More than thirty years of publishing research articles related to the theory and applications of psychological type and the Myers-Briggs Type Indicator® instrument.

73

Journal of Psychological Type®

Issue 2 | 2013

"Comparison of results from the MBTI® and the MMTIC® instruments in adolescents (age appropriate for both instruments) revealed high correlations of type scores. The MMTIC assessment was perceived as easier to read and understand."

Convergent Validity of the MBTI® and MMTIC® Assessments

Robert W. McPeek[1] and **Judith Breiner**
Center for Applications of Psychological Type, Inc.

ABSTRACT

Teenagers (*N* = 123, mean age = 16.54) were administered both the 2008 revision of the Murphy-Meisgeier Type Indicator for Children® and Form M of the Myers-Briggs Type Indicator® assessments on two occasions. This design allowed both a comparison of results of the two instruments to each other and test-retest reliability computations for each. Continuous scores of the four scales of the two instruments correlated strongly (mean *r* = .75, range .65 to .83). Category result agreement was less consistent, with significant differences in the S–N, T–F, and J–P results. Test-retest continuous score correlations were very good for both instruments, with MMTIC® scale results ranging from *r* = .77 to .83 and MBTI® results ranging from *r* = .83 to .90. Self-estimates of type preferences on a 5-point Likert scale also correlated well with results from both instruments, and descriptions matched to assessment results were rated as more accurate than mismatched descriptions, even when only one preference of four was mismatched. In conclusion, the results (using continuous scores) showed high agreement of the two measures of type in an age group where both instruments are appropriate, as well as high test-retest reliability and good agreement with self-estimates of type for both measures. The ability of subjects to distinguish matched from mismatched type descriptions is important evidence that such descriptions are not generic or horoscope-like. Type constructs

Published by the Center for Applications of Psychological Type
Thomas G. Carskadon, Ph.D., Founding Editor
ISSN 1938-3290

measured by the two instruments have convergent validity, and thus results using the two instruments contribute to a common understanding of psychological type. The MMTIC assessment, however, was rated as easier to read and understand than the MBTI assessment and thus may be a better choice for younger children or less accomplished readers.

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 most commonly used assessment of psychological type, the Myers-Briggs Type Indicator® assessment (Myers, McCaulley, Quenk, & Hammer, 1998), is designed for use with individuals whose reading level meets or exceeds eighth grade, approximately age 14 (see Quenk, 2000). However, the constructs measured by the MBTI assessment are presumed to be innate, and children are believed to show evidence of them at ages younger than the instrument is targeting. The Murphy-Meisgeier Type Indicator for Children, or MMTIC, assessment was therefore designed for use with younger clients, using language appropriate for children (Murphy & Meisgeier, 2008). The core constructs of the two type instruments are intended to be the same, with interchangeable definitions for the four preference domains (Extraversion–Introversion, Sensing–Intuition, Thinking–Feeling, and Judging–Perceiving). The common underlying theory presumes that these domains dynamically interact with one another to influence behavior, following the same rules.

The *MMTIC® Manual* (Murphy & Meisgeier, 2008) indicates that the children's instrument is appropriate and validated for children and adolescents from elementary (grade 2) to high (grade 12) schools, or approximately ages 7 to 18. There is thus a user overlap (ages 14–18) between the intended audiences for the MBTI and MMTIC instruments. However, no studies have yet compared results from the most recent version of the MBTI instrument to those of the most recent version of the MMTIC instrument. A demonstration of convergent validity—evidence that the two instruments produce highly correlated results on matched scales and are thus measuring identical or similar constructs—is an essential requirement if research findings from either instrument are to have implications for age groups outside their targets. Both tools are important in the quest to gain insight into how individuals develop over the course of a lifetime.

The MMTIC assessment was first published in 1987 (Meisgeier & Murphy, 1987) and then revised 21 years later (Murphy & Meisgeier, 2008). In the interim, at least two research versions of the instrument were sometimes used, one which expanded the original 70 items to 93 (e.g., Lang, 1999) and one using 64 items (from which items were chosen for the revision published in 2008). The current version of the MMTIC assessment consists of 43 items. Internal consistency and test-retest reliabilities reported in the two manuals for the four scales (Extraversion–Introversion, Sensing–Intuition, Thinking–Feeling, and Judging–Perceiving) for both MMTIC versions range from .57 to .78, with most values in the .60's. These slightly low values are perhaps attributable to the emergent, developmental nature of type preferences in particular and children's nascent self-awareness and cognitive skills in general. Thus, reliabilities for older children are generally higher (Murphy & Meisgeier, 2008) than those of younger children and reliabilities for adults are generally higher than corresponding values for children (see Myers & McCaulley, 1985, pp. 166–167). Also of note is that the reduction of items to 43 in the revised MMTIC assessment had little effect on internal consistency reliability and in fact produced improvements in test-retest correlations (from .61, .69, .58, and .68 in the original MMTIC assessment to .78, .72, .71, and .69 in the revision, for E–I, S–N, T–F, and J–P respectively)[2].

Besides a reduction in length, the revised MMTIC instrument also eliminated the practice of labeling midzone scores for any of the four preference domains as "undetermined," or U for short. The original test sample reported in the 1987 manual (Meisgeier & Murphy, 1987) classified anywhere from 19% (T–F and J–P) to 25% (EI) of students in the "U-band" of different scales (with the S–N rate of 22%). The revised MMTIC instrument did away with U-scores by assigning midzone scores to one preference or the other. (Note that the percentages of adults with midzone, or "low preference clarity," scores on the MBTI and MMTIC instruments are similar—see Myers et al., 1998, p. 122.)

Three studies, two of them unpublished doctoral dissertations, administered the MBTI and MMTIC instruments to a group of students and correlated the results. One of the studies, Gilbert (1998), gave both instruments to the same students twice, separated by a 2-year interval. MBTI-MMTIC correlational results are summarized in **Table 1** for all these studies.

These values indicate a strong relationship between

Table 1. MBTI®-MMTIC® continuous score correlations in previous studies.

Study	Sample	EI	SN	TF	JP
Gilbert (1998)	*N* = 252, grades 6–10, private religious school, Ohio	.63	.60	.52	.61
Gilbert (1998) retest	*N* = 252, grades 8–12, private religious school, Ohio	.67	.61	.62	.64
Parker & Mills (1998)	*N* = 152, grades 5–7, gifted program, Johns Hopkins University	.59	.54	.62	.64
Lang (1999)	*N* = 220, grades 7–8, public school, Texas	.69	.54	.54	.47
Mean of above		.65	.57	.58	.59

Note: for all values, $p < .001$.

the two instruments, though not as high as might be desired. As suggested previously, the slightly low values may be attributable to developmental emergence of self-awareness and the less established personality qualities of children. This interpretation is consistent with the higher reliability values for Gilbert's retest, involving the same students two years older than at the time of the initial measurement.

Of course, there may also be differences between the two measurements that contribute to divergent results. Such differences are evident upon tabulating preference results for the two measurements. In all four of the samples shown in **Table 1**, the MBTI instrument indicated a significantly higher proportion of preferences for T and J than did the MMTIC instrument. The average percentage of Ts measured by the MMTIC assessment across the four assessments was 33%, whereas the MBTI assessment average was 48%, $\chi^2(1, N = 1672) = 36.20$, $p < .0001$ for the combined sample. The MMTIC assessment produced a J result only 25% of the time, compared to 37% of the MBTI results, $\chi^2(1, N = 1689) = 26.50$, $p < .0001$ for all four samples combined. In the Lang sample only, the MBTI assessment indicated a significantly higher percentage (31%) of Introverts than did the MMTIC assessment (21%), $\chi^2(1, N = 409) = 4.87$, $p = .027$. Finally, 73% of the Parker and Mills sample of gifted students tested as *N* on the MBTI assessment, compared to only 58% on the MMTIC assessment, $\chi^2(1, N = 263) = 6.26$, $p = .012$.[3]

With significantly different preference percentages for three of the four type scales using the two instruments, Parker and Mills concluded that "this study casts doubt on the MMTIC assessment's ability to identify a child's MBTI type accurately" (p. 20). This conclusion could just as easily be worded to question the MBTI assessment's ability to accurately identify a child's MMTIC type—the point being that, in the absence of validity evidence confirming a true preference, either or both instruments may be inaccurate. Thus, the evidence shows that the MMTIC consistently identifies more F and P preferences than the MBTI instrument does, but does not tell us which instrument, if either, is more accurate.

We also know very little about developmental influences on the measurement of type preferences. If attributes such as "Sensing" or "Thinking" are nascent or undeveloped in a middle school child, forced choice self-report questionnaires (the MBTI and MMTIC formats) may elicit low confidence, tenuous responses that may readily change when asked again.

The final caveat is that all of these studies used not only an older MMTIC form, but also an earlier MBTI form. Thus, the results are out of date, and any potential improvements (or detriments) resulting from more recent versions of the instruments have not been evaluated.

To address these concerns, the present research compared measurements of type using both the MBTI and MMTIC instruments (in their current versions) with individuals from an age group appropriate for both assessments. The primary focus examined whether the instruments' results were sufficiently similar to support convergent validity. A secondary interest was to examine users' perceptions of the instruments.

METHOD

Subjects. The study collected data from students in a high school dual enrollment program at a community college. This program is comprised of high school juniors and seniors who apply and are accepted. Their course work includes both high school and college level classes. Subjects were recruited on a voluntary basis (with both parental and teacher consent required as well for adolescents younger than 18), after the students

were notified of the project, presented as "self-discovery research." Subjects were drawn from courses in English, American Government, and Economics. Students were offered free feedback on their type assessment results (in a group session lasting approximately 75 minutes) and $25 for full participation.

Assessments. Students took both the MMTIC (including the current form's 43 items plus an additional 19 research items) and MBTI (Form M) instruments during their regularly scheduled classroom periods. The MMTIC items were presented in paper form, and students answered by circling one of two alternative answers. The MBTI items were presented in booklet form and required completion of a separate answer sheet. Students completed both instruments in a single class period (50 minutes long). Two to three weeks later they completed both instruments a second time, again during class time. The order in which the assessments were administered was determined randomly the first time and reversed for the second.

The final phase of the study involved the delivery of feedback to small groups of students, ranging in size from 6 to 30 participants. A trained type professional delivered the feedback using a standard PowerPoint slide show as a guide. The session lasted approximately 75 minutes and included brief background information on the history of type and its assessment, detailed descriptions of the type domains and preferences, exercises illustrating each domain, and distribution of two type reports (based on the first administration of both the MBTI and MMTIC instruments) to each participant. Subjects were asked to estimate their own type preferences three times during the feedback session—first, and one preference at a time, after hearing each corresponding preference domain description; second, after reading the two reports; and third, after additional discussion and questioning. The first two ratings were done using a 5-point Likert scale with verbal anchors ranging from (for example) "I definitely prefer E" to "I definitely prefer I" at opposite ends. The midpoint option was "I'm not sure which I prefer," and the two other choices were "I lean towards E (I)." These were coded as 1–5 points, with scores above 3 indicative of preference for the second letter pole (I, N, F, P), scores below 3 indicative of a first letter preference (E, S, T, J), and a score of 3 indicating uncertainty about preference. The final self-estimate of type consisted simply of self-identifying a four-letter type without any rating of certainty.

Ratings of the instruments and reports. After the first completion of each assessment, students were asked the following questions:

1) How easy were the questions to understand? (using a 5-point scale anchored with "very difficult" and "very easy")
2) Were you confused by any of the questions? (a 5-point scale anchored with "never" and "very often")
3) Did you have trouble understanding the wording of any questions? (a 5-point anchored with "not at all" and "very often")

Upon completion of the feedback session, students were asked to rate (using 5-point Likert scales) the accuracy of both the MBTI and MMTIC reports they received, as well as the age appropriateness of both (choosing among "appropriate for someone younger," "appropriate for my age," or "appropriate for someone older"). They also rated how interesting they found the information they received, also using a 5-point scale.

Type description ratings. Prior to receiving type reports during the feedback session, each participating student was given two lists of 12 adjectives and short phrases and asked to rate how well each list item matched their own personalities. There were 16 different sets of adjectives (though some adjectives were common to types with shared preferences) corresponding to 16 type descriptions drawn from the handout *Descriptions of the Sixteen Types* (Lawrence, 1998) and the book *Introduction to Type®* (Myers, 1980). Students were instructed to select one of four possible answers to each description: "not at all like me," "only a little like me," "mostly like me," and "very much like me." Each rating was awarded 1 to 4 points and the 12 items were totaled. The highest possible similarity point total would thus be 48 and the lowest 12.

Different students received two different sets of descriptions, following these rules:

1) When the whole type results from the two instruments disagreed, one description corresponded to each of the two results.
2) When the instruments' results agreed, one type description corresponded to the consensual indicated whole type (congruent descriptors). The second description (incongruent descriptors) was randomly selected to be one of the following:
   a. Half were assigned to receive a description of a polar opposite type, i.e., all four letters disagreed.
   b. The other half were assigned to receive a description one letter different. Half of these

Table 2. MMTIC® internal consistency for all scales.

	Administration 1 (*N* =123)		Administration 2 (*N* = 110)	
MMTIC® Scale	Cronbach's alpha	Split half	Cronbach's alpha	Split half
E–I	.73	.77	.77	.78
S–N	.75	.75	.78	.75
T–F	.77	.78	.78	.78
J–P	.72	.73	.76	.73

differed on the perceiving function (e.g., ISTP instead of INTP) and the other half on judging function (e.g., ISTP instead of ISFP).

RESULTS

Subject demographics. We collected at least one complete assessment response from 123 students, of whom 115 completed at least one set of answers for both the MMTIC and MBTI assessments. One hundred and eleven completed the MBTI assessment a second time, 110 of whom also completed the MMTIC assessment twice. Eighty-one of these 110 subjects also attended a feedback/verification session.

Of the original 123 subjects, 83 (67%) were female. The average age at time of the first assessment was 16.54 years (*SD* = .63, range from 15 to 19). The average reading grade level equivalent was 12.18 (*SD* = .89, range from grade 9 to 12.9). Eighty-seven percent of students were rated at or above grade level for reading ability.

Of the 81 subjects who completed the entire study, 60 were female and 21 male, meaning disproportionately more males (32% vs. 22%) failed to complete at least one part of the study.

Students were generally very positive about their participation, with a mean rating of 4.26 on a 5-point scale of how interesting they found the research project. Sixty-one percent of students took the opportunity to provide optional written comments; of these, 91% were positive, with the remaining 9% offering mild suggestions for process improvements. The positive comments primarily pointed out beneficial insights and self-understanding.

MMTIC® internal consistency. Table 2 shows the results for the four MMTIC scales and two internal reliability measures, Cronbach's alpha and split-half, for each of the two times students took the MMTIC instrument.

MBTI®-MMTIC® agreement: continuous scores. Both instruments generate continuous scores as well as dichotomous category results for each of the four scales measured. The MBTI continuous scores used for analysis were the preference clarity index (pci) values given in MBTI reports, adjusted to be negative values for preferences for E, S, T, and J. Thus, a pci score of 19 for a result indicating a preference for Extraversion (E) was changed to -19, whereas a preference score of 19 for Introversion remained positive 19. These scores, which range from -30 to +30, are a near perfect substitute (correlations of .99+) for MBTI theta scores derived from IRT scoring, which are available only by request from the MBTI publisher, CPP. MMTIC continuous scores[4] are not included in MMTIC reports, but were provided by the MMTIC publisher, CAPT. These are based on scoring weights derived from Latent Class Analysis (see Murphy & Meisgeier, 2008) and range from approximately -1000 to +1000, again following the type convention of making scores for E, S, T, and J negative and scores for I, N, F, and P positive. Results are shown in Table 3. All values are highly significant and indicative of substantial correlation.

MBTI®-MMTIC® agreement: categories. Table 4

Table 3. MBTI®-MMTIC® continuous score correlations for two administrations.

Preference Pair	Administration 1 (*N* = 115)	Administration 2 (*N* = 110)
E–I	.81	.83
S–N	.78	.83
T–F	.65	.71
J–P	.74	.75

Note: $p < .001$ for all correlations

Table 4. MBTI®-MMTIC® preference agreements for two administrations.

Number of preferences in agreement	Administration 1 (*N* = 115)	Administration 2 (*N* = 110)
4	42 (36.5%)	46 (41.8%)
3	47 (40.9%)	40 (36.4%)
2	21 (18.3%)	20 (18.2%)
1	5 (4.3%)	4 (3.6%)
0	0 (0.0%)	0 (0.0%)

shows the percentage of students with 0, 1, 2, 3, or 4 identical preferences identified by the two type measures. The results for the two administrations were very similar: 77.4% of students agreed on 3 or 4 preferences for the first administration, and 78.2% on the second administration. No students disagreed on all four preferences as indicated by the two measures.

Table 5 shows the frequency counts for different preference results as measured by the two instruments for each of the two administrations. The column to the right of each set of preference pairs is the chi-square test of significance for differences in the resulting proportions.

The clearest difference is the greater incidence of a preference for S when the MMTIC instrument is the measure, 22% higher than the MBTI instrument for the first administration and 15% higher for the second. The MMTIC assessment also reported a significantly higher percentage of Ts, though only marginally so for second administration. The MMTIC assessment also generally reported a higher incidence of J, significantly for the second administration but not the first.

Table 6. MBTI®-MMTIC® continuous score test-retest correlations.

Instrument	EI test-retest	SN test-retest	TF test-retest	JP test-retest
MMTIC®	.83	.82	.77	.83
MBTI®	.90	.88	.83	.85

Note: *N* = 111 for the MBTI assessment and *N* = 110 for the MMTIC assessment. All values $p < .001$.

Test-retest agreement: MMTIC® and MBTI® continuous scores. Table 6 shows the correlations of continuous scores between the two administrations of each instrument, separated by 2–3 weeks. All the test-retest reliability results are highly significant. While the MBTI correlations are slightly higher, none of the differences from corresponding MMTIC reliabilities is significant.

Test-retest agreement: MMTIC® and MBTI® category results. Table 7 shows the number of preference agreements (out of four possible) for the two administrations of each instrument.

The number of preference measurements with perfect retest agreement was proportionately higher for the MBTI instrument (70 of 111, or 63%) than the MMTIC instrument (55 of 110 cases, 50%), $\chi^2(1, N = 221) = 3.84$, $p = .05$. The mean number of preference agreements was also significantly higher for the MBTI instrument (3.54 vs. 3.30 for the MMTIC instrument), $t(109) = 2.59$, $p = .011$. In addition, for individual preferences, the MBTI percentage of test-retest agreement was higher for 7 of the 8 preferences (only the J preference had lower test-retest agreement on the MBTI assessment than the MMTIC assessment).

Agreement with self-estimates of type. Table 8 shows the correlations of MBTI® and MMTIC® continu-

Table 5. MBTI®-MMTIC® individual preference results for two administrations.

	E	I		S	N		T	F		J	P	
MMTIC® #1	60 (52%)	55 (48%)	χ^2= 3.01	77 (67%)	38 (33%)	χ^2= 11.03	48 (42%)	67 (58%)	χ^2= 4.91	48 (42%)	67 (58%)	χ^2= 1.18
MBTI® #1	73 (63%)	42 (37%)	*p* = .08	52 (45%)	63 (55%)	*p* < .001	32 (28%)	83 (72%)	*p* = .03	40 (35%)	75 (65%)	*p* = .28
MMTIC® #2	57 (52%)	53 (48%)	χ^2= 1.49	66 (60%)	44 (40%)	χ^2= 4.67	45 (41%)	65 (59%)	χ^2= 3.38	48 (44%)	62 (56%)	χ^2= 5.03
MBTI® #2	66 (60%)	44 (40%)	*p* = .22	50 (45%)	60 (54%)	*p* = .03	32 (29%)	78 (71%)	*p* = .07	32 (29%)	78 (71%)	*p* = .02

Table 7. Number of preference agreements for two MMTIC® and MBTI® administrations.

Number of test-retest preference agreements	MMTIC® test-retest	MBTI® test-retest retest	Individual categories	MMTIC®	MBTI®
4	55 (50.0%)	70 (63.1%)	E on both	46 (80.7%)	62 (87.3%)
3	35 (31.8%)	33 (29.7%)	I on both	42 (79.2%)	36 (90.0%)
2	18 (16.4%)	5 (4.5%)	S on both	59 (80.8%)	44 (86.3%)
1	2 (1.8%)	3 (2.7%)	N on both	30 (81.1%)	53 (88.3%)
0	0 (0%)	0 (0%)	T on both	35 (77.8%)	25 (80.6%)
			F on both	55 (84.6%)	72 (90.0%)
			J on both	39 (84.8%)	30 (78.9%)
			P on both	55 (85.9%)	70 (95.9%)
Mean (*SD*) number of preference agreements	3.30* (.81)	3.54* (.71)	Mean percentage of individual agreement	81.79%	87.16%

Note: $N = 110$. Means differ $p = .015$

Table 8. MMTIC® and MBTI® continuous score correlations with Likert-scale self-estimates of preferences.

	Self-estimate 1 (5-point Likert scale)				Self-estimate 1 (5-point Likert scale)			
	E–I	S–N	T–F	J–P	E–I	S–N	T–F	J–P
MBTI®	.76	.63	.61	.61	.76	.66	.65	.67
MMTIC®	.71	.57	.48	.71	.73	.59	.60	.72

Note: N = for the first self-estimate = 81; N for the second = 78. All values $p < .001$

ous scores (first administration) with the two 5-point Likert scale self-estimates of each type preference made by the students, one before receiving feedback about their scores on both the MBTI and the MMTIC instruments (but after being introduced to the concepts being measured) and one after receiving feedback.

E–I self-ratings agree the most, and T–F the least, of the preference domains. The MBTI self-correlations are slightly higher than the MMTIC self-correlations for E–I, S–N, and T–F, but slightly lower for J–P. Post-feedback values are equal to or slightly higher than pre-feedback results (with MMTIC T–F showing the greatest difference from pre to post). None of these differences is statistically significant, however.

The third and final self-rating, occurring at the end of the feedback session and further discussion and questioning, represents what type practitioners commonly call "best fit" or "verified" type. The agreement of each of these self-estimates of preference letter with the results of the first administration of the MBTI and MMTIC instruments is shown in **Table 9**.

Further analyses revealed a significantly lower rate of agreement of the final self-estimate with either an MBTI or MMTIC S rather than an N result. For the MBTI assessment, 6 of 36 S results were self-confirmed as N, compared to only 1 of 37 N results, Fisher's exact $p = .056$. For the MMTIC assessment, there were 17 of 48 MMTIC Ss who self-confirmed N, compared to no switches for any of the 25 MMTIC Ns, $\chi^2(1, N = 73) = 11.54$, $p < .001$. There was also a significantly greater likelihood of agreement with the MMTIC instrument (but not the MBTI instrument) of self-estimates of F

Table 9. Percentage of agreement of the final self-estimate of type preferences with MBTI® and MMTIC® results (first administration).

	E	I	S	N	T	F	J	P	Mean
MBTI®	85.4%	92.3%	83.3%*	97.3%*	85.0%	88.7%	85.2%	83.0%	87.5%
MMTIC®	92.5%	82.4%	64.6%**	100%**	65.5%**	90.9%**	84.4%	90.5%	83.8%

*difference in agreement rate between opposite preferences $p = .056$
**difference in agreement rate between opposite preferences $p < .01$

than with T, $\chi^2(1, N = 73) = 7.27$, $p = .007$.

Preference changers' continuous scores. Both the MMTIC and the MBTI instruments produce scores that represent the consistency with which a respondent chooses the item response pointing to one preference rather than its opposite. The higher the preference clarity index (MBTI) or percentage scores (MMTIC)[5], the more the responses for a given student consistently indicated the same preference. Lower scores indicate that opposing preferences were chosen more equally.

Any two measures or estimates of a type preference for the same individual will either agree or disagree, allowing the pci and percentage scores of "changers" to be compared to those of "non-changers." Analyses across a wide variety of measurement combinations (two MBTI and two MMTIC measures of each preference domain, compared to each other, compared for test and retest, and compared to the final self-estimate of preference) invariably showed a higher score (more consistent) for students whose preferences agreed on the two relevant measures. These results are summarized in **Table 10**.

In every case in **Table 10**, the consistency scores for students whose preferences agree on any two measures (including self-estimates) are higher than those who disagree. Of the 40 comparisons of means, all but six differ significantly. Most of the mean differences are large, close in magnitude to a standard deviation.

Student self-ratings of type descriptors. Two different sets of 12 adjective/phrase type descriptors were distributed to students for self-ratings of similarity. The content of these sets varied, depending on degree of agreement of a student's MBTI and MMTIC results. Ratings were collected from 80 subjects; of these, 51 differed on one, two, or three preferences as measured by the MBTI and MMTIC assessments and were given the descriptor lists corresponding to the two indicated types. Ratings of descriptions congruent with the MBTI results were marginally significantly higher than those congruent with MMTIC results, $t(49) = 1.97$, $p = .055$. Results are summarized in **Table 11**. (See page 16.)

Twenty-nine of the 80 students had identical MBTI®-MMTIC® results. These subjects were given one descriptor list congruent with the indicated type and an alternative list corresponding to a type either one letter or all letters different. Results are shown in **Table 12**. (See page 16.).

The incongruent descriptors were rated as much less self-descriptive than the congruent lists, $t(28) = 4.55$, $p < .001$. The paired comparison of similarity ratings for congruent vs. incongruent lists was significant even when the incongruence involved a single letter, $t(12) = 2.80$, $p = .016$, though the difference was much larger when all four letters differed, $t(15) = 5.05$, $p < .001$. The similarity ratings for 4-letter incongruent descriptors ($M = 28.69$, $SD = 5.62$) were also very much lower than the 1-letter incongruent descriptors ($M = 39.62$, $SD = 3.40$), $t(27) = 6.15$, $p < .001$. (Note that this is a between subjects comparison—those who received a 1-letter discrepant report and those who received a 4-letter discrepant one—rather than a within subject, paired comparison as the other *t*-tests are.)

Student perceptions of the instruments. The MMTIC ($M = 4.14$) and MBTI ($M = 4.09$) reports were perceived by their students as approximately equal in accuracy, $t(76) = 0.62$, $p = .54$. The modal response to the accuracy question, endorsed by 38 of 77 students who answered it, rated the two instruments as equally accurate. Of the remaining 39, one indicated both were "equally inaccurate," and the MBTI and MMTIC instruments were each rated as more accurate than the other instrument by two sets of 19 students.

There was little difference in the perceived age appropriateness of the two instruments, with the overwhelming majority of students rating both the MMTIC (89.6%) and the MBTI (93.5%) instruments as age appropriate. Only one student rated the MMTIC assess-

Table 10. Consistency measure differences for changers and non-changers for two measures of each preference.

Consistency measure	First administration MBTI®-MMTIC® agreement			Second administration MBTI®-MMTIC® agreement			MBTI® and MMTIC® Test-Retest			First administration MBTI® and MMTIC® agreement with final self-estimate			Second administration MBTI® and MMTIC® agreement with final self-estimate		
	Mean/ *SD*	*N*	*t*-test	Mean/ *SD*	*N*	*t*-test	Mean/ *SD*	*N*	*t*-test	Mean/ *SD*	*N*	*t*-test	Mean/ *SD*	*N*	*t*-test
EI agree MBTI® pci	16.02 (9.18)	92	t (113) = 3.51 p = .001	16.38 (9.08)	89	t (108) = 4.72 p < .001	15.97 (8.93)	97	t (109) = 4.89 p < .001	15.09 (9.44)	64	t (72) = 2.51 p = .014	17.36 (8.70)	58	t (72) = 5.09 p < .001
EI disagree MBTI® pci	8.83 (6.97)	23		6.67 (5.17)	21		4.21 (2.16)	14		7.50 (2.92)	10		5.69 (5.38)	16	
EI agree MMTIC® pct	91.72 (11.62)	92	t (113) = 3.18 p = .002	93.37 (11.19)	89	t (109) = 4.56 p < .001	92.74 (10.14)	88	t (109) = 5.53 p < .001	90.68 (11.56)	65	t (72) = 2.59 p = .012	92.18 (12.41)	59	t (71) = 2.65 p = .01
EI disagree MMTIC® pct	82.65 (14.52)	23		79.57 (17.31)	21		77.92 (15.01)	22		79.54 (15.70)	9		81.80 (16.96)	14	
SN agree MBTI® pci	12.03 (8.05)	86	t (113) = 3.20 p = .002	14.12 (7.75)	86	t (109) = 5.66 p < .001	11.66 (7.71)	97	t (109) = 3.44 p = .001	11.70 (7.92)	66	t (71) = 2.21 p = .031	14.54 (8.38)	57	t (71) = 4.35 p < .001
SN disagree MBTI® pci	6.90 (5.35)	29		4.92 (3.99)	24		4.43 (3.96)	14		5.00 (3.37)	7		5.13 (3.90)	16	
SN agree MMTIC® pct	93.92 (10.53)	86	t (113) = 3.84 p < .001	93.51 (10.74)	86	t (109) = 5.44 p < .001	93.60 (10.97)	89	t (108) = 4.49 p < .001	93.30 (12.12)	56	t (71) = 1.56 p = .12	92.76 (12.41)	58	t (70) = 1.67 p = .10
SN disagree MMTIC® pct	84.30 (14.60)	29		77.67 (18.43)	24		81.01 (13.86)	21		88.19 (10.83)	17		86.06 (17.17)	14	
TF agree MBTI® pci	12.95 (8.03)	87	t (113) = 4.21 p < .001	15.33 (8.87)	85	t (109) = 3.57 p = .001	11.92 (8.06)	97	t (109) = 4.77 p < .001	12.05 (8.52)	64	t (71) = 2.62 p = .01	14.80 (9.32)	65	t (71) = 1.33 p = .19
TF disagree MBTI® pci	6.14 (5.15)	28		8.32 (7.63)	25		6.64 (6.21)	14		4.44 (3.91)	9		10.25 (7.54)	8	
TF agree MMTIC® pct	94.10 (9.81)	87	t (113) = 1.46 p = .148	94.54 (12.01)	85	t (109) = 3.29 p = .001	94.72 (9.30)	90	t (108) = 2.78 p = .006	95.72 (8.38)	59	t (71) = 3.15 p = .002	97.04 (8.70)	49	t (70) = 4.11 p < .001
TF disagree MMTIC® pct	90.76 (12.66)	28		85.18 (14.66)	25		87.92 (12.35)	20		86.49 (14.74)	14		86.35 (13.12)	23	
JP agree MBTI® pci	14.54 (8.59)	91	t (113) = 4.96 p < .001	16.70 (8.49)	88	t (109) = 3.69 p < .001	13.41 (8.36)	100	t (109) = 3.67 p < .001	13.58 (8.78)	62	t (72) = 2.60 p = .01	16.90 (8.51)	59	t (72) = 3.45 p = .001
JP disagree MBTI® pci	5.71 (2.58)	24		9.32 (6.64)	22		4.09 (2.47)	11		6.75 (5.03)	12		8.93 (5.24)	15	
JP agree MMTIC® pct	90.09 (14.32)	91	t (113) = 1.48 p = .14	93.88 (10.15)	88	t (109) = 4.40 p < .001	90.89 (13.46)	94	t (108) = 4.02 p < .001	90.26 (14.69)	65	t (72) = 1.30 p = .20	94.26 (10.11)	59	t (72) = 2.99 p = .004
JP disagree MMTIC® pct	85.18 (14.91)	24		81.19 (18.44)	22		75.89 (15.72)	16		83.47 (15.01)	9		83.32 (15.09)	9	

Table 11. Self-ratings of descriptors corresponding to differing MBTI® and MMTIC® type results.

MBTI®-MMTIC® results	*N*	Descriptor Ratings MBTI® Match Mean (*SD*)	Descriptor Ratings MMTIC® Match Mean (*SD*)	Mean Difference
3 letters different	3	36.33 (7.10)	33.67 (3.51)	2.67
2 letters different	13	39.46 (4.05)	37.54 (4.65)	1.92
1 letter different	34	38.74 (4.22)	37.88 (5.09)	0.85
Any or all letters different	50	38.78 (4.31)	37.54 (4.92)	1.24*

*p = .055

Table 12. Self-ratings of descriptors corresponding to congruent and incongruent type results.

Alternative Descriptors Difference from Congruent Descriptors	*N*	Descriptor Ratings Congruent Match Mean (*SD*)	Descriptor Ratings Incongruent Match Mean (*SD*)	Mean Difference
SN different	9	41.89 (4.01)	39.78 (3.27)	2.11*
TF different	4	40.25 (4.35)	39.25 (4.19)	1.00
1 letter different	13	41.38 (4.01)	39.62 (3.40)	1.76*
All 4 letters different	16	40.38 (5.27)	28.69 (5.62)	11.69**
1 or 4 letters different	29	40.83 (4.69)	33.59 (7.24)	7.24**

*p < .05; **p < .001

ment as "appropriate for someone younger."

The MMTIC instrument itself, as distinct from the report, was perceived as significantly easier to read (M = 4.68) than the MBTI instrument (M = 4.42), $t(108)$ = 3.99, p <.001. It was also seen as less confusing (M = 1.41) and as having less problematic wording (M = 1.29) than the MBTI instrument (M = 1.61 and 1.69, respectively), $t(108)$ = 2.34, p =.02 and $t(107)$ = 4.03, p <.001.

DISCUSSION

The primary question and finding of this research was a validity question: the consistency of results using two measures of type, the MBTI and MMTIC assessments, with adolescents. Correlations of scores from the two instruments were all above .70, with the exception of the first administration of the two instruments for the T–F scale (r = .65). These results are an improvement over previous studies (Gilbert, 1998; Lang, 1999; Parker & Mills, 1998), all of which found much lower correlations (ranging from .47 to .67). These studies used the previous versions of both the MBTI and MMTIC instruments, but also (with the exception of some of Gilbert's subjects) worked with children younger than the high school students in the current research. However, the correlations for Gilbert's high school students were also lower than the current results, supportive of an improvement in agreement with the two revised instruments.

The category agreement results also differed from studies using the previous versions of the two instruments. Prior results showed a consistent difference on the T–F and J–P scales, with the MMTIC instrument indicating more Fs and Ps and the MBTI instrument more Ts and Js. In the current study, with the new versions of the instruments, the MMTIC assessment indicated more Ts and Js relative to the MBTI assessment, a reversal of the previous findings. The present results also showed a higher proportion of S results with the MMTIC instrument than the MBTI instrument—a finding consistent with Parker & Mills (1998). Further research will be necessary to determine if and when the MMTIC assessment measures more Ss than the MBTI assessment, as the other previous studies failed to find such a difference.

A related validity question concerns the agreement of self-ratings of type with the results of the instruments. As seen in **Table 8**, instrument scores and self-ratings correlated significantly (ranging from .44 to .76), though not as well as scores between the two instruments. The two type scales considered attitude scales, J–P and especially E–I, had higher instrument-self correlations than the function scales (S–N and T–F) did. Perhaps these constructs are more easily understood or observed by children as they mature. Extraversion–Introversion was the first construct that Jung extracted from his observations of his clients, and indeed is arguably the most consistently implicated construct in a wide variety of personality theories, from Eysenck (1967) to the Five Factor Model (McCrae & Costa, 1989).

Test-retest agreement is concerned with reliability rather than validity, but is also an essential goal for accurate measurement. Both the MMTIC and MBTI instruments showed high test-retest reliabilities for all scales. The MMTIC results with this sample were an improvement upon the results presented in the *MMTIC® Manual* (Murphy & Meisgeier, 2008), which were based upon the same item set and scoring as the current study, and much higher than the results reported by Gilbert (1998), using an earlier MMTIC version. This better result may reflect both improvements in the MMTIC measure and the shorter test-retest interval (2–3 weeks) employed in the current study, contrasted to the 3–6 months of the data in the *MMTIC® Manual* and the two years in Gilbert's study.

On most of the yardsticks used to measure reliability and validity in this study, the MBTI results showed a slight advantage over the MMTIC results. This is not surprising in light of the fact that the MBTI instrument is more than twice as long as the MMTIC instrument (93 items versus 43), and psychometric indicators like reliability are calculated with formulas that give a benefit to higher item counts. Additionally, the MMTIC items were selected using Latent Class Analysis, a method that eliminates redundantly worded items that otherwise enhances internal consistency/reliability indicators such as Cronbach's alpha (see Murphy & Meisgeier, 2008). Nonetheless, the reliability and validity measures for the MMTIC scales were respectable, and the fact that students found it easier to read but not so easy as to be age unsuitable suggests the MMTIC assessment is appropriate for young children and adolescents, particularly those who are less developed readers.

Another important finding of the current research is a refutation of the criticism of type descriptions as being indiscriminately generic and universally endorsable (e.g., Long, 1992; Paul, 2004; Zemke, 1992). The Forer or Barnum effect was first documented by Forer (1949), who found that subjects described the same vague personality description, ostensibly unique to them and based on a personality test (which was in fact never scored), as highly accurate. In the current study, students clearly rated descriptions based on their measured type results as more accurate than those descriptions that were mismatched. This was true if just one of four type results was mismatched, and the effect increased the further the bogus descriptors deviated from actual results. The current results corroborate past research (Carskadon, 1982; Carskadon & Cook, 1982) which also found strong alignment of matched reports with perceived accuracy.

A more persistent criticism of type theory and measurement, often singled out as "the central question in the evaluation of the instrument" (McCrae & Costa, 1989, p. 20), is the assumption that the constructs being measured are categorical (specifically, bimodal) rather than "a continuous, normally distributed psychological dimension or trait" (Devito, 1985, p. 1032). If this objection is accurate, one consequence is a loss of resolution and measurement accuracy from the reduction of multiple scores to two categories (Boyle, 1995; Harvey & Murry, 1994; McCrae & Costa, 1989; Pittenger, 1993, 2005). A corollary implication is that relatively minor differences in scores near the dividing point between preferences, as opposed to shifts at the extremes of a normal distribution, will often result in a change of preference from one instrument to another (or between two administrations of the same instrument). There is ample evidence of such an effect in the data in **Table 10**, in which type preferences changed significantly more often (across different measurements or time) for people with lower continuous scores (i.e., closer to the cut point). With four measurement results from each administration of a type measure (one for each scale), the likelihood of at least one low clarity score with either the MBTI or MMTIC instruments is much greater than with a single scale measure. Thus, the practice of categorizing type measurement may inherently limit the test-retest reliability of such an instrument.

Whether type measurement results are best conceptualized as dichotomous or continuous is ultimately an empirical question. Even some harsh critics of the

MBTI instrument recognize its potential and call for research to elucidate the controversial tenets of the theory (Barbuto, 1997; Coan, 1978; Druckman & Bjork, 1991; Thayer, 1988; Thompson & Ackerman, 1994; Zemke, 1992). Rigorous research regarding many of the theoretical underpinnings of the MBTI and MMTIC instruments, however, is still relatively rare.

FOOTNOTES

[1] Address correspondence to Robert McPeek, Director of Research, Center for Applications of Psychological Type, 2815 NW 13th Street, Suite 401, Gainesville, FL 32609. Bob@capt.org. The authors gratefully acknowledge financial support offered by a donation from the Arizona chapter of the *Association of Psychological Type*, which helped fund this study.

[2] Selection of items for the MMTIC revision was based largely on results of latent class analysis (Magidson & Vermunt, 2002; McCutcheon, 1987). Because LCA, unlike traditional psychometric analysis, seeks to identify scale items that minimize redundancy, internal consistency (Cronbach's alpha) is sometimes lowered as a consequence. LCA advocates counter that alpha is artificially inflated by overly similar items that elicit overly similar responses.

[3] These values were not always explicitly presented in the sources cited but in such cases could be extracted from tables. Note that this effort uncovered an abundance of inconsistencies, generally minor, in Gilbert's tables. Additionally, corresponding "error" counts in tables 5 and 7 in Parker and Mills (1998) did not match as they should, suggesting transcription errors in one or both tables. The MMTIC U-band results also were treated differently: Lang (1998) dropped them from analysis, Parker and Mills (1998) analyzed data including or excluding U-band scores, and Gilbert (1999) assigned U-band scores to one preference or another depending on which side of the midpoint the score fell.

[4] These Latent Class Analysis-based continuous scores correlate highly with response consistency scores (presented as percentages) included in the reports, provided that the percentage values are made negative for E, S, T, and J preference results. Correlations of corresponding LCA-based scores and such scaled percentage scores in the present study were all .89 or higher.

[5] MBTI pci scores are always positive and may be the same value for opposite preferences—for example, a Judging and Perceiving pci value may be the same even though the categorical result is opposite. In previous analyses, pci scores for E, S, T, and J were made negative to distinguish them from I, N, F, and P scores. MMTIC percentage scores were not used in previous analyses, but are appropriate here (without changing the valence for E, S, T, and J scores) because the intent in this analysis is to measure the consistency of responses in relationship to preference categorical measurement consistency.

REFERENCES

Barbuto, J. E., Jr. (1997). A critique of the Myers-Briggs Type Indicator and its operationalization of Carl Jung's psychological types. *Psychological Reports, 80*(2), 611–625.

Boyle, G. J. (1995). Myers-Briggs Type Indicator (MBTI): Some psychometric limitations. *Australian Psychologist, 30*(1), 71–74.

Coan, R. W. (1978). Review of the Myers-Briggs Type Indicator. In O. K. Buros (Ed.), *The eighth mental measurements yearbook*. Highland Park, NJ: The Gryphon Press.

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

Harvey, R. J., & Murry, W. D. (1994). Scoring the Myers-Briggs Type Indicator: Empirical comparison of preference score versus latent-trait methods. *Journal of Personality Assessment, 62*(1), 116–129.

Carskadon, T. G. (1982). Myers-Briggs Type Indicator characterizations: A Jungian horoscope? *Research in Psychological Type, 5*, 87–88.

Carskadon, T. G., & Cook, D. D. (1982). Validity of MBTI descriptions as perceived by recipients unfamiliar with type. *Research in Psychological Type, 5*, 89–94.

Devito, A. J. (1985). Review of Myers-Briggs Type Indicator. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 1030–1032). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Thomas Publishing.

Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44*, 118–121.

Gilbert, A. P. (1998). *A test-retest study of the Myers-Briggs Type Indicator (MBTI) and the Murphy-Meisgeier Type Indicator for Children (MMTIC)* over a two year time period. Doctoral dissertation, Ohio State University). Dissertation Abstracts International, 59(05), 1457A (University Microfilms No. AAG98-33981)

Lang, M. L. S. (1999). *A concurrent validity study of the MBTI, MMTIC-R and SSQ with middle school students*. Doctoral dissertation, Texas Woman's University. Dissertation Abstracts International, 60(09), 3271A. (University Microfilms No. AAI99-44500)

Lawrence, G. D. (1998). *Descriptions of the sixteen types.* Gainesville, FL: Center for Applications of Psychological Type.

Long, T. G. (1992). Myers-Briggs and other modern astrologies. *Theology Today,* 291–295.

Magidson, J. & Vermunt, J. K. (2002). A nontechnical introduction to latent class models. Retrieved 9-12-13 from http://statisticalinnovations.com/technicalsupport/articles.html.

McCrae, R. R. & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the Five Factor Model of personality. *Journal of Personality, 57*(1), 17–40.

McCutcheon, A. L. (1987). *Latent class analysis*. Quantitative Applications in the Social Sciences Series No. 64. Thousand Oaks, California: Sage Publications

Meisgeier, C. H., & Murphy, E. A. (1987). *Murphy-Meisgeier Type Indicator for Children: Manual*. Palo Alto, CA: Consulting Psychologists Press.

Murphy, E. A., & Meisgeier, C. H. (2008). *MMTIC Manual: A guide the development and use of the Murphy-Meisgeier Type Indicator for Children* (2nd ed.). Gainesville, FL: Center for Applications of Psychological Type.

Myers, I. B. (1980). *Introduction to type* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Myers, I. B., McCaulley, M. H., Quenk, N. L. & Hammer, A. L. (1998). *MBTI Manual: A guide to the development and use of the Myers-Briggs Type Indicator* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Parker, W. D., & Mills, C. J. (1998). An examination of the Murphy-Meisgeier Type Indicator for Children with a sample of academically talented children. *Journal of Psychological Type, 44*, 20–25.

Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research, 63*(4), 467–488.

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57*(3), 210–221.

Thayer, P. W. (1988). *The Myers-Briggs Type Indicator and enhancing human performance*. A report prepared for the Committee on Techniques for the Enhancement of Human Performance of the National Academy of Sciences. Unpublished manuscript, North Carolina State University.

Thompson, B., & Ackerman, C. (1994). Review of the Myers-Briggs Type Indicator. In J. Kapes, M. Mastie, & E. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (3rd ed., pp. 283–287). Alexandria, VA: American Counseling Association.

Zemke, R. (1992). Second thoughts about the MBTI. *Training, 29*(4), 43–47.

ROBERT MCPEEK, Ph.D., is Director of Research at the Center for Applications of Psychological Type (CAPT).

JUDITH BREINER, Ph.D., is Director of Research Operations at CAPT.

CONTACT

Robert W. McPeek
2815 NW 13th Street, Suite 401
Gainesville, Florida 32609
ph: 352.375.0160 x 106
Email: bob@capt.org

This *Journal* is being made available through the collaborative efforts of Dr. Tom Carskadon, Founding Editor of the *Journal of Psychological Type*®, and the Center for Applications of Psychological Type, Inc., CAPT, worldwide publisher. The Journal of Psychological Type's editorial staff includes Dr. Robert W. McPeek, Editor, and Dr. Judith F. Breiner, Managing Editor.

CAPT is a not-for-profit organization dedicated to the meaningful application and ethical use of psychological type as measured through the Myers-Briggs Type Indicator instrument.

ISSN 1938-3290