

Testing the Test

Affirmative: In Support of Researching the Myers-Briggs Type Indicator

JOHN G. CARLSON

In a previous review of recent clinical, counseling, and related research literature, I examined the reliability and validity of the Myers-Briggs Type Indicator MBTI (Carlson, 1985). It was concluded that the MBTI has yielded generally satisfactory split-half and test-retest reliabilities and similarly favorable (although unsystematic) validity measurements across a variety of studies in these areas, a conclusion that paralleled that of earlier reviews (e.g., Carlyn, 1977). However, I also emphasized that considerably more research is needed to ensure the reliability of the instrument, particularly across long test-retest intervals, and especially to provide more systematic construct validation.

This review was initiated somewhat critically, owing to my skepticism concerning typologies in general. I gather that I am not alone. In the words of Hicks (1984), "It would be safe to venture that most American behavioral scientists are strongly disposed to reject typological approaches" (p. 1127). Therefore, the relatively favorable pattern of findings that emerged was as much enlightening to me as it may be to others with a bias against either the typological approach or to psychodynamic theory—in particular, Carl Jung's, from which the MBTI is derived. Owing to this positive outcome, coupled with personal experience with the administration of the MBTI on myself and others, I have become increasingly interested in what we might term the *psychological validity* of the dimensions of the MBTI and in the utility of the instrument in certain applications. Therefore, I agreed for the sake of what I hope will be a spirited and heuristic debate to present in this shortened format some evidence in support of the instrument. Following a brief comment on the psychological validity of the dimensions of the instrument (not to be confused with the face validity of its items), I will highlight a few of the previously reviewed results of studies of the instrument and select a few recent ones to document the reliability and validity that has been observed, emphasizing clinical and counseling settings. However, in this brief forum I have made no effort to summarize data relative to the uses of the MBTI in career counseling (cf. Lynch, 1985).

The use of an information-processing analogy in the following interpretation of the dimensions of the MBTI is a personal preference and does not owe directly to theoretical or assessment roots of the instrument. The dimensions of the MBTI are four in number: The EI (extraversion-introversion index), the SN (sensation-intuition index), the TF (thinking-feeling index), and the JP (judging-perceiving index). The EI index has many parallels in the literature (Morris, 1979) and refers to a commonly recognized and well-researched aspect of personality that owes its formal theoretical origins to Jung. Here we might say that

the dimension refers to an individual preference for obtaining information directly from others (extraversion) or from oneself, such as through reading or reflection (introversion). The remaining indices are less common, although at least some of their relatives do appear under other labels in other instruments. The SN scale relates to a person's preference either for empirical, sense-based data (sensation) or for self-generated information, hunches, or the "sixth-sense" (intuition). The TF scales attempt to measure the difference between the respective tendencies of some people to prefer logical, synthetic, or analytical approaches to information (thinking) and the preferences of other people for more personal, subjective, and evaluative assessments of information (feeling). All three of these dimensions stem from Jung's (1923) theory of personality types. The fourth dimension, distinguishing between a preference for evaluating the moment-to-moment and day-to-day influx of information (judging) versus merely gathering and storing data for use (perceiving), is an innovation of the MBTI, which doubles the possible combinations of personality types from Jung's 8 types to 16. In addition, continuous scores for each dimension may be obtained and employed in correlational or regression analyses.

The psychological reality of these dimensions becomes especially apparent to the taker of the MBTI when it is interpreted using the descriptions provided by Myers (1962/1975). The parallels between the descriptions of the individual from the test and the individual's own self-perceptions are often nothing short of uncanny to some clients on first assessment. In a study by Carskadon and Cook (1982), 50% of the subjects rated their type descriptions as "very true of me," and very low percentages of subjects rated artificially reversed descriptions as characteristic. Cohen, Cohen, and Cross (1981) also found significant correlations between ratings of spouses and the subject's MBTI profiles (except on the JP scale). A recent report by Ware and Yokomoto (1985) using a different set of descriptions from those of Myers confirms these results. Subjects were asked to rate in percentages the extent to which each profile described them. The average rating of a type description as similar to one's own MBTI type was 61.7%, while the average rating of an opposite type was 28.8%.

The point of this mention of the apparent psychological reality of MBTI dimensions is clearly *not* that this serves to empirically validate the test. However, from the standpoint of both the user and taker of the MBTI, the credibility of the test is greatly enhanced by this aspect. If it is useful to promote (a) a client's self-awareness (such as in career counseling, cf. Pinkney, 1983) or (b) his or her awareness of the preferences and actions of

someone else (such as one's spouse in couple's therapy, cf. Sherman, 1981), the perceived veridicality of MBTI-generated descriptions is a feature of considerable importance. To the extent that attitudinal and behavioral change in a client is a positive function of the credibility of the therapist and his or her instruments, with certain therapeutic problems the MBTI may be a genuine assist in this respect.

With respect to the empirical issue of reliability, among the relatively few reports on the MBTI, two of the more recent studies I previously reviewed are characteristic in terms of coefficient levels. Howes and Carskadon (1979) examined scores on the MBTI across 5-week intervals under conditions of artificially induced mood changes. Some of the subjects were given questionnaires that induced either mood elevation or depression, reversed across the two testings. Reliability coefficients for the MBTI continuous scores ranged from .78 to .87 across the 5 weeks, irrespective of the subjects' mood states, an innovative demonstration of the stability of MBTI measures. Similarly, McCarley and Carskadon (1983) reported 5-week test-retest reliabilities of continuous scores ranging from .77 to .89, depending upon the scale.

In a recent study, Leiden, Veach, and Herring, 1986, compared scores of medical students on the original (166 item) form of the MBTI and an abbreviated (50 item) form across a period of 9–21 months. Considering both the longer period of retesting and the change of forms, it could be expected that r values would be lower than usual for reliability, but they were significant nevertheless: .41 (TF scale), .63 (SN scale), .64 (EI scale), and .66 (JP scale). On the other hand, the median percentage of subjects receiving the same typology on the second administration was high, 79%. Leiden et al. (1986) also allude to satisfactory reliability of the long form over 3½ years but fail to document the report. With respect to reliability within the MBTI, Inclan (1986) administered different language versions of the instrument to a sample of bilingual subjects. The analyses yielded split-half reliability coefficients on continuous scores of .77 to .97 (English version) and .81 to .88 (Spanish version), values within the ranges reported earlier in Carlyn's (1977) review of the literature.

Therefore, despite the fact that there are still very few studies on reliability of the MBTI, especially test-retest reliability, recent reports remain favorable. Nevertheless, if MBTI type is to be regarded as a consistent predictor of behavior, the need remains for reliability assessment over somewhat longer periods of time.

Two recent studies have yielded conflicting results on the item structure of the MBTI. In one study, a factor analysis of the items revealed four factors that matched the instrument's scales (Tseng, Outcalt, Boyer, Ware, & Landis, 1984). In the other study, six factors were found, four corresponding to the MBTI scales, and among the latter a number of items failed to load significantly on their respective factors (Sipps, Alexander, & Friedt, 1985). Clearly, more research on the convergent validity of the MBTI is in order.

Turning to the issue of interest and criterion validity, owing to the theoretically unsystematic way in which data has been gathered on the MBTI, it is difficult to conveniently summarize the literature in an effort to make a supportive case. Also, as previously reported (Carlson, 1985), there has been a tendency to emphasize the EI scale in research with the MBTI, leaving the issue of validity even less certain for the remaining scales. In the category of interest validity, various of the dimensions of the MBTI have been related to Eysenck's Personality Questionnaire, Rotter's Locus of Control Scale, Harvey's "This I Believe" test, Kelley's Role Construct Repertory, and Bem's Sex Role Inventory, among other instruments. In general, significant correlations were reported where there were theoretical reasons for overlapping constructs (Carlson, 1985). The pattern of re-

ports thus appears to validate the scales, albeit in a somewhat haphazard fashion.

With respect to criterion studies, the previously reviewed one by Sherman (1981) in a counseling setting is especially pertinent to our current concern. She observed that clients who were similar or completely opposite on all four MBTI dimensions reported the fewest relationship problems but that certain patterns of type differences were especially problematic. For example, differences on the judging-perceiving dimension were correlated with the greatest number of reported problems. Again, a more systematic look at the variables would be helpful. For instance, a larger variety of MBTI types in combination could be examined, along with attempts to categorize problems by content as well as frequency.

Several recent studies also relate to the issue of criterion validity of the MBTI. In one that is directly relevant in the clinic, Garden (1985) observed that feeling-oriented respondents on the MBTI who were experiencing "burnout" as measured by a questionnaire were likely to manifest negative reactions to others. By contrast, thinking-oriented subjects experiencing burnout exhibited mixed or more positive reactions. The author argues that these results have implications for conceptual treatments of burnout that have focused on human services professionals, a group that is overrepresented by feeling types.

In another recent study of the TF dimension, Ware, Wilson, and Yokomoto (1986) hypothesized that, since thinking types prefer analysis and logic in their "weighing" of the facts, they should spend more time evaluating incoming stimuli than feeling types, who rely more on their sheer "likes and dislikes" for information to be evaluated. The hypothesis was tested experimentally through presentation of a variety of selected photographs and observation of time spent looking at each photo. Thinking types viewed all types of photos significantly longer than feeling types. Moreover, the extraversion-introversion dimension failed to predict viewing time, as predicted from some earlier studies of this dimension. In addition to validity assessment, this study demonstrates a potentially useful aspect of the MBTI. In a therapeutic setting the differences between thinking and feeling types might be apparent, say, in the time a client needs to spend processing interpretive material. An effective therapist, knowledgeable of the client's preferences for information processing, would make this allowance.

An article by Hicks (1984) provides some limited criterion validity data plus considerable insight into the nature of the MBTI typological analysis. The study conducted by Hicks related "bookishness," defined as number of books read in the preceding 12 months (as recalled by the subjects), to the sensation-intuition scale of the MBTI. The subjects were a broad sample of adults of both sexes, half of whom were S types and the other half N types, with no effort made to otherwise select on MBTI dimensions. As expected from type descriptions of the SN scale, a difference between subjects representing the two poles in terms of bookishness appeared, with the intuitive subjects (as defined by either preference or continuous score) reporting having read significantly more books than the sensing subjects. The remaining MBTI scales were not predictive, either singly or in interaction with the SN scale. However, in terms of library participation (holding a library card), both the SN and EI scales were predictive with one important adjustment. Within the introversion end of the scale (above the midpoint), the higher the continuous score, the greater the proportion of library participation. The extraversion end of the continuous scores was not predictive. In other words, the midpoint of the scale itself was predictive on this criterion.

The point of Hicks's demonstration has to do more with the particular nature of the dichotomous scales of the MBTI than with the relatively trivial demonstration of reading behavior in

sensing and intuitive types. In Hicks's view the MBTI scales are at their predictive best when "discontinuities" at the midpoint (or other points) of the continuous scores on the scale are shown relative to the criterion. In Hicks's (1985a) defense of dichotomous scaling and his methods in particular:

To propose that psychological dichotomies may exist is to assault one of the most unquestioned of the articles of faith in contemporary psychological theory: the continuity hypothesis. . . . When the skeptical believer in continuity is introduced to the four variables composing the Jungian system, the first question is often, "Are the distributions bimodal?" My study (Hicks, 1984) agrees with Myers (1962/1975) that the answer is no, they are not, and they probably won't ever be, but this null finding is not important evidence against the theory. I argue, agreeing with Myers, that a much more telling requirement is that there be a midpoint discontinuity when a type variable is graphically plotted against a second criterion. (pp. 11-12)

In other words, if values on the criterion variable show abrupt changes at the midpoint of the continuous variable, there is evidence for a dichotomous variable. Normally it has been assumed that distributions must be bipolar before the predictive utility of the scales is established. Hicks suggests this alternative form of analysis. Unfortunately, the MBTI has been shown to possess this discontinuous characteristic on only two of its scales, EI and SN (citations in Hicks, 1985a). Most investigations have failed to check for midpoint discontinuity. It is conceivable that a much broader range of positive validity assessments would now be available if this discontinuity analysis had been applied.

Finally, Hicks (1984) points out that other forms of analysis that have not been systematically applied to MBTI data are those that are "interaction-sensitive." That is, if it can be shown that the MBTI scales *depend* on one another rather than that they simply *combine* in an additive fashion, somewhat more effective and innovative uses of the instrument may evolve. For example, an interesting interaction between the SN and TF scales was reported by Hicks (1985b). In this study the tendency to commit fundamental attribution errors—to "overattribute causality to dispositions when there are obvious situational constraints"—was found to be significantly lower in intuitive-thinking subjects than in other combinations of the SN and TF dimensions. In Hicks's view these data demonstrate the importance of "cognitive processing style" (the specific roles of the sensation-intuition and thinking-feeling "functions") in attribution errors. Other dimensions and another instrument reflecting subject sensitivity to the environment were not predictive. This is a sophisticated and theoretically important application of the MBTI showing not only its predictive power in relation to attribution theory but also the discriminant validity of its dimensions.

In summary, since my 1985 review, recent criterion-based assessments of the MBTI remain largely unsystematic theoretically but generally positive. In some instances, more methodologically sophisticated approaches have developed. Also, more studies have focused on dimensions of the instrument other than the EI scale, which was a traditional favorite for study.

New statistical approaches to the MBTI may have the potential for disclosing uses and relationships not hitherto revealed. The MBTI may lend itself more to the study of potentially useful dichotomies for the clinician or counselor than many other in-

struments owing to its breadth and its founding in a general theory. To me it seems premature simply to reject the MBTI owing either to its dichotomous approach to personality variables or to its foundations in analytical theory. Ultimately the most effective strategy will be more and careful assessment of the strengths and weaknesses of the instrument with proper methods and statistics, reserving judgment on its potential utility in understanding and treatment.

REFERENCES

- 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. *Journal of Personality Assessment*, 41, 461-473.
- Carskadon, T.G., & Cook, D. (1982). Validity of MBTI type descriptions as perceived by recipients unfamiliar with type. *Research in Psychological Type*, 5, 89-94.
- Cohen, D., Cohen, M., & Cross, H. (1981). A construct validity study of the Myers-Briggs Type Indicator. *Educational and Psychological Measurement*, 41, 883-891.
- Garden, A.M. (1985). The effect of Jungian type on burnout. *Journal of Psychological Type*, 10, 3-10.
- Hicks, L.E. (1984). Conceptual and empirical analysis of some assumptions of an explicitly typological theory. *Journal of Personality and Social Psychology*, 46, 1118-1131.
- Hicks, L.E. (1985a). Dichotomies and typologies: Summary and implications. *Journal of Psychological Type*, 10, 11-13.
- Hicks, L.E. (1985b). Is there a disposition to avoid the fundamental attribution error? *Journal of Research in Personality*, 19, 436-456.
- 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.
- Inclan, A.F. (1986). The development of the Spanish version of the Myers-Briggs Type Indicator, Form G. *Journal of Psychological Type*, 11, 35-46.
- Jung, C.G. (1923). *Psychological types*. London: Routledge & Kegan Paul.
- Leiden, L.I., 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.
- Lynch, A. (1985). The Myers-Briggs Type Indicator: A tool for appreciating employee and client diversity. *Journal of Employment Counseling*, 22, 104-109.
- McCarley, N., & Carskadon, T.G. (1983). Test-retest reliabilities of scales and subscales of the Myers-Briggs Type Inventory and of criteria for clinical interpretive hypotheses involving them. *Research in Psychological Type*, 6, 24-36.
- Morris, P.V. (1979). *Extraversion and introversion*. Washington, DC: Hemisphere Myers, I.B. (1975). *Manual: The Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press. (original work published 1962)
- Pinkney, J.W. (1983). The Myers-Briggs Type Indicator as an alternative in career counseling. *Personnel and Guidance Journal*, 62, 173-177.
- Sipps, G.J., Alexander, R.A., & Friedt, L. (1985). *Educational and Psychological Measurement* 45, 789-796.
- Sherman, R.G. (1981). Typology and problems in intimate relationships. *Research in Psychological Type*, 4, 4-23.
- Tseng, 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, 255-56.
- Ware, R., Wilson, T.J., & Yokomoto, C. (1986). Looking time at selected photographs by Jungian personality type. *Journal of Psychological Type*, 11, 59-63.
- Ware, R., & Yokomoto, C. (1985). Perceived accuracy of Myers-Briggs Type Indicator descriptions using Keirsey profiles. *Journal of Psychological Type*, 10, 27-31.

John G. Carlson is a professor and chair of the Department of Psychology, University of Hawaii, Honolulu. Correspondence regarding this article should be sent to John G. Carlson, Chair, Dept. of Psychology, University of Hawaii, 2430 Campus Road, Honolulu, HI 96822.