

Measuring the MBTI... And Coming Up Short

*by David J. Pittenger**

Some research has shown that the Myers-Briggs Type Indicator test doesn't really measure what it purports to measure. The author, too, has his reservations as to its reliability and validity.

Imagine a test that would allow you to predict the type of career for which a person is best suited. For example, Mary comes to you for career counseling. She presents you with her education and work history, and an outline of her career objectives. Specifically, she says that she would like to find an entry-level position where she can help people, and is interested in a career in the social services.

You then administer a personality inventory. The scores suggest to you that Mary is a logical person who is achievement oriented, quick to identify flaws in others, and values truth over tact in herself and others. Based on these observations, you conclude that Mary may not be well suited for work that requires empathy for others' feelings and tolerance for ambiguity. Instead, you suggest to Mary that her strong points are her ability to focus her attention upon objective information and to make rational decisions. You then advise Mary to consider alternative careers that match her education, abilities, values, and personality.

Many claim that such a test is available with the Myers-Briggs Type Indicator (MBTI).¹ Recently, Paul D. Tieger and Barbara Barron-Tieger described how the MBTI can be used by career counselors to help clients find jobs for which they are best suited and with which they will be most satisfied.² Indeed, they claim that understanding a person's type is one of the most important factors to consider when helping that person make career decisions. The Tiegiers also provide a brief summary of the MBTI test and review how it can be used in job counseling.

The MBTI is a very popular test of personality. Each year millions of copies of the test are administered in the workplace, schools, churches, community groups, management workshops, and counseling centers. Many people see the MBTI as an invaluable tool that helps them understand their own behavior as well as the behavior of others.

In spite of the popularity of the MBTI, there are many problems with its use. There is a large body of research that suggests that the claims made about the MBTI cannot be supported. In other words, although the MBTI appears to measure something, many psychologists are not convinced that any significant conclusions can be based on the test. In this article I will review the basic research that questions the validity of the MBTI.

*David J. Pittenger is assistant professor and Chair of the Department of Psychology at Marietta College. He has written extensively in his field, and has received a Teaching Excellence Award, Early Career from Division Two of the American Psychological Association.

Dr. Pittenger earned his Ph.D. in psychology from the University of Georgia, his M.S. from Texas A&M University, and his B.A. from the College of Wooster.

A Brief History of the MBTI

The MBTI was developed by Isabel Briggs Myers and her mother, Katherine Briggs. Katherine Briggs became interested in type theory after reading Carl Jung's book, *Psychological Type*. Isabel Briggs Myers shared her mother's interest in type theory and began to create the MBTI in the early 1940s as a test to be used for personnel selection.³ Myers believed that different occupations favored different personality orientations, and that Jung's theory provided a theoretical link between personality and job performance.

In 1957, the Educational Testing Service (ETS) began to distribute the MBTI for research purposes. ETS spent considerable time and resources in deciding whether the MBTI should have been adopted as a part of its vast library of proprietary tests. After an unfavorable internal review of the test, ETS chose not to pursue the further development and ended its relationship with Myers.⁴ In 1975, Consulting Psychologists Press acquired the right to sell the MBTI. Since then, the test has been successfully marketed to an extremely wide audience.

The test is now available to licensed counselors and psychologists, to college instructors and personnel who have had graduate training in the theory of testing, and to individuals who have completed short courses on the administration and interpretation of the MBTI. There are currently many professional organizations that support the study of type. *The Journal of Psychological Type* is a scholarly periodical that publishes original research and reviews of research on type theory. There are also several professional organizations for individuals who use the MBTI as a part of their work. The Center for Applications of Psychological Type provides training on the administration and interpretation of the test, offers scoring services, and maintains a data base of MBTI profiles. The Association of Psychological Type (APT) represents the interests of professionals who use the MBTI. The APT also provides workshops that qualify non-psychologists to purchase and administer the MBTI in nonclinical settings.

A Brief Theory of Type

The primary feature of the theory behind the MBTI is that each person's personality fits into only one of 16 types. These categories are based on four features of personality, each consisting of two opposite preferences. According to the theory, all people have an innate preference that determines how they will behave in all situations. The four dimensions are:⁵

Extroversion (E) vs Introversion (I). This dimension reflects the perceptual orientation of the individual. Extroverts are said to react to immediate and objective conditions in the environment. Introverts, however, look inward to their internal and subjective reactions to their environment.

Sensing (S) vs Intuition (N). People with a sensing preference rely on that which can be perceived and are considered to be oriented toward that which is real. People with an intuitive preference rely more on their nonobjective and unconscious perceptual processes.

Thinking (T) vs Feeling (F). A preference for thinking indicates the use of logic and rational processes to make deductions and decide upon action. Feeling represents a preference to make decisions that are based on subjective processes that include emotional reactions to events.

Judgment (J) vs Perception (P). The judgment-perception preferences were invented by Briggs and Myers to indicate if rational or irrational judgments are dominant when a person is interacting with the environment. The judgmental person uses a combination of thinking and feelings when making decision, whereas the perception person uses the sensing and intuition processes.

Because the MBTI is a theory of types, a person can have only one preference. Although it is possible for people to develop the complimentary style (an introvert, for example, could learn to be more extroverted when speaking in groups) the primary preference will always dominate the person's personality. A person's MBTI score determines his or her type, a label based on his or her dominant preference for each of the four dimensions. Since there are two preferences within each dimension, there are 16 potential personality types.

Each personality type is said to be different from the others. That is, an ESTJ is a different person than an ISTJ. Many books and other printed materials about the MBTI provide descriptions of each type. Tieger and Barron-Tieger's article provides examples of these summaries as they are applied to career planning.

Assessment of the MBTI

Given this short introduction to the MBTI and its theory, we can ask a very basic question: Does the MBTI measure what it claims to measure? To answer this question we must examine the basic issues concerning the foundations of any psychological test: its statistical structure, its reliability, and its validity.

Statistical Structure

Because the MBTI is a typology, we would expect that its scores would be distributed bimodally and not be normally distributed. Let me give an analogy. If you randomly selected 500 people between the ages 18 and 25, measured their heights, and then drew a graph of the results, you would probably have a normal or bell-shaped distribution. Most people would have a height close to the mean, say 5'8". Of course, some people would be very short, and others would be very tall, but these extreme scores would be rare. Now, imagine what would happen if you divided your sample by sex. When you redraw the data you should get a bimodal distribution. Women, on average, are shorter than men; but within each sex there will be a normal distribution of heights. The same thing should happen for the MBTI.

We would expect that since people are either introverts or extroverts, the test results should yield two different curves. One curve would represent all the introverts, the other, all the extroverts. True, some people may be more extroverted than others, but we would expect that all the extroverts would be different from all the introverts. What we should find is that there are two normal curves representing the two preferences, and that there is little or no overlap of the curves.

The data indicate that there is no evidence of bimodal distributions for the MBTI.⁶ Instead, most people score between the two extremes. This means that although one person may score as an E, his or her test results may be very similar to those of another person's, who scores as an I.

Reliability

Reliability refers to the consistency in measurement of a test. Tests that are highly reliable are preferred because we can be sure that we will get the same result each time we measure the same thing. If the test is not reliable, we do not know if the changes in the score are due to changes in the person we are measuring or to some type of error in the testing process.

It is important that the MBTI be reliable for many reasons. As Tieger and Barron-Tieger note in their article, "The Type to which you are born will be the one you take to your grave." In other words, once an INTJ, always an INTJ. Therefore, we would expect the reliability of the MBTI to be extremely high and that people's type will not change.

The primary method for testing reliability is to give the test to a person on two occasions. This procedure is known as "test-retest reliability." Typically, the test-retest interval can range from several weeks to more than a year. Because type is said to be a constant characteristic, we would expect that people's personality would not change over time. Several studies, however, show that even when the test-retest interval is short (e.g., 5 weeks), as many as 50 percent of the people will be classified into a different type.⁷

The reliability data of the MBTI bring into question the stability of the test. How is it possible that there can be a change in personality, across a short interval, when such a change should not occur? The reliability data also bring into question whether there are meaningful differences across the preference categories.

Standard Error of Measurement. This testing concept is really a statistic that psychologists use to decide when the difference between two test scores is meaningful and when the difference is trivial. For example, two people could take the same test. One receives a score of 100, the other a score of 105. The standard error measurement helps us decide whether that 5-point difference represents a substantial difference between the two people or if the difference reflects simply an error in measurement.

There are two factors that influence the standard error of measurement: the standard deviation and the test-retest reliability of the test. If the standard deviation of the test is small and the reliability is high, it is possible that small differences among scores can represent significant differences among the items measured, which, in this case, are individuals' personalities. If, however, the standard deviation is large and the reliability is low, then large differences among scores must be found before we can assume that there are meaningful differences among the individuals.

The standard error of measure for each of the four dimensions is fairly large.⁸ Unfortunately, the MBTI method of scoring obscures this important distinction. It classifies people into a rigid dichotomy. Thus, two people could have raw scores that are close to one another but that define different classifications. This occurs because there are cutoff points that divide the dimensions. When the score is above the cutoff, one classification is given; if the score is below the cutoff, the opposite classification is given. Although some users of the MBTI try to interpret how close the score is to the cutoff, this practice is inconsistent with the theory of the MBTI. For example, Carskadon argues that the raw scores are overused and contends that "it is probably better to use dichotomous classification."

In summary, the differences between the two-letter categories are not as sharp and clear cut as it would appear. Because the MBTI uses an absolute classification scheme for people, it is possible for people with relatively similar scores to be labeled with much different personalities.

Validity

As the degree to which a test measures what it is supposed to measure, validity is a difficult property to evaluate in a test. Consider tests of intelligence. Many people are skeptical of the results of these tests. Some people are concerned that the tests measure only "book learning" and do not test "common sense." Other people feel that intelligence tests have cultural, racial, and gender biases. Therefore, to conclude that a test is a valid measure of intelligence, it must be shown that the test measures intelligence independent of the testee's education, culture, race, and sex.

There are many ways to evaluate the validity of the MBTI test. I will examine two important pieces of evidence. First, we can determine if the four dimensions described in the MBTI theory really exist. This is accomplished by using a statistical procedure known as "factor analysis." Secondly, we can determine whether knowing a person's MBTI type really allows us to predict how that person will perform under different circumstances. The importance of this question of validity is obvious. It must be shown that there is a consistent and meaningful relation between MBTI results and success in career placement.

Factor Analysis. The factor analysis is a type of statistic procedure that consists of making an analysis of the correlations among the questions in the test. If the MBTI theory is correct, three results should come from the factor analysis. First, the results should show that there are four clusters, or factors, of questions. Each of the questions within a factor will be highly correlated with the other questions in the factor. Moreover, the questions within the factor should be related to the MBTI dimension that is measured. For example, a question like "I like to be the life of a party" should be in the factor related to extroversion-introversion.

Secondly, we would expect each factor to be independent of the other factors, inasmuch as the MBTI theory states that each of the four preference dimensions stands alone. That is, questions within one factor should not correlate with questions in the other factors. If two factors are correlated, it means they are probably measuring the same thing.

Finally, we would expect that the factors would account for most of the differences among individuals. Here is an example that will illustrate this point. A test consisting of many unrelated questions would produce no consistent pattern in the differences among the people tested, and we would say that there is a large amount of measurement error. If, however, the questions are highly related, there should be consistent patterns that can be accounted for by the test, and the measurement error would be relatively small.

Research on the factor analysis of the MBTI has not produced convincing results. In one study, based on the results of 1,291 college aged students, six different factors were found.¹⁰ In addition, the study authors found a high level of measurement error. Specifically, 83 percent of the differences among the students could not be accounted for by the MBTI. The results led the authors to conclude that the factors found in the statistical analysis were inconsistent with the MBTI theory. In other studies, researchers found that the JP and the SN scales are correlated with one another.¹¹ In sum, the statistical analysis of the test does not support the theory used to describe the MBTI.

Relation Between MBTI Type and Occupation. Many people have examined the relation between type and occupation by examining the proportions of type within each profession. For example, one might observe that many elementary teachers are ESTJs and conclude that ESTJs prefer to be elementary school teachers or to work in a related occupation. Although it sounds appealing, such a conclusion runs into many fundamental problems.

First, we need to examine the normative data to judge the relation between type and profession. For example, the proportion of ESTJs in the teaching profession is the same as the proportion of ESTJs in the general population, or 12 percent. This similarity suggest that there is nothing special about the type of person who becomes an elementary school teacher.

Another problem stems from jobs that are dominated by men or women. Nursing is a good example. If we compare the distribution of type for nurses against managers, there appears to be a different pattern of type. We could conclude that certain types are more likely to enter nursing while other types are more likely to become managers. There is, however, an alternative interpretation. Nursing has been and remains a profession dominated by women. There is a high correlation ($r = .91$) between the percentages of types for all women and people in nursing. The correlation between all men and people in nursing is, by contrast, small ($r = .21$). In a male dominated profession such as management, there is a high correlation between types in management positions and men in general ($r = .92$), but a smaller correlation for women ($r = .60$).¹² If it is true that certain types are attracted to certain professions, then these correlations should be much smaller. Instead, these data suggest that the proportion of MBTI types within each occupation is equivalent to that within a random sample of the population.

Finally, there is no evidence to show a positive relation between MBTI type and success within an occupation. That is, there is nothing to show that ESFPs are better or worse salespeople than INTJs are. Nor is there any data to suggest that specific types are more satisfied within specific occupations than are other types, or that they stay longer in one occupation than do others.

In summary, it appears that the MBTI does not conform to many of the basic standards expected of psychological tests. Many very specific predictions about the MBTI have not been confirmed or have been proved wrong. There is no obvious evidence that there are 16 unique categories in which all people can be placed. There is no evidence that scores generated by the MBTI reflect the stable and unchanging personality traits that are claimed to be measured. Finally, there is no evidence that the MBTI measures anything of value.

Conclusions

In a recent review of the MBTI, commissioned by the Army Research Institute, it was concluded that the instrument should not be used for career planning counseling.¹³ The Institute's analysis of the available research showed no evidence for the utility of the test. Indeed, with respect to career planning they note that "the types may simply be an example of stereotypes." I agree.

The MBTI reminds us of the obvious truth that all people are not alike, but then claims that every person can be fit neatly into one of 16 boxes. I believe that MBTI attempts to force the complexities of human personality into an artificial and limiting classification scheme. The focus on the "typing" of people reduces the attention paid to the unique qualities and potential of each individual.

Many readers may be surprised by my interpretation and objections to such a popular test. It has been my experience that this reaction stems from how they view the MBTI. In many cases, the popularity of the instrument is interpreted as an indication of its accuracy and utility, which then leads to wider use and less inclination to question the foundations of the test. As a consequence, the MBTI has become a popular instrument for reasons unrelated to its reliability and validity.

The publishers do a very good job of promoting the test and providing support for its users. The MBTI also has much intuitive appeal. The descriptions of each type are generally flattering and sufficiently vague so that most people will accept the statements as true of themselves. If you tell people that they are "innovative thinkers and good problem solvers, and good at understanding and motivating people, but may have trouble following through on details of a project," they will believe that the statement is an accurate description of themselves regardless of the truth of the statement. This phenomenon is known as the "Barnum Effect," named in honor of the great entertainer.¹⁴

Because of its apparent simplicity, the MBTI may be misused unintentionally by some people. A manager, for example, may come to believe that only certain personality types are appropriate for specific jobs. After learning about type, such a manager may conclude that only ISTJs make good accountants whereas the best people for the sales force will be the ESFJs.¹⁵ Thus, the type label may bias a manager's decisions on hiring, firing, evaluating, and promoting. Similarly, employees may use type labels inappropriately. Thus, one might feel that "She's an INFP, so I will never be able to work with her on an assignment," or that "I'm an ESTP and don't do well when it comes to details."

It has been my intention here to raise questions about the fundamental concepts that underlie the MBTI, and to caution against undue reliance upon its use without fully investigating the accuracy of its test results. There is considerable more research available than I have cited that supports my allegations. My hope is that career counselors and recruiters who use or plan to use the MBTI will review this research and take a long look at the value of using personality type labels in their work.

Endnotes

- (1) The Myers-Briggs Type Indicator and MBTI are Registered Trademarks of Consulting Psychologists Press.
- (2) Tieger, Paul, D. and Barbara Barron-Tieger. "Personality Typing: A First Step to a Satisfying Career." *Journal of Career Planning & Employment*, Vol. 53, No. 2 (January 1993), pp. 50-56.
- (3) For a complete history of the development of the MBTI, see Saunders, F. W. *Katherine and Isabel: Mother's Light, Daughter's Journey*. Palo Alto, CA: Consulting Psychologists Press, 1991.
- (4) Stricker, L. J. And J. Ross. *A Description and Evaluation of the Myers-Briggs Type Indicator* (Research Bulletin #RB-62-6). Princeton, NJ: Educational Testing Service, 1962.
- (5) A more complete account of the theory behind the MBTI can be found in: Myers, I. B. and M. H. McCaulley. *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, Palo Alto, CA: Consulting Psychologists Press, 1985. And: Myers, I. B. *Gifts Differing*. Palo Alto, CA: Consulting Psychologists Press, 1980.
- (6) See Stricker and Ross (1962). Also: Stricker, L. J. & J. Ross. "An Assessment of Some Structural Properties of the Jungian personality Typology." *Journal of Abnormal and Social Psychology*, Vol. 68 (1964), pp. 62-71.
- (7) Howes, R. J. and T. G. Carskadon. "Test-Retest Reliabilities of the Myers-Briggs Type Indicator as a Function of Mood Changes." *Research in Psychological Type*, Vol. 2, No. 1 (1979), pp. 67-72.
- (8) Based on a short interval (5-week) test-retest reliability of 0.82 and a standard deviation for the E1 scale of 25 (Howes & Carskadon, 1979), The standard error of measurement is approximately $10.6 = 25 \sqrt{1 - 0.82^2}$ points. This means that raw scores with a 21-point difference are considered statistically significant.
- (9) Carskadon, T. G. "Clinical and Counseling Aspects of the Myers-Briggs Type Indicator: A Research Review." *Research in Psychological Type*, Vol. 2, No. 4 (1979a), pp. 2-31.
- (10) Sippes, G. J., R. A. Alexander, and L. Friedt. "Item Analysis of the Myers-Briggs Type Indicator." *Educational and Psychological Measurement*, Vol. 45, No. 4 (1985), pp. 789-796.
- (11) "McCrae, R. R. and P. T. Costa. "Reinterpreting the Myers-Briggs Type Indicator from the Perspective of the Five-Factor Model of Personality." *Journal of Personality*, Vol. 57, No. 1 (1989), pp. 12-40.
- (12) These analyses are based on information in Myers and McCaulley.
- (13) Druckman, D. and R. A. Bjork, Eds. *In the Mind's Eye: Enhancing Human Performance*. Washington, DC: National Academy Press, 1991.
- (14) Dickson, D. H. and I. W. Kelly. "'The Barnum Effect' in Personality Assessment: A Review of the Literature." *Psychological Reports*, Vol. 57, No. 2 (1985), pp. 367-382.
- (15) Auerbach, E. "Not Your Type, But Right for the Job." *The Wall Street Journal*, January 6, 1992, editorial, p. 11.

Reprinted from the Fall 1993 issue of the *Journal of Career Planning & Placement*, with permission of the College Placement Council, Inc., Copyright holder.