Extrovert-Introvert, (2) Sensing-Intuition, (3) Thinking-Feeling, and (4) Judging-Perceiving. Responses to items on this assessment will categorise individuals into one of the preference types in each dichotomy. Based on these results, individuals will be classified into one of the sixteen possible types in the instrument (eg, ENFJ). The resulting personality type will be used as a reference in evaluating individual talents, interests, and potential in various contexts, such as work, education, and personal development.

Data Collection

This study uses secondary data with a documentation method in the form of MBTI test participant answer sheets that have been examined by the Applied Psychology Institute X. The answer sheets are the results of the 60-item MBTI test version that has been filled in with raw scores.

Data Analysis

The psychometric analysis in this study was conducted using the R programming language (v4.4.3; R Core Team, 2025) in combination with Positron (Posit Software, 2025). Various R packages were used to support the analysis, such as *readxl* (Wickham et al., 2023a) to import Excel files, *jmv* (Selker et al., 2023) for descriptive analysis, *eRm* (Mair & Hatzinger, 2007; Mair et al., 2024) to conduct a more comprehensive evaluation of the scale and general modeling in the calibration of Rasch modeling with the Conditional Maximum Likelihood (CML; Padgett & Morgan, 2020) estimator, *psych* (Revelle, 2024) to measure internal consistency using Cronbach's alpha and omega coefficients, and *MplusAutomation* (Hallquist et al., 2024) to conduct item factor analysis using the Mplus software function on big data (Hallquist & Wiley, 2018). In addition, for data visualization, several packages such as *ggplot2* (Wickham et al., 2023b) and *ggstatsplot* (Patil & Powell, 2024) were used for plotting, and *tidyverse* (Wickham et al., 2023c) for the initial plot setup, and *patchwork* (Pedersen, 2024) was used to display multiple plots in one layout. The *qgraph* package (Epskamp et al., 2023) was used to create correlation network plots, while the *Matrix* package (Bates et al., 2024) was used to support the creation of these plots. All stages of the analysis are further explained in the following four points.

Item Factor Analysis (IFA)

Item factor analysis (IFA) is a factor analysis method designed for ordinal scale data, such as Likert scale item scores with response categories lower than five (Rhemtulla et al., 2012) or scales with a score typology of 0 and 1. This method is suitable for handling Likert data that is not normally distributed or highly skewed (Muthén & Kaplan, 1985). IFA is also known as Ordinal Confirmatory Factor Analysis (CFA; Bock et al., 1988; Hayat et al., 2021; Rahayu et al., 2021) or Categorical CFA (Clark & Bowles, 2018; Chen & Zhang, 2021). IFA includes factor analysis for categorical item-level data, using both exploratory and confirmatory approaches (Cai, 2010; Rifenbark et al., 2021). In confirmatory IFA, a specifically proposed factor structure (including correlations between factors) is statistically evaluated. If the estimated model fits the data, then the researcher can conclude that the factor structure can be replicated (Reise et al., 2000). This study used IFA to test the fit of the MBTI dimension factor structure to empirical data with a limited-information IFA approach based on a tetrachoric correlation matrix (Wirth & Edwards, 2007). The model fit was evaluated using several criteria, such as an insignificant Chi-square ($\chi 2$) value ($p > 0.05$), Standardized Root Mean Square Residual (SRMR) less than 0.08, Root Mean Square Error of Approximation (RMSEA) less than 0.08, and Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values of more than 0.90 (Brown, 2015; Hair et al., 2019; Wang & Wang, 2020; Kline, 2023).

After the model is stated to be in accordance with the data, the validity of each item is evaluated. An item is declared valid if it has a positive factor loading and a significant z value (>1.960) with a p-value less than 0.05 (Brown, 2015; Bakker et al., 2018; Hair et al., 2019; Umar & Nisa, 2020; Garcia et al., 2022). In addition, internal consistency is reported using the omega coefficient (ω), where a value greater than 0.70 indicates good reliability (Viladrich et al., 2017). Model estimation was performed using the weighted least square mean and variance adjusted (WLSMV) method implemented in the R programming language. The results of the analysis were reported according to APA's Journal Article Reporting Standards (JARS) for quantitative research (Appelbaum et al., 2018), including descriptive statistics, correlation matrix, overall model fit test (goodness-of-fit), item-level fit test, and residual analysis as an in-depth analysis of the results.

Rasch Model

The Rasch model was designed to ensure that measurements in the social sciences have equivalent units to those in physics (Bond et al., 2021). Using the same logit metric, this model estimates individual and item parameters simultaneously (Saggino et al., 2020), with the aim of estimating individual ability while assessing item difficulty. Person measurement in the Rasch Model illustrates that individuals with higher ability always have a greater chance of succeeding on an item than individuals with lower ability. In contrast, item calibration ensures that each item in a test or instrument is arranged according to its level of difficulty (Bond et al., 2021). In addition, in its early development, the Rasch model was designed to analyse dichotomous data (Rasch, 1960). Meanwhile, the Rasch model for dichotomous data has the following basic equation (de Ayala, 2009):

$$P\left(x_j = 1 \middle| \theta, \delta_j\right) = \frac{e^{(\theta-\delta_j)}}{1 + e^{(\theta-\delta_j)}}$$

where $P\left(x_j = 1 \middle| \theta, \delta_j\right)$ is the probability of answering item j correctly (score 1), θ is the test taker's ability, and δ_j is the difficulty level of item *j*. The item difficulty and personal ability were expressed in logit scales. The Rasch model analysis was conducted using the *eRm* package in the R programming language (Mair & Hatzinger, 2007), which is suitable for analyzing data with large numbers of respondents and complex item scales.

In this study, the researchers used the Conditional Maximum Likelihood (CML; Padgett & Morgan, 2020) estimator for the Rasch model analysis. This approach was chosen because the CML estimator can overcome the limitations of the Joint Maximum Likelihood (JML) estimator, such as inconsistent estimates and the inability to handle individuals with a total score of zero or maximum (Andersen, 1972; Haberman, 1977). In addition, CML can also produce consistent parameter estimates by separating item and individual estimates through conditioning on person adequacy statistics (Padgett & Morgan, 2020), and is efficient for use in dichotomous and polytomous Rasch models to allow stable item parameter estimates before estimating individual parameters through a two-step procedure (de Ayala, 2009). Mair (2018) stated that in using the Rasch model, three assumptions need to be met, namely: (1) unidimensionality, namely the test only measures one trait, (2) local independence, namely the response given by test takers to one item must be statistically independent of the response to other items in a test, and (3) parallel item characteristic curves (ICC), which means that the items have the same discriminating power. In the Rasch measurement model, the fit index used is the mean square-based statistic called infit (unweighted) and outfit (weighted) mean square (Wright & Stone, 1979). The expected value of these two statistics is 1, where values in the range of 0.5–1.5 indicate that the data is in accordance with the Rasch model (Boone & Staver, 2020).

Differential Item Functioning (DIF)

The ability of items in the Rasch Model not to show differential item functioning (DIF) is one of the most important psychometric aspects (Christensen et al., 2013). DIF occurs when two individuals with the same ability level, but from different groups, have unequal opportunities to answer an item correctly, as explained by Hambleton et al. (1991). This indicates that the difference in results is not caused by the ability being measured, but by group factors such as gender, age, or education level (Rahayu et al., 2024). In the Rasch Model, the presence of DIF can be analysed using methods such as the Wald Test and the Likelihood Ratio Test (LR Test) (Mair & Hetzagon, 2007; Padgett & Morgan, 2020). This approach provides deeper insight into differences in response patterns between groups, thereby helping to ensure that the measurement instrument is more valid and fair to all respondents. Meanwhile, to identify DIF between student groups (male vs female), substantial differences between groups in item estimate scores of ± 0.50 logits (Bond & Fox, 2015; Wu et al., 2016).

Results

Descriptive Statistics of Items

Before conducting the item factor analysis (IFA) of the hypothetical model, descriptive statistics were examined to provide an overview of item characteristics. The results (see Appendix 1) showed that twelve items (EI15, EI31, SI16, SI46, TF9, TF32, TF42, TF49, TF58, JP12, JP47, JP56) exhibited non-normal response distributions, with skewness values exceeding the acceptable range of –2 to +2 (Muthén & Kaplan, 1992; Kim, 2013; Privado et al., 2024). Given the presence of non-normality and the dichotomous nature of the MBTI item responses, the IFA with an ordinal approach was employed. Accordingly, the tetrachoric correlation matrix (see Appendix 2) was used as the input data, as it assumes that each dichotomous response reflects an underlying continuous latent variable, enabling more accurate estimation of inter-item correlations.

Item Factor Analysis (IFA)

Model fit based on IFA

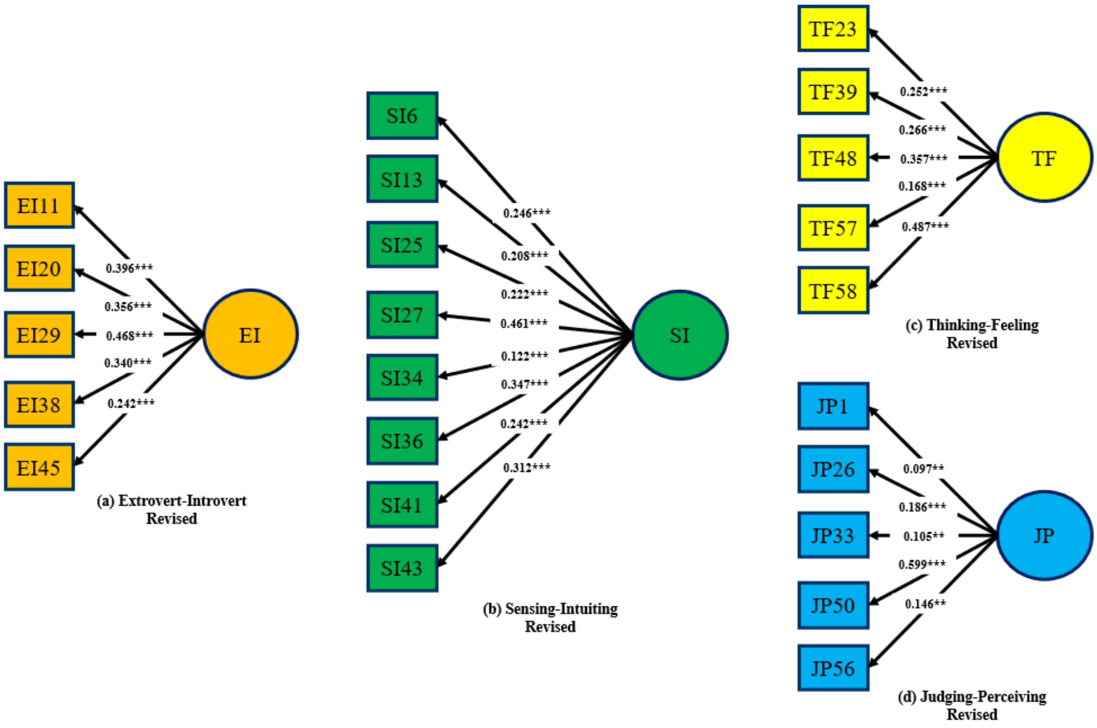
Based on the theoretical formulation, the 1-factor model was applied to each MBTI dimension (Table 1). However, the results of the hypothetical model testing with IFA indicated that none of the four MBTI dimensions (extrovert–introvert, sensing–intuitive, thinking–feeling, and judging–perceiving) showed adequate model fit, as reflected in low CFI/TLI values despite acceptable SRMR and RMSEA indices.

The results of the goodness-of-fit index on four dimensions involving fifteen items for each dimension indicated a suboptimal model, thus requiring model re-specification (Anderson & Gerbing, 1988). In this study, model re-specification was carried out by removing items with the lowest factor loadings in each dimension, following the recommendations of Jöreskog and Sörbom (1993) as well as Landells and Albrech (2019). After dropping these items, the optimal model for each dimension was obtained. The results of the analysis showed that the revised 1-factor extrovert-introvert model obtained a model fit to the data ($\chi 2$ = 9.831, df = 5, p = 0.080, SRMR = 0.013, RMSEA = 0.011 [0.000, 0.022], CFI = 0.989, TLI = 0.979). The same thing is also shown in the revised 1-factor sensing-intuition model that has fit the data ($\chi 2$ = 22.510, df = 20, p = 0.313, SRMR = 0.013, RMSEA = 0.004 [0.000, 0.011], CFI = 0.995, TLI = 0.992). Likewise, the revised 1-factor thinking-feeling model shows a fit to the data ($\chi 2$ = 8.820, df = 5, p = 0.116, SRMR = 0.012, RMSEA = 0.010 [0.000, 0.021], CFI = 0.986, TLI = 0.971). Meanwhile, the 1-factor model for judging-perceiving revised has fit the data ($\chi 2$ = 8.384, df = 5, p = 0.136, SRMR = 0.013, RMSEA = 0.009 [0.000, 0.020], CFI = 0.951, TLI = 0.902). The 1-factor model for each dimension of the MBTI test can be seen in Figure 2.

Table 1. Model comparison based on IFA

Model	$\chi 2$	df	p	SRMR	RMSEA	CFI	TLI
Extrovert-introvert	470.988	90	<0.001	0.050	0.024 [0.022, 0.026]	0.758	0.718
Sensing-intuiting	567.420	90	<0.001	0.041	0.027 [0.024, 0.029]	0.510	0.428
Thinking-feeling	390.335	90	<0.001	0.040	0.021 [0.019, 0.023]	0.601	0.534
Judging-Perceiving	533.112	90	<0.001	0.052	0.026 [0.024, 0.028]	0.403	0.304
Extrovert-Introvert Revised	9.831	5	0.080	0.013	0.011 [0.000, 0.022]	0.989	0.979
Sensing-Intuiting Revised	22.510	20	0.313	0.013	0.004 [0.000, 0.011]	0.995	0.992
Thinking-Feeling Revised	8.820	5	0.116	0.012	0.010 [0.000, 0.021]	0.986	0.971
Judging-Perceiving Revised	8.384	5	0.136	0.013	0.009 [0.000, 0.020]	0.951	0.902

Sources: Research data (2019).



Sources: Personal data (2019).

Figure 2. One-factor model IFA for each MBTI dimension

Item-level fit based on IFA

Table 2 showed the item-level findings, indicating that the factor loadings for all items had positive values for each dimension, such as extrovert–introvert revised (λ = 0.242–0.468), sensing–intuition revised (λ = 0.122–0.461), thinking–feeling revised (λ = 0.168–0.487), and judging–perceiving revised (λ = 0.105–0.599). In addition, all item factor loadings were significant ($z > 1.96$, $p < 0.05$), confirming that each item was valid in measuring the intended construct. The tetrachoric correlations among items within each MBTI dimension are illustrated in the network visualization (see Appendix 2), where thicker blue lines represent stronger associations. Notably, no negative correlations were observed, supporting the internal consistency of items within each dimension.

Residual analysis based on IFA

In addition, the results of the IFA indicated that the revised model, obtained by dropping items with the lowest factor loadings in each dimension, fit the data without requiring further modifications. However, further study is needed to examine whether item wording similarities or construct overlap exist. To address this, residual correlation analysis for each item was carried out. As shown in Figure 3, thicker blue lines indicate higher positive residual correlations between item pairs, while red lines indicate negative residual correlations (e.g., I34 and I43 on the sensing–intuition dimension, r = –0.046). Although some item pairs (e.g., I1 and I26 in judging–perceiving, r = 0.037; I6 and I34 in sensing–intuition, r = 0.042) showed relatively high positive residual correlations, these values were not strong enough to threaten the validity of the 1-factor model in each dimension, which was overall supported by the IFA results.

Table 2. Item parameters based on IFA

Dimension	Item	Std. Est.	S.E.	95%C.I.		*z*	*p*
				Lower	Upper		
Extrovert-Introvert Revised	EI11	0.396	0.029	0.399	0.454	13.439	<0.001
	EI20	0.356	0.029	0.299	0.413	12.333	<0.001
	EI29	0.468	0.032	0.405	0.532	14.407	<0.001
	EI38	0.340	0.029	0.282	0.398	11.551	<0.001
	EI45	0.242	0.031	0.182	0.302	7.876	<0.001
Sensing-Intuiting Revised	SI6	0.246	0.028	0.191	0.301	8.800	<0.001
	SI13	0.208	0.032	0.145	0.271	6.467	<0.001
	SI25	0.222	0.028	0.167	0.278	7.840	<0.001
	SI27	0.461	0.032	0.398	0.525	14.271	<0.001
	SI34	0.122	0.027	0.069	0.175	4.529	<0.001
	SI36	0.347	0.029	0.290	0.403	12.061	<0.001
	SI41	0.242	0.028	0.188	0.297	8.736	<0.001
	SI43	0.312	0.029	0.256	0.368	10.869	<0.001
Thinking-Feeling Revised	TF23	0.252	0.031	0.192	0.312	8.232	<0.001
	TF39	0.266	0.032	0.204	0.328	8.398	<0.001
	TF48	0.357	0.035	0.288	0.426	10.125	<0.001
	TF57	0.168	0.030	0.108	0.227	5.535	<0.001
	TF58	0.487	0.043	0.403	0.571	11.329	<0.001
Judging-Perceiving Revised	JP1	0.097	0.036	0.027	0.168	2.697	0.007
	JP26	0.186	0.053	0.081	0.291	3.480	0.001
	JP33	0.105	0.038	0.031	0.180	2.770	0.006
	JP50	0.599	0.158	0.290	0.908	3.799	<0.001
	JP56	0.146	0.047	0.054	0.239	3.093	0.002

Sources: Research data (2019).

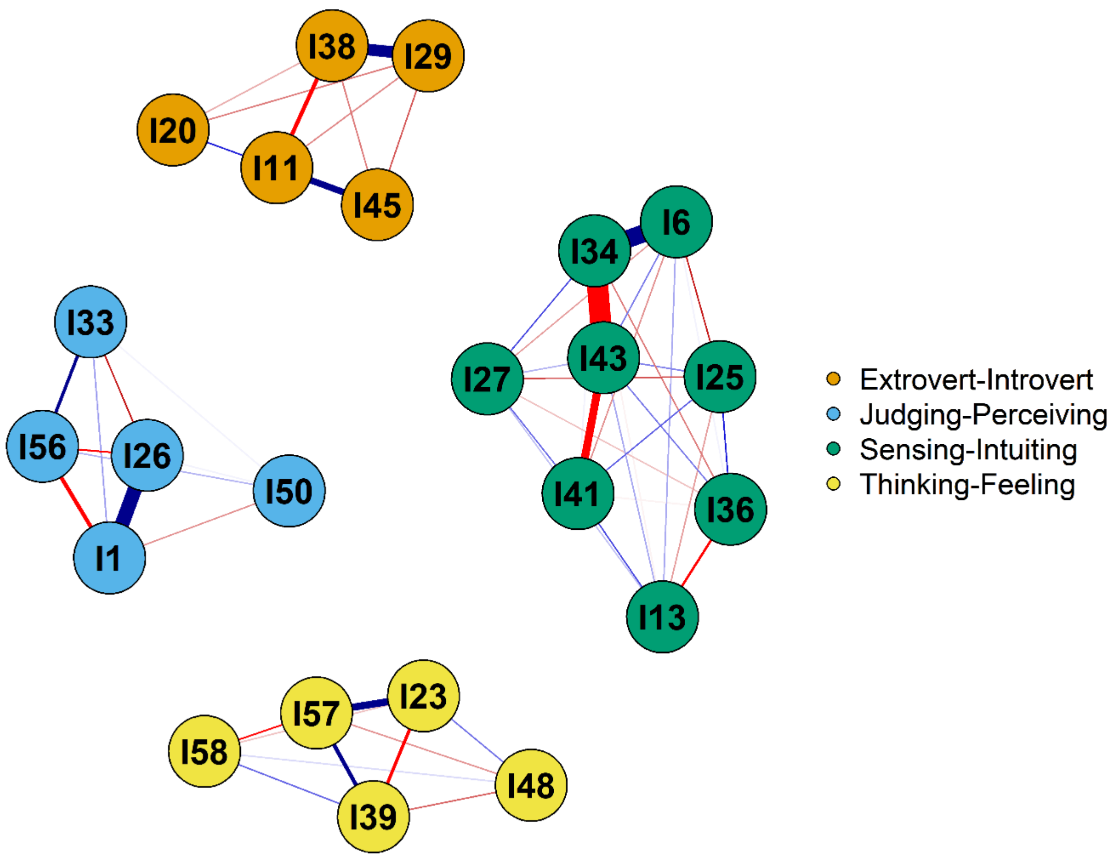
Rasch Model

Unidimensionality

Rasch modelling assumes that the measuring instrument must be unidimensional, meaning that the construct actually measures what it is supposed to measure (Mair, 2018). Researchers conducted a unidimensional analysis in Rasch modelling based on Principal Component Analysis on Residuals (PCAR; Smith, 2002) to ensure that the measuring instrument only assesses one attribute, namely digital leadership. The results of the analysis showed that the MBTI test for each dimension had a raw variance explained by the measurement of more than 30% (e.g., the extrovert-introvert dimension = 31.648%, and the judging-perceiving dimension = 34.785%). The variance explained by more than 30% indicates that each subscale only measures one characteristic (Linacre & Wright, 2012; Azizah et al., 2022). In addition, these results have also been strengthened by the results of the IFA fit model that have been reported previously. Thus, the unidimensionality assumption has been met.

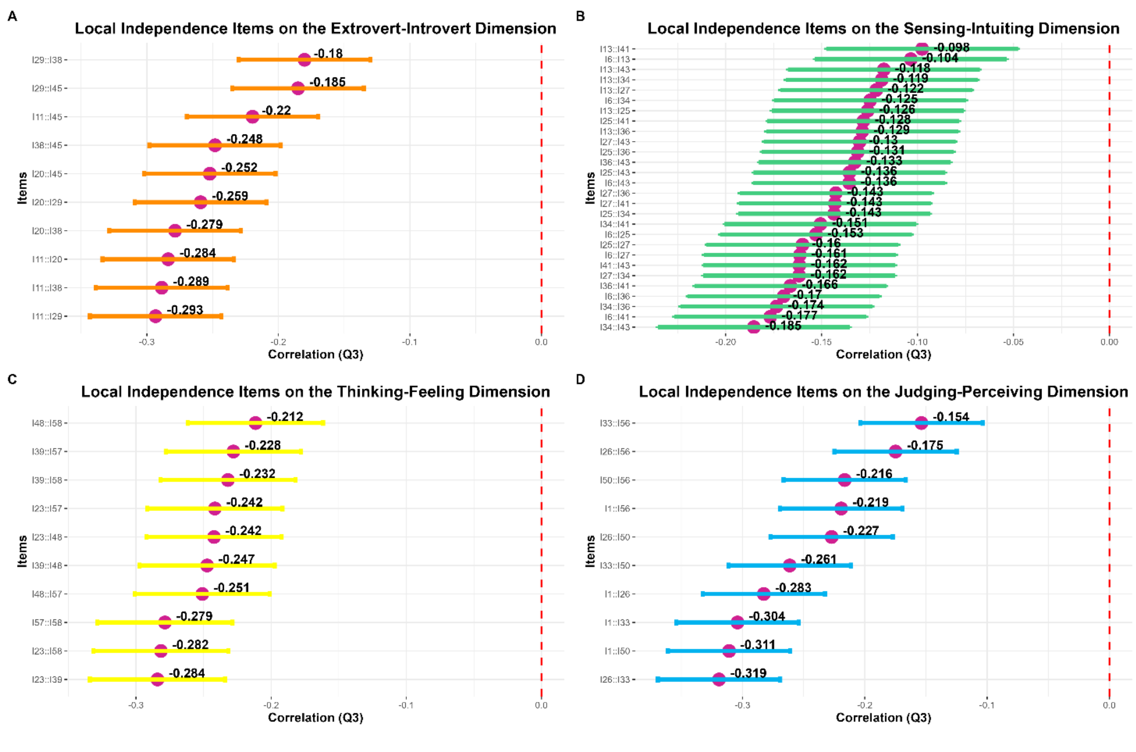
Local independence

The local independence was tested using the Q3 statistic (Yen, 1984), with violations indicated when residual correlations exceed 0.30 (Chen & Thissen, 1997; Nair et al., 2011; Roe et al., 2014; Christensen et al., 2016). The analysis showed that all residual correlations were below this threshold, indicating that no item pairs exhibited local dependence. Thus, the assumption of local independence was supported for all MBTI dimensions (see Figure 4).



Sources: Personal data (2019).

Figure 3. Inter-item residual correlation plot for each MBTI dimension



Sources: Personal data (2019).

Figure 4. Plot correlation Q3 between all pairs of items in each dimension

Item fit statistics

After ensuring that the assumptions of unidimensionality and local independence were met, the research scale retained 23 valid items. Statistical evaluation of the items was conducted to assess the fit between the expected responses based on the model and the observed responses. The range of MNSQ values considered appropriate, both for Infit and Outfit, is 0.5 to 1.5 (Boone et al., 2014; DiStefano et al., 2019). The analysis showed that all items met these criteria with no Infit or Outfit MNSQ values falling outside the specified limits (Table 3). Therefore, all items in each dimension have good statistical fit and can be relied on to optimally measure the dimensions of the MBTI test. Item-level properties, including item characteristic curves (ICC), person–item map, and test information function (TIF), are provided in the Supplementary Materials (see Figure S1a–S1c).

Table 3. Rasch analysis results for item measure and fit statistics

Dimension	Item	Measure	S.E.	95%C.I.		$\chi 2$	p	Infit	Outfit
				Lower	Upper			MNSQ	MNSQ
Extrovert-Introvert	EI11	0.334	0.022	0.290	0.377	6219.156	1.000	0.939	0.923
	EI20	-0.111	0.023	-0.155	-0.066	6300.527	1.000	0.948	0.935
	EI29	1.458	0.025	1.409	1.508	5422.492	1.000	0.885	0.804
	EI38	-0.506	0.024	-0.652	-0.558	6069.559	1.000	0.938	0.900
	EI45	-1.076	0.026	-1.128	-1.024	6243.710	1.000	0.934	0.926
Sensing-Intuiting	SI6	-0.937	0.023	-0.982	-0.891	7149.309	0.960	0.976	0.971
	SI13	1.285	0.029	1.229	1.341	6928.727	1.000	0.922	0.941
	SI25	0.597	0.024	0.550	0.645	7157.211	0.954	0.973	0.972
	SI27	0.021	0.023	-0.023	0.066	6690.725	1.000	0.932	0.909
	SI34	0.229	0.023	0.183	0.274	7614.497	0.019	1.026	1.034
	SI36	-0.630	0.023	-0.674	-0.586	6968.105	0.999	0.964	0.947
	SI41	-0.969	0.023	-1.015	-0.924	7177.666	0.934	0.986	0.975
	SI43	0.404	0.024	0.358	0.450	6954.876	1.000	0.964	0.945
Thinking-Feeling	TF23	-0.638	0.022	-0.680	-0.595	6491.327	0.214	1.014	1.014
	TF39	0.202	0.023	0.157	0.246	6327.208	0.745	0.989	0.988
	TF48	0.342	0.023	0.296	0.387	5999.953	1.000	0.956	0.937
	TF57	0.263	0.023	0.218	0.307	6505.205	0.181	1.012	1.016
	TF58	-0.168	0.022	-0.211	-0.125	6054.929	0.999	0.954	0.946
Judging-Perceiving	JP1	-0.339	0.022	-0.382	-0.296	6832.479	0.998	0.963	0.953
	JP26	-1.021	0.024	-1.067	-0.975	6332.392	1.000	0.926	0.884
	JP33	-0.898	0.023	-0.943	-0.853	6656.277	1.000	0.954	0.929
	JP50	0.533	0.023	0.488	0.578	6062.576	1.000	0.881	0.846
	JP56	1.726	0.030	1.667	1.784	5584.099	1.000	0.832	0.779

Sources: Research data (2019).

Reliability

Rasch analysis evaluates reliability using the person separation reliability (PSR) coefficient, which reflects the instrument's ability to differentiate participants based on the measured trait. The PSR ranges from 0 to 1, with values ≥ 0.70 indicating good internal consistency (Wright & Stone, 1979; Geldenhuys & Bosch, 2019; Bond et al., 2021). Higher PSR values suggest greater measurement precision, as they account for both observed variance and measurement error. In this study, the PSR reliability for each dimension was measured, and the results showed that the four dimensions analysed had low or negative PSR values. The extrovert-introvert dimension has a PSR value of -0.120, which indicates very low reliability. This negative PSR value indicates that measurement error is more dominant than observed variance, meaning that the instrument is less effective in differentiating participants based on this dimension. This could be due to the mismatch between items and participants in this dimension, leading to less consistent measurements. In addition, the results of the analysis also showed that the reliability values of Cronbach's Alpha (α = 0.290 <0.70; Nunnally &

Bernstein, 1994) and McDonald's omega (ω = 0.30 <0.70; Viladrich et al., 2017) on the extrovert-introvert dimension of the MBTI test were very weak.

In the sensing-intuition dimension, the PSR value of 0.204 indicates low reliability. Although the result is positive, the value is still far below the recommended minimum threshold of 0.70. This means that the instrument has not been able to effectively differentiate participants based on their level of the Sensing-Intuition dimension, indicating potential problems in item targeting or inappropriate distribution of participants' traits. In addition, the analysis results also show that the CTT reliability value (α = 0.260; ω = 0.410) in the sensing-intuition dimension of the MBTI test is very weak. Meanwhile, the thinking-feeling dimension shows a PSR value of -0.337, indicating negative reliability. With a negative PSR value, this instrument fails to differentiate participants well in the thinking-feeling dimension, which may be due to a mismatch between the items given and the traits being measured. This is in line with the CTT reliability (α = 0.230; ω = 0.250) on the sensing-intuition dimension of the MBTI test which showed very weak results. Likewise with the judging-perceiving dimension, which showed a PSR of -0.104. This value indicates that the instrument failed to effectively differentiate participants in the judging-perceiving dimension, with measurement error being more dominant than the observed variance. The CTT reliability results (α = 0.110; ω = 0.140) on the judging-perceiving dimension of the MBTI test were also very weak. Overall, the results of the analysis showed that the PSR reliability for the four dimensions was below the recommended value. Low and negative PSR values indicate that the instrument used in this study was not effective in differentiating participants based on the traits measured. This finding aligns with Wihardini (2020), who suggested that low response variability, limited sample size, and diversity in item format and response types can contribute to lower reliability scores. To improve reliability, it is recommended to revise the items used, improve item targeting, and ensure a more appropriate distribution of participant traits (Linacre, 2018).

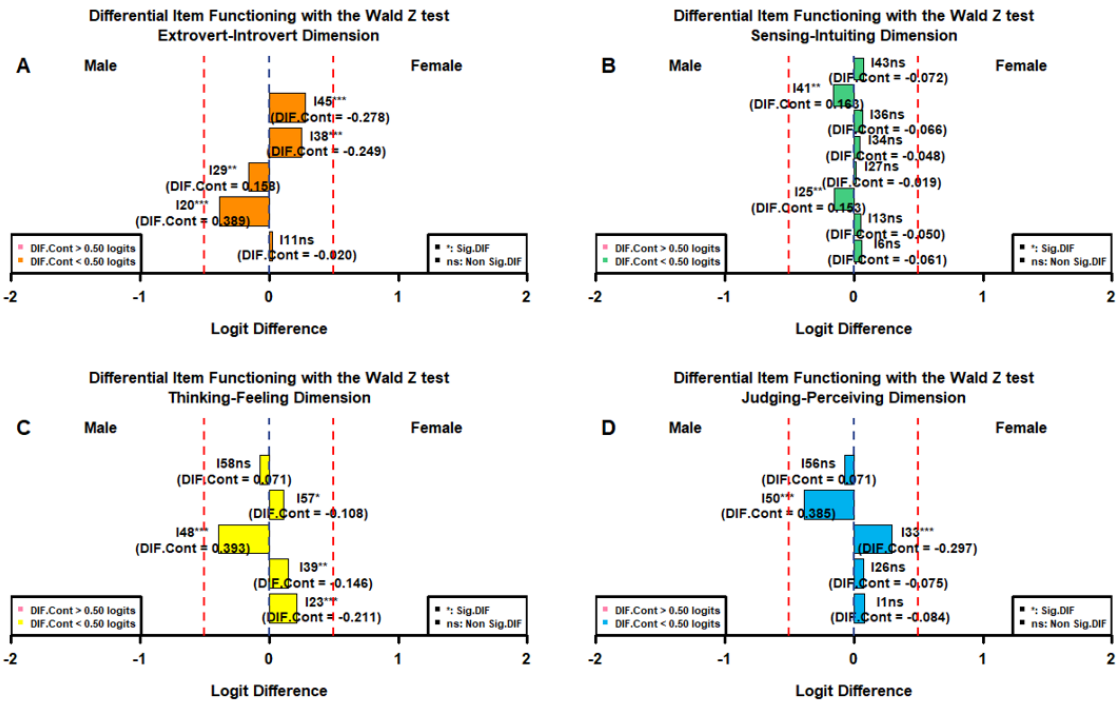
Differential Item Functioning (DIF)

Each dimension of the MBTI test was examined for potential item-level bias using DIF analysis. In the context of this study, although male and female students have the same basic trait levels, they may respond differently to certain items, which may indicate item bias between groups (Saggino et al., 2020). To detect item bias, the item-trait Chi-square method is used. The presence of significant chi-square differences between groups (male and female students) and DIF effect sizes greater than 0.50 are considered as indications of item bias (Yan & Mok, 2012). The results of the Chi-square significance difference analysis between groups have been presented in Table 4. Through DIF analysis, the researcher found that there were twelve statistically significant items, namely: extrovert-introvert dimensions (EI20, EI29, EI38, EI45), sensing-intuiting dimensions (SI25, SI41), thinking-feeling dimensions (TF23, TF39, TF48, TF57), and judging-perceiving dimensions (JP33, JP50). Although statistically significant, no item in each dimension had a DIF effect size exceeding 0.50 (see Figure 5). The largest DIF effect sizes were found in items EI20 (DIF contrast = 0.389) and JP50 (DIF contrast = 0.385). This finding indicates that male students (as the focal group coded as 1) tend to endorse the two items more easily than other groups. The results also show strong positive correlations for all dimensions, with coefficients of 0.928 for extrovert–introvert (EI), 0.560 for sensing–intuition (SI), 0.839 for thinking–feeling (TF), and 0.983 for judging–perceiving (JP). These findings suggest that items perceived as difficult by male students tend to be similarly difficult for female students, supporting the assumption of measurement equivalence between gender groups. The consistent item difficulty patterns across genders demonstrate that the MBTI dimensions function similarly for both male and female respondents (see Appendix 3).

Table 4. DIF based on gender for each item in the MBTI dimension

Dimension	Item	Male		Female		DIF			
		Measure	S.E.	Measure	S.E.	DIF Contrast	Joint S.E.	*z*	*p*
Extrovert-Introvert	EI11	0.329	0.026	0.349	0.043	-0.020	0.050	-0.401	0.688
	EI20	-0.014	0.026	-0.403	0.046	0.389	0.053	7.345	0.000
	EI29	1.504	0.030	1.346	0.047	0.158	0.056	2.829	0.005
	EI38	-0.670	0.028	-.0421	0.046	-0.249	0.054	-4.607	0.000
	EI45	-1.148	0.031	-0.871	0.050	-0.278	0.059	-4.703	0.000
Sensing-Intuiting	SI6	-0.953	0.027	--0.892	0.045	-0.061	0.053	-1.164	0.244
	SI13	1.271	0.034	1.322	0.055	-0.050	0.064	-0.779	0.436
	SI25	0.640	0.029	0.487	0.046	0.153	0.054	2.830	0.005
	SI27	0.016	0.027	0.035	0.044	-0.019	0.051	-0.363	0.717
	SI34	0.216	0.027	0.264	0.045	-0.048	0.052	-0.921	0.357
	SI36	-0.648	0.026	-0.582	0.044	-0.066	0.051	-1.287	0.198
	SI41	-0.928	0.027	-0.090	0.046	0.163	0.054	3.029	0.002
	SI43	0.384	0.028	0.456	0.046	-0.072	0.053	-1.349	0.177
Thinking-Feeling	TF23	-0.696	0.026	-0.485	0.042	-0.211	0.049	-4.294	0.000
	TF39	0.159	0.027	0.305	0.043	-0.146	0.051	-2.874	0.004
	TF48	0.458	0.028	0.065	0.042	0.039	0.051	7.749	0.000
	TF57	0.230	0.027	0.338	0.043	-0.108	0.051	-2.104	0.035
	TF58	-0.151	0.026	-0.222	0.042	0.071	0.049	1.452	0.147
Judging-Perceiving	JP1	-0.363	0.026	-0.279	0.042	-0.084	0.049	-1.700	0.089
	JP26	-1.043	0.027	-0.968	0.046	-0.075	0.053	-1.412	0.158
	JP33	-0.979	0.027	-0.682	0.044	-0.297	0.051	-5.767	0.000
	JP50	0.641	0.027	0.255	0.042	0.385	0.050	7.647	0.000
	JP56	1.745	0.035	1.674	0.055	0.071	0.066	1.081	0.280

Sources: Research data (2019).



Sources: Personal data (2019).

Figure 5. Differential item functioning with the wald z test

Discussion

The main objective of this study was to assess the psychometric properties (factor structure, invariance, item quality) of the 60-item version of the MBTI test in Indonesian students using modern methods (IFA, Rasch), and to conduct DIF analysis to evaluate whether the MBTI test works well in both male and female student groups. This study is the first construct validity test of the MBTI test in the context of Indonesian students, with an approach that emphasises the analysis of each dimension independently. The results showed that a one-factor model for each MBTI dimension had a good fit with the data, which supports the replication of the factor structure found in previous studies using the MBTI instrument. In previous studies, conducted in South Africa (De Bruin, 1996; van Zyl & Tylor, 2012), the exploratory factor analysis (EFA) method provided strong evidence for the theoretical structure of the MBTI.

This study confirms the theoretical factor structure reported in previous studies, which applied the four-factor oblique model (Tzeng et al., 1984; Thompson & Borrello, 1986; Tischler, 1994; Harvey et al., 1995; van Zyl & Tylor, 2012). The approach in previous studies assumed that each MBTI dimension (extrovert-introvert, sensing-intuiting, thinking-feeling, judging-perceiving) is part of a complementary factor framework, but analysed in one comprehensive model. However, the approach in this study is different, as it applies an orthogonal model to each MBTI dimension independently. This approach is based on the basic assumption of the MBTI that each dimension is typological or bipolar, where there is no linear relationship between dimensions. For example, an increase in the extrovert-introvert dimension is not assumed to have an impact on other dimensions, such as sensing-intuiting. With the orthogonal model, psychometric analysis is conducted separately for each dimension, ensuring more accurate results in reflecting the construct validity and item quality of each dimension. This approach not only supports the basic principles of MBTI but also provides an essential contribution to ensuring that each dimension functions according to its theoretical nature. These findings indicate that each MBTI dimension works independently, without influencing the others, and provides significant potential to improve accuracy in identifying respondents' typology preferences. The results of this study provide a strong foundation for the development of MBTI instruments in the context of education in Indonesia.

The differences between this study and previous studies are not problematic, but rather complementary. The current study highlights the theoretical nature of the MBTI, which is typological, with an emphasis on the independence of each dimension, while previous studies describe the relationship between dimensions as a whole. Thus, this study deepens the understanding of the construct validity of the MBTI through a more specific approach to each dimension. Although the 60-item version of the MBTI used in this study has a different format compared to Form G (De Bruin, 1996) and Form M (van Zyl & Taylor, 2012), the results of the study indicate that the application of the orthogonal model, where each dimension is analysed with a one-factor model, is a valid representation for testing the factor structure of the MBTI. This approach is not only in line with the theoretical principles of the MBTI, but also confirms that each dimension can function independently without any interrelationships between dimensions. By focusing on the construct validity of each MBTI dimension separately, this study complements previous studies that emphasise the overall analysis. These findings provide an important contribution to the development of the MBTI, especially in ensuring that this instrument remains relevant and accurate in the context of students in Indonesia.

Meanwhile, further discussion of the results of item-level validation revealed that all items are valid to measure each dimension of MBTI. No items have negative or insignificant factor loads. However, this study shows that the factor loading value is lower than in previous studies (De Bruin, 1996; Van Zyl & Taylor, 2012). The low factor loading value can be associated with several factors, including differences in cultural context between Indonesia and the country of origin of the MBTI development, which can affect how respondents understand and respond to items in the test. In addition, the item content in the 60-item version of the MBTI may require further evaluation to ensure that each item is relevant and appropriate to the culture of respondents in Indonesia. These cultural differences include language, social norms, and views on psychological preferences that may not be fully reflected in the current MBTI items. Therefore, although all items are declared valid, there is an opportunity to improve the quality of the items through cultural adaptation or revision of item content to be more informative and accurate in identifying respondents' preferences in the local context. These findings indicate the need for further development to ensure that the MBTI remains relevant and can be used effectively in various cultures.

Rasch analysis for each item on each dimension of the MBTI test showed advantages over previous psychometric validation methods (Periantalo & Azwar, 2017), such as those based on classical test theory (CTT). Rasch analysis provides more detailed item-level information, complementing the results of CTT-based analysis (Hayat et al., 2021). In this study, dichotomous Rasch analysis was used because the unidimensionality requirement was met. Significant evidence related to one latent trait measuring each MBTI dimension was confirmed through principal component analysis of residuals (PCAR). Based on the Rasch model, unidimensionality on each dimension of the 60-item MBTI version was successfully confirmed, supporting the results also obtained through IFA analysis. In addition, local independence tests did not show any items with standardised residual correlations greater than 0.30 (e.g., Saggino et al., 2020), so all items on each dimension were included in the Rasch analysis (Yan & Mok, 2012). The level of item fit in the Rasch Model is measured using Infit and Outfit Mean Square Error (MNSQ). The results of the analysis show that all items in each MBTI dimension have met the Infit and Outfit MNSQ criteria, with no items exceeding the tolerance limits of the item location values (Bond et al., 2021; Rahayu et al., 2024). This consistency is also reflected in the local independence analysis, where no violations of the assumptions were found. In addition, the Rasch Model allows for item and person fit analysis, which has been carried out previously in the study to detect participants who do not fit the model. This approach increases the reliability of the analysis results and minimises the potential for type II errors (Price, 2017).

The results of this study are consistent with previous findings, which showed the expected value of Infit and Outfit MNSQ of 1.00, and the value of ZSTD Infit of -0.1 and ZSTD Outfit of 0.0 (Susanto & Mudaim, 2017). In addition, other studies reported that Rasch analysis confirmed the accuracy of the items in each MBTI dimension, ensuring that each item in a particular dimension measures the same construct without redundancy (van Zyl & Taylor, 2012). These results, together with the findings of factor analysis, provide strong evidence that each item in the MBTI dimension does measure similar psychological attributes, making it valid to be used as one scale in each dimension. However, the internal consistency in this study showed suboptimal results, with Pearson and item reliability below the criteria and negative values. This is different from previous studies that obtained an overall instrument reliability of 0.71 (sufficient), and an overall respondent reliability of 0.68, which is also included in the sufficient category (Susanto & Mudaim, 2017). In addition, another study by van Zyl and Taylor (2012) revealed that reliability analysis showed that the MBTI instrument could be used reliably in the South African population, and overall, the reliability of the MBTI in South Africa was in line with international results obtained in North America, Australia, Asia, the Middle East, Europe, and Latin America (Schaubhut et al., 2009).

Low reliability in terms of internal consistency in a test instrument, such as the MBTI, can be caused by various factors. One significant factor is the low number of items in a scale, which causes the alpha value to decrease because the correlation between items is not strong enough to support internal consistency (Nunnally & Bernstein, 1994). In addition, the dichotomous answer choices in this study limit the variation of participants' responses compared to polytomous tests, resulting in differences in accuracy between the two (Jiao et al., 2012). Low internal consistency is also seen in the low omega coefficient, which is directly related to the low factor loadings produced by the IFA estimation. Low factor loadings indicate that the items may not adequately represent the underlying construct, which negatively impacts the omega coefficient (Wen & Ye, 2011). When items do not have good factor loadings on the construct being measured, this indicates that the items are less able to represent the underlying trait, further reducing internal consistency.

In this case, this study found that there were twelve items that were statistically significant in the DIF test. However, no items on each dimension had a DIF effect size exceeding 0.50 logits (Bond & Fox, 2015). This finding indicates that there is no item bias, and male and female students have equivalent scores. This result is in line with the findings in the study of Susanto and Mudaim (2017), which showed that all MBTI items had probability values above 5% in the DIF analysis (Sumintono & Widhiarso, 2013). Another study by van Zyl and Taylor (2012) also used Rasch to evaluate the scoring function related to DIF. Overall, the results of the uniform DIF analysis revealed that items on each of the four scales were responded to consistently by men and women. Items identified by gender and ethnic group did not overlap, and there was no clear pattern regarding the direction of DIF. That is, items that were easier or harder to endorse did not show a particular tendency for all groups in the analysis.

According to Linacre (2010), if the uniform DIF obtained in Rasch is mainly not directed towards one group, its impact on measurement is generally small. Meanwhile, the correlation between item locations (measures) for male and female students of each dimension shows a positive and strong relationship (e.g., the correlation of male and female measures on the extrovert-introvert dimension = 0.928 [95%CI: 0.250, 0.955]), indicating that items that are more easily endorsed by male students are easily endorsed by females. This is in line with the results of previous studies showing a strong correlation (r = 0.88; Dancey & Reidy, 2020) between male and female item locations (van Zyl and Taylor, 2012). Overall, the MBTI assessment does not appear to show a consistent bias towards one group based on gender, although different responses at the item level were found. To further improve the assessment, it is recommended to delete or rewrite items that may function somewhat differently across relevant demographic groups. Further investigation into the specific content of the items is also needed to identify the sources of differential responses among the groups.

Conclusion

Overall, the findings in this study support the psychometric validity of the 60-item MBTI in the context of students in Indonesia. The factor structure of the test is consistent with previous studies, indicating that the MBTI test can be replicated in Indonesia with adequate construct validity. This study also found that latent variance invariance was maintained for each test dimension, except for the Judging–Perceiving dimension, which showed residual variance invariance. The quality of the test items was also well maintained, as evidenced by the appropriate MNSQ infit and outfit values. In addition, no item bias was found based on the DIF effect size (DIF Contrast), ensuring that the MBTI test is equivalent for male and female student groups without disadvantaging one group. These results provide strong evidence regarding the psychometric properties of the 60-item MBTI in Indonesia and contribute to the development of this instrument in accordance with legislative requirements regarding the use of psychometric tests in Indonesia.

Although this study makes a contribution, there are several limitations that need to be considered. First, although this study used data from a large sample (7,526 students), the data used were secondary and derived from a convenience sample. This may affect the generalizability of the findings, as the sample may not be fully representative of the overall student population in Indonesia. Furthermore, although the majority of individuals who completed the 60-item version of the MBTI were represented in this study, demographic information was only available for a small portion of the sample because many respondents chose not to indicate their ethnic group. The absence of complete demographic data limits more in-depth analysis of differences based on these variables. Another limitation is the imbalance in the distribution of student groups, such as the underrepresentation of minority groups in the population, which may lead to bias in the findings.

Finally, the results of the reliability analysis showed that internal consistency, as measured using separation alpha and omega, was very low and even negative. This indicates the need to add more items to the MBTI test to improve its reliability. The addition of these items is expected to improve the quality of measurement and increase the internal consistency of the instrument. This study is in line with the Rasch results, which show good item quality, but need to be improved in terms of reliability to ensure more stable and consistent results. Therefore, further research is expected to continue to explore the potential for ethnic bias and refine this instrument by increasing the number of items to make it more relevant and applicable to a broader range of educational contexts.

Acknowledgment

The author would like to express gratitude to all the research collaborators who have contributed to the completion of this study.

Conflict of Interest

The author affirms that there are no conflicts of interest regarding the publication of this study. All data and findings presented in this research have been analyzed and reported objectively, without any external influence or bias. The study was conducted exclusively for academic and scientific purposes.

Authors Contribution

DYN: Conceptualization; Data Curation; Formal Analysis; Methodology; Writing Original Draft; Visualization. W: Validation; Writing, Review & Editing. AR: Methodology; Validation; Writing, Review & Editing.

References

Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(1), 42–54. https://doi.org/10.1111/j.2517-6161.1972.tb00887.x

Anderson, J. C., &Gerbing, D. W. (1988). Structural equation modelling in practice: a review and recommended two-step approach. *Psychol. Bull. 103*, 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Andrich, D., & Marais, I. (2019). *A course in rasch measurement theory: Measuring in the educational, social and health sciences*. Springer Nature Singapore Pte Ltd.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25. https://doi.org/10.1037/amp0000191

Aryadoust, V., Tan, H. A. H., & Ng, L. Y. (2019). A scientometric review of Rasch measurement: The rise and progress of a specialty. *Frontiers in Psychology, 10*(2197), 1–16. https://doi.org/10.3389/fpsyg.2019.02197

Azizah, N., Suseno, M., & Hayat, B. (2023). Item analysis of the rasch model items in the final semester exam Indonesian language lesson. *World Journal of English Language, 12*(1), 15–26. https://doi.org/10.5430/wjel.v12n1p15

Bakker, A. B., Ficapal Cusí, P., Torrent Sellens, J., Boada Grau, J., & Hontangas Beltrán, P. M. (2018). The Spanish version of the job crafting scale. *Psicothema, 30*(1), 136–142. https://doi.org/10.7334/psicothema2016.293

Bates, D., Maechler, M., Jagan, M., Davis, T. A., Karypis, G., Riedy, J., Oehlschlägel, J., & R Core Team. (2024). *Matrix: Sparse and dense matrix classes and methods* [Computer software]. Retrieved from https://CRAN.R-project.org/package=Matrix

Baylor, C., McAuliffe, M. J., Hughes, L. E., Yorkston, K., Anderson, T., Kim, J., & Amtmann, D. (2014). A differential item functioning (DIF) analysis of the communicative participation item bank (CPIB): Comparing individuals with parkinson's disease from the United States and New Zealand. *Journal of Speech, Language, and Hearing Research, 57*(1), 90–95. https://doi.org/10.1044/1092-4388(2013/12-0414)

Bess, T. L., Harvey, R. J., & Swartz, D. (2003). Hierarchical confirmatory factor analysis of the Myers-Briggs Type Indicator. In *Annual Conference of the Society for Industrial and Organizational Psychology, Orlando*, 1-13. https://doi.org/10.1037/E518712013-042

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*(3), 261–280. https://doi.org/10.1177/014662168801200305

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd Ed.). Routledge.

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the rasch model fundamental measurement in the human sciences* (4th Ed.). Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Boone, W. J., & Staver, J. R. (2020). *Advances in rasch analyses in the human sciences*. Springer.

Borsboom, D., & Molenaar, D. (2015). Psychometrics. In J. O. Ramsay (eds.). *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed.). (vol. 18, pp. 12416–12422). Elsevier Ltd. https://doi.org/10.1016/B978-0-08-097086-8.43079-5

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications, Inc.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a metropolis-hastings robbinsmonro algorithm. *Psychometrika, 75*(1), 33-57. https://doi.org/10.1007/s11336-009-9136-x

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265–289. https://doi.org/10.3102/10769986022003265

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, Y., & Zhang, S. (2021). Estimation methods for item factor analysis: An overview. In Y. Zhao & D. Chen (Eds.), *Modern statistical methods for health research* (pp. 329-350). Springer. https://doi.org/10.1007/978-3-030-72437-5_15

Christensen, K. B., Makransky, G., & Horton, M. (2016). Critical values for yen's identification of local dependence in the rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194. https://doi.org/10.1177/0146621616677520

Clark, D. A., & Bowles, R. P. (2018). Model fit and item factor analysis: Overfactoring, underfactoring, and a program to guide interpretation. *Multivariate Behavioral Research, 53*(4), 544–558. https://doi.org/10.1080/00273171.2018.1461058

Comrey, A.L. (1983). An evaluation of the myers briggs type indicator. *Academic Psychology Bulletin, 5*, 115–129. https://psycnet.apa.org/record/1984-19392-001

Dancey, C., & Reidy, J. (2020). *Statistics without maths for psychology* (8th ed.). Pearson.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

De Bruin, G. P. (1996). *An item factor analysis of the Myers-Briggs Type Indicator® (Form G Self-scorable) for a South African sample [Technical report]*. Jopie van Rooyen and Partners.

De Jong, M. G., Steenkamp, J-B. E. M., Fox, J-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research, 45*, 104–115. https://doi.org/10.1509/jmkr.45.1.104

DiStefano, C., Greer, F. W., & Dowdy, E. (2019). Examining the BASC-3 BESS parent form–preschool using rasch methodology. *Assessment, 26*(6), 1162–1175. https://doi.org/10.1177/1073191117723112

Edelsbrunner, P. A., & Dablander, F. (2019). The psychometric modeling of scientific reasoning: A review and recommendations for future avenues. *Educational Psychology Review, 31*(1), 1–34. https://doi.org/10.1007/s10648-018-9455-5

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organisational surveys. *European Journal of Psychological Assessment, 16*, 20–30. https://doi.org/10.1027//1015-5759.16.1.20

Engelhard, G., & Wind, S. A. (2021). A history of rasch measurement theory. In B. E. Clauser & M. B. Bunch (Eds.), *The history of educational measurement* (pp. 343–360). Routledge.

Epskamp, S., Costantini, G., Haslbeck, J., Isvoranu, A., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2023). *qgraph: Graph plotting methods, psychometric data visualization, and graphical model estimation* [Computer software]. Retrieved from https://CRAN.R-project.org/package=qgraph

Furr, R. M. (2022). *Psychometrics: An introduction* (4th ed.). SAGE publications.

Garcia, D., Björk, E., & Kazemitabar, M. (2022). The A (ffect) B (ehavior) C (ognition) D (ecision) of parasocial relationships: A pilot study on the psychometric properties of the Multidimensional Measure of Parasocial Relationships (MMPR). *Heliyon, 8*(e10779), 1–8. https://doi.org/10.1016/j.heliyon.2022.e10779

Geldenhuys, M., & Bosch, A. (2019). A rasch adapted version of the 30-item bem sex role inventory (bsri). *Journal of Personality Assessment, 102*(3), 428–439. https://doi.org/10.1080/00223891.2018.15 27343

Haberman, S. J. (1977). Log-linear models and frequency tables with small expected cell counts. *The Annuals of Statistics, 5*(6), 1148–1169. https://doi.org/10.1214/aos/1176344001

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

Hallquist, M. N., & Wiley, J. F. (2018) MplusAutomation: An r package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 621–638. https://doi.org/10.1080/10705511.2017.1402334

Hallquist, M., Wiley, J., van Lissa, C., & Morillo, D. (2024). *MplusAutomation: An r package for facilitating large-scale latent variable analyses in Mplus* [Computer software]. Retrieved from https://CRAN.R-project.org/package=MplusAutomation

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory library*. SAGE Publications Inc.

Harvey, R. J., Murry, W. D., & Stamoulis, D. (1995). Unresolved issues in the dimensionality of the myers-briggs type indicator®. *Educational and Psychological Measurement, 55*, 535–544. https://doi.org/10.1177/0013164495055004002

Hayat, B., Hidayat, R., & Putra, M. D. K. (2021). Exploring the factor structure of environmental attitudes measure in a sample of Indonesian college students. *Revista CES Psicologia, 14*(1), 112-129. https://doi.org/10.21615/cesp.14.1.9

Horton, M., Marais, I. & Christensen, K.B. (2013). Dimensionality. In Christensen, K. B., Kreiner, S., & Mesbah, M. (Eds.). *Rasch models in health*. Hoboken, NJ: JohnWiley & Sons Inc.

Jiao, H., Liu, J., Haynie, K., Woo, A., & Gorham, J. (2012). Comparison between dichotomous and polytomous scoring of innovative items in a large-scale computerized adaptive test. *Educational and Psychological Measurement, 72*(3), 493–509. https://doi.org/10.1177/0013164411422903

Jöreskog, K. G., & Sörbom, D. (1993). *Lisrel 8: Structural equations modeling with simplis command language*. Hillsdale, NJ: Erlbaum.

Kaplan, R.M., & Saccuzzo, D.P. (2009). *Psychological testing: Principles, applications, and issues*. (7th ed.). Belmont, CA: Wadsworth Cengage Learning.

Kaplan, R. M., & Saccuzzo, D. P. (2018). *Psychological testing: Principles, applications, and issues* (9th ed.). Cengage.

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38*(1), 52–54. https://doi.org/10.5395/rde.2013.38.1.52

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th Ed). The Guilford Press.

Landells, E. M., &Albrecht, S. L. (2019) Perceived organizational politics, engagement, and stress: The mediating influence of meaningful work. *Front. Psychol. 10*(1612), 1–12. https://doi.org/10.3389/fpsyg.2019.01612

Linacre, J. M. (2010). Two perspectives on the application of Rasch models. *European Journal of Physical and Rehabilitation Medicine, 46*(2), 301–302. Retrieved from https://pubmed.ncbi.nlm.nih.gov/20485233/

Linacre, J. M., & Wright, B. D. (2012). *A user's guide to BIGSTEPS: Rasch-model computer program*. Mesa Press.

Linacre, J. M. (2018). *Winsteps® Rasch measurement computer program user's guide* [Computer software]. Beaverton, OR: Retrieved from https://www.winsteps.com/winsteps.htm

Mair, P., & Hatzinger, R. (2007). Extended rasch modeling: The eRm package for the application of irt models in r. *Journal of Statistical Software*, *20*(9), 1–20. https://doi.org/10.18637/jss.v020.i09

Mair, P. (2018). *Modern psychometrics with R*. Springer International Publishing AG.

Mair, P., Rusch, T., Hatzinger, R., Maier, M. J., Debelak, R. (2024). *eRm: Extended rasch modelling* [Computer software]. Retrieved from https://CRAN.R-project.org/package=eRm

Mesbah, M., & Kreiner, S. (2012). Rasch models for ordered polytomous items. In K. B. Christensen, S. Kreiner, & M. Mesbah (eds.), *Rasch models in health* (pp. 27–42). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118574454.ch2

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of nonnormal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*(2), 171–189. https://doi.org/10.1111/j.2044-8317.1985.tb00832.x

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *Journal of Mathematical and Statistical Psychology, 45*(1), 19–30. https://doi.org/10.1111/j.2044-8317.1992.tb00975.x

Myers, I. B. (1998). *Introduction to type : A guide to understanding your results on the MBTI instrument* (6th ed.). Consulting Psychologists Press, Inc.

Myers, I. B., McCaulley, M. H., Quenck, N. L., & Hammer, A. L. (1998). *MBTI® manual*. Palo Alto, CA: CPP, Inc.

Nair, R., Moreton, B., & Lincoln, N. (2011). Rasch analysis of the nottingham extended activities of daily living scale. *Journal of Rehabilitation Medicine, 43*(10), 944–950. https://doi.org/10.2340/16501977-0858

Naisaban, L. (2003). *Psikologi Jung: Tipe kepribadian manusia dan rahasia sukses dalam hidup (Tipe kebijaksanaan Jung).* PT Grasindo.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Osterlind, S. J. (1983). *Test item bias*. Sage Publication.

Padgett, R. N., & Morgan, G. B. (2020). Using the eRm package for Rasch modeling. *Measurement: Interdisciplinary Research and Perspectives, 18*(3), 163–176. https://doi.org/10.1080/15366367.2020.1732155

Patil, I., & Powell, C. (2024). *ggstatsplot: 'ggplot2' based plots with statistical details* [Computer software]. Retrieved from https://cran.r-project.org/web/packages/ggstatsplot/index.html

Pedersen, T. L. (2024). *patchwork: The composer of plots* [Computer software]. Retrieved from https://CRAN.R-project.org/package=patchwork

Periantalo, J., & Azwar, S. (2017). Pengembangan skala kepribadian siswa SMA dari tipologi kepribadian Jung dan Myers-Briggs. *Jurnal Sains Sosio Humaniora*, *1*(2), 191–207. https://doi.org/10.22437/jssh.v1i2.4301

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research*, *57*(3), 210-221. https://doi.org/10.1037/1065-9293.57.3.210

Posit Software, PBC. (2025). *Positron* [Computer software]. https://positron.posit.co/

Price, L. R. (2017). *Psychometric methods: Theory into practice.* Guilford Press.

Privado, J., Sanchis Sanchis, E., Sancho-Cantus, D., Cubero-Plazas, L., Navarro-Illana, E., & de la Rubia Ortí, J. E. (2024). Prediction of caregiver psychological distress in amyotrophic lateral sclerosis: A cross-sectional study. *Rehabilitation Psychology, 69*(4), 364–374. https://doi.org/10.1037/rep0000554

Quenk, N. L. (2009). *Essentials of myers-briggs type indicator assessment* (2nd ed.). John Wiley & Sons, Inc.

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Rahayu, W., Putra, M. D. K., Faturochman, Meiliasari, Sulaeman, E., & Koul, R. B. (2021). Development and validation of Online Classroom Learning Environment Inventory (OCLEI): The case of Indonesia during the COVID-19 pandemic. *Learning Environments Research, 25*(1), 97–113. https://doi.org/10.1007/s10984-021-09352-3

Rahayu, W., Hayat, B., & Putra, M. D. K. (2024). *Analisis rasch: Aplikasi dan interpretasi*. UNJ Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainments tests*. Danish Institute for Educational Research, University of Chicago Press.

Read, H., Fordham, M., Adler, G., & Mcguire, W. (1974). *Psychological types*. Princeton, N.Y: Princeton Press.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*(3), 287-297. https://doi.org/10.1037/1040-3590.12.3.287

Revelle, W. (2024). *psych: Procedures for psychological, psychometric, and personality research* [Computer software]. Retrieved from https://CRAN.R-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Rifenbark, G., Lombardi, A., & Freeman, J. (2021). A confirmatory item factor analysis of a school climate measure for adolescents with and without disabilities. *Educational Assessment, 26*(1), 52-68. https://doi.org/10.1080/10627197.2020.1841625

Roe, C., Damsgaard, E., Fors, T., & Anke, A. (2014). Psychometric properties of the pain stages of change questionnaire as evaluated by rasch analysis in patients with chronic musculoskeletal pain. *BMC Musculoskeletal Disorders, 15*(95), 1–9. https://doi.org/10.1186/1471-2474-15-95

Saggino, A., Molinengo, G., Rogier, G., Garofalo, G., Loera, B., Tommasi, M., & Velotti, P. (2020). Improving the psychometric properties of the dissociative experiences scale (DES-II): A Rasch validation study. *BMC Psychiatry, 20*(8), 1–10. https://doi.org/10.1186/s12888-019-2417-8

Schaubhut, N. A., Herk, N. A., & Thompson, R. C. (2009). *MBTI® form m manual supplement*. Palo Alto, CA: CPP Inc.

Selker, R., Love, J., Dropmann, D., Moreno, V., & Agosti, M. (2023). *jmv: The 'jamovi' analyses* [Computer software]. Retrieved from https://cran.r-project.org/web/packages/jmv/index.html

Sipps, G.J., Alexander, R.A., & Friedt, L. (1985). Item analysis of the myers-briggs type indicator®. *Educational and Psychological Measurement, 45*, 789–796. https://doi.org/10.1177/0013164485454009

Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231. Retrieved from https://www.rasch.org/rmt/rmt161l.htm

Stein, R., & Swan, A. B. (2019). Evaluating the validity of myers-briggs type indicator theory: A teaching tool and window into intuitive psychology. *Social and Personality Psychology Compass, 13*(11), 1–11. https://doi.org/10.1111/spc3.12434

Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial*. Trim Komunikata Publishing House.

Suryadi, B., Hayat, B., & Putra, M. D. K. (2020). Evaluating psychometric properties of the Muslim daily religiosity assessment scale (MUDRAS) in Indonesian samples using the rasch model. *Mental Health, Religion & Culture, 23*(3–4), 331–346. https://doi.org/10.1080/13674676.2020.1795822

Susanto, E., & Mudaim. (2017). Pengembangan inventori MBTI sebagai alternatif instrumen pengukuran tipe kepribadian. *Journal of Educational Counseling*, *1*(1), 41–52. https://doi.org/10.30653/001.201711.5

Temel, G. Y., Rietz, C., Machunsky, M., & Bedersdorfer, R. (2022). Examining and improving the gender and language DIF in the VERA 8 Tests. *Psych, 4*(3), 357–374. https://doi.org/https://doi.org/10.3390/psych4030030

Thompson, B., & Borrello, G. M. (1986). Construct validity of the myers-briggs type indicator®. *Educational and Psychological Measurement, 46*, 745–752. https://doi.org/10.1177/0013164486463032

Tischler, L. (1994). The MBTI® factor structure. *Journal of Psychological Type, 31*, 24–31. Retrieved from https://psycnet.apa.org/record/1995-19764-001

Tzeng, O.C.S., Outcalt, D., Boyer, S.L., Ware, R., & Landis, D. (1984). Item validity of the myers-briggs type indicator®. *Journal of Personality Assessment, 48*, 255–256. https://doi.org/10.1207/s15327752jpa4803_4

Umar, J., & Nisa, Y. F. (2020). Uji validitas konstruk dengan CFA dan pelaporannya. *JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia), 9*(2), 1–11. https://doi.org/10.15408/jp3i.v9i2.XXXXX

van Zyl, C. J. J., & Taylor, N. (2012). Evaluating the MBTI® form m in a South African context. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 38*(1), 1–15. https://doi.org/10.4102/sajip.v38i1.977

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicología, 33*(3), 755–782. https://doi.org/10.6018/analesps.33.3.268401

Wang, J., & Wang, X. (2020*). Structural equation modeling: Applications using Mplus* (2nd Ed.). Hoboken, NJ: Wiley.

Wen, Z.-L., & Ye, B.-J. (2011). Evaluating test reliability: From coefficient alpha to internal consistency reliability. *Acta Psychologica Sinica, 43*(7), 821–829. https://journal.psych.ac.cn/acps/EN/Y2011/V43/I07/821

Wickham, H., Bryan, J., Kalicinski, M., Valery, K., Leitienne, C., Colbert, B., Hoerl, D., & Miller, E. (2023a). *readxl: Read excel files* [Computer software]. Retrieved from https://cran.rproject.org/web/packages/readxl/index.html

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2023b). *ggplot2: Create elegant data visualisations using the grammar of graphics* [Computer software]. Retrieved from https://cran.rproject.org/web/packages/ggplot2/index.html

Wickham, H., & RStudio Team. (2023c). *tidyverse: Easily install and load the 'Tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wihardini, D. (2020). Assessing lecturer's understanding of plagiarism: A case in Indonesia. *JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia, 9*(2), 12–43. https://doi.org/10.15408/jp3i.v9i2.16043

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.

Wu, M., Tam, H. P., & Jen, T. (2016). *Educational measurement for applied researchers: Theory into practice*. Springer Nature. https://doi.org/10.1007/978-981-10-3302-5

Yan, Z., & Mok, M. M. C. (2012). Validating the coping scale for Chinese athletes using multidimensional Rasch analysis. *Psychology of Sport and Exercise, 13*(3), 271–279. https://doi.org/10.1016/j.psychsport.2011. 11.013

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125–145. https://doi.org/10.1177/014662168400800201

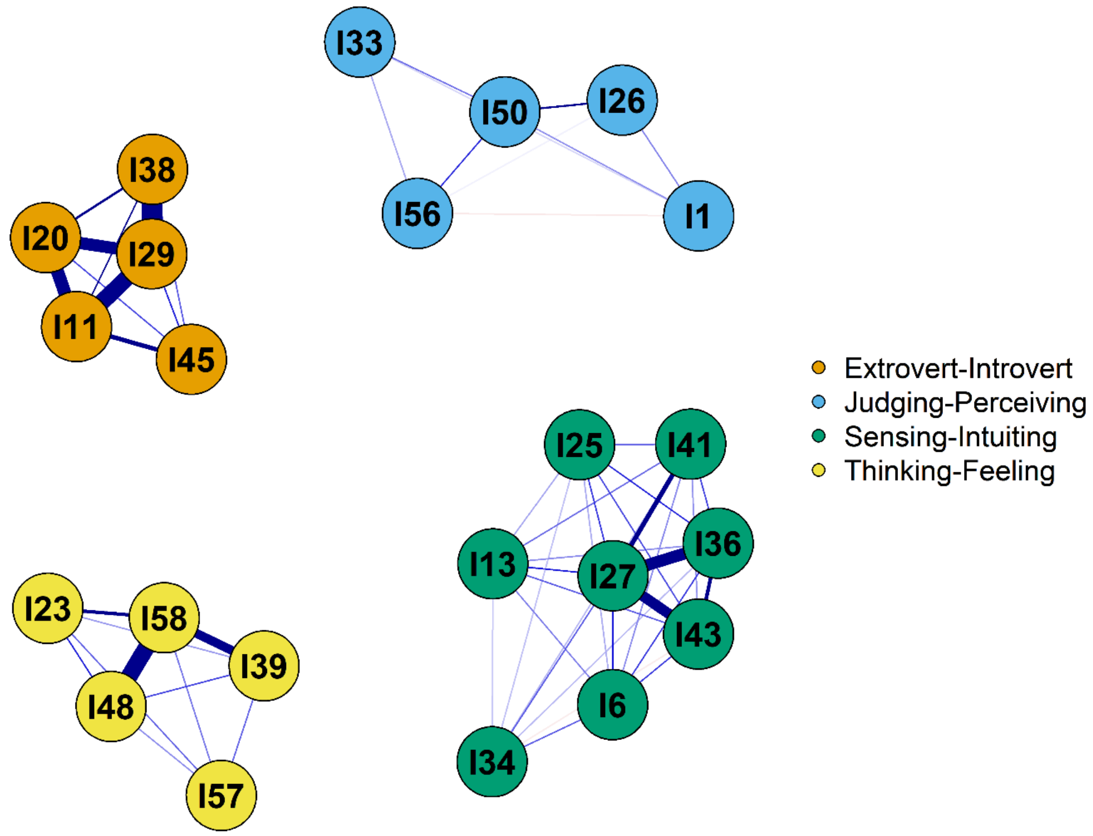
Appendix

Appendix 1. Descriptive statistics of items

Item	M	SD	Skewness	Kurtosis	Item	M	SD	Skewness	Kurtosis
EI2	0.173	0.378	1.730	0.993	TF4	0.693	0.461	-0.839	-1.296
EI5	0.828	0.377	-1.739	1.023	TF9	0.137	0.343	2.116	2.478
EI7	0.290	0.454	0.924	-1.147	TF14	0.313	0.464	0.805	-1.352
EI10	0.599	0.490	-0.403	-1.838	TF17	0.409	0.492	0.372	-1.862
EI11	0.482	0.500	0.072	-1.995	TF23	0.497	0.500	0.013	-2.000
EI15	0.028	0.166	5.673	3.019	TF30	0.686	0.464	-0.802	-1.357
EI20	0.383	0.486	0.483	-1.767	TF32	0.972	0.166	-5.673	3.019
EI28	0.743	0.437	-1.111	-0.767	TF37	0.725	0.447	-1.006	-0.988
EI29	0.721	0.449	-0.985	-1.030	TF39	0.685	0.465	-0.795	-1.368
EI31	0.972	0.165	-5.702	3.052	TF42	0.880	0.325	-2.342	3.487
EI35	0.793	0.405	-1.445	0.088	TF48	0.713	0.453	-0.940	-1.116
EI38	0.282	0.450	0.969	-1.061	TF49	0.896	0.306	-2.586	4.690
EI45	0.202	0.401	1.485	0.207	TF55	0.737	0.440	-1.075	-0.844
EI52	0.244	0.429	1.193	-0.577	TF57	0.697	0.460	-0.857	-1.265
EI60	0.541	0.498	-0.163	-1.974	TF58	0.605	0.489	-4.294	-1.816
SI6	0.368	0.482	0.549	-1.698	JP1	0.456	0.498	0.177	-1.969
SI8	0.605	0.489	-0.428	-1.817	JP3	0.835	0.371	-1.808	1.270
SI13	0.830	0.376	-1.753	1.075	JP12	0.042	0.201	4.558	1.878
SI16	0.968	0.176	-5.303	2.612	JP19	0.371	0.483	0.533	-1.716
SI18	0.588	0.492	-0.358	-1.872	JP21	0.192	0.394	1.563	0.442
SI22	0.367	0.482	0.552	-1.696	JP24	0.319	0.466	0.776	-1.397
SI25	0.715	0.452	-0.951	-1.096	JP26	0.302	0.459	0.863	-1.256
SI27	0.591	0.492	-0.369	-1.864	JP33	0.328	0.469	0.734	-1.461
SI34	0.638	0.481	-0.573	-1.672	JP40	0.208	0.406	1.440	0.074
SI36	0.437	0.496	0.252	-1.937	JP44	0.739	0.439	-1.089	-0.814
SI41	0.360	0.480	0.582	-1.662	JP47	0.021	0.143	6.681	4.264
SI43	0.675	0.468	-0.749	-1.439	JP50	0.661	0.473	-0.682	-1.536
SI46	0.884	0.320	-2.400	3.760	JP54	0.500	0.500	-0.001	-2.000
SI51	0.422	0.494	0.317	-1.900	JP56	0.863	0.343	-2.116	2.478
SI53	0.531	0.499	-0.125	-1.985	JP59	0.256	0.436	1.120	-0.746

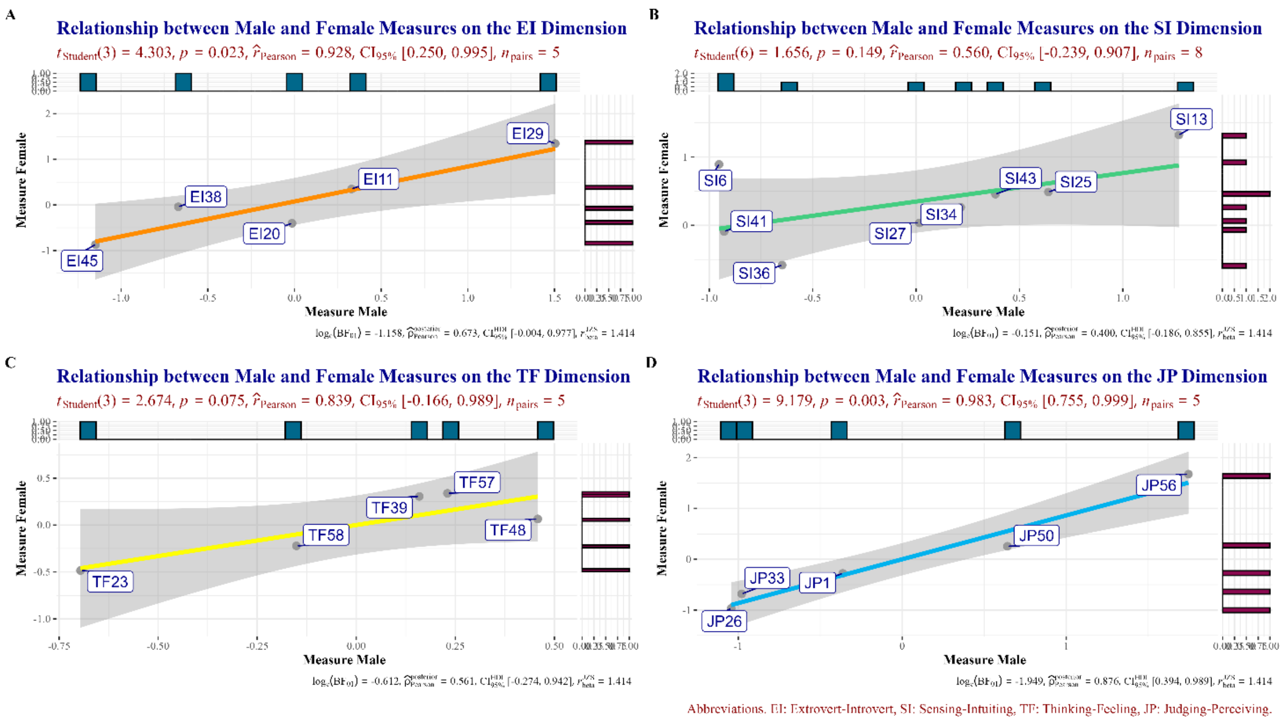
Sources: Research data (2019). Note: EI = Extrovert–Introvert; SI = Sensing–Intuiting; TF = Thinking–Feeling; JP = Judging–Perceiving.

Appendix 2. Tetrachoric correlation network between all pairs of items in each dimension



Sources: Personal data (2019).

Appendix 3. Scatter plot of correlation between male and female measures for each MBTI dimension



Sources: Personal data (2019).