

## Reliability of the MBTI® Form Q Assessment

**Nancy A. Schaubhut & Nicole A. Herk  
CPP, Inc.**

In light of recent criticism of the MBTI assessment from researchers and academics, this study examines the psychometric properties, namely internal consistency and test-retest reliability of the MBTI Form Q assessment using a modern data set. Comparisons to other personality assessments are made. The aim of the current paper is to refute claims that the MBTI is an unreliable assessment.

The Myers-Briggs Type Indicator® (MBTI®) is one of the most widely used personality assessments in the world, often used in the workplace for training and development. It is a self-report questionnaire developed by Katharine Cook Briggs and Isabel Briggs Myers to make Carl Jung's theory of psychological types understandable and useful in everyday life. By completing the assessment and learning about the preferences it identifies, respondents gain an understanding of basic patterns of behavior. These patterns, according to Jung, are a result of people's inborn tendency to use their minds in different ways (Myers, McCaulley, Quenk, & Hammer, 1998). The preferences identified by the MBTI tool are divided into four pairs, or dichotomies: Extraversion-Introversion (E-I), Sensing-Intuition (S-N), Thinking-Feeling (T-F), and Judging-Perceiving (J-P).

The E-I dichotomy is where people focus their attention and how they get their energy. Extraverts direct their attention to and receive energy from the outer world of people and activities. Introverts direct their attention to and get energy from their own inner world. The S-N dichotomy is how people take

in information. People who prefer Sensing use their senses to take in information about things that are happening, while those who prefer Intuition take in information through hunches, seeing the big picture and possibilities. The T-F dichotomy is how people make decisions. Those who prefer Thinking make decisions through logic, cause-and-effect reasoning, and objectivity; those who prefer Feeling consider the effect of decisions on other people before making a decision. Judging-Perceiving is how people deal with the external world. People who prefer Judging prefer to be planful, scheduled, and orderly, and those who prefer Perceiving prefer to be flexible, unscheduled, and spontaneous.

The first three dichotomies were part of Jung's original theory, and the fourth was added by Briggs and Myers. Although everyone uses all eight of the preferences in different situations, each individual is likely to use one preference or the other of each dichotomy first, most often, and most comfortably. The MBTI assessment helps people determine which four preferences (one from each dichotomy) they are more likely to use. These preference

combinations form the 16 unique personality types identified by the MBTI assessment. The MBTI Form Q assessment identifies these four preferences, as well as preferences for 20 different facets that provide a more richly textured picture of type and behavior. There are five facets, or subscales for each of the preferences.

The MBTI assessment has been heavily criticized by both academics (Bess & Harvey, 2002; Cowan, 1989; McCrae & Costa, 1989, Pittenger, 2005) and popular press (Long, 1992; Paul, 2004; Zemke, 1992). Many of the criticisms concern the assessment's psychometric properties. For example, Pittenger (2005) cautions that poor test-retest reliability coefficients contradict the theory that personality preferences become and remain stable early in life. In doing so the author references the work of others to support this notion (Howes & Carskadon, 1979; McCarley & Carskadon, 1983; Stricker & Ross, 1962) and states that their work suggests that types can change at each testing. Psychometric criticisms of the assessment exist although the MBTI Manuals demonstrate the assessment has good psychometric properties. For example, the *MBTI® Step II Manual* reports internal consistency reliabilities for the facets ranging from .57-.85 in a national sample; and test-retest reliabilities ranging from .56-90 in an

adult sample and .55-.78 in a student sample (Quenk, Hammer, & Majors, 2001). The *MBTI® Manual* also shows excellent reliabilities for the four preference pairs. The internal consistency reliabilities in a national sample range from .91-.92, while the test-retest reliabilities range from .83-.97 in three different samples (Myers, et al., 1998). This paper will examine some of psychometric properties of the MBTI assessment, specifically, internal consistency reliability and test-retest reliability. By presenting reliability statistics using a contemporary dataset, this study aims to add to the existing literature by clearing up misconceptions regarding the reliability of the MBTI assessment.

## Method

### *Participants*

The sample used in this paper consists of 409 individuals who each completed the MBTI Form Q assessment twice between January 2004 and September 2008. There were an equal number of male (49%) and female participants (49%), with 2% not reporting their gender. See Table 1 for a full demographic description of the sample. All demographics are reported from the time of the first assessment.

Table 1. Demographic Characteristics of Sample at Time 1

	M	SD	n	%
Males			201	49
Females			201	49
Age	37.2	13.6		
Ethnicity				
Caucasian			134	33

African American	10	2
Latino/Latina/Hispanic	7	2
Asian or Pacific Islander	11	3
Indian	3	1
American Indian/Alaskan Native	0	0
Middle Easterner	3	1
Multi-ethnic	3	1
Not reported	238	58
Employment status		
Working full-time	168	41
Working part-time	3	1
Not working for income	1	<1
Full-time student	15	4
Not reported	191	47
Current work years	8.2	10.1
Education completed		
High school diploma	9	2
Trade/technical training	1	<1
Some college - no degree	21	5
Associate degree	9	2
Bachelor's degree	70	17
Masters degree	53	13
Professional degree – e.g., MD	8	2
Doctorate – e.g., PhD	6	1
Not reported	232	57
Reason for completing assessment		
Training	115	28
Employment testing	4	1
Career counseling	10	2
Education	11	3
Personal growth	34	8
Not reported	235	57

## Analyses and Results

The test-retest reliability of the preferences was evaluated by correlating the continuous scores from time one and time two. The period of time between the first and second completion of the assessment ranged from less than one week to more than

four years. The test-retest correlations are presented in Table 2, with five different time intervals – 3 weeks or less, 1-6 months, 6-12 months, more than 1 year, and all combined. The

correlations of the preferences range from .80 (Judging-Perceiving, 1-6 month interval) to .90 (Thinking-Feeling, 3 weeks or less interval), indicating good reliability for each preference pair over long periods of time. The test-retest correlations in the present study are consistent with those in the *MBTI Manual*, which reports test-retest correlations for the preferences ranging from .83-.97 in three different samples over a one-month time period (Myers et al., 1998).

The preference test-retest correlations are superior to those of other

personality assessments such as the NEO PI-R™, which range from .75 to .83 for a 3-month time interval (Costa & McCrae, 1992). The correlations are also higher than those of Jungian type inventories such as the Gray-Wheelwright Jungian Type Survey, which has test-retest reliabilities range from .61 to .89 with an 8-week time period (Davis & Mattoon, 2006); and the Singer-Loomis Type Deployment Inventory™ test-retest reliabilities, which range from .62 to .80 for raw scores and .28 to .52 for ipsatised percentage scores, both a two-week time period (Arnau, Rosen, & Thompson, 2000).

The facet correlations in the current study range from .44 (Questioning-

Accommodating, more than 1 year interval) to .90 (Expressive-Contained and Early Starting-Pressure Prompted, 3 weeks or less interval; Expressive-Contained, more than 1 year interval). Most of the facet reliabilities are very good over a long period of time. They are similar to those reported in the *MBTI® Step II Manual*. The MBTI Form Q facet correlations in the manual range from .56-.90 for an adult sample with a test-retest time period of 30 days and from .55-.78 for a college student sample also with a 30 day test-retest time period (Quenk et al., 2001). The facet test-retest correlations here are also similar to the NEO PI-R facet correlations (.66-.92) reported in a study by McCrae and Costa (1983).

Table 2. MBTI® Form Q Test-retest Correlations

MBTI Preference or Facet	All	<=3 weeks	1-6 months	6-12 months	>1 year
Extraversion-Introversion	.87	.84	.89	.84	.88
Initiating-Receiving	.82	.80	.83	.80	.86
Expressive-Contained	.79	.75	.85	.75	.88
Gregarious-Intimate	.71	.73	.72	.61	.77
Active-Reflective	.77	.80	.75	.77	.75
Enthusiastic-Quiet	.81	.83	.84	.80	.77
Sensing-Intuition	.85	.84	.82	.85	.88
Concrete-Abstract	.79	.79	.73	.80	.85
Realistic-Imaginative	.77	.77	.81	.72	.79
Practical-Conceptual	.70	.80	.72	.65	.65
Experiential-Theoretical	.78	.71	.79	.75	.84
Traditional-Original	.78	.80	.83	.68	.79
Thinking-Feeling	.86	.90	.84	.84	.88
Logical-Empathetic	.81	.88	.79	.79	.81
Reasonable-Compassionate	.77	.83	.73	.76	.80
Questioning-Accommodating	.50	.49	.51	.53	.44
Critical-Accepting	.63	.73	.58	.62	.61
Tough-Tender	.76	.75	.75	.74	.82
Judging-Perceiving	.83	.87	.80	.83	.85
Systematic-Casual	.70	.75	.64	.72	.75
Planful-Open-ended	.72	.79	.72	.72	.68
Early starting-Pressure-prompted	.79	.88	.74	.80	.78

Scheduled-Spontaneous	.77	.78	.73	.84	.73
Methodical-Emergent	.68	.78	.60	.68	.75

Note. <=3 weeks n = 70, 1-6 months n = 139, 6-12 months n = 115, >1 year n = 85.

Internal consistency reliabilities, or coefficient alphas, for time 1 and time 2 are shown in Table 3. Overall, these reliability estimates are very good. The alphas for MBTI preferences (.91-.93) are consistent with those presented in the *MBTI Manual*, which range from .91-.92 for the national sample (Myers et al., 1998). The coefficient alphas for the MBTI preferences are superior or comparable to other personality and Jungian type assessments. For example, they are similar to the coefficient alphas for the NEO PI-R™ five factors, which range from .86-.92 for self-reports (Costa & McCrae, 1992). Next, they are superior to the DiSC®, .69-.85 (Watson & Klassen, 2004). Finally, they are also superior to those reported for other Jungian type assessments. For example, one study reported internal consistency reliabilities from .38 to .72 for the Gray-Wheelwright Jungian Type Survey (Davis & Mattoon, 2006). Arnau et al. (2000) reported internal consistency reliabilities for the Singer-Loomis Type Deployment Inventory raw scores ranging from .65 to .91.

Most of the coefficient alphas for the facets (.38-.88) are good, and are consistent with those in the *MBTI Step II Manual*, which range from .57 to .85 for the national sample (Quenk et al., 2001). The lowest alpha reported in the *MBTI Step II Manual* is the Questioning-Accommodating facet, which is also the lowest alpha in the present study. However, the alpha for this facet reported here is slightly lower. One

possible reason for the lower reliability of the Questioning-Accommodating facet is that it has only five items. Internal consistency reliability, measured by coefficient alpha, increases as the number of items on a scale increases (Schwab, 2005).

Coefficient alphas for NEO PI-R (Costa & McCrae, 1992) facets range from .56 to .81 for self-reports. These alphas are similar to those for the facets in Table 3. Regarding the NEO PI-R alphas, Costa & McCrae (1992, p. 44) note that, "These values are acceptable for scales with only eight items." Several of the MBTI Form Q facet scales have fewer than eight items, namely Gregarious-Intimate, Active-Reflective, Enthusiastic-Quiet, Realistic-Imaginative, Practical-Conceptual, Experiential-Theoretical, Questioning-Accommodating, Critical-Accepting, Systematic-Casual, Planful-Open Ended, Early Starting-Pressure Prompted, and Methodical-Emergent.

Table 3. MBTI® Form Q Internal Consistency Reliabilities for Time 1 and Time 2

	Time 1 coefficient alpha	Time 2 coefficient alpha
Extraversion-Introversion	.92	.93
Initiating-Receiving	.87	.88
Expressive-Contained	.81	.84
Gregarious-Intimate	.67	.73
Active-Reflective	.60	.68
Enthusiastic-Quiet	.74	.78
Sensing-Intuition	.91	.93
Concrete-Abstract	.79	.82
Realistic-Imaginative	.78	.82
Practical-Conceptual	.52	.58
Experiential-Theoretical	.83	.85
Traditional-Original	.75	.78
Thinking-Feeling	.91	.92
Logical-Empathetic	.81	.83
Reasonable-Compassionate	.76	.79
Questioning-Accommodating	.38	.48
Critical-Accepting	.58	.62
Tough-Tender	.80	.84
Judging-Perceiving	.92	.93
Systematic-Casual	.77	.81
Planful-Open-ended	.83	.86
Early starting-Pressure-prompted	.78	.81
Scheduled-Spontaneous	.84	.87
Methodical-Emergent	.67	.72

### Discussion

The results presented in this paper provide support for the reliability of the MBTI Form Q assessment. Specifically, internal consistency reliability estimates and test-retest correlations were consistent with those presented in both the *MBTI Manual* (Myers, et al., 1998) and the *MBTI Step II Manual* (Quenk, et al., 2001). Additionally, these results are superior or comparable to reliability coefficients and test-retest correlations reported for other well-known personality instruments, e.g., the NEO PI-R and the DiSC, as well as similar measures of Carl Jung's theory of psychological types, e.g., Gray-

Wheelwright Jungian Type Survey and the Singer-Loomis Type Deployment Inventory.

The reliabilities for some facets, such as Questioning-Accommodating were slightly low. However, acceptable alpha levels depend on whether an assessment's results are being used for high-stakes decisions (e.g., employment) or low stakes decisions that have minor consequences to a test-taker (Wasserman & Bracken, 2004). The MBTI Form Q assessment should not be used for employment decisions such as selection; rather it is used for personal growth or training and

development purposes. Therefore, somewhat lower levels of alpha are acceptable in such circumstances. Thus, the findings of the current study refute previous arguments (e.g., Pittenger, 2005) made against the MBTI assessment and provide further support for the reliability of the instrument.

Future studies may examine whether the reliabilities reported in this paper hold up with different populations, such as ethnic groups, age groups, international samples, as well as translated versions of the MBTI assessment.

## References

- Arnau, R. C., Rosen, D. H., & Thompson, B. (2000). Reliability and validity of scores from the Singer-Loomis Type Deployment Inventory. *Journal of Analytical Psychology, 45*, 409-426.
- Bess, T. L., & Harvey, R. J. (2002). Bimodal score distributions and the Myers-Briggs Type Indicator: Fact or artifact? *Journal of Personality Assessment, 78*, 176-186.
- Costa, P. T., Jr. & McCrae, R. R. (1991). *NEO PI-R™ professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Cowan, D. A. (1989). An alternative to the dichotomous interpretation of Jung's psychological functions: Developing more sensitive measurement technology. *Journal of Personality Assessment, 53*, 459-471.
- Davis, M. F., & Mattoon, M. A. (2006). Reliability and validity of the Gray-Wheelwrights Jungian Type survey. *European Journal of Psychological Assessment, 22* (4), 233-239.
- Howes, R. J., & Carskadon, T. G. (1979). Test-retest reliabilities of the Myers-Briggs Type Indicator® as function of mood changes. *Research in Psychological Type, 2*, 67-72.
- Long, T. G. (1992). Myers-Briggs and other modern astrologies. *Theology Today, 49*, 295.
- McCarley, N. G., & Carskadon, T. G. (1983). Test-retest reliabilities of scales and subscales of the Myers-Briggs Type Indicator® and of criteria for clinical interpretative hypotheses involving them. *Research in Psychological Type, 6*, 24-36.
- McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator® from the perspective of the five-factor model of personality. *Journal of Personality, 57*, 17-40.
- McCrae, R. R., & Costa, P. T., Jr. (1983). Joint factors in self-reports and ratings: Neuroticism, extraversion, and openness to experience. *Personality and Individual Differences, 4*, 245-255.
- Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI® Manual: A guide to the development and use of the Myers-Briggs Type Indicator®* (3<sup>rd</sup> ed.). Mountain View, CA: CPP, Inc.
- Paul, A. M. (2004). *The cult of personality: How personality tests are leading us to miseducate our children, mismanage our companies, and misunderstand ourselves*. New York, NY: Free Press, Simon and Schuster.
- Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*, 210-221.
- Quenk, N. L., Hammer, A. L., & Majors, M. S. (2001). *MBTI® Step II Manual: Exploring the next level of type with the Myers-Briggs Type*

- Indicator Form® Q*. Mountain View, CA: CPP, Inc.
- Schwab, D. P. (2005). *Research methods for organizational studies* (2<sup>nd</sup> ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Stricker, L. J., & Ross, J. (1964). An assessment of some structural properties of the Jungian personality typology. *Journal of Abnormal and Social Psychology*, 68, 62-71.
- Wasserman, J. D., & Bracken, B. A. (2004). Psychometric characteristics of assessment procedures. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (eds.), *Handbook of psychology: Assessment psychology*, Volume 10, pp. 43-66. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Watson, R. J., Klassen, P. T. (2004). *Style Insights – DiSC® instrument validation manual*. Scottsdale, AZ: Target Training International – Performance Systems, Ltd.
- Zemke, R. (1992, April). Second thoughts about the MBTI. *Training*, 29 (4), 43-47.