Myers-Briggs Type Indicator (MBTI): Some psychometric limitations

Article *in* Australian Psychologist · March 1995

DOI: 10.1111/j.1742-9544.1995.tb01750.x · Source: OAI

1 author:

Gregory J. Boyle
University of Melbourne
221 PUBLICATIONS **4,519** CITATIONS

SEE PROFILE

3-1-1995

Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations

Gregory J. Boyle
Bond University, Greg_Boyle@bond.edu.au

Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations

GREGORY J. BOYLE
Bond University

The present paper critically reviews the psychometric adequacy of the Myers-Briggs Type Indicator (MBTI). Although the instrument is extremely popular in applied settings, there is an urgent need for the development of valid and comprehensive local norms, in order to increase its predictive validity and utility within the Australian context. In addition, there are a number of psychometric limitations pertaining to the reliability and validity of the MBTI, which raise concerns about its use by practitioners. In view of these serious limitations, routine use of the MBTI is not recommended, and psychologists should be cautious as to its likely misuse in various organisational and occupational settings.

The Myers-Briggs Type Indicator or MBTI (Briggs-Myers & Briggs, 1985) is an extremely popular personality inventory which has received widespread use over the last 30 years (Carlyn, 1977). The MBTI is a self-report questionnaire designed to quantify non-psychopathological personality types as postulated in Jung's psychodynamic type theory (see Myers & McCaulley, 1985).

Each individual's personality type is described in terms of a four-letter code, a brief descriptive interpretation of which is provided on the back of the report form. For a simple (four-dimensional), straightforward description of one's personality make-up, use of the MBTI would seem an appropriate choice. The instrument is ipsatively scored, and predominantly utilises forced-choice (true/false) items. Four dichotomous dimensions classify individuals either as extraverted (E) or introverted (I), sensing (S) or intuitive (N), thinking (T) or feeling (F), and judging (J) or perceiving (P). Combinations of the four preferences determine personality types. Each individual is classified in terms of one of 16 possible four-letter codes (such as ESFJ, ENFP, INTP, and ISFJ). Each type is said to define a specific set of behavioural tendencies, reflecting differences in attitudes, orientation, and decision-making styles. All materials, including the manual, test booklets, answer sheets, and score keys, are professionally produced.

The Jungian construct of extraversion is embedded within a somewhat different conceptual framework than either Cattellian or Eysenckian interpretations. The E-I dimension does not pertain to shyness versus gregariousness, but focuses on whether one's general attitude towards the world is actively oriented outward to other persons and objects, or is internally oriented (Sipps & Alexander, 1987). The S-N dimension describes the individual's characteristic perceptual style. Sensing is viewed as attending to sensory stimuli, whereas intuition involves a more detached, insightful analysis of stimuli and events. Somewhat reminiscent of the field dependence-independence distinction, for the T-F dimension, thinking involves logical reasoning and decision processes, while feeling entails a more subjective, interpersonal approach. Thomas (1983) reported a correlation between field independence and thinking of 0.37, suggesting a small amount of commonality. The J-P dimension distinguishes between the judging attitude associated with prompt decision making (often before all facts are at hand), while perception involves greater patience and waiting for more information, before making decisions (cf. Thomas, 1984). The individual is typed as either one or the other (each dimension is discontinuous, rather than continuous). Willis (1984, p. 483) indicated that the J-P dimension determines which of two function preferences is dominant and which is auxiliary (S or N versus T or F). The auxiliary style is used only in unusual situations; such as when an introverted child is required by the school system to "play the role" of an extraverted individual. DeVito (1985, p. 1030) asserted that"the dominant and auxiliary function is not well developed in

Jung's writings (see McCaulley, 1981, pp. 301-302) and is the most controversial aspect of Myers' interpretation of Jung". To date, the validity of the dominant and auxiliary functions has not been tested empirically.

The four main MBTI dimensions can be placed within the context of modern thinking about personality theory concerning the major dimensions of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (see Digman, 1990; Goldberg, 1992). Additionally, in the most comprehensive higher-order factor analytic study of personality dimensions to date, Krug and Johns (1986) carried out a scale factoring of the 16PF on well over 17,000 individuals, and crossvalidated their findings across sex. They reported six second-order factors labelled: Extraversion, Neuroticism, Conscientiousness, Tough Poise, Control, and Intelligence. These dimensions have been verified independently in a higher-order factor analysis of the Cattell, Comrey, and Eysenck scales by Boyle (1989). The E-I dimension of the MBTI relates directly to the Extraversion-Introversion higher-stratum factor; the S-N dimension appears to relate to Tough Poise – Sensitivity (i.e., sensing individuals may exhibit scepticism and tough poise, whereas intuitive persons appear to be more sensitive); and both the T-F and J-P dimensions appear to relate somewhat to Conscientiousness, Control, and Intelligence, and to the Openness dimension in the Norman Big Five, as implemented in say the NEO-PI (Costa & McCrae, 1992). The MBTI appears to measure about 30-35% of the normal personality trait variance. However, it clearly places a proportionately greater emphasis on cognitive styles than do most other personality instruments.

Current Forms of the MBTI

The 126-item Form G which is now the standard version of the MBTI (cf. Leiden, Veach, & Herring, 1986) was constructed from the 166-item Form F, with nine items slightly reworded to reduce reading level, and with 38 experimental items and two other items removed. Only the first 95 items need be answered to score the instrument. As Coan (1978, p. 973) pointed out, inclusion in Form F of the experimental items (not used for scoring purposes) serves only to increase testing time, and provides no relevant information for standard use of the instrument. Misunderstanding of items is potentially problematic. As well, an abbreviated version (A V) also exists, comprising only 50 items. The manual provides normative data for high school and university student samples, but data for the general adult population is notably lacking. There is little empirical information on minorities or working-class populations.

Scoring of the Instrument

Scoring of the MBTI is either in terms of preference or continuous scores. According to the manual, continuous scores are appropriate for research purposes only.[1] As stated by DeVito (1985), use of continuous scores is not emphasised. Jungian theory underlying construction of the MBTI asserts that an individual's dichotomous preference scores symbolise fundamental differences between types (e.g. introverts versus extraverts). However, most psychometricians regard personality dimensions such as extraversion-introversion as continuous, and normally distributed. According to Wiggins (1989, p. 538), "The principal stumbling block to more widespread acceptance of the MBTI lies in the structural model of bipolar discontinuous types to which the test authors are firmly committed." Consequently, use of dichotomous forced-choice items greatly limits both the theoretical and statistical import of the MBTI. Wiggins argued that there is no evidence to support Jung's theory of bimodal distributions of preference scores, and that evidence of stability of types is lacking. Thus, as Wiggins pointed out, the four-variable types remaining stable across measurement occasions exceeds 50% only rarely. Another problem is "considerable redundancy in the scoring operation because the two contrasting scales of each pair are scored on the basis of essentially the same items". (Coan, 1978, p. 974).

Item Homogeneity

Use of preference scores (with concomitant restriction in variance) requires computation of phi coefficients if both variables represent true dichotomies (Gorsuch, 1983, p. 296). According to Willis (1984), median phi estimates of item homogeneity for Form Fare .60 (E-I), .69 (S-N), .59 (T-F), and .71 (J-P), and median tetrachoric correlations (if the dichotomous variables actually represent continuous dimensions) are .76 (E-I), .87 (S-N) , .78 (T-F) , and .80 (J-P). The inconsistency of these dichotomous preference score estimates has been alluded to by Carlyn (1977).
Item homogeneity estimates based on the continuous scores are .79 (E-I), .81 (S-N), .78 (T-F), and .82 (J-P). Nevertheless, item homogeneity estimates based on continuous scores may be inappropriate given the assumption underlying the MBTI of dichotomous dimensions (even though continuous scores provide greater precision than do dichotomous scores). These estimates suggest moderately high levels of item homogeneity, and the possibility of associated item redundancy and narrow breadth of measurement (Boyle, 1987, 1991; Cattell, 1978).

Test-Retest Reliability

McCaulley (1981) has investigated the test-retest reliability (stability) of the MBTI. Estimates of the proportion of preferences reclassified into the same categories ranged from 61 % to 90%. For continuous scores, median stability coefficients are .78 (E-I), .78 (S-N), .69 (T-F), and .74 (J-P),over intervals from five weeks to 21 months. For Form G, the corresponding Spearman-Brown corrected estimates (cf. Ferguson, 1981) are .73 (E-I), .69 (S-N), .64 (T-F), and .69 (J-P). Likewise, for the abbreviated version (Form A V), the corresponding Spearman-Brown corrected estimates are .52 (E-I), .47 (S-N), .42 (T-F), and .42 (J-P). However, little direct information for Forms G and A V is available (Carskadon, 1979, 1982). Clearly, Form G is less reliable than Form F (because of the reduction in number of items), and Form A V is considerably less reliable than Form G. These test-retest estimates for the current forms of the MBTI indicate some instability. For enduring personality dispositions, stability estimates should be in the 0.8 to 0.9 range (see Boyle, 1985).

Validity

Reviews supportive of the validity of the MBTI have been provided, for example, by Carlyn (1977), and Carlson (1985). According to Carlson (p. 357), "literature on the scale (one bibliography lists approximately 700 references) reflects largely successful efforts to apply it in a large variety of educational, clinical, counselling, business, and research settings." Despite these claims, the psychometric limitations of the MBTI raise concerns about the validity of the instrument.

Using continuous scores, the instrument has been correlated with several well known psychological inventories(Corman & Platt, 1988; Schurr, Ruble, & Henriksen, 1988). Yet there has been little attempt to ascertain inter-battery relationships via multiple regression procedures. Evidence concerning the interrelationships of the MBTI with other well known personality inventories remains mostly at the simple correlational level. Interpretations of concurrent validity based solely on correlations are likely to be biased due to the unreliability of correlation coefficients (Detterman 1979).

There have been several factor analyses of the MBTI (e.g. Sipps, Alexander, & Friedt, 1985; Thompson & Borrello, 1986a, 1986b; Tzeng, Outcalt, Boyer, Ware, & Landis, 1984) - (see Carlyn, 1977, for a summary of earlier factor analytic studies). In general, these factor analyses have followed variants of the "Little Jiffy" procedure (principal components plus orthogonal varimax rotation, an inadequate method of exploratory factor analysis; see Boyle, 1988, pp. 742-745; Gorsuch, 1983; McDonald, 1985). Therefore, the results of these analyses which claim to support the independence of the four MBTI dimensions are unsatisfactory, and in part may be a statistical

artifact due to (a) extraction of at four factors, and (b) use of inappropriate orthogonal rotation. More appropriate confirmatory factor analyses of the MBTI dimensions remain to be undertaken (e.g. via LISREL, COSAN, EQS). The question is not whether other structures can be found for the MBTI (a common result from exploratory factor analyses), but whether the purported structure of the instrument is valid-a confirmatory rather than an exploratory issue (cf. Joreskog & Sorbom, 1989). Undue reliance on exploratory factor analyses of the MBTI to the exclusion of confirmatory methods has undoubtedly resulted in theory conflation, rather than more appropriately discriminating between hypotheses.

Regarding predictive validity, Myers (1962, p. 77) recommended that the MBTI is best viewed "as affording hypotheses for further testing and verification rather than infallible expectations of all behaviors". Since the MBTI types are not "source traits" verified factor analytically (i.e., "causal" psychological dimensions), predictions based on these "surface traits" (discontinuous types) are inevitably less powerful and remain somewhat speculative. On a different note, there are no scales built into the MBTI to detect the effects of random responding, response sets such as social desirability, or either conscious or unconscious response distortion. Social desirability response set appears to influence scores on the EI and JP scales (McCaulley,1981, p. 339). Also, there is no control for the mood of the respondent, which may greatly affect responses (Howes & Carskadon, 1979). Consequently, the issue of motivational distortion in the MBTI responses needs to be addressed.

Conclusions

The MBTI is one of the most frequently used instruments for personality assessment. However, as Bjork and Druckman (1991) pointed out, the instrument's popularity is not consistent with research evidence. Furthermore, the MBTI manual does not provide norms based on continuous scores. Much of the supporting evidence provided in the manual is of questionable validity (Coan, 1978). Reliance on dichotomous preference scores rather than continuous scores unduly restricts the level of statistical analysis (such as assigning frequencies to the 16 types). As DeVito (1985, p. 1032) indicated, "The issue regarding type vs. continuous scores will probably remain most unsettling for those espousing traditional test construction standards and procedures." Empirical evidence does not strongly support the Jungian notion of discrete or "true" dichotomies.

As well, there are problems in using MBTI preference scores to predict behavioural or occupational outcomes. In addition, test-retest estimates raise doubts about the stability of MBTI-type scores. Some investigators (e.g. Anastasi, 1990) have suggested that all personality questionnaires have dubious psychometric standing. Certainly, the problems associated with item transparency, and concomitant response distortion ranging all the way from lack of self-insight to deliberate faking, as well as the effects of response sets in general, apply universally to sef- report questionnaires (Boyle, 1985).

Bjork and Druckman (1991) asserted that most of the extant studies of the MBTI are defective and that there is insufficient research into the utility of the MBTI in organisational settings. They further argued (p. 99) that, "At this time, there is not sufficient, well-designed research to justify the use of the MBTI in career counseling programs. Much of the current evidence is based on inadequate methodologies. "

Overall, the MBTI provides a psychometrically simple description of Jungian personality types. Although this brief characterisation may be useful in some applied contexts (such as in predicting an individual's characteristic style of behaviour, intellectually and interpersonally), there are evident

psychometric limitations of the instrument. With further research and refinement, the MBTI may serve a more useful role in applied psychological assessment. Certainly, development of valid and comprehensive local norms (including relevant motivational distortion scales, e.g. Faking Good; Faking Bad) should increase its predictive validity within the Australian context. Given the lack of appropriate local norms, it would seem prudent for practitioners to be alert to its possible misuse, and to be cautious in undertaking personality assessments with the instrument. The current enthusiasm for the METI is certainly not warranted on psychometric grounds.

Footnote

1. This viewpoint is contentious. Psychometrically, a continuous scoring system should be used with at least four to five response options per question (cf. Joreskog & Sorbom, 1988).

References

Anastasi, A. (1990). *Psychological testing* (6th ed.). New York: Macmillan.

Bjork, RA., & Druckman, D. (1991). *In the mind's eye: Enhancing human performance*. Washington, DC: National Academy Press.

Boyle, G.J. (1985). Self-report measures of depression: Some psychometric considerations. *British Journal of Clinical Psychology*, 24, 45-59.

Boyle, G.J. (1987). Review of the (1985) "Standards for educational and psychological testing: AERA, APA, and NCME". *Australian Journal of Psychology*, 39, 235-237.

Boyle, GJ. (1988). Elucidation of motivation structure by dynamic calculus. In J.R Nesselroade & R.B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 737-787). New York: Plenum.

Boyle, GJ. (1989). Re-examination of the major personality-type factors in the Cattell, Comrey, and Eysenck scales: Were the factor solutions by Noller et al. optimal? *Personality and Individual Differences,* 10, 1289-1299.

Boyle, GJ. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, 12, 291-294.

Briggs-Myers, I., & Briggs, K.C. (1985). *Myers-Briggs Type Indicator (MBTI).* Palo Alto, CA: Consulting Psychologists Press.

Carlson, J.G. (1985). Recent assessments of the Myers-Briggs Type Indicator. *Journal of Personality Assessment,* 49, 356-365.

Carlyn, M. (1977). An assessment of the Myers-Briggs Type Indicator. *Joumal of Personality Assessment,* 41, 461-473.

Carskadon, T.G. (1979). Test-retest reliabilities of continuous scores on Form G of the Myers-Briggs Type Indicator. *Research in Psychological Type*, 2, 83-84.

Carskadon, T.G. (1982). Sex differences in test-retest reliabilities of continuous scores on Form G of the Myers-Briggs Type Indicator. *Research in Psychological Type*, 5, 78-79.

Cattell, R.B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum.

Coan, RW. (1978). Review of the Myers-Briggs Type Indicator. *Eighth Mental Measurements Yearbook*, 1, 973-975.

Corman, L.S., & Platt, R.G. (1988). Correlations among the Group Embedded Figures Test, the Myers-Briggs Type Indicator and demographic characteristics: A business school study. *Perceptual and Motor Skills*, 66, 507-511.

Costa, P.T., & McCrae, R.R. (1992). NEO Personality Inventory. *Psychological Assessment*, 4, 5-13.

Detterman, D.K. (1979). Detterman's laws of individual differences research. In Sternberg, R. & Detterman, D.K. (Eds.), *Human intelligence: Perspectives on its theory and measurement* (pp. 165-175). Norwood, NJ: Ablex.

DeVito, A.J. (1985). Review of the Myers-Briggs Type Indicator. *Ninth Mental Measurements Yearbook,* 1, 1030-1032.

Digman, R.M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.

Ferguson, G.A. (1981). *Statistical analysis in psychology and education* (5th ed.). Singapore: McGraw-Hill.

Goldberg, L.R. (1992). The development of markers for the Big Five factor structure. *Psychological Assessment*, 4, 26-42.

Gorsuch, R.L. (1983). *Factor analysis* (5th ed.). New York: Plenum.

Howes, R.J., & Carskadon, T.G. (1979). Test-retest reliabilities of the Myers-Briggs Type Indicator as a function of mood changes. *Research in Psychological Type*, 2, 67-72.

Jareskog, K.G., & Sarbom, D. (1988). PRELIS: *A program for multivariate data screening and data summarization: A pre-processor for LISREL*. Mooresville, IN: Scientific Software.

Jareskog, K.G., & Sorbom, D. (1989). LISREL 7: *User's reference guide*. Mooresville, IN: Scientific Software.

Krug, S.E., & Johns, E.F. (1986). A large scale cross-validation of second-order personality structure defined by the 16PF. *Psychological Reports*, 59, 683-693.

Leiden, LJ., Veach, T.L., & Herring, M.W. (1986). Comparison of the abbreviated and original versions of the Myers-Briggs Type Indicator personality inventory. *Journal of Medical Education*, 61,319-321.

McCaulley, M.H. (1981). Jung's theory of psychological types and the Myers-Briggs Type Indicator. In P. McReynolds (Ed.), *Advances in Personality Assessment* (Vol. 5, pp. 294-352): San Francisco: Jossey-Bass.

McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.

Myers, LB. (1962). Manual: *The Myers-Briggs Type Indicator*. Princeton, NJ: Educational Testing Services.

Myers, LB., & McCaulley, M.H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.

Schurr, K.. T., Ruble, V.E., & Henriksen, L.W. (1988). Relationship of Myers-Briggs Type Indicator personality characteristics and self-reported academic problems and skill rating with Scholastic Aptitude Test scores. *Educational and Psychological Measurement,* 48, 187-196.

Sipps, G.J., & Alexander, R.A. (1987). The multifactorial nature of extraversion-introversion in the Myers-Briggs Type Indicator and Eysenck Personality Inventory. *Educational and*

Psychological Measurement, 47, 543-552.

Sipps, G.J., Alexander, R.A., & Friedt, L. (1985). Item analysis of the Myers-Briggs Type Indicator. *Educational and Psychological Measurement,* 45, 789-796.

Thomas, C.R. (1983). Field independence and Myers-Briggs thinking individuals. *Perceptual and Motor Skills,* 57, 790.

Thomas, C.R. (1984). Regression of Myers-Briggs Type scales. *Psychological Reports*, 55, 568.

Thompson, B., & Borrello, G.M. (1986a). Construct validity of the Myers Briggs Type Indicator. *Educational and Psychological Measurement*, 46, 745-752.

Thompson, B., & Borrello, G.M. (1986b). Second-order factor structure of the MBTI: A construct validity assessment. *Measurement and Evaluation in Counselling and Development,* 18,148-153.

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.

Wiggins, 1.S. (1989). Review of the Myers-Briggs Type Indicator. *Tenth Mental Measurements Yearbook,* 1, 537-538.

Willis, C.G. (1984). Review of the Myers-Briggs Type Indicator. *Test Critiques*, 1, 482-490.