



## Recent Assessments of the Myers-Briggs Type Indicator

John G. Carlson

To cite this article: John G. Carlson (1985) Recent Assessments of the Myers-Briggs Type Indicator, *Journal of Personality Assessment*, 49:4, 356-365, DOI: [10.1207/s15327752jpa4904\\_3](https://doi.org/10.1207/s15327752jpa4904_3)

To link to this article: [http://dx.doi.org/10.1207/s15327752jpa4904\\_3](http://dx.doi.org/10.1207/s15327752jpa4904_3)



Published online: 10 Jun 2010.



Submit your article to this journal [↗](#)



Article views: 199



View related articles [↗](#)



Citing articles: 61 View citing articles [↗](#)

## Recent Assessments of the Myers-Briggs Type Indicator

JOHN G. CARLSON  
University of Hawaii

**Abstract:** The present paper focuses on approximately two dozen recent published studies that examined reliability and validity of the Myers-Briggs Type Indicator (MBTI) in clinical, counseling, and research settings. Several assessments of split-half and test-retest reliability of the standard Form F and shorter Form G of the Inventory have yielded generally satisfactory correlations for all four scales. A larger number of studies of construct validity of the MBTI have yielded support for research hypotheses in situations ranging from correlations of the MBTI with a personality inventory, to couples problems in a counseling setting, to line judgment in groups, and others. Therefore, the applications of the MBTI have been broad, although somewhat unsystematic, and with generally favorable validity assessment. Continued attempts to validate the instrument in a variety of settings are needed.

In the words of Isabel Briggs Myers, the Myers-Briggs Type Indicator (MBTI) is a test designed "to implement Jung's theory of type" (1962, p. 1). Therefore, like the projective techniques, the MBTI is closely allied with psychodynamic thought, at least in its original conception. However, unlike the projectives, the MBTI is a highly quantified device with distinct methods of administration, objective scoring, and for which data exists concerning norms, reliability, and validity.

Following a brief description of the theoretical underpinnings of the MBTI, this paper will review very recent assessments of reliability and validity of the scales within the Indicator. Somewhat more extensive reviews of the earlier literature relevant to the MBTI may be found in Myers (1962) and Carlyn (1977). The present paper focuses on published studies from 1975 that examined reliability and validity of the MBTI in clinical, counseling, and research settings.

In one view, Carl Jung's theory of types (1923) is based upon distinctions with respect to the sources from which information is derived, the ways that people perceive information, and the ways that information is dealt with in reaching conclusions. In Jung's view, psychic energy, or libido, may flow out-

ward, towards other people or external objects in the "attitude" labeled *extraversion*. By contrast, the libido may flow inward, towards one's own conscious self, defining the attitude, *introversion*. The distinction essentially determines one's primary direction for mental functioning, either predominately upon the external world as a source of factual material (the extraverted personality) or upon the internal world as a source of ideas and concepts (the introverted personality).

Besides these two fundamental attitudes, each of us is said to perform mental "functions" along two dimensions. The first dimension, perceiving, describes two polar opposites in terms of the ways that information is received. In *sensation*, data takes the form of acts or sense impressions, that is, information from the five senses. In *intuition*, one's perceptions are indirect, distorted by the unconscious, and the sources of one's information are not very clear, as in "hunches" or the "sixth sense." The intuitive-dominant personality will tend to rely more upon the latter mode for receipt of significant data; just as the sensation-dominant individual prefers to utilize the basic senses for receipt of information and not to go beyond the objective, empirical world for facts.

The other dimension of mental functioning in Jung's theory of types is the judging function, the function that processes in some "rational" way the information that is perceived, irrespective of

This is to express my sincere thanks to my dear friend and colleague, Dr. Ruth Sherman, for my first introduction to the MBTI and assistance during preparation of the manuscript.

its source. The *thinking*-dominant type prefers to analyze, synthesize, and determine the truth or falseness of information in an impersonal fashion. By contrast, the *feeling*-dominant type evaluates incoming data in terms of their goodness or badness, a process that is personal, subjective, and non-analytical (but nonetheless rational, in Jung's terms).

In any individual, one of the polar opposites of each function (sensing vs. intuition, thinking vs. feeling) and one of the two attitudes (extraversion vs. introversion) are possible. Therefore, in this view there are eight possible combinations or personality types in terms of mental attitudes and functions.

The Myers-Briggs formulation essentially parallels Jung's theory in terms of these three dimensions of personality plus a fourth dimension that was "implicit" in Jung's system (Matson, 1981). An individual's type is assessed by means of the MBTI which uses scales or indices for determining an individual's "preferences" for a given attitude or function. Thus:

the EI index is designed to reflect whether the person is an extravert or an introvert in the sense intended by Jung. . . . The SN index is designed to reflect the person's preference as between two opposite ways of perceiving, i.e., whether he relies primarily on the familiar process of sensing. . . or primarily on the less obvious process of intuition. . . . The TF index is designed to reflect the person's preference as between two opposite ways of judging, i.e., whether he relies primarily upon thinking. . . or primarily upon feeling (Myers, 1962, pp. 1-2).

In addition to the EI, SN, and TF indices, the MBTI introduces the JP (judging-perceiving) index. While not a direct application of Jung's type theory, the index does relate to Jung's notion that one end of a dimension of personality will tend to dominate in one's everyday interactions with the environment. The Myers-Briggs JP index assigns a preference to one or the other two mental functions themselves. That is, either the perceiving (SN) function or the judg-

ing (TF function) is said to be dominant in one's dealings with the world. Thus, with this additional dimension, 16 possible personality types are generated by combinations of preferences.

Form F of the Indicator, apparently the most commonly used version, contains 166 forced-choice items. Form G, a more recent innovation, makes use of 126 items and appears to be a reliable and faster alternative to Form F (Carskadon, 1979). Each scale of either form is a subset of items designed "to reflect . . . a habitual choice between opposites, analogous to right- or left-handedness" (Myers, 1962, p. 2).

For purposes of computing correlations and for regression analyses of the MBTI, a continuous score may be computed (or found in a table in Myers, 1962) that converts type and preference score into a single value. But more commonly, preference scores on the MBTI are expressed as types in terms of the closest pole on a given scale.

#### Reliability and Validity of the MBTI: 1975-83

The following is a selective discussion of recent attempts to establish various types of reliability and validity of the MBTI. The recent voluminous literature on the scale (one bibliography lists approximately 700 references) reflects largely successful efforts to apply it in a large variety of educational, clinical, counseling, business, and research settings. Perhaps the widest usage of the test is in educational areas, several reviews of which are contained in Carskadon (1981). The present review focuses on studies in clinical, counseling, and research settings.

#### Reliability

The original reliability studies, reported in the Myers-Briggs Manual, yielded split-half reliability coefficients (Pearson *rs*) commonly exceeding .80. More recent studies, including assessment of test-retest reliabilities, have also yielded favorable correlations.

Most assessments of the Indicator

have examined characteristics of each of the four scales separately. Carlyn's review of the MBTI (Form F) literature through 1975 reported tetrachoric coefficients for split-half reliability ranging from .66 to .92 and concluded that "Estimated reliabilities of type categories appear to be satisfactory in most cases" (Carlyn, 1977, p. 465) irrespective of the scales examined. Computations based on continuous scores are higher, as might be expected, since the conversion of score to type categories reduces information. Carlyn (1977) also 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 appear to be even more stable. Again in this case, analyses of continuous scores have typically yielded higher coefficients, with Pearson  $r$ s ranging from .69 to .83 across two month periods.

Since Carlyn's review, several other reliability studies have appeared. However, reports of test-retest reliabilities are sparse. Steele and Kelley (1976), as part of a study reviewed below, reported test-retest reliabilities of the *EI*, *TF*, and *SN* scales on the MBTI ranging from .86 to .89. Carskadon (1977) reported reliabilities of continuous scores on Form F across an eight week interval. For females, test-retest  $r$ s ranged from .73 to .87 on the four scales; for males,  $r$ s ranged from .56 to .79. In both cases, the lowest coefficients were obtained on the *TF* scale, a result that had been reported by others earlier (Stricker & Ross, 1964), but not obtained in a recent study (Carskadon, 1982).

Carskadon (1979c) also reported test-retest reliabilities for the short Form G of the MBTI. Across seven week intervals, the four scales yielded  $r$ s ranging from .48 to .84 for continuous scores. Carskadon also compared these coefficients with those obtained in his study using Form F (1977) and found no differences.

A study of reliability of a still shorter version of the MBTI, consisting of the first 50 items of Form G, yielded mixed

results. Kaiser (1981) compared category types based on the shorter form to those on Form G and found that type classification was unchanged in 91% to 96% of the cases using the shorter version. However, when scores themselves were examined, the only dimension that did *not* show significant changes was the *EI* scale. Therefore, Kaiser concludes that for nominal type the shorter version of the Form G is congruent with the standard Form G for all categories; but for actual scores reflecting strength of type, the shortened version of the MBTI is unreliable in at least three categories.

In a nicely conceived study, Howes and Carskadon (1979) examined the reliability of Form G as a function of both time and mood changes. In each of two administrations at five week intervals, subjects were given a bogus "Self-report Inventory (SRI)" designed actually to produce either mood elevation, mood depression, or no mood changes (control). Among other questionnaires, the subjects were also administered the MBTI. The SRI content from first to second administrations for the two groups was reversed from elevation to depression, and vice-versa. Significant changes in moods on a "mood adjective questionnaire" were obtained consistent with the SRI manipulations. However, reliabilities of test-retest continuous scores on the MBTI ranged from .78 to .87 and showed no significant differences as a function of the mood manipulations. An additional analysis of type reliability showed that 19% of subjects changed type on the *EI* scale, 11% on *SN*, 17% on *TF*, and 16% on *JP*, and that the greater the preference score, the lower was the likelihood of type change upon retest.

Very recently, McCarley and Carskadon (1983) again examined reliabilities for the four subscales of Form G. Across all items (five week retest period), reliability coefficients for continuous scores were high, ranging from .77 (*TF* scale) to .89 (*JP* scale). Interestingly, however, the percentage of subjects who retained their specific dichotomous type preferences across all four scales

was only 47%. In other words, a subject who had, say, an ESTJ preference on first testing had only a 50-50 chance of maintaining the identical preference on every one of the four subscales upon retesting. It was additionally found that these results applied equally well to type of item, that is, the phrased questions or the word pair items used in the MBTI, as well as to split-half assessments ("X-half" and "Y-half" forms within the MBTI). The authors regard the former as important since they view word pair items to be more valid than phrased questions, whereas the X-half, Y-half comparison is aimed simply at consistency of the test.

Finally, the reliability of a Spanish translation of the MBTI, Form G, was recently reported by Levy and Padilla (1982). Beyond the fact that the test was in Spanish, the population of Puerto Ricans from which the subjects were drawn is also unusual in research with the MBTI. Test-retest reliabilities were high, ranging from .79 (*TF* scale) to .89 (*JP* scale). The *r* values may have been positively influenced by the short, two week, test-retest interval in this study.

In summary, published studies on reliability of the MBTI are relatively few in number. However, the 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. The latter generalization is particularly true of nominal type. In one study, test-retest reliability was unaffected by verbally induced mood changes. At this time, there is a need for more reliability assessment, especially across longer periods of time and a wider variety of test conditions. In addition, since most studies have sampled only college students, there is a need for assessment with a wider range of subject populations.

#### Construct Validity

By far, the largest number of recent reports on use of the MBTI have been with respect to its construct validity, that is, the extent to which the test mea-

sures "some attribute or quality (construct) that people are presumed to possess" (Brown, 1976, p. 128). It seems premature to consider research on the MBTI as more than just tentative beginnings on the establishment of the "nomological net" (Cronbach & Meehl, 1955) that underlies the test. Jung's theory itself is often vague and concepts ill-defined. Moreover, translations of the theory as well as attempts to operationalize it in the form of the MBTI often leave out conceptual steps that would allow for clear empirical assessment of related constructs. Nevertheless, the research reviewed here can be seen as important in the early stages of development of useful measurement of Jung's constructs as translated by the MBTI. The degree to which attempts to empirically validate the test are successful will no doubt influence the likelihood that the nomological net will gain wider acceptance (Brown, 1976).

A brief summary of evidence with regard to content and predictive validity of the MBTI was included in the paper by Carlyn (1977). In addition, approximately 50 papers, many of them unpublished, in the application of the MBTI in clinical and counseling settings were summarized by Carskadon (1979b). The following review focuses on published studies relating the MBTI to other assessment devices and to a variety of behaviors in various clinical, counseling and research contexts.

#### Intertest correlations

Wakefield, Sasek, Brubaker, and Friedman (1976) correlated the MBTI with Eysenck's Personality Questionnaire (EPQ). The latter contains scales for psychoticism, neuroticism, extraversion, and lying. The only significant correlation obtained was between the *EI* scale on the MBTI and the extraversion scale on the EPQ. The authors also reported nonsignificant but moderately positive *rs* between the *TF* scale and neuroticism (a result paralleling that of McCary & Grant, 1976, using another anxiety test).

Steele and Kelley (1976) speculate that considerable overlap exists between

the personality dimensions assessed by the Eysenck extraversion scale and the *EI* index of the MBTI. In this study, as in that by Wakefield et al. (1976), the *EI* scale was found to correlate significantly with Eysenck's extraversion factor ( $r = .74$ ). In addition, Steele and Kelley argue that their data provide an assessment of the discriminant validity of the other scales on the MBTI. That is, the remaining scales failed to correlate significantly with one another or with any of the scales of the EPQ.

A result that may not be expected on intuitive grounds was reported by Eliot and Hardy (1977) in a study of the relationship between *EI* preference on the MBTI and "internality" on Rotter's (1966) locus of control scale. An internal orientation on the locus of control test is said to be characteristic of individuals who attribute control of significant outcomes of behavior to their own behavior. An external orientation typifies an individual who attributes control of reinforcers and punishers to outside forces or people (Carlson, 1982). Eliot and Hardy argue that internals should tend to be more extraverted since, theoretically, they have experienced more success in attaining reinforcers as a function of their behavior in social contexts. The results of the study support this analysis since extraverts on the MBTI obtained significantly more internal scores on Rotter's test.

Apparently, very few attempts to correlate the other three scales on the MBTI with other tests have been made. In one exception, an application of the MBTI to modes of conflict handling was proposed by Killman and Thomas (1975). College students were administered three tests designed to assess an individual's methods for handling conflict (the dimensions being competing, collaborating, compromising, avoiding, or accommodating). When the MBTI was administered, a preference for feeling on the *TF* scale was generally significantly related to accommodation in a conflict situation as measured by the conflict-mode tests. In addition, extraverted individuals on the *EI* index tended to be

those who, when combining scores on the conflict-mode tests, were integrative, assertive, and cooperative. Neither the *SN* or *JP* scales were significantly correlated with the conflict tests in this analysis. The authors conclude that the Jungian dimensions are useful "in documenting and explaining psychological bases of interpersonal behavior" in conflict situations (Killman & Thomas, 1975, p. 979).

Carskadon and Knudson (1978) reported correlations between the MBTI and O. J. Harvey's "This I Believe" test. The latter is used to classify individuals primarily with respect to emphases upon concreteness and autonomy in behavior. At one end of a four-step continuum, System I individuals have "lower" conceptual systems, tend to agree with external standards, and tend to be relatively resistant to expressions of personal autonomy. At the other end of the continuum, System IV individuals use "higher" conceptual systems are open to expressions of autonomy and/or multiple alternatives for action (Carskadon & Knudson, 1978). The researchers speculated that the *SN* scale also assesses the tendency not to go beyond the objective facts (sensing types) or to trust largely indirect and personal sources of information (intuitive types). The results favored this analysis since sensing types were overrepresented in the System I category and intuitive types were overrepresented in the System IV category.

Kelley's Role Construct Repertory Test and the MBTI were administered by Carlson (1980) to adult subjects in one of three studies (also discussed in the next section). The former test is designed to assess "individuals' construction of their personal worlds." Carlson hypothesized that the *SN* scale on the MBTI would differentiate subjects in terms of responses on Kelly's test because intuitive types are more likely than sensing types to develop personal constructs that are "inferential" and go beyond the basic concrete facts. The results strongly favored this analysis. "Sixteen of the 20 intuitives, compared with only 3 of the 14 sensing types, offered a

preponderance of inferential constructs" on Kelly's tests, a highly significant effect (Carlson, 1980, p. 808). Other MBTI scales were not predictive in this study.

Finally, Padgett, Cook, Nunley, and Carskadon (1982) generated a number of clear predictions concerning the relationship between androgyny, as measured by the Bem Sex Role Inventory, and type preferences on the MBTI. Three of their predictions were confirmed in a group of male and female college students. It was found that androgynous males tended to prefer feeling, whereas androgynous females preferred thinking on the *TF* dimension. The *EI* scale predicted significantly for women only; more androgynous women were extraverts than introverts. Also, in men both *SF* and *NF* preferences were more closely associated with androgyny, whereas *ST* and *NT* preferences were associated with sex-typed responses. The authors conclude that the MBTI may add information to sex typing that will aid clinicians in identifying the flexible and adaptive behaviors that characterize both androgyny and mental health. Clearly, however, considerably more data is needed before the value of the MBTI in making such predictions is established.

#### *Criterion-Related Studies*

The remaining studies in this review have attempted to relate types on the MBTI to a variety of behaviors in treatment and research settings. It may be noted that following Carlyn's (1977) general review of a number of earlier published studies it was concluded that "The Indicator appears to be a reasonably valid instrument which is potentially useful for a variety of purposes" (p. 471). A more negative conclusion was more recently reached in Carskadon's (1979b) review of the MBTI as applied in clinical and counseling settings in which problems in statistics, interpretation, magnitude of differences, and failures to replicate are cited. However, it should be noted that this review included a large number of unpublished theses and dissertations fraught with

methodological problems. In the present paper, only recent published studies are cited and, in the main, these are more positive in outcome as well as methodologically sound.

*The MBTI in treatment settings.* Among studies of the MBTI in clinical settings, Dewinne and Johnson (1976) suggested that in the Jungian-defined extravert there is an attempt to suppress infantile and primitive demands due to cultural pressures. In addition, the extravert is said to be impulsive and to be inattentive to the possible effects of behavior on the body. Therefore, there is some theoretical basis for the likelihood that the extravert may turn to drug abuse for the relief of cultural pressures. Dewinne and Johnson assessed the type characteristics of drug addicts and did find a significantly larger percentage of extraverts than introverts in their small sample.

This outcome might lead to a number of hypotheses concerning various other indices of psychopathology and type on the MBTI. For example, in Dewinne and Johnson's interpretation of Jung, introverts are said to be shy, distrustful, defensive, suspicious, and more likely than extraverts to be subject to neuroses characterized by fear and anxiety, obsessive-compulsive reactions, feelings of inadequacy, and the like. In one related study, however, a negative outcome was obtained in an attempt to predict neurosis and psychosis from the MBTI. No significant correlation between introversion and extraversion on the *EI* index and clinically diagnosed neuroticism or psychoticism was found by Hughes and Johnson (1975). Note that this result also conflicts with the Wakefield et al. (1976) report of a correlation between the *EI* scale and test neurosis, cited in an earlier section of this paper.

In a very large study comparing psychiatric patients with earlier studies on "normal" populations, Bisbee, Mullaly, and Osmond (1982) found many significant type preferences. The researchers compared patients in four groups: depression, schizophrenia, substance abuse, and bipolar-manic disorder. Across all

four groups, ISFJ, ISFP, and ISTJ types were overrepresented relative to the normal populations. As a group, substance abusers tended to be more extraverted than the other patient groups. In the schizophrenics, no types were significantly overrepresented by comparison with the patient groups, suggesting to the authors that this disorder occurs without reference to type and thus is biological in origin. Based upon these and a number of other results in this study, the authors conclude that knowledge of type is essential to providing appropriate treatment. However, it is clear that considerable research yet remains in order to clarify the actual utility of the MBTI in the treatment of psychiatric disorders. The Bisbee et al. results merely demonstrate potential differences among disorders in terms of type preferences. The differences have no clear theoretical bases and, therefore, no particular usefulness in the establishment of a nomological net. Moreover, while the comparisons between patient groups were sound, those between the groups and earlier "normal" populations lack the credibility of more traditional control methods, that is, control groups within the study itself.

In a counseling setting, Sherman (1981) demonstrated several clear correlations between MBTI-assessed type and problems in intimate relationships in an impressive analysis of 167 couples. Not surprisingly, the fewest relationship problems were reported by individuals similar to one another on all four MBTI scales. However, the next lowest category of couples' problems was among individuals with completely opposite typology (ESTJ males paired with INFP females). Among other findings, all males (extraverted or introverted) reported fewer relationship problems with introverted females. However, extraverted females reported significantly fewer problems with extraverted males than with introverted males. For either sex, sensing types reported significantly fewer couples problems when paired with sensing types. Intuitive types, however, reported more problems when paired

with intuitive types than when paired with sensing types. Similarly, on the JP scale, judging types reported fewer problems when paired with judging types and significantly greater satisfaction in marriage than all other combinations of type. By contrast, perceiving types paired with judging individuals reported the most problems. These and a number of other relationships found by Sherman demonstrate the clear, although complex, utility of the MBTI in one aspect of counseling work — identifying sources of relationship problems.

*The MBTI in research settings.* In this research context, five recent reports on the MBTI are notable. Among them, Carlson (1980) outlines studies with college students typed on the MBTI and subjected to various tests of memory and social perception. (See also the discussion of one of these studies in the reliability section above.) In the first study of the series, subjects were asked to describe personal and critical life incidents relevant to seven basic emotions. Essentially, extraverts remembered more incidents that occurred in social contexts, whereas introverts remembered more emotional events occurring in individual contexts. Also, feeling types reported significantly more intense emotions of joy, excitement, and shame than thinking types. In another study, adult subjects were asked to imagine that they were writing a letter to an unknown person in another country and attempting to provide a self-description. Most of the intuitive subjects but hardly any of the sensing subjects made reference to the imagined other person. Conversely, a significantly higher proportion of sensing types made reference to the physical appearance of the other person. Carlson concludes that the *SN* scale successfully differentiates subjects who will directly participate in imagined events from those who do not, and it differentiates subjects who use concrete self-description from those who do not. In passing, it may also be noted that a high degree of discriminant validity was also established in these studies since the only predictive MBTI scale in each study was the one

originally hypothesized to be relevant to the manipulations.

In one creative experiment, Carskadon (1979a) gave subjects who were differentiated as extraverts and introverts on the MBTI an opportunity to present a five-minute talk in front of judges on the subject of political candidacy. As hypothesized, extraverts stood closer to the judges, spent less time in silence during the talk, and remembered more of the judges' names later than did introverts. However, extraverts did not use more gestures and expressive behavior in their presentations.

In a recent report, Matthews, Miller, and Carskadon (1981) examined conformity in judgment of line length in group contexts as a function of type on the MBTI. It was hypothesized from earlier studies (e.g., Cooper & Scalise, 1974) that extraverts would be more likely to conform to group influences than introverts. Also, prior evidence suggested that feeling types would be more conforming than thinking types. Confederates of the experimenters participated in line-length judgments (paralleling the classic Asch conformity procedures) in an attempt to influence statements concerning comparison lines relative to standard lines. The results were complex but essentially supported the hypotheses: *IN* types conformed the least of all of the groups and *EF* and *ES* types together tended to conform more than *IT* and *IN* types.

Finally, in two studies investigators have compared types as measured by the MBTI with subjects' and others' description of their type. Carskadon and Cook (1982) examined the ability of their subjects to describe their own type in terms of their selection of four one-page type descriptions. Based upon the results of the MBTI administered two months before, the subjects were each given packets of self-descriptions that were either their actual type, their type with *SN* and *TF* scales reversed, or the completely opposite type. Fifty percent of the subjects rated the actual type description as "very true of me," whereas only 10% and 13%, respectively, ranked

the *SN*, *TF* reversal and the completely opposite type as "very true of me." In short, subjects were generally good at selecting the type description that matched their MBTI preferences and at recognizing that the opposite type was unlike them.

In a somewhat more complex study, Cohen, Cohen, and Cross (1981) compared self-ratings and ratings of spouses on Likert scales corresponding to definitions of type (Myers, 1962) to actual MBTI types. By contrast with Carskadon and Cook's (1982) results, the authors found no relationship between self-ratings of type and the MBTI preferences. Similarly, the subjects' ratings of their perception of their "ideal" selves showed no relationship to MBTI type. However, except on the *JP* scale, the spouses' ratings of the subjects' types corresponded significantly with the MBTI types. The authors argue that the failure to obtain a relationship between subjects' ratings of ideal selves and MBTI preferences was an indication of the "divergent validity" of the MBTI. That is, the MBTI does not measure ideal or perceived socially desirable self-perceptions but rather actual type preferences. Further, the fact that spouses recognized the subjects' type on most scales demonstrates the construct validity of the MBTI: that is, its measures correspond with the psychological reality of others' assessments of the individual. This experiment should weigh heavily with those who are concerned with the fundamental psychological usefulness of the Indicator.

### Conclusions

As noted earlier, both internal and test-retest reliability of both forms F and G of the MBTI have proven satisfactory in recent assessments, with *r* values of individual scales often exceeding .80. The greatest need for reliability assessment at this time appears to be studies with a greater diversity of test conditions and populations since many of the existing studies utilized testing environments in university settings and students as subjects. Moreover, longer test-

retest intervals should be sampled in future research.

Validity of the MBTI remains in greater question than reliability. Although relationships between the Indicator and other tests have generally supported hypotheses concerning underlying theoretical overlap, five of the eight studies reviewed in the intertest correlations section concentrated upon the *EI* scale of the instrument. However, it is notable that the three more recent studies cited have found evidence of validity of some of the remaining scales when correlations with other tests were performed. Future studies should give relatively greater attention to these other scales.

Eight other reports of attempts to correlate the MBTI with a variety of behaviors in treatment and research settings were also generally positive with respect to the hypotheses tested, with some exceptions that were noted in the criterion-related studies section. 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. However as in the case of correlations with other tests, the emphasis in several of the studies cited in this section was upon the *EI* scale, leaving the question of validity of the bulk of the items on the MBTI still open to more thorough examination. Fortunately, there appears to be a trend in more recent studies to examine the validity of the other scales. Perhaps the present review will encourage more research of this type in treatment and research settings. Also a greater variety of validity research with less restrictive populations, that is, outside of the university, and by investigators with little vested interest in the instrument would be a great advantage. The current tendency for a relatively small number of researchers to conduct a large proportion of the research in this area leaves the generalizability of some of the findings yet to be established.

Finally, while for construct validation a wide variety of research has clear utility, at the same time there has been a notable lack of systematic programs of

research on the Indicator, such as has characterized, for instance, development of the MMPI. Perhaps future efforts to validate the Indicator in treatment and research settings will incorporate more systematic approaches and thereby help tighten the nomological net.

#### References

- Bisbee, C., Mullaly, R., & Osmond, H. (1982). Type and psychiatric illness. *Research in Psychological Type*, 5, 49-68.
- Brown, F. G. (1976). *Principles of psychological testing*. New York: Holt, Rinehart, and Winston.
- Carlson, J. G. (1982). Some concepts of perceived control and their relationship to bodily self-control. *Biofeedback and Self-Regulation*, 7, 341-375.
- Carlson, J. R. (1980). Studies of Jungian typology: II. Representations of the personal world. *Journal of Personality and Social Psychology*, 38, 801-810.
- Carlyn, M. (1977). An assessment of the Myers-Briggs Type Indicator. *Journal of Personality Assessment*, 41, 461-473.
- Carskadon, T. G. (1977). Test-retest reliabilities of continuous scores on the Myers-Briggs Type Indicator. *Psychological Reports*, 41, 1011-1012.
- Carskadon, T. G. (1979a). Behavioral differences between extraverts and introverts as measured by the Myers-Briggs Type Indicator: An experimental demonstration. *Research in Psychological Type*, 2, 78-82.
- Carskadon, T. G. (1979b). Clinical counseling aspects of the Myers-Briggs Type Indicator. *Research in Psychological Type*, 2, 83-84.
- Carskadon, T. G. (1979c). 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. (Ed.). (1981). *Research in Psychological Type*, 3.
- 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.
- 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.
- Carskadon, T. G., & Knudson, M. L. (1978). Relationship between conceptual systems and psychological types. *Psychological Reports*, 42, 483-486.
- Cohen, D., Cohen, M., & Cross, H. (1981). A construct validity study of the Myers-Briggs Type Indicator. *Educational and Psychological Measurement*, 41, 883-891.
- Cooper, J., & Scalise, C. J. (1974). Dissonance produced by deviations from life styles: The

- interaction of Jungian typology and conformity. *Journal of Personality and Social Psychology*, 29, 566-571.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Dewinne, R., & Johnson, R. W. (1976). Extraversion-introversion: The personality characteristics of drug abusers. *Journal of Clinical Psychology*, 32, 744-746.
- Eliot, J., & Hardy, R. C. (1977). Internality and extraversion-introversion. *Perceptual and Motor Skills*, 45, 430.
- Eysenck, H. J. (1973). *Eysenck on extraversion*. New York: Wiley.
- 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.
- Hughes, R. C., & Johnson, R. W. (1975). Introversion-extraversion and psychiatric diagnosis: A test of Eysenck's hypothesis. *Journal of Clinical Psychology*, 31, 426-427.
- Jones, J. H., & Sherman, R. G. (1979). Clinical uses of the Myers-Briggs Type Indicator. *Research in Psychological Type*, 2, 32-45.
- Jung, C. G. (1923). *Psychological types*. London: Routledge & Kegan Paul.
- Kaiser, K. M. (1981). Use of the first 50 items as a surrogate measure of the Myers-Briggs Type Indicator. *Research in Psychological Type*, 4, 55-61.
- Killman, R. H., & Thomas, K. W. (1975). Interpersonal conflict-handling behavior as reflections of Jungian personality dimensions. *Psychological Reports*, 37, 971-980.
- Levy, N., & Padilla, A. (1982). A Spanish translation of the Myers-Briggs Type Indicator Form G. *Psychological Reports*, 51, 109-110.
- Matthews, S. K., Miller, D. I., & Carskadon, T. G. (1981). Conformity behavior in an experimental situation as a function of psychological type. *Research in Psychological Type*, 4, 62-70.
- Mattoon, M. (1981). *Jungian psychology in perspective*. New York: The Free Press.
- 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.
- McCary, P. W., & Brant, W. H. (1976). The relationship between trait anxiety and Myers-Briggs personality — a question of validity. *Colorado Journal of Educational Research*, 15, 28-30.
- Myers, I. B. (1962). *Manual: The Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Padgett, V., Cook, D., Nunley, M., & Carskadon, T. (1982). Psychological type, androgeny, and sex-typed roles. *Research in Psychological Type*, 5, 69-88.
- Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80 (1) Whole No. 609.
- Sherman, R. G. (1981). Typology and problems in intimate relationships. *Research in Psychological Type*, 4, 4-23.
- Steele, R. S., & Kelly, T. J. (1976). Eysenck personality questionnaire and Jungian Myers-Briggs Type Indicator correlation of extraversion-introversion. *Journal of Consulting and Clinical Psychology*, 44, 690-691.
- 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.
- Wakefield, J. A., Sasek, J., Brubaker, M. L., & Friedman, A. F. (1976). Validity of Eysenck's Personality Questionnaire. *Psychological Reports*, 39, 115-120.

John G. Carlson  
Dept. of Psychology  
University of Hawaii  
2430 Campus Rd.  
Honolulu, HI 96822

Received: March 24, 1984  
Revised: January 23, 1985