

Investigating the Psychometric Properties of the Myers-Briggs Type Indicator

Kritika Rajeswari S.¹, Surej Unnikrishnan², Vrinda Kamath³

¹8th Semester Integrated B.Sc. - M.Sc. Psychology (Specialization in Clinical Psychology)

²Assistant Professor, Post Graduate Department of Psychology, Jyoti Nivas College Autonomous

³8th Semester Integrated B.Sc. - M.Sc. Psychology (Specialization in Clinical Psychology)

Abstract

The Myers-Briggs Type Indicator (MBTI) is a largely used personality assessment tool based on the psychological type theory proposed by Carl Jung. It bifurcates people into 16 personality types across four bipolar dimensions: Extraversion–Introversion, Sensing–Intuition, Thinking–Feeling, and Judging–Perceiving. Reliability refers to the consistency of a test over time, and Validity, measures how well a test assesses what it claims to measure. This study aims to investigate the psychometric properties of reliability and validity of the Myers-Briggs Type Indicator (MBTI), a widely used assessment. A systematic review of journal articles, thesis material published between 2017 and 2025 was conducted using platforms such as Google Scholar, Research Gate and Wiley Online Library. The results of the study show inconsistent test-retest reliability, with 50% of participants receiving different type results on repeated testing. Certain MBTI domains such as Extraversion-Introversion, displayed stronger reliability while others, like Judging-Perceiving and Thinking-Feeling show weaker psychometric properties. Additionally, the MBTI falls short on predictive validity and has been criticized for its binary typological model, which fails to include the diverse continuum of personality traits. The study findings highlight the need for application with caution of the MBTI, in both clinical and organizational settings due to its limitations in reliability and validity.

Keywords: Myers Briggs Type Indicator (MBTI), Reliability, Validity

Introduction

According to Allport (1961), “Personality is the dynamic organization within the individual of those psychophysical systems that determine his characteristic behavior and thought” (Allport, 1961, p. 28). Personality refers to the enduring characteristics and behavior that comprise a person’s unique adjustment to life, including major traits, interests, drives, values, self-concept, abilities, and emotional patterns. (American Psychological Association, 2018) The Myers-Briggs Type Indicator (MBTI) is developed by Katharine Cook Briggs and Isabel Briggs Myers. It is based on the psychological theories of Carl Gustav Jung, particularly his work on psychological types (Jung, 1921). The MBTI assessment is designed to help people identify and understand how they perceive information and make decisions, as well as their judging and social interaction preferences. The MBTI categorizes personality based on four

opposing dichotomies. The test evaluates personality in four key areas: energy, perception, decisions, and orientation.

Extraversion (E) vs. Introversion (I)

Energy focuses on opposite ways of receiving and directing oneself and can be broadly classified into introversion and extraversion. People may be extroverted (E), or Introverts (I). The former focus on the outer world of people and activity vs. the latter tend to focus on the inner world of thoughts and reflections. According to De Boer and Bothma (2020), extroverts try things out, focus on the outer world of people. On the other hand, introverts generally think things thoroughly and are more introspective in nature.

Sensing (S) vs. Intuition (N)

Perception describes how people prefer to take in information. A person may be sensing (S), meaning they prefer to use their senses to perceive the external world, or intuitive (N), meaning they rely on instincts, and noticeable patterns. Preference for concrete, factual information vs. abstract concepts and patterns. Sensors pay more attention to details, and focus on facts and procedures, while “intuitors” are imaginative, creative, concept-oriented, and focus on possibilities.

Thinking (T) vs. Feeling (F)

This dimension measures whether a person relies more on thinking (T) or feeling (F) to evaluate decisions which lead to conclusions. The thinking type people prefer rationale, while feeling refers to those who seek consonance and use their emotions. This dichotomy talks about decision-making based on objective logic vs. personal values and emotional impact. Thinkers are usually more guarded, and careful, making decisions based on logic, rules and factual information. Feelers are accommodative, and make decisions based on personal and humanistic considerations, therefore appearing to be more empathetic in their decision making.

Judging (J) vs. Perceiving (P)

Orientation describes how people prefer to approach the outside world. On a scale from judging (J) to perceiving (P). Judging type people work better with order and stability, while perceiving types thrive on flexibility. Preference for structure and planning vs. spontaneity and adaptability. Judgers usually set and follow schedules and require closure, even with incomplete data. Perceivers are more adaptive to changing circumstances and do not seek closure without the required data. (De Boer and Bothma, 2020)

These dimensions combine to form 16 personality types, such as INTJ, ESFP, INFP, ISTJ, ISFJ, INFJ, ISTP, ISFP, INTP, ESTP, ENFP, ENTP, ESTJ, ENFJ, ESFJ, ENTJ, each offering a distinct profile of behavior and thinking styles.

Methods

The investigators used search engines like *Google Scholar*, *Research Gate*, and *Wiley Online Library* to identify the relevant studies for the systematic review. The keyword search used the following terms in various combinations: reliability, validity, and Myers-Briggs Type Indicator. A date limit of 7 years before 2025 was set to ensure that only the latest findings were included.

Studies were selected according to specified inclusion criteria, which included the following: primary research, indexed in Scopus or peer-reviewed journals, thesis content and analyses post 2017, since there was a meta-analysis review paper of the same by Randall (2017). Papers were excluded on the following basis: books or book chapters, studies with only abstracts available, published in languages other than English, and those not indexed in Scopus or UGC Care Lists. Two independent researchers (KS and VK) conducted the search process. In addition, the authors also searched through references of each article for further research content. The titles and abstracts were reviewed to check whether the articles were relevant.

Study Selection and Categorization

Studies were initially checked for eligibility by reading the abstract. If most of the inclusion criteria were met, the full article was retrieved and rechecked to confirm eligibility. In case it was evident from the abstract that the article was not suitable, then the article was excluded without further inquiry. The studies were then categorized into two broad emerging themes, namely Reliability and Validity, post which the interrelatedness of the themes was analyzed using the gathered data.

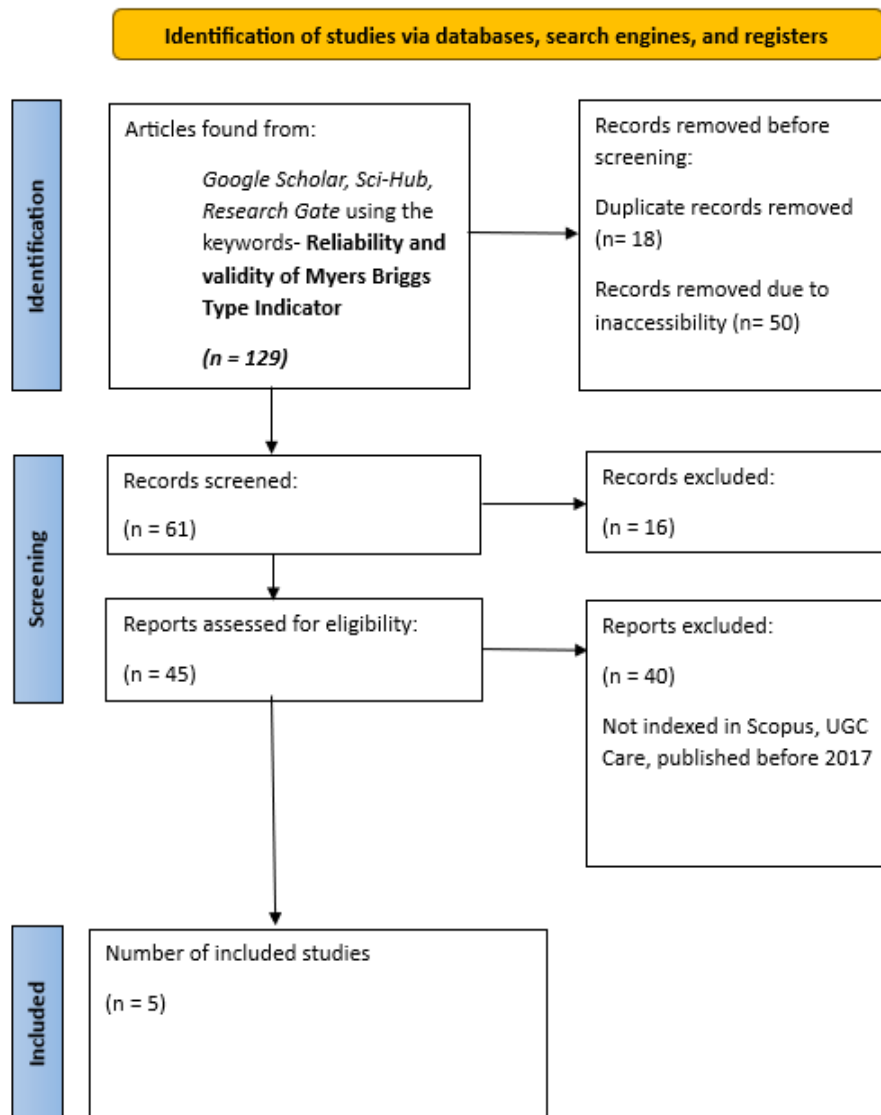
Data Extraction

Information was extracted and entered into Microsoft Excel under the following fields: Title of the Study, Year, Author, Journal Name, Sample, and Findings. Finally, the PRISMA Flow diagram was used to show the process of selection and elimination of papers.

Results

The initial search with the keyword's reliability and validity of Myers-Briggs Type Indicator on Google Scholar showed 129 studies (refer to Appendix B), but not all of them were relevant to the topic. 18 records were removed as duplicates, while 50 were removed due to inaccessibility, and 16 were deemed irrelevant based on the inclusion criteria. Of these, 45 studies were checked for eligibility, and 5 studies were included in the systematic review. The search ended on April 29th, 2025. All included studies and their characteristics are listed in Table 1 (refer to Appendix C).

Figure 1. PRISMA Flow diagram



Discussion

Theme 1- Reliability

According to a study done by Francis J. and Village A. (2022), the reliability of the MBTI has been assessed in two main ways: through test-retest reliability based on stability of categorizations into the sixteen types, and through examining the continuous scale scores. Test-retest reliability studies have been conducted over various time intervals, with reported correlations ranging from .56 to .89 for different scales and sample groups. Thus, while the scale shows relative unreliability in sorting individuals into discrete type categories, it demonstrates relative reliability in bifurcating individuals based on assessing perceiving, judging, orientation, and attitude towards the outer world.

Mikayla A. Burton (2022) states that to examine the test-retest reliability of the MBTI, participants were asked to complete the self-questionnaire once, and then do it again post five weeks. In just over a month, approximately 50% of participants received a different personality type (Pittenger, n.d.). Thus, the Myers Briggs Type indicator was found to have poor test-retest reliability, since nearly 75% of retests resulted in the individual getting a new personality type than the previous time it was conducted.

Tomat, Trkman, and Manfreda (2021) found that with 452 participants tested for eligibility for various types of information systems (IS) professions using the MBTI was a reliable and trustworthy indicator of candidates' personalities. They stated that when the MBTI was interpreted properly, the test can give a reasonably reliable result for selecting an employee for a particular information systems job.

According to Stein R. and Swan A. (2019), the reliability of the MBTI has been a significant point of criticism in the literature. The MBTI manual reports acceptable levels of internal consistency; the test-retest reliability, particularly over longer intervals, shows greater variability. For instance, the study highlights findings where up to 50% of respondents change their type upon retesting within 5 weeks, to question the stability of the MBTI's claim to identify enduring personality "types." Given the polar and categorical nature of MBTI scoring, even small shifts in responses can lead to type changes, especially for individuals near the midpoint of a dimension. The study also found that while the MBTI can show moderately reliable continuous scale scores, its dichotomous type classifications are far less stable, which affects both personal and organizational decisions based on MBTI typing.

Theme 2- Validity

Criterion-Related Validity

Helen Farabee Centers (2022) states that in assessing the criterion-related validity of the test, the National Academy of Sciences (1991) compared each of the four dimensions of the MBTI to similar scales, designed to measure similar constructs. While the introversion-extraversion domain of MBTI showed a strong correlation with tests that measured similar concepts, and a weak correlation with tests that measured different ones, the thinking-feeling, sensation-intuition, judging-perceiving domains had weak criterion-related validity. (Burton, 2022)

Predictive Validity

Stein, R., and Swan, A. (2019), in their study, mention that a significant issue with the MBTI is its lack of testability. The theory does not generate empirical predictions that can be tested and verified, which is an important criterion for evaluating any scientific theory. Although the MBTI categorizes individuals in ways that may seem useful, it lacks predictive validity. This means that the categories it creates do not reliably predict important outcomes, such as job satisfaction or performance in various roles. Thus, the paper argues that while the MBTI is widely popular and used, it falls short on several critical criteria for scientific validity,

including agreement with known data, internal consistency, and testability. Its continued use may reflect more about public perception than its actual scientific merit.

Interrelatedness of Themes

The studies above pointed to a conclusion that the introversion-extraversion domain of the MBTI shows higher test-retest reliability than the other 3 dichotomies of sensing-intuition, thinking-feeling, and judging-perceiving. Papers that derived the conclusion that MBTI had poor test-retest reliability also noted that there was poor concurrent and predictive validity.

A major critique of the MBTI pertains to its binary classification approach, where individuals are assigned to one of two polar categories within each of the four dichotomies. Stein and Swan (2019) have critically examined the Jungian theoretical foundations of the MBTI, arguing that the concept of individuals possessing a "true type" is overly simplistic and lacks empirical support. They claim that the binary nature of MBTI does not capture the continuum of personality traits, potentially neglecting those whose characteristics fall under the dichotomous categorizations.

The MBTI has received criticism from trait theorists stating that people cannot be classified dichotomously (King and Mason, 2020). Despite this, MBTI researchers have shown different studies with acceptable indexes of reliability and validity. (King and Mason, 2020)

Limitations

There are several limitations to a systematic review that have to be acknowledged. This review relies solely on published studies, especially some without a sample size, which makes the study lean towards a qualitative one. Studies that were not indexed in Scopus or the UGC Care List were not included, which limited the possibilities of finding more valuable data that matched the inclusion criteria. The studies included vary in quality, methodology, and psychometric characteristics. This hampers the validity of the consequent findings as the data is not homogeneous. There may also be bias by the researchers in the method of selection, as well as analysis, which does not allow for complete generalization, making room for error. Keeping these limitations in mind, the authors concede that while the conclusions drawn are coherent and theoretically efficient, they may not always hold.

Conclusions

The results of this review paper show that the test-retest reliability of the Myers-Briggs Personality Indicator is inconsistent, unless used and interpreted with precision. While continuous scale scores show moderate reliability, categorical type assignments are unstable. Studies (e.g., Burton, 2022) show that up to 50% of participants receive different MBTI types within a 5-week gap, raising concerns about its temporal stability. Tomat et al. (2021), however, found MBTI to be reasonably reliable for screening IS professionals if interpreted correctly.

The study revealed that the introversion-extraversion scale aligns strongly with similar psychological constructs, while the other three dimensions of sensing-intuition, judging-perceiving, and thinking-feeling lack criterion concurrent validity with scales like FPTS and NEO-5 Inventory. Predictive validity of the MBTI is also lacking as it does not reliably predict outcomes like job performance or satisfaction, in workplace settings for which it is so popularly used. In conclusion, the MBTI's binary structure is widely criticized for oversimplifying personality, ignoring trait continuums as well as having an inconsistent test-retest reliability, predictive validity, and criterion concurrent validity, as of studies from 2017 to 2025.

Critiques emphasize that small response changes can shift personality type, especially for individuals near the dichotomy midpoints, like that of introversion and extraversion. Further research for more reliable scales like the NEO-5 Factor Inventory (Costa & McCrae, 1992) and the HEXACO Personality Inventory (Ashton & Lee, 2009) can be used, which is beyond the scope of this particular study's objective.

Conflict of Interest

There is no conflict of interest among the authors.

References

1. Francis, L. J., & Village, A. (2022). The Francis Psychological Type Scales (FPTS): factor structure, internal consistency reliability, and concurrent validity with the MBTI. *Mental Health, Religion & Culture*, 25(9), 931–951. <https://doi.org/10.1080/13674676.2022.2041584>
2. 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(2), e12434. <https://doi.org/10.1111/spc3.12434>
3. Burton, M. (n.d.). *Evaluating validity and deviant personality: What the Myers Briggs type indicator overlooks*. Scholars Crossing. <https://digitalcommons.liberty.edu/honors/1208/>
4. Tomat, L., Trkman, P., & Manfreda, A. (2021). Personality in information systems professions: identifying archetypal professions with suitable traits and candidates' ability to fake-good these traits. *Information Technology and People*, 35(8), 52–73. <https://doi.org/10.1108/itp-03-2021-0212>
5. King, S. P., & Mason, B. A. (2020). Myers-Briggs Type Indicator. *The Wiley Encyclopedia of Personality and Individual Differences*, 315–319. <https://doi.org/10.1002/9781119547167.ch123>

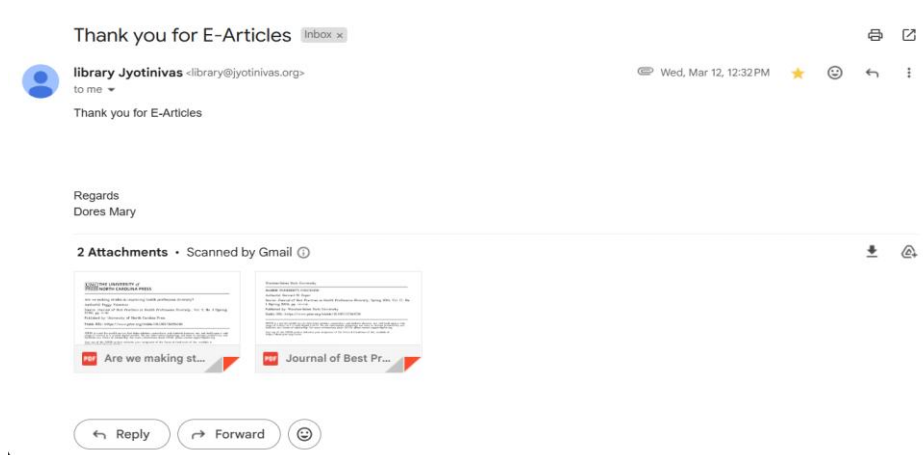
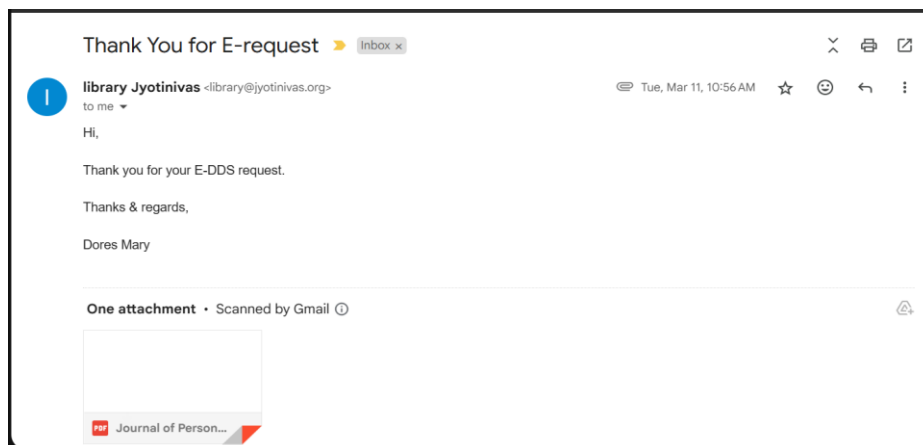
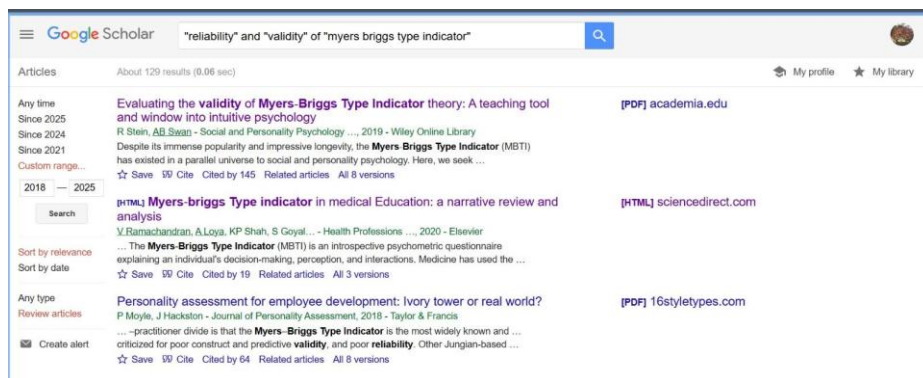
Appendix

A- Contributions

Kritika- Introduction, Study selection and Screening, PRISMA Flowchart, formatting, methods, conclusion, limitations, Excel sheet.

Vrinda- Introduction, Abstract, Study selection and Screening, results, methods, limitations, Table indexed.

B- Proof of Initial Pool



C. Studies Included in the Review

Table 1 shows the final studies included in the systematic review

Sl no.	Year	Author	Sample	Findings	Journal
1	2022	Burton M.	N/A	The MBTI was found to have poor test-retest reliability, with some researchers estimating that nearly 75% of test-takers will receive a different result each time they take the test. It also had poor criterion concurrent validity	Thesis Submission, Liberty University
2	2022	Francis J., Village A.	165 participants (65 male, 120 female)	The scale shows relative unreliability in sorting individuals into discrete type categories, it demonstrates relative reliability in grading individuals on the four continua assessing orientation, perceiving, judging, and attitude toward the outer world.	Mental Health, Religion, Culture (Taylor & Francis online)
3	2021	Tomat, Trkman, and Manfreda	452 participants	When the MBTI was interpreted correctly, the test can give a reasonably reliable result when selecting an employee for a particular IS job.	Information Technology & People, Vol. 35, Issue 8
4	2020	King S.P., Mason B.		The MBTI has received criticism from trait theorists stating that people cannot be classified dichotomously. Despite this, MBTI researchers have shown different studies with acceptable indexes of reliability and validity.	Wiley Encyclopedia of Personality and Individual Differences (315–319)
5	2019	Stein R., Swan A.	N/A	MBTI can show moderately reliable continuous scale scores, its dichotomous type classifications are far less stable, which affects both personal and organizational decisions based on MBTI typing. It shows poor predictive validity.	Social and Personality Psychology Compass, Vol. 13, Issue 2