

Assessment Critique of Meyers-Briggs Type Indicator (Form M)

LIU CW Post



### **General Features**

The Myers-Briggs Type Indicator (Form M) instrument, authored by Katharine Briggs and Isabel Briggs, is one of the most commonly used tools in assessing personality characteristics. Over two million assessments have been administered since its creation. It is used throughout the world and has been translated into many different languages. The instrument was published in 1998, by Consulting Psychologists Press, Inc. (CPP). The test is intended to measure personality traits in order to give a person insight into his overall personality strengths. Each personality trait is assigned a letter code. At the end of the test, an overall personality profile combines four of the traits based on test scores. The instrument is written at a 7<sup>th</sup> grade level, so it is recommended for children age 14 through adults. When discussing test user qualifications it is important to note that the test is most effective when administered by a person who is educated in how to administer the MBTI as well as how to score it, interpret it, and make recommendations (Turner, DeMers, Fox, and Reed, 2001). Turner et. al. (2001) also note that APA guidelines state, "the phrase test user qualifications refers to the combination of knowledge, skills, abilities, training experience and where appropriate, practice credentials that the APA considers desirable for the responsible use of psychological testing" (p. 1099).

### **Description of Instrument**

The purpose of the instrument is to sort people into groups by personality characteristics while giving a person insight into how he sees the world as well as how he sees others in the world. The test measures 4 sets of traits, which are dichotomous. Each personality trait is assigned a letter code. For example, E=extraversion, I=introversion, S=sensing, I=intuition, T=thinking, F=feeling, J=judging, and P=perceiving. The result is a four-letter code, which can yield 16 different possibilities, describing on overall personality type. Capraro & Capraro (2002)

state, "the purpose of the MBTI is to identify, from self-report of easily recognized reactions, the basic preferences of people in regard to perception and judgment, so that the effects of each preference, singly and in combination, can be established by research and put to practical use" (p. 593).

The test is designed to be used with high school and college aged students as well as any person over the age of 14. It can be administered to people who are interested in finding information about their personality traits that could help them with career counseling, team building, leadership, life coaching, marriage counseling or even just for personal understanding, growth and development.

The test can be administered by a trained professional or self-administered. If a trained professional is administering the test, he would read the printed directions and explain the importance of confidentiality. The test can be done online where the test taker enters the answers and the online site provides scores immediately following completion. The test can also be done on a score sheet with pencil where the answers are marked on the score sheet and the test administrator manually scores the answers upon completion. If the test is self-administered, there are easy to follow directions and the test can be self-scored or scored online. Sewall (1996) explains,

"the MBTI can be hand scored or processed by computer on a dual-purpose answer sheet. Answer keys for hand-scoring are easy to use and contain explicit instructions. Since a complete set of instructions is printed on each key, scoring can begin with any key and proceed in any order. Tables for converting raw scores to linearly derived preference scores are also printed on each answer key for ease of scoring" (p. 9-10).

However, it is not recommended that a person self-administer because the test taker will not fully understand how to interpret the scores and therefore, may make incorrect assumptions about his personality characteristics.

The scores obtained are raw scores for each polarity (E-I, S-N, T-F, J-P) and a preference score (number), which shows the strengths of the polarities. "While the letters indicate the direction of the preference, the numbers reflect strength of an individual's identification with their preference. A person who is more developed and self-knowing will test more convincingly towards one pole under each of the four indexes" (Barbuto, 1997, p. 616). For example a preference score of N9 and N41 shows a difference in preference in intuition, but not necessarily that the N41 score means a person is more sure than the N9.

### **Psychometric Properties**

A norm group is defined as the group of individuals who took the test on whom the test was standardized. The norm group used for the MBTI was

"drawn from hundreds of thousands of respondents who have completed the MBTI assessment. Participants who fit the demographic profile needed for each analysis were selected from the commercial database. Participants within each of those groups were then randomly selected to create an analysis sample with an appropriate size and equal numbers of men and women, when possible" (Schaubhut et al., 2009, p. 4).

The MBTI is available in over 21 languages, including for example, Spanish, German, French, Chinese and many more (Schaubhut et al., 2009). "Both Jung and Myers felt that psychological type is universal. Therefore, the implications of promoting understanding between cultures and increasing appreciation of diversity within a culture are significant" (Ring, 2008, p. 10). There

seems to be research that shows significant success with using the MBTI across multiple cultures.

Reliability is defined as consistency of scores or measurement. An assessment is considered reliable when it proves consistent results over time. There are two widely used types of reliability measures, which are test-retest reliability and internal consistency reliability. Test-retest measures and compares the test takers' scores from one test administration to the next. Internal consistency tells how well each item relates to the other items on a test and how the items relate to the overall score. (Schaubhut et al., 2009)

As we have studied in class, a good reliability score is a score greater than .80. In the MBTI manual supplement, Schaubhut et al. (2009) state their findings for test-retest reliability as strong. Test takers completed the test two times between January 2004 and September 2008. "The correlations with the MBTI dichotomies range from .57 (S-N, 6-12 months interval) to .81 (T-F, < 3 weeks interval), indicating good reliability for each preference over long periods of time" (Schaubhut et al., 2009, p. 7). In another study, Capraro & Capraro (2002) found that

"test-retest reliabilities for MBTI scores suggest score consistency over time.

Test-retest coefficients from 1 week to 2.5 year intervals ranged from .93 to .69 on the SN scale, .93 to .75 on the EI scale, .89 to .64 on the JP scale and .89 to .48 on the TF scale" (p.594).

Carlson (1985) also noted that he "reviewed test-retest reliability studies and concluded that scores of college students appear to be reasonably stable over time in terms of type category, while older subjects' scores have typically yielded higher coefficients with Pearson rs ranging from .69 to .83 across two month periods" (p. 358).

In general, the internal consistency of the MBTI is more reliable when the assessment is studied across several different groups of people that share common demographics. The MBTI

manual, studied groups based on employment status, age, ethnicity and a person's country of origin. The internal consistency reliability examining employment shows "The reliabilities for all five employment status categories are high, ranging from .86 to .92. These results indicate that the MBTI instrument can be used reliably to assess type preferences across a variety of employment situation" (Schaubhut et al., 2009, pp.4-5). The results for the study based on age and ethnicity were similar. The reliabilities were good therefore indicating the MBTI can be used with all age groups and ethnic groups reliably. Finally, Schaubhut et al. found that the internal consistency reliability was good, ranging from .81 to .91, when the test was administered to people around the world encompassing many different countries.

In addition, Thompson and Borrello (1986) state that, "in various studies the internal stability reliability coefficients for the four scales have tended to vary between .80 and .90" (p.746). Likewise, Carlson (1985) said, "studies that are available show satisfactory internal consistency of each of the four scales and, with the possible exception of the TF scale, satisfactory stability of scores across several months" (p.359). Ring (2008), also states, that "results of meta-analytic studies, using generally accepted standards applied to instruments with continuous scores, show the reliabilities of the MBTI continuous score to be quite good—average overall reliabilities of .84 and .86 for internal consistency measures. These compare quite well with reliabilities of even the most well-established and respected trait-based instruments" (p.6).

Validity is defined as the accuracy to measure what the instrument is designed to measure. Validity can be measured through behavioral observations, factor analysis, and correlations with related instruments. In order for a test to show strong validity a score of .50 or greater should be obtained. Schaubhut et al. (2009) concluded that,

“Validity was established in several ways. First, are correlations of the MBTI Form M assessment with six other assessments. The correlations show expected relationships with these other instruments. Next, results of best-fit type analyses are shown to be similar to results from previous research, with high rates of agreement between reported and best-fit types and discrepancies occurring more frequently for those with preference clarity indexes in the slight range. Finally, factor analysis shows the expected four-factor structure of the assessment” (p. 17).

Content validity involves how adequately the test represents the content domain of the identified construct. This can be done by consulting expert’s opinions from the field. The MBTI was developed using Jung’s theory of personality therefore using similar content. Many experts agree that the MBTI is one of the most important tools in measuring personality traits as evidenced by the measurement of E-I, S-N, T-F, and J-P traits.

Construct validity pertains to the appropriateness of inferences drawn from test scores on the construct being measured. “Convergent validity can be demonstrated when results of an assessment are related to results of other similar measures” (Schaubhut et al., 2009, p.9). The MBTI Form M manual supplement illustrates the correlations between the MBTI and several tests such as the CPI 260, FIRO-B, Adjective Check List, Strong Interest Inventory, Thomas-Kilmann Conflict Mode Instrument, and Birkman Method. Overall, most instruments showed a strong correlation to the MBTI instrument. In addition, Capraro & Capraro (2002) state, “Validity of MBTI scores is typically established by correlating the scores with findings from various personality instruments and inventories of interest. Statistically significant correlations have been found between MBTI scores, behaviors reflective of MBTI constructs, and persons’ self-assessment of their own MBTI type” (p. 594). Using factor analysis, Thompson and

Borrello (1986) found that the items had coefficients that were higher than .30, which shows a good measure of reliability.

Criterion-related validity refers to the relationship between test results and some external variables associated with the identified construct. Carlson (1985) noted, "eight reports of attempt to correlate the MBTI with a variety of behaviors in treatment and research setting were generally positive with respect to the hypotheses tested, with some exceptions. It is to the credit of the MBTI that the instrument successfully predicted behaviors as far ranging as couples problems to story imagery and group conformity" (p. 364). A study by Dewinne and Johnson, examined the how personality traits results from the MBTI may predict drug abuse behavior. They used the definition of an extrovert as someone who is impulsive and inattentive to possible effects to their body. They found in their research studying drug abusers, that there was a significantly larger percentage of extroverts than introverts in the sample they studied. Therefore, finding that the personality trait of extrovert may predict future drug use (Carlson, 1985). While there is a large volume of research that shows strong validity psychometrics for the MBTI Form M, there are also many studies that find inconclusive or contradictory results. However, the MBTI is still the most widely used personality instrument.

### **Practical Aspects**

When administering or taking the MBTI, there is no time limit. It can take an average of 20-30 minutes to complete the test. The scoring or interpretation will take extra time. It is easy to administer the test because the directions are clearly stated and the questions are 2-choice forced answers. Therefore, it is not a complicated process. If it is completed with paper and pencil, there is a score sheet provided for recording answers. The administrator tallies the number of responses corresponding with each letter to obtain a raw score. That information is

entered onto a score sheet indicating the strength of each polarity. A final 4 letter score is obtained based on the 4 highest scores within the polarities. A chart is provided, illustrating a detailed explanation of the person's 4-letter code. For example, a score of ISTJ would indicate a person was quiet, serious, earned success by thoroughness and dependability, practical, realistic and responsible, among other characteristics.

If the test is taken online, it is easier because all of the answers are recorded directly on the website page and the test is automatically scored. The results are given immediately following completion, and a description of the personality profile is also given. There are also additional links offered to obtain further information and insight into the personality description.

There is a cost associated with the use of the MBTI. The complete manual currently costs \$142. The test booklets and answer sheet come in a set. There are two options. The basic self-scoring test booklet and answer sheet costs \$14.95 each. The MBTI complete test booklet includes an additional supplement of assessment and interpretation information of the results, which costs \$39.95 each. The computer generated instrument costs \$49.95.

The MBTI manual contains 440 pages and is available in paperback. It's most recent release date was 1/1/98. The manual is well organized and easy to read. It has an introduction that gives an overview of the important features of the MBTI, reviews the history of it's conception and development and describes the reasoning for the form M addition. The manual gives an explanation behind the test and describes the 16 personality types. It gives a guide to the administration of the test and the interpretation of the results. There's an extensive section covering the psychometric features of the MBTI, including reliability and validity. Finally, it provides a guide to practical uses of the results (ie. career counseling, leadership, etc.) (Sewall, 1986). There is also a manual supplement available, which was copyrighted in 2009. It

is 18 pages and gives a condensed version of the manual. It is easily accessed online.

### **Summary**

Although the MBTI is widely used and renowned for its high standard of reliability and validity, there are some researchers who have found contrary evidence. In addition, while there are many strengths associated with the MBTI, researchers have also criticized the instrument for some of its weaknesses.

The MBTI has many strengths. It is very useful in helping people discover personality information about themselves, which allows them to become more aware of not only themselves but others as well. The test avoids negative stigmas and focuses on personality strengths. Also, it is an easy test to take, administer, score and interpret. The test booklet and the manual give clear information describing how to take and score the test, as well as how to interpret the scores. A detailed explanation of the personality characteristics is given so that the test taker can fully understand the results. As a result, test takers generally will learn about themselves and will be able to use that information to further develop and grow in many areas.

The MBTI manual supplement touts the psychometric strengths of the instrument. Schaubhut, Herk and Thompson state, "the internal consistency and test-retest reliabilities of the MBTI Form M assessment are as good as or superior to those reported for the other personality assessments" (2009, P. 8) such as NEO, Birkman Method, DiSC, BarOn EQ-I and 16PF. In addition Ring (2008) noted strength in reliability and validity of the MBTI. Ring (2008) found that, "the MBTI appears to be at least comparable to other personality measures currently available. It is strongly supported by research, and is continually being reviewed and revised for psychometric improvement" (p. 11).

The MBTI also exhibits several weaknesses. Since it is a fairly simple test to administer

and interpret, people may self-administer and self-interpret the information without proper training. It is important to understand the theory and background information about the test to get the proper results. In addition, some test takers may use the results to make excuses about their shortcomings instead of looking for areas of positivity and areas of growth. The test taker may also use the results to enter a career or relationship solely based on the results of the personality findings without taking other influences into account. For example, a person may have a result suggesting that their personality fits a career as a flight attendant, but the person may be afraid of flying. Consequently, that person would need to take other factors into account and not rely only on the findings of the test. Pittenger (1993) described this as self-fulfilling prophecies whereby,

“The MBTI provides individuals with a method for interpreting their own behavior and the behavior of others. The concern here is the degree to which this system predisposes individuals to make misattributions about their own and others’ behavior while ignoring other conditions, such as the environmental context that contribute to a person’s behavior” (p. 482).

Additionally, some researchers have noted some psychometric weaknesses of the MBTI.

Carlson (1985) states that while “test-retest reliability has proven satisfactory, with  $r$  values of individual scales often exceeding .80, there is a need for reliability assessment in studies with a greater diversity of test conditions and populations” (p.363). In addition, when looking at the extroversion variable, Pittenger (1993) found,

“data suggests that extroversion is a ubiquitous personality dimension that may be found in many tests of personality and that the MBTI’s theory of extroversion cannot be supported. If all tests including the MBTI defined extroversion similarly,

then these correlations would provide evidence of convergent validity. The MBTI, however, offers a theoretical account of extroversion that is different from other accounts of extroversion, suggesting that the MBTI's account of extroversion may be incomplete" (p.475).

Capraro and Capraro (2002), suggest a few weaknesses. For example, "they postulated that the MBTI did not adequately represent the Jungian theory on which it was presumably based" (p.595). They also had concerns about gender weighting. They felt that, "different weights are applied for men and women on the 'thinking-feeling' scale based on socialization effects, leading to difficulty in comparing men and women on this scale" (Capraro & Capraro, 2002, p. 595). Finally, Barbuto (1997) raised the question of an individual who does not express a strong preference for either pole. He pointed out that, "the current scoring system does not recognize respondents who have strong preferences toward both poles of a preference index" (Barbuto, 1997, p. 618). He felt there should be a third category type that could represent a person who was non-dichotomous. For example if a person did not show a strong preference for sensing or intuitive, that person could consider either ESFP or ENFP.

In conclusion, the MBTI Form M was developed from Jung's personality type theory in an effort to make studying personality characteristics more understandable and useful in a person's life. With over two million tests used world wide over the past 50 years, it has become one of the most widely used and most trusted instrument measuring personality characteristics. The information gained from using the MBTI has been useful in many ways including counseling, psychotherapy, management and leadership in the field of business, health services, education, and career counseling (Ring, 2008). Research throughout the last 50 years has proven the MBTI to be psychometrically sound both in reliability and validity measurements. As stated

above, there have been some researchers who have called for further research to address certain weaknesses, but overall, the MBTI remains a psychometrically sound instrument as well as a very useful instrument in applying information learned about personality characteristics. Ring (2008) states,

“Reliability and validity are at the heart of the MBTI, for it promises great usefulness if it is, indeed, a dependable instrument. Because personality type pertains to every human interaction, applications for the MBTI are limited only by understanding of the underlying theory and the desire to implement it in practical situations” (p.8).

## References

- Barbuto, John, E. Jr. (1997). A Critique of the Meyers-Briggs Type Indicator and It's Operationalization of Carl Jung's Psychological Types. *Psychological Reports*, 80, 611-625.
- Capraro, Robert M. & Capraro, Mary Margaret (2002). Myers-Briggs Type Indicator Score Reliability Across Studies: A Meta-Analytic Reliability Generalization Study. *Educational and Psychological Measurement*, 62:4, 590-602.
- Carlson, John G. (1985). Recent Assessments of the Myers-Briggs Type Indicator. *Journal of Personality Assessment*, 49:4, 356-365, DOI: 10.1207/s15327752jpa4904\_3
- Pittenger, David J. (1993). The Utility of the Myers-Briggs Type Indicator. *Review of Educational Research*, 63:4, 467-488.
- Ring, Barbara P. (2008). Myers-Briggs Type Indicator: A Research Report. Website: <http://personality/Institute.tripod.com/mbtiresearchreport.htm>.
- Schaubhut, Nancy A., Herk, Nicole A. & Thompson, Richard C. (2009). MBTI Form M Manual Supplement. Website: [https://www.cpp.com/pdfs/MBTI\\_formM\\_supp.pdf](https://www.cpp.com/pdfs/MBTI_formM_supp.pdf).
- Sewall, Timothy J. (1986). The Measurement of Learning Style: A Critique of Four Assessment Tools, website: <files.eric.ed.gov/fulltext/ED267247.pdf>.
- Thompson, B. & Borrello, G. M. (1986). Construct Validity of the Myers-Briggs Type Indicator, *Educational and Psychological Measurement*, 46, 745-752.
- Turner, Samuel M., DeMers, Stephen T., Fox, Heather R., & Reed, Geoffrey M. (2001). APA Guidelines for Test User Qualifications. *American Psychologist*, 56 (12), 1099-1113. DOI: 10.1037//0003-066x.56.12.1099