Results from three different measures of type (the Implicit Association Test, the MBTI® instrument, and best-fit type) indicate an effect for social desirability as measurement becomes more reliant on self-report. The type of the person conducting feedback/best-fit discussions also influences best-fit type decisions, though not exactly as originally reported by Bathurst (2000).

Implicit, Explicit, and Best-Fit Assessments of Psychological Type: Explorations in “Shoes off” Measurement

Robert W. McPeek and Judith Breiner
Center for Applications of Psychological Type, Inc.

ABSTRACT
As with any personality measure dependent on self-report items, the Myers-Briggs Type Indicator® or MBTI® instrument is subject to unknown self-presentation bias reflecting presumed social desirability of responses and the ability to answer candidly and/or accurately. One strategy to improve accuracy employed in psychological type practice allows MBTI users to review their results and agree with or alter them to determine a “best-fit” type. This research compares three different measures of psychological type: best-fit results, MBTI results, and scores from a type measure based on the Implicit Association Test, or IAT (Greenwald, McGhee, & Schwartz, 1998), in order to illuminate potential bias in type measurement. Study 1 compared standard MBTI results for each of the four preference domains (E–I, S–N, T–F, and J–P) to results for the corresponding IAT measures. Study 2 compared reported type (MBTI results) on the four preferences to best-fit type, using two large datasets. All four IAT preference measures showed solid evidence of internal consistency (split-half correlations and Cronbach’s alphas). Compared to IAT scores, MBTI results showed a shift towards E, N, T, and J, preferences previous research has suggested are more...
socially desirable than I, S, F, and P. Best fit relative to reported preferences showed more S-to-N shifts than N-to-S and more E-to-I than I-to-E changes. One dataset showed a shift from P to J, but the second did not. Females were significantly less likely than males to change from F to T when comparing IAT to MBTI results or MBTI to best-fit results. A reanalysis of an expanded dataset first reported in Bathurst (2000) suggested that the preference of the MBTI professional conducting type verification (determining best fit in consultation with clients following presentation of MBTI results) led to switches from the opposite user preference to the same as the administrator for individuals with E, N, and P MBTI results, and for females with F results.

Note: For the Myers-Briggs Type Indicator® (MBTI®) instrument, the eight preference categories are the following: Extraversion (E) versus Introversion (I), Sensing (S) versus Intuition (N), Thinking (T) versus Feeling (F), Judging (J) versus Perceiving (P).

INTRODUCTION

The study of human personality is an investigation into an invisible domain full of fun house mirrors, trap doors, and mazes with dead-ends. From the largely discounted psychodynamic theories of Freud (1920/2010) to present day research (e.g., Gilbert, 2005; Wilson, 2002), personality theory is rife with stories of and evidence for unseen forces influencing behavior, often in ways outside of the awareness of the actor. Yet despite how poor human self-understanding often is, asking people to answer questions about themselves remains a primary tool in personality assessments, many of which treat self-report as not just useful but valid information (see Buros Institute, 2010).

The Myers-Briggs Type Indicator instrument (Myers, McCaulley, Quenk, & Hammer, 1998) is one such instrument, based in part on deliberative responses to queries about behavioral tendencies (“Do you usually get along better with a) imaginative people, or b) realistic people?”) or inclinations (“If you were a teacher, would you rather teach a) fact courses, or b) courses involving theory?”). Based largely on the work of Jung (1921/1971), an early associate of Freud’s who broke away to develop his own dynamic theories of personality, the MBTI instrument measures four different personality dimensions, described in the MBTI literature as follows:

Extraversion–Introversion (E–I) is concerned with the source and focus of “energy” for an individual, which can either be outwardly directed towards people and objects in the environment, or inwardly towards concepts, ideas, and internal experiences. An Extravert (E) is described by adjectives such as active, expressive, or sociable, whereas an Introvert is typically depicted as calm, reflective, or quiet. (Martin, 1997, is one of many possible sources for these and the following descriptions.) Preference for E and I are about evenly divided in adult samples drawn from the United States (Martin, 2003; Myers et al., 1998).

Sensing and Intuition (S–N), the perceiving functions, describe two opposing modes of taking in information. Sensing is based on observable, tactile, or other sensory data, whereas Intuition attends to meanings, patterns, and implications of such information. Sensing (S) individuals are described as factual, detailed, or practical, and Intuitive (N) individuals are described as interested in possibilities, abstract, or even unconventional. (N is used instead of I to avoid confusion with Introversion.) Sensing is estimated to be the preference of 70–75% of the US population (Martin, 2003; Myers et al., 1998).

Thinking and Feeling (T–F), the two opposing judging functions, describe how information drawn from the world or from internal sources is evaluated, either on the basis of impartial logic (Thinking) or personal or social values (Feeling). A Thinking (T) individual is described as tough-minded, logical, or impersonal, and a Feeling (F) person as warm, compassionate, or tender-hearted. T–F is the only type dimension with an important gender component. Two-thirds to three-quarters of U.S. females prefer F and up to two-thirds of males prefer T (Martin, 2003; Myers et al., 1998).

Judging–Perceiving (J–P) describes an individual’s preferred method of interacting with the outer world. A person with a Judging preference relies more on the use of his or her Judging function (Thinking or Feeling) and tends to prefer order and closure. A person who prefers Perceiving relies more on the Perceiving function (Sensing or Intuition) and is drawn to exploration, spontaneity, and flexibility. A preference for J (approximately 55%) is slightly more normative in the United
States than a preference for P (Martin, 2003; Myers et al., 1998).

The term “preference” (one letter or its opposite), used to denote results for each of the four measures, has an important quality: while the preferred result is more descriptive and habitual, the opposing letter may also (though less frequently) inform an individual’s behavior. Thus, while the four MBTI preference results point to only one letter for each of the four dimensions (for example, E, N, F, and P), or domains or preferences pairs as they are also called, behaviors consistent with the unreported letters (in this case, I, S, T, and/or J) may also be part of an individual’s repertoire, albeit less preferred.

The four type dimensions combine into one of 16 possible four-letter “types,” such as ESTJ or INFP. Use of the term “type” has some controversial implications for the theory. Chief among them is the practice of categorical measurement, in contrast to most personality measures, which characterize measurement as a point along a scale. Instead, the MBTI results make four dichotomous classifications, though the assignment to E or I, T or F, etc., is in fact based on an underlying scale. These continuous scores are typically used to calculate test-retest reliability or correlations with other measures. A particularly notable result of the continuous score calculations is a moderate to strong correlation (ranging from .45 to .70) of MBTI scores with similar scales of the NEO-PI, a prominent Five Factor Model personality measure (Johnson, 1995; McCrae & Costa, 1989). Given the very different lexical-based origins of the NEO-PI, in contrast to the MBTI roots in Jung’s clinical observations, these results are interpretable as evidence of convergent validity for both instruments.

In recognition of the potential bias in self-report, instructions for completing the MBTI instrument have long stated “there are no ‘right’ or ‘wrong’ answers” and advised users to “not spend too much time thinking about any one question” (e.g., Consulting Psychologists Press, 1998). The recommended response set is a “frame of reference . . . termed the ‘shoes-off’ self” (Myers & McCaulley, 1985), a candid, natural mindset intended to avoid results distorted by either internal (e.g., reporting an ideal rather than an actual response) or external pressures (such as one’s work environment).

There is, however, no method of gauging whether such instructions truly elicit more accurate results, given the unavailability of criteria for evaluating whether a given response is “true.” Some self-report measures have incorporated scales (e.g., the MMPI Lie scale) intended to identify respondents who appear to be misrepresenting themselves if not outright lying. But even if such scales can detect deception, a sincere self-report may still be “wrong” if the respondent lacks access to or awareness of his or her true nature. Such intentional self-misrepresentation, typically presumed to cater to social desirability, is not easily distinguished from outright self-deception (e.g., McPeek, 1976; Paulhus, 1991). Yet there is clear evidence that both forms of misrepresentation bias self-report results. These biases arguably carry more import when the constructs being measured are core personality components, such as the MBTI assessment, as opposed to, say, self-report attitude surveys of less significant or personally relevant issues.

The MBTI assessment employs a second accuracy-improving strategy: use of “word pair” test items that make up 47 of the 93 MBTI questions (with the remaining 46 classified as “phrase” questions). Word pairs require a forced choice of “which word in each pair appeals to you more,” including options such as compassion vs. foresight or sensitive vs. just. While the creators of the MBTI instrument have speculated that the word pair items are “less susceptible [than phrase items] to . . . personal reticence, and conscious and unconscious censorship” (Myers et al., 1998, p. 129), these speculations have received little research attention. A notable exception is a master’s thesis by Nechworth (1977; see also Nechworth & Carskadon, 1979), in which college students were asked to complete the MBTI assessment twice, once with standard instructions and once as their “ideal” selves. “Ideal” phrase scores changed significantly in the direction of the MBTI poles of Extraversion, Intuition, Thinking, and Judging, whereas word pair scores did not significantly differ. Nechworth interpreted these results to “indicate that the respondent is consciously or unconsciously trying to modify his [results] . . . either in response to influences of social desirability or ideal self image” (Nechworth, 1977, p. v). The inference that preferences for E, T, and J are perceived as more ideal than their I, F, and P counterparts agrees with the MBTI® Manual’s assertion that “social desirability may affect a person’s responses in the direction of E, S, T, or J” (Myers et al., 1998, pp. 12–13). Consistent with the manual, a survey taken in England found that workers reported greater pressure to exhibit E, S, T, and J behaviors within their organizations (OPP, 2009).

These sources agree that preferences for E, T, and J are more highly valued in both academia and the work-
place. The desirability of the S–N scale, however, is inconsistent, with Nechworth’s data pointing towards N and the manual and English survey indicating S as more desirable. The discrepancy is perhaps best explained by cultural context. At higher levels of education (university and graduate) and for higher level (upper management and executive) employees in organizations, the N preference is not only more common (Macdaid, McCaulley, & Kainz, 2005; OPP, 2009; Quenk, Hammer, & Majors, 2004; Reymierse, 1993) but is linked to better performance (Myers & McCaulley, 1985). However, in the population of adults at large, in certain occupations, and at certain job positions or hierarchical levels in the workforce, the S preference is more likely to predominate and find favor. Thus, S or N may be more desirable depending on context.

Also worth nuanced consideration is the T–F type dimension, which may be subject to gender bias, as women are “socialized towards F” and men are “socialized towards T” (Myers et al., 1998), and the two genders do differ accordingly in their T–F preference distributions.

A third strategy designed to improve the accuracy of self-report is unique to the MBTI instrument. The administration of the assessment employed is typically followed by a feedback session, during which the meaning of results is explained. MBTI takers are encouraged to “verify” their results, a process of evaluating the degree of “fit” and confirming or overriding assessment results to make the final determination of personal four-letter type (called “verified,” “best fit,” or even “true” type). The MBTI® Manual recommends such verification as “a first check on the accuracy of the reported type” and a method to “enable identification of a different, better-fitting type” (Myers et al., 1998, p. 117). The median percentage of individuals who verify the same type indicated by the instrument across 13 different studies reported by Myers et al. (1998) is 70.5%. Thus, a sizable percentage of individuals disagree with their MBTI assessed results.

The verification practice should be examined critically. Methods for conducting verification are not only non-standardized (Kummerow, 1986; Myers et al., 1998) but also unmonitored and largely untested. Some research suggests that the process may be compromised by spurious influences such as the type of the person conducting the verification (Bathurst, 2000). These results raise the possibility that the consciously deliberative, non-standardized, non-monitored, process of verifying a type preference of assessing type preferences with a single direct question (“are you an Extravert or an Introvert?”) may be even more subject to social desirability or other forms of bias. Nonetheless, type practitioners place great faith in the resultant self-determined type, which is sometimes referred to as “true” type in the MBTI literature (e.g., Hammer & Yeakley, 1987; Walck, 1992).

To type theorists and practitioners, correct determination of type is a paramount concern, not just an academic subtlety. When environmental pressures (from family, culture, or the workplace) distort results of assessment, inaccuracies may “[rob its] victims of their real selves and [make] them into inferior, frustrated copies of other people” (Myers & Myers, 1980, p. 181). At the extreme, the result may be a “falsification” of type, which “often proves exceedingly harmful to the physiological well-being” (Jung, 1921/1971, p. 333). At a minimum, the more accurate the results, the better any personality measure, including the MBTI instrument, will serve a fundamental purpose of such assessments: improving self-awareness and self-management.

In short, like all self-report measures, the MBTI instrument is prey to self-report biases that likely distort accurate measurement of its constructs. Yet the tools and strategies devised to improve accuracy are largely untested.

**IMPLICIT MEASURES OF COGNITION**

Over the last 15 years developments in the field of implicit cognition have proposed alternative assessment tools that “are not direct, deliberate, controlled, and intentional” (Nosek, Hawkins, & Frazier, 2011, p. 153). One measure in particular, the Implicit Association Test, or IAT, developed by Greenwald and his associates (Greenwald, McGhee, & Schwartz, 1998; Lane, Banaji, Nosek, & Greenwald, 2007) is not only the most commonly used, but also has been shown to be consistently reliable and valid. In fact, in socially sensitive arenas such as prejudice, the IAT outperforms direct self-report as a valid predictor of behavior (Greenwald, Poehlmann, Uhlmann, & Banaji, 2009). Studies typically find a moderate correlation of explicit (self-report) and implicit measures, consistent with the idea that the two are measuring (to a limited extent) the same or similar underlying construct, but also tapping distinct processes and/or motivations (Greenwald et al., 2009; Hofmann, Gawronski, Gschwender, Le, & Schmitt, 2005; Nosek et al., 2011).
The fact that the IAT “may resist masking by self-presentation strategies” (Greenwald et al., 1998, p. 1465) has made it an attractive tool used in hundreds of published studies assessing a wide variety of attitudes. More recently, its use has spread to include implicit measurement of personality variables, notably the “Big Five” of Extraversion, Openness, Conscientiousness, Agreeableness, and Neuroticism (e.g., Back, Schmukle, & Egloff, 2009; Schnable, Banse, & Asendorpf, 2006; Steffens & Schultze-König, 2006). This research typically finds a low-to-moderate correlation of explicit and implicit measures, but has identified some conditions (such as spontaneous behavior, presumably less modified by intentionality) in which implicit measures are better than explicit ones in predicting behavior.

HOW THE IAT WORKS

The IAT is built upon the premise that objects, people, attitudes, and other constructs are uniquely associated in the cognitive structure and processes of a given individual, rooted in “traces of past experience” (Greenwald & Banaji, 1995, p. 5). The degree of association is operationalized by reaction time (latency) required to sort different words or visual images into different categories. Stimuli to be sorted appear in randomized order consecutively on a computer monitor and require a particular key response linked to one or two broader categories that the stimuli belong to. This is a relatively simple task when there is only one category to sort, such as negative (press the “a” key for words like nasty, sad, stupid) or positive (press the “k” key for words like pleasant, happy, smart) adjectives. The focus of the IAT is on latencies when a second category sorting task is mixed with the first—for example, self (words like me, mine, I, self) or other (words like them, their, other). The degree to which latencies are impacted by the way responses to stimuli of different categories are paired indicates the degree to which the categories are implicitly associated. Thus, a person with positive associations of him/herself will sort more quickly when self words and positive words require the same response (“a” key), thus assigning other and negative words to the “k” key, than when self/negative and other/positive stimuli require the same response.

This self/other, positive/negative scenario describes one common area of IAT research, intended as an implicit measure of positive associations with one’s self, or implicit self-esteem. When the category self/other is replaced by race (e.g., black/white) or gender, the IAT is presumed to measure implicit racial or gender attitudes respectively. (Note that neither implicit nor explicit results for such measures tell the whole story. Implicit results do not in any sense trump explicit.) The category of positive/negative adjective may also be replaced by a descriptive personality category—for example, extraversion/introversion. In such a design, the IAT would be presumed to measure implicit extraversion or introversion (i.e., the degree to which a person associates herself with extraversion words like talkative or sociable or introversion words like quiet or private). Similar substitutions would create implicit measures of the other MBTI preference scales (Sensing–Intuition, Thinking–Feeling, and Judging–Perceiving).

The research reported here (in two studies) looks at three different measures of psychological type. Measure #1 is the first attempt at implicit measurement of type preferences. Measure #2 is the Myers-Briggs Type Indicator assessment (the 93-item Form M). Measure #3 is best-fit type, the four preferences chosen by individuals who have completed the MBTI instrument and received feedback about type constructs and their MBTI scores. These three different measurements may be conceptualized as increasingly explicit. Measure #1, the IAT, is of course most implicit. The MBTI is an explicit self-report measure, using items relevant to type constructs and behavioral expectations, but never overtly linking type constructs to the items. The third measure, best-fit type, is a very explicit, direct question posed during the verification process (e.g., “based on what you have learned about yourself and these constructs, do you believe your preference to be Thinking or Feeling?”).

Based on the findings of similar research using the Five Factor Model of personality and other relevant past studies, the research will examine and evaluate the following hypotheses:

1) IAT and MBTI measures (continuous scores) of type preferences will correlate significantly but imperfectly; i.e., the two measures both share variance and contribute distinct information.

2) Given differences in IAT and MBTI measurements of type preferences, the MBTI assessment will indicate a more socially desirable result, as follows:
   a. More Extraversion (E) than Introversion (I)
   b. More Intuition (N) than Sensing (S)
   c. More Thinking (T) than Feeling (F) for men
   d. More Feeling (F) than Thinking (T) for women
   e. More Judging (J) than Perceiving (P)
3) When MBTI and best-fit measures of type preferences differ, the best-fit measure will indicate a more socially desirable result, exhibiting the same pattern as hypotheses 2a–2e above. These will be designated hypotheses 3a–3e.

STUDY ONE: IAT AND MBTI 
MEASURES OF TYPE

METHOD
Subjects were recruited from current and past attendees to MBTI Certification courses offered by the Center for Applications of Psychological Type (CAPT). The courses were attended primarily by a mixture of HR and organizational development consultants or staff from small, medium, and large organizations, typically from the United States. Attendees to the July 2012 and September 2012 courses were invited during the first class session to participate in a research project and given instructions on accessing and completing the type IAT, using their own laptop computers. In a few cases, CAPT provided computers to facilitate completion. Participation was voluntary, and anyone completing the IAT was provided lunch in a group session during which the IAT research was explained and personal results presented. Additional subjects were recruited via an e-mail invitation sent to approximately 3600 attendees to MBTI Certification courses from the previous three years. Participation was again voluntary, with a small discount for purchases from CAPT as well as a summary of research findings upon completion offered as incentives.

Subjects gave permission to the researchers to access their records to learn their MBTI results and scores. Taking the MBTI instrument was part of the course participation and was completed online, using the publisher’s (CPP) MBTI®Complete web platform. The IAT consisted of four parts, one for each of the four type measurements. The measures were constructed and administered using Inquisit software, designed for collecting IAT (and other forms of) data. The order of presentation of each preference was counterbalanced to control for order effects. Each of the four measures used a 7-block structure. Block 1 consisted of 20 practice trials sorting words relevant to the type dimension. Block 2 included 20 practice trials sorting self vs. other words. Blocks 3 (20 trials) and 4 (40 trials) mixed the stimuli, requiring sorting for two categories. Block 5 (20 trials) was another practice block (single category), essentially a repeat of block 1 with the computer key responses reversed. Blocks 6 (20 trials) and 7 (40 trials) repeated blocks 3 and 4, respectively, except that the pairing of self was reversed for blocks 6 and 7. Thus, if self and extraversion words were paired (i.e., required the same key to be pushed) in blocks 3 and 4, self and introversion words were paired in blocks 6 and 7.

Scoring of the IAT. IAT scoring used the D-score protocol (Greenwald, Nosek, & Banaji, 2003), which subtracts the mean of latencies for trials in blocks 3 and 4 from the mean for trials in blocks 6 and 7, divided by the standard deviation for all trials in those four blocks. This process thus used only trials in which two categories were sorted, when association-based delay or facilitation would be operative. The resulting scores may range from -2 to +2. Results were scaled so that a positive value indicated an IAT preference for I, N, F, and P, whereas a negative value pointed towards E, S, T, and J. This valencing is consistent with the conventions of type measurement.

Words and labels. The self category words used in the IAT consisted of I, me, mine, my, myself, and self. The other category words included other, their, them, and they.

Due to potential misinterpretations of some of the type preference terminology, personality category labels believed to be more familiar and less susceptible to confusion due to preexisting connotations were used for some of the preferences. Extraversion and Introversion, words familiar to most people (and certainly to all attending CAPT courses), were not changed. The alternative category labels Practical–Imaginative, Logical–Warm, and Organized–Spontaneous were substituted for Sensing–Intuition, Thinking–Feeling, and Judging–Perceiving, respectively. The original labels all have meanings, connotations, and therefore associations that differ from their more precise definitions in type theory.

The personality category words presented to be sorted included the following:

- Active, Enthusiastic, Expressive, Extroverted, Gabby, Gregarious, Loud, Showoff, Talkative (Extraversion)
- Cautious, Discreet, Intimate, Introspective, Modest, Private, Quiet (Introversion)
- Detailed, Factual, Literal, Practical, Proven, Realistic, Traditional (Practical)
- Abstract, Creative, Imaginative, Innovative, Intuitive, Theoretical, Unconventional, Visionary (Imaginative)
- Analytical, Critical, Frank, Logical, Objective, Skeptical, Straightforward, Tough-minded (Logical)
- Accepting, Agreeable, Compassionate, Emotional,
Empathetic, Generous, Harmonious, Sentimental, Sympathetic, Warm (Warm)

Conscientious, Disciplined, Methodical, Meticulous, Orderly, Planful, Prudent, Systematic (Organized)

Casual, Changeable, Easy going, Flexible, Impulsive, Informal, Relaxed, Spontaneous (Spontaneous)

These selected words have not only been used frequently in type descriptions (e.g., Myers et al., 1998), but have also been verified as appropriately and selectively descriptive by type experts (McPeek & Martin, 2012), observer ratings (Harker, Reynierse, & Komisn, 1998; Thorne & Gough, 1991), and/or self-descriptions (Thorne & Gough, 1991). Given the exploratory nature of the research, more words than have commonly been used in IAT personality research (e.g., Steffens & Schulze-König, 2006) were included. The mean likability/desirability valence of each set of opposing adjectives was made as equivalent as possible, based on past research quantifying the desirability of such descriptors (Anderson, 1968; Dumas, Johnson, & Lynch, 2002; Norman, 1967).

Prior to each type IAT series of trials, these words were presented in lists for each category.

RESULTS

Subject participation. As expected, the response rate to CAPT's invitation to participate in the IAT research was much higher for in-person MBTI Certification course attendees (24 of 43 attendees, or 56%, participated) than for past course attendees invited by e-mail (140 of 3,663 invitees, or 3.8%, participated). Data from five CAPT employees or interns, who completed the IAT without awareness of its purpose, were also included. Nineteen of these 169 subjects failed to complete all four type IATs; any data completed were used. Data from two subjects who responded with an excessive number (> 10%) of very fast response latencies were excluded following recommendations drawn from Greenwald et al. (2003). Also, following these authors' recommendations, any trials with latencies > 10 seconds were deleted before the calculation of D-scores (fewer than 1 in 1,000 trials were excluded).

Aside from the MBTI measure, the only other data collected from the volunteer subjects was gender. The sample, like the CAPT Certification classes in general, was overwhelmingly female (129 of 169 subjects, or 76%).

IAT reliability. The internal reliability of each of the four (E–I, S–N, T–F, and J–P) IAT scores was estimated using two methods. The first of these calculated two separate D-scores using reaction time latencies for every other scored trial, producing split half (odd-even) correlations. These results provided evidence of good internal consistency: for E–I, \( r(151) = .84, p < .001 \); for S–N, \( r(152) = .71, p < .001 \); for T–F, \( r(153) = .78, p < .001 \); and for J–P, \( r(153) = .75, p < .001 \).

The second method computed Cronbach's alpha for a composite score for each adjective used in a given preference IAT. The average raw latencies (in milliseconds) for all trials for each individual adjective were computed separately for both preferences (e.g., extraversion and introversion) paired with self words, and then one mean was subtracted from the other. This created a latency difference for each adjective, analogous to a single test item contributing to a composite scale score. The scores for each item were then used to calculate alpha values for each of the four preference scales. For the 26 different individual E–I words (including self-other words), \( \alpha = .89 \); for the 26 S–N words, \( \alpha = .68 \); for the 28 T–F words, \( \alpha = .85 \); and for the 26 J–P words, \( \alpha = .82 \).

IAT-MBTI correlations. The correlations of continuous scores for each preference IAT with the MBTI preference clarity scores are shown in bold in Table 1.

The results shown in Table 1 clearly confirmed hypothesis 1, that corresponding IAT and MBTI scores will correlate significantly. The four correlations ranged from .47 for J–P to .61 for E–I, all highly significant. There were also modest correlations of IAT scores with one another (IAT S–N with IAT T–F, S–N with J–P, and T–F with J–P) as well as MBTI internal correlations (S–N with T–F and J–P; T–F with J–P). These are all below .30, with the exception of the MBTI S–N: J–P correlation (\( r = .32, p < .001 \)). The moderate S–N: J–P correlation is typical of other MBTI results (see Myers et al., 1998, p. 153), and this value is in fact on the lower side of the norm.

Changes in category results. As both the MBTI and IAT measures produced categorical results (either E or I, S or N, etc.), the differences in classifications for each preference IAT from the corresponding MBTI category were compared in order to test the second set of hypotheses.

Hypotheses 2a, 2b, and 2e, predictions of a greater explicit than implicit proportion of scores indicating E, N, and J, were all confirmed, as shown in Table 2.

The proportion of subjects who shifted from E on the IAT to I on the MBTI instrument was significantly lower than those showing the opposite shift, from I to E, \( \chi^2(1, N = 143) = 9.98, p = .002 \). The S-to-N change...
was significantly more common than N-to-S, $\chi^2(1, N = 145) = 7.98$, $p = .005$, and P-to-J cases outnumbered J-to-P, $\chi^2(1, N = 146) = 9.68$, $p = .002$.

Hypotheses 2c and 2d are gender specific. The former was strongly confirmed—a much higher proportion (73%) of males with F results on the IAT were classified as Ts on the MBTI self-report than IAT T males (only 13%) who changed to F on the MBTI self-report, Fisher’s exact $p = .003$. Hypothesis 2d, however, was not confirmed: females were in fact more likely to change from IAT F to MBTI T than vice versa, $\chi^2(1, N = 116) = 3.99$, $p = .05$. Males (11 of 15 possible changes, or 73%), however, showed this F-to-T pattern more significantly than females (24 of 86 possible changes, or 28%), $\chi^2(1, N = 101) = 11.64$, $p = .001$. The corresponding analysis for T-to-F changes revealed no significant differences by gender, with males changing 13% of the time and females 10%. In sum, both genders showed very high MBTI agreement with T results on the IAT, but males were much more likely to change to T than females when answering a self-report questionnaire.

Changes in continuous scores. Both the IAT and MBTI measures of type preferences produce continuous scores with zero as the dividing point between preferences. In order to perform a within subjects comparison of implicit and explicit scores, the respective scales were equated by dividing each result by the standard deviation of all scores on that measure. This is akin to a $z$-score transformation, omitting resetting the zero point to the mean of all scores, in order to retain the conceptual significance of the raw score zero point. Setting the mean to zero with a full $z$-score transformation would eliminate the very difference in deviation from the “true” zero under investigation.

Paired comparison $t$-tests for implicit/explicit (IAT/MBTI) scores for EI, SN, and JP were performed to

| Table 1. IAT and MBTI® Correlations for Each Preference Domain. |

<table>
<thead>
<tr>
<th></th>
<th>EI_IAT</th>
<th>SN_IAT</th>
<th>TF_IAT</th>
<th>JP_IAT</th>
<th>EI_MBTI</th>
<th>SN_MBTI</th>
<th>TF_MBTI</th>
<th>JP_MBTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>EI_IAT</td>
<td>.152</td>
<td>.182*</td>
<td>-.034</td>
<td>.609***</td>
<td>.176*</td>
<td>.152</td>
<td>-.110</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n=154$</td>
<td>$n=153$</td>
<td>$n=154$</td>
<td>$n=141$</td>
<td>$n=141$</td>
<td>$n=141$</td>
<td>$n=141$</td>
<td></td>
</tr>
<tr>
<td>SN_IAT</td>
<td>.287***</td>
<td>.279**</td>
<td>.052</td>
<td>.493***</td>
<td>.075</td>
<td>.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n=155$</td>
<td>$n=154$</td>
<td>$n=142$</td>
<td>$n=142$</td>
<td>$n=142$</td>
<td>$n=142$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF_IAT</td>
<td>.238**</td>
<td>.031</td>
<td>.170*</td>
<td>.504***</td>
<td>.110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n=155$</td>
<td>$n=143$</td>
<td>$n=143$</td>
<td>$n=143$</td>
<td>$n=143$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JP_IAT</td>
<td>-.198*</td>
<td>.183*</td>
<td>.065</td>
<td>.473***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n=143$</td>
<td>$n=143$</td>
<td>$n=159$</td>
<td>$n=143$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EI_MBTI</td>
<td>-.004</td>
<td></td>
<td>-.065</td>
<td>-.151</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n=159$</td>
<td></td>
<td>$n=159$</td>
<td>$n=159$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN_MBTI</td>
<td></td>
<td>.231**</td>
<td></td>
<td>.322***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n=159$</td>
<td></td>
<td>$n=159$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF_MBTI</td>
<td></td>
<td></td>
<td>.214**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$n=159$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *$p < 0.05$ level (2-tailed); **$p < 0.01$ level (2-tailed); ***$p < 0.001$ level (2-tailed).

| Table 2. Category Shifts from IAT to MBTI® Results for Each Preference. |

<table>
<thead>
<tr>
<th>IAT Result</th>
<th>MBTI® Result Same</th>
<th>Opposite</th>
</tr>
</thead>
<tbody>
<tr>
<td>E I</td>
<td>51 (89.5%)</td>
<td>6 (10.5%)</td>
</tr>
<tr>
<td>S N</td>
<td>39 (60.0%)</td>
<td>26 (40.0%)</td>
</tr>
<tr>
<td>T female F female</td>
<td>27 (90.0%)</td>
<td>3 (10.0%)</td>
</tr>
<tr>
<td>T male F female</td>
<td>13 (86.7%)</td>
<td>2 (13.3%)</td>
</tr>
<tr>
<td>J P</td>
<td>55 (82.1%)</td>
<td>12 (17.9%)</td>
</tr>
</tbody>
</table>

Journal of Psychological Type®, Volume 74, 2014
test hypotheses 2a, 2b, and 2e. A within-subjects 2 (IAT or MBTI type measure) \times 2 (male or female) ANOVA was performed to test the gender hypotheses (2c and 2d) for TF scores.

As Table 3 shows, the three simple t-tests all confirmed their respective hypotheses. Relative to implicit scores, MBTI scores shifted towards E, N, and J.

Figure 1 shows the results of a 2 \times 2 ANOVA testing hypotheses 2c and 2d. Evidence in favor of these two hypotheses would take the form of a significant interaction, with explicit scores lower (more T) for males and higher (more F) for females. This interaction was non-significant (F < 1). Instead, there were two significant main effects: male scores were significantly lower than females on both the IAT and MBTI measures, F(1, 141) = 12.56, p = .001; and explicit scores were more T than implicit scores, F(1, 141) = 24.29, p < .001. Both males and females scored as more T on the explicit MBTI measure. The estimated effect size (partial eta squared = .15) for the implicit-explicit effect was larger than the effect size for gender (partial eta squared = .08), noteworthy given the strong, consistent, and well-documented gender difference on MBTI T–F scores.

**STUDY TWO: MBTI® AND BEST-FIT MEASURES OF TYPE**

**INTRODUCTION**

One of the unique aspects of MBTI usage involves a process known as type verification, which involves review of MBTI results and information by the user, who is given the opportunity to confirm or alter any or all of the preferences indicated to arrive at a “best-fit type.” Quenk (2000) summarizes the practice and presumed benefits:

Interpretation of the Myers-Briggs Type Indicator personality inventory differs from the interpretation of other assessment instruments because the initial step is to have the client verify the accuracy of the results obtained . . . that self-report results are subject to error is also the underlying motivation for verifying type results (pp. 53–54).

Some of the “factors influencing self-report on the MBTI” Quenk lists are “pressure to conform to expectations . . . gender biases, especially affecting male Feeling types and female Thinking types . . . [and] perception that a particular kind of person is desired, such as in employment or team situations” (p. 54).

These are the very same factors that would produce the kinds of differences observed when comparing implicit results to the MBTI (self-report) results in the first study. Thus, the intent of type verification and implicit measurement of type is the same: greater accuracy. With regard to best-fit type, however, the methodology still relies on the same fallible self-evaluation and self-

---

**Table 3. Comparison of IAT and MBTI® Continuous Scores (Within Subject) in SD Units.**

<table>
<thead>
<tr>
<th>Preference Pairs</th>
<th>N</th>
<th>IAT mean</th>
<th>MBTI® pci mean</th>
<th>t test</th>
</tr>
</thead>
<tbody>
<tr>
<td>EI</td>
<td>141</td>
<td>.36</td>
<td>-.13</td>
<td>6.36***</td>
</tr>
<tr>
<td>SN</td>
<td>142</td>
<td>.07</td>
<td>.41</td>
<td>4.08***</td>
</tr>
<tr>
<td>JP</td>
<td>143</td>
<td>.20</td>
<td>-.27</td>
<td>5.41***</td>
</tr>
</tbody>
</table>

Note: ***p < 0.001

**FIGURE 1. Transformed IAT and MBTI® T–F Scores as a Function of Gender.**

![Transformed Scores (higher scores = F)](image)

Transformed Scores (higher scores = F)
report that produced the MBTI results. In fact, the “expectations” and “biases” Quenk mentions may be more salient when answering a direct question to determine best fit (which amounts to asking whether you an extravert or an introvert, etc.) as compared to the more varied, less obvious MBTI items. From a psychometric perspective, substituting the results of a single question for the 20+ carefully honed, tested, and vetted MBTI items for each preference domain can only negatively impact internal consistency (in fact, without at least a second question, “consistency” has no meaning).

Studies comparing indicated to best-fit results, as summarized on page 197 of Myers et al. (1998), have found complete (all four letters) agreement ranging from 53% to 85% of cases. At least some of this variability is likely attributable to different methods and procedures used in providing feedback and assessing best fit. Some studies (e.g., Walck, 1992) used extensive discussion and essay writing exercises to help understand type and as an aid to self-assessment. Some (e.g., Kummerow, 1988) employed a comprehensive 4-step process involving visual aids, anecdotes, descriptive adjectives, considerations of work activities, long and short descriptions of each type, and the opportunity to ask questions of the feedback provider.

One particularly important variation on best fit methodology involves the timing of best fit assessment in relationship to learning the results of the MBTI scoring. Quenk (2000) clearly recommends that MBTI results precede best fit determination. Despite their use of agreement with best-fit type as a criterion for evaluating scoring methods in the MBTI revision, Myers et al. (1998) indicated that because best fit was sometimes measured before and sometimes after individual MBTI results were presented, unknown biases may have affected the results. More recent work has found evidence of a “prompting” effect—that presentation of MBTI results influences best fit decisions in the direction of greater agreement with reported preferences in an estimated 10–25% of cases (M. Morris & R. Thompson, personal communication, March 20, 2013).

An additional influence upon best fit was identified by Bathurst (2000). He surveyed the changes from reported to best-fit type for 61 different MBTI administrators who conducted feedback sessions “mostly as recommended in the APT Qualifying Training Program” (Bathurst, 2000, p. 10). His data (collected using MBTI Forms F and G from participants in MBTI courses given in Australia and New Zealand) suggested that reported to best fit preference switches were disproportionately influenced in the direction of administrators with preferences for S, N, T or F. A plausible explanation is that administrators may, even unintentionally, overemphasize the strengths and attractions of their own preferences (e.g., a positive Thinking preference description might emphasize its accuracy and high standards rather than its critical and impersonal qualities).

Taking all this research into account presents an unclear picture: type verification is an unmonitored practice that evidence suggests is potentially biased, yet it is presented as a standard practice in determining type. With this in mind, Study 2 looked for patterns in the changes of reported (MBTI) results during determination of best-fit type. The investigation began with data from the MBTI® Complete online assessment, which incorporates a verification process that is both standardized (using an online interactive branching series of questions) and less likely to be influenced by variations in administrator type (as the administration is automated). Additional advantages of the MBTI® Complete data include the size (over 34,000 cases) and the ethnic diversity of the database, a better representation of male users (43% of the cases of known gender), and the fact that the data are both recent (within the past four years) and based on the current MBTI Form M.

Data from a second large database provided by John Bathurst that expanded upon his previously published study (Bathurst, 2000) were also analyzed. As his research involved face-to-face verification, differences between Bathurst’s data and the MBTI® Complete data will allow examination of the possible influence of administrator preference on best-fit results.

**METHOD**

Chi-square tests (as the relevant data existed only in categorical form) were used to compare the number of individuals whose preferences changed or did not change from reported to best fit measures, creating a 2 x 2 grid. For the T–F preferences, males and female data were analyzed separately.

An important note is that this analysis differs from the method employed by Bathurst (2000), which did not take into consideration some important data. He simply compared the number of changes from one preference to another to the number of changes in the opposite direction, ignoring the essential information of how many cases could have changed preferences. For example, knowing that 10 people changed from T to F but only
five changed from F to T is not sufficient information to determine direction-of-switch trends. On the face the raw change numbers suggest a much stronger T-to-F switch. If, however, the 10 T-to-F changers are 10% of 100 people who tested as T on the MBTI assessment, and the five F-to-T changers are 100% of those whose MBTI results were F; a better conclusion is a strong trend for an MBTI F to verify T. New data and these new analyses allowed a test of not only the study's hypotheses, but whether the preferences of administrators resulted in different rates of switching from reported to best-fit results. As shall become evident, this reanalysis of Bathurst's data led to conclusions that differ from his originally published report, though in both sets of results administrator preference did in fact appear to influence best-fit type.

**RESULTS**

*Table 4* shows the results of chi-square tests on the CPP MBTI® Complete data.

There were clear differences in the direction of best fit changes. Four (hypotheses 3b to 3e) of the five predictions were confirmed (switches from S to N, P to J, T to F for females, and F to T for males). The sole exception was an E-to-I shift rather than the opposite direction predicted by hypothesis 3a.

*Table 5* shows the same analyses for the older Australian/New Zealand data provided by Bathurst.

The E-to-I, S-to-N, and female T-to-F switches were significant, replicating the findings from the MBTI® Complete data. Males did not show significantly more frequent F-to-T switches, though the very strong T-to-F switch tendencies of females did not hold for males. There was also no significant P-to-J switch as was seen in the MBTI® Complete data. In summary, hypotheses 3b and 3d were confirmed, and hypothesis 3a showed a pattern opposite to prediction. Hypotheses 3c and 3e were not confirmed (no significant differences).

**Gender differences on T–F changes.** *Table 6* demonstrates a consistent gender difference in the direction of reported-to-best fit changes for the Thinking and Feeling preferences. These changes were consistent with the gender hypotheses (3c and 3d): a higher percentage of females switched from reported T to verified F, and the trend was reversed for males.

**Effect of administrator preferences on verification.** *Table 7* shows that the number of cases whose best fit preference was the same as or different from their reported (MBTI) preference as a function of the associated preference of the MBTI administrator providing feedback and conducting the verification process. These data, from the Bathurst database, contained results from at least 61 different administrators. There were no available details documenting how the verification process was conducted.

The top line of each paired comparison corresponds to conditions in which the administrator preference was the same as the reported preference of the user. Comparing the second line change rate to the line above is thus suggestive of the power of opposite-to-reported administrator preferences to “attract” changes in best fit preferences. There were three clear effects and

<table>
<thead>
<tr>
<th>Reported MBTI® preference</th>
<th>same</th>
<th>Best-fit preference change</th>
<th>change rate</th>
<th>Chi-square (df = 1)</th>
<th>Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported E both genders</td>
<td>13,126</td>
<td>1,497</td>
<td>10.2%</td>
<td>230.21</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported I both genders</td>
<td>12,380</td>
<td>694</td>
<td>5.3%</td>
<td>230.21</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported S both genders</td>
<td>13,117</td>
<td>2,516</td>
<td>16.1%</td>
<td>252.48</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported N both genders</td>
<td>10,910</td>
<td>1,154</td>
<td>9.6%</td>
<td>252.48</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported T females</td>
<td>5,212</td>
<td>741</td>
<td>12.4%</td>
<td>84.36</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported F females</td>
<td>6,346</td>
<td>522</td>
<td>7.6%</td>
<td>84.36</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported T males</td>
<td>6,600</td>
<td>634</td>
<td>8.8%</td>
<td>12.37</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported F males</td>
<td>2,400</td>
<td>299</td>
<td>11.1%</td>
<td>12.37</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported J both genders</td>
<td>14,506</td>
<td>1,070</td>
<td>6.9%</td>
<td>149.79</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>Reported P both genders</td>
<td>10,782</td>
<td>1,339</td>
<td>11.0%</td>
<td>149.79</td>
<td>p &lt; .0001</td>
</tr>
</tbody>
</table>
one weak effect for administrator preference. I administrators were associated with E-to-I changes ($p < .0001$), J administrators with P-to-J changes ($p = .016$), and T administrators with F-to-T changes for females only ($p < .001$). There was a weaker effect for S administrators to produce more changes from N-to-S than N administrators did ($p = .05$).

**DISCUSSION**

This research looked at three different assessments of type preferences, representing three levels of increasing explicitness, from the IAT to the MBTI to best-fit results. Predictions were that greater explicit measurement would produce results that moved towards greater social desirability at each step, due to the influence of self-presentation bias in self-reports. Past research has suggested that E, N, and J are more socially desirable than their I, S, and P opposites, with a gender factor at work with the T–F preference. T has been described as more consistent with male gender roles men and F more consistent with females’.

These predictions were mostly confirmed, though some results varied across the three sets of data. The most consistent finding was a movement from a Sensing to an Intuitive preference with increasing explicitness, both from the IAT to the MBTI results and from the MBTI to best-fit results. Also, although females showed more F-to-T than T-to-F shifts in Study 1 (contrary to hypothesis) and more T-to-F shifts in Study 2 (as predicted), their F-to-T shifts were consistently less likely than
**Table 7. Effect of administrator preference on best fit changes (Bathurst data).**

<table>
<thead>
<tr>
<th>Administrator preference</th>
<th>MBTI® preference (user)</th>
<th>Chi-square ((df = 1))</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E–E</strong></td>
<td><strong>E–I</strong></td>
<td><strong>change rate</strong></td>
<td><strong>16.26</strong> ( p &lt; .0001)</td>
</tr>
<tr>
<td>E</td>
<td>2054</td>
<td>191</td>
<td>8.5%</td>
</tr>
<tr>
<td>I</td>
<td>1743</td>
<td>244</td>
<td>12.3%</td>
</tr>
<tr>
<td>I–I</td>
<td>I–E</td>
<td>change rate</td>
<td><strong>0.02</strong> ( p = .89)</td>
</tr>
<tr>
<td>I</td>
<td>2491</td>
<td>161</td>
<td>6.1%</td>
</tr>
<tr>
<td>E</td>
<td>2298</td>
<td>151</td>
<td>6.2%</td>
</tr>
<tr>
<td><strong>S–S</strong></td>
<td><strong>S–N</strong></td>
<td>change rate</td>
<td><strong>0.70</strong> ( p = .40)</td>
</tr>
<tr>
<td>S</td>
<td>526</td>
<td>58</td>
<td>9.9%</td>
</tr>
<tr>
<td>N</td>
<td>3674</td>
<td>458</td>
<td>11.1%</td>
</tr>
<tr>
<td>N–N</td>
<td>N–S</td>
<td>change rate</td>
<td><strong>3.90</strong> ( p = .05)</td>
</tr>
<tr>
<td>N</td>
<td>3644</td>
<td>298</td>
<td>7.6%</td>
</tr>
<tr>
<td>S</td>
<td>609</td>
<td>66</td>
<td>9.8%</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td><strong>T–T</strong></td>
<td><strong>T–F</strong></td>
<td>change rate</td>
</tr>
<tr>
<td>T</td>
<td>716</td>
<td>206</td>
<td>22.3%</td>
</tr>
<tr>
<td>F</td>
<td>979</td>
<td>258</td>
<td>20.9%</td>
</tr>
<tr>
<td><strong>F–F</strong></td>
<td><strong>F–T</strong></td>
<td>change rate</td>
<td><strong>11.34</strong> ( p &lt; .001)</td>
</tr>
<tr>
<td>F</td>
<td>2402</td>
<td>102</td>
<td>4.1%</td>
</tr>
<tr>
<td>T</td>
<td>1502</td>
<td>103</td>
<td>6.4%</td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td><strong>T–T</strong></td>
<td><strong>T–F</strong></td>
<td>change rate</td>
</tr>
<tr>
<td>T</td>
<td>753</td>
<td>96</td>
<td>11.3%</td>
</tr>
<tr>
<td>F</td>
<td>908</td>
<td>115</td>
<td>11.2%</td>
</tr>
<tr>
<td><strong>F–F</strong></td>
<td><strong>F–T</strong></td>
<td>change rate</td>
<td><strong>0.00</strong> ( p &gt; .95)</td>
</tr>
<tr>
<td>F</td>
<td>584</td>
<td>62</td>
<td>9.6%</td>
</tr>
<tr>
<td>T</td>
<td>494</td>
<td>53</td>
<td>9.7%</td>
</tr>
<tr>
<td><strong>J–J</strong></td>
<td><strong>J–P</strong></td>
<td>change rate</td>
<td><strong>0.78</strong> ( p = .38)</td>
</tr>
<tr>
<td>J</td>
<td>2568</td>
<td>284</td>
<td>10.0%</td>
</tr>
<tr>
<td>P</td>
<td>2306</td>
<td>235</td>
<td>9.2%</td>
</tr>
<tr>
<td><strong>P–P</strong></td>
<td><strong>P–J</strong></td>
<td>change rate</td>
<td><strong>5.80</strong> ( p = .016)</td>
</tr>
<tr>
<td>P</td>
<td>1943</td>
<td>162</td>
<td>7.7%</td>
</tr>
<tr>
<td>J</td>
<td>1654</td>
<td>181</td>
<td>9.9%</td>
</tr>
</tbody>
</table>

*Additional chi-square tests reveal that there are significantly more E-to-I than opposite changes, \(\chi^2(1, N = 4,694) = 9.51, p = .002\), even for E administrators analyzed separately, as well as more T-to-F than F-to-T switches, \(\chi^2(1, N = 2527) = 138.38, p < .0001\), for T administrators working with female subjects. The E-to-I and T-to-F effects are larger, however, for I and F administrators, respectively.
males’ (notably 28% compared to 73% of possible shifts in Study 1). These results are consistent with an overall social desirability bias towards T, modified or counteracted by a female gender bias towards F. This might describe the culture of many workplaces. For example, Walck (1996) has summarized evidence that T is favored in the workplace, as many sources (Kirby, 1992; Macdaid et al., 2005) have shown that “the tendency in organizations of all kinds is for leaders to be predominantly Thinking and Judging” (Kirby, 1992, p. 13).

The E–I pattern was the least consistent. There was a clear shift from I to E from the IAT to the MBTI measures, but the reverse was true for shifts from reported (MBTI) type to best fit. While reasons for this reversal are not clear, consider that the process of verifying type is an inwardly focused self-examination. That meets the definition of introverted activity as “directing energy mainly toward the inner world of experiences and ideas” (Myers et al., 1998, p. 6) and involving a process of reflection, depth, “and reactions that are a part of [the] inner word” (Martin, 1997). Given the well-known research by Bem (1967, 1972) and other evidence suggesting that people draw conclusions about themselves based on self-observation, the act of observing oneself engage in introverted behavior may provide immediate evidence that such activity is more self-descriptive than it would appear at a different time.

Evidence for more common P-to-J shifts emerged clearly in both Study 1 and the MBTI® Complete data from Study 2. In the Bathurst data this shift was more pronounced when the administrator preferred Judging rather than Perceiving, consistent with an “attractor” effect. Further analysis of Bathurst’s data offers additional evidence that changes from indicated to best fit preferences are in the direction of the preference of the person conducting the feedback session. In addition to J preferences, administrators whose preferences for I and S opposed their MBTI clients’ corresponding reported preferences (E and N) produced significantly greater changes to I and S than did administrators with the same preferences as their clients. Compared to F administrators, T administrators with female clients whose MBTI results indicated Feeling preferences also led to more client changes towards T. (Note that these results are based on a fairer test of differences than originally conducted by Bathurst, 2000, and differ from his findings on some specifics, though both analyses arrive at the conclusion that administrators can bias verification.)

The evidence for both an administrator preference effect and a social desirability effect suggests that self-report based type measurement can be influenced by environmental forces. Social desirability is presumably reflective of cultural influences, whereas the administrator effect may reflect the influence of a more temporary mini-cultural environment whose tone is set by the “expert.” The viability of this explanation is bolstered by the fact that psychological type practitioners in particular, and psychologists (whether clinical, counseling, or industrial/organizational), consultants, and counselors in general, overwhelmingly prefer Intuition to Sensing (Macdaid, McCaulley, & Kainz, 2005; Schaubhut & Thompson, 2008). So, though a shift from S to N was the most consistent in the present research, reanalysis of data from other research (Kummerow, 1988; see footnote 4) showed an N to S shift among bank and retail managers, occupations which predominantly prefer S (Macdaid, McCaulley, & Kainz, 2005; Schaubhut & Thompson, 2008) and potentially create an S-friendly culture.

Kummerow’s different results call attention to a major limitation of this study: the fact that subjects were drawn from non-representative populations. The sample in Study 1 (which is further compromised by self-selection bias in the volunteer participants) was comprised entirely of people interested enough to seek type training. In Study 2, the sample was more diverse, including “public, church, and corporate settings” (Bathurst, 2000, p. 5). There is no means of determining the balance of such sources, but it is fair to speculate that in Study 2, many or most of the subjects were interested in personality measurement in general if not type in particular.

Future studies using the IAT should thus employ different sample sources in order to test the generalizability of present findings. Additional research may also explore the utility of implicit type preference measurement in testing some of the tenets of type theory. For example, type theory suggests that the S–N and T–F preferences are organized into a hierarchy of “dominance,” with one of these four preferences “likely to be used most enthusiastically, most often, and with the greatest confidence” (Myers et al., 1998, p. 22). This “dominant function” is presumed to be “most conscious . . . that is, [it] has the greatest amount of conscious energy at its command” (Myers et al., 1998, p. 23). IAT researchers caution against equating implicit responses to unconscious ones (Nosek et al., 2012), but definitions of implicit cognition as “introspectively unidenti-
The practice of treating type preferences as discrete categories rather than as points along a continuum (likely normally distributed) is a target of long-standing criticism of the MBTI instrument, dating at least to Stricker and Ross (1962). Full test-retest congruence of all four categories of MBTI results, though well above chance level, is disappointingly low; the median full agreement reported by Myers and McCaulley (1985) is 41%, suggesting that whole type disagrees more often than it agrees upon retest. Thus, reporting categorical type results facilitates objections to the MBTI instrument’s test-retest reliability (e.g., Pittenger, 2005). Continuous MBTI score test-retest correlations, in contrast, are robust and meet or exceed those of other instruments, including the NEO-PI (Schaubhut, Herk, & Thompson, 2009). This paper takes a neutral stance and reports both categorical and, when available, continuous score results.

The MBTI®Complete content was primarily written by a single author, leaving open the possibility that type bias may still exist, stemming from the limited perspective of one person. The present study has not systematically examined this possibility.

The same concern applies to the report by Kummerow (1988). A reanalysis of her data results in conclusions that differ from hers, as follows: she reported a significant E-to-I trend which was non-significant in the reanalysis, \(\chi^2(1, N = 487) = 0.93, p = .35\); she reported no significant S–N changes, whereas the reanalysis indicated a significantly greater proportion of N-to-S than S-to-N shifts, \(\chi^2(1, N = 487) = 10.13, p = .001\); she reported a significant J-to-P switch which was non-significant in the reanalysis, \(\chi^2(1, N = 487) = 1.19, p = .28\). Her significant T-to-F switch (gender combined) was confirmed, although the \(\chi^2\) value dropped markedly, from 42.26 (\(p < .001\)) to 5.84 (\(p = .02\)).

Note that because best fit preferences were collected as categorical data only, analyses were limited to chi-squares (or, in some cases with smaller \(n\), Fisher’s exact test).

This was presumably the best fit preference as determined by each administrator for him/herself. No information is available regarding how many of these preferences differed from reported preferences.

FOOTNOTES

1 Address correspondence to Robert McPeek, Ph.D., Center for Applications of Psychological Type, 2815 NW 13th Street, Suite 401, Gainesville, FL 32609. E-mail: bob@capt.org. The authors express gratitude to Anthony Greenwald, Konrad Schnabel, and Charles Martin for helpful comments along the way and to Rich Thompson, Mike Morris, and John Bathurst for providing data and support.

2 The practice of treating type preferences as discrete categories rather than as points along a continuum (likely normally distributed) is a target of long-standing criticism of the MBTI instrument, dating at least to Stricker and Ross (1962). Full test-retest congruence of all four categories of MBTI results, though well above chance level, is disappointingly low; the median full agreement reported by Myers and McCaulley (1985) is 41%, suggesting that whole type disagrees more often than it agrees upon retest. Thus, reporting categorical type results facilitates objections to the MBTI instrument’s test-retest reliability (e.g., Pittenger, 2005). Continuous MBTI score test-retest correlations, in contrast, are robust and meet or exceed those of other instruments, including the NEO-PI (Schaubhut, Herk, & Thompson, 2009). This paper takes a neutral stance and reports both categorical and, when available, continuous score results.

The MBTI®Complete content was primarily written by a single author, leaving open the possibility that type bias may still exist, stemming from the limited perspective of one person. The present study has not systematically examined this possibility.

4 The same concern applies to the report by Kummerow (1988). A reanalysis of her data results in conclusions that differ from hers, as follows: she reported a significant E-to-I trend which was non-significant in the reanalysis, \(\chi^2(1, N = 487) = 0.93, p = .35\); she reported no significant S–N changes, whereas the reanalysis indicated a significantly greater proportion of N-to-S than S-to-N shifts, \(\chi^2(1, N = 487) = 10.13, p = .001\); she reported a significant J-to-P switch which was non-significant in the reanalysis, \(\chi^2(1, N = 487) = 1.19, p = .28\). Her significant T-to-F switch (gender combined) was confirmed, although the \(\chi^2\) value dropped markedly, from 42.26 (\(p < .001\)) to 5.84 (\(p = .02\)).

5 Note that because best fit preferences were collected as categorical data only, analyses were limited to chi-squares (or, in some cases with smaller \(n\), Fisher’s exact test).

6 This was presumably the best fit preference as determined by each administrator for him/herself. No information is available regarding how many of these preferences differed from reported preferences.
REFERENCES


Journal of Psychological Type, 25, 11–23.


