Personality Type Estimator (PIE): Development of Instrument to Identify Personality Types

Gary J. Conti
Professor of Adult Education, Retired
Oklahoma State University

[Note: This is the methodology section of a manuscript on the development of the Personality Type Estimator (PIE). It describes the development of PIE and the procedures used to establish PIE's validity and reliability.]

Purpose and Methodology

The purpose of this study was to develop an instrument to estimate personality type which was easy to administer, which could be completed rapidly, and which could be used immediately by both facilitators and learners. The instrument which was created was named the Personality Identity Estimator (PIE). Numerous steps were undertaken to establish its validity and reliability.

Results from the analyses of responses to the Myers-Briggs Type Indicator (MBTI) were used for constructing PIE. Using the 94-item form of the MBTI (Form G Self-Scorable), data were collected from 553 volunteers in Alberta, Canada, and in the states of Montana, Nebraska, New Mexico, Oklahoma, and Texas. This group was composed of Adult Basic Education teachers, public school teachers, professionals who teach adults in various agencies, adult students returning to a nontraditional college credit program, fire fighters, students in continuing education classes, community college students, and college students.

Respondents provided information concerning their age, gender, ethnicity, and educational level and then completed the 94-item version of the MBTI. The sample consisted of 321 females (58.2%) and 231 males (41.8%). The average age of the group

was 30.8 with a range from 18 to 90. The ethnic make-up of the group was as follows: White--83.9%, Native American--6%, African American--4.9%, Hispanic--4.2%, and Other--1%. The educational level of the respondents varied as follows: Less than a high school diploma--.7%, high school diploma--37%, vocational or educational certificate--11.5%, associates degree--24%, bachelors degree--13.9%, and graduate degree--13.8%.

This data set was subjected to multi-variate analyses to identify the concepts for the items for PIE. Additional data were then collected to establish the validity and reliability of PIE. Data were collected from 174 participants using the PIE items to establish criterion-related validity and from 89 participants to establish reliability. These items were subjected to tests for validity and reliability.

Construct Validity

Validity is concerned with what a test actually measures (Wiersma & Jurs, 2005, p. 362). The three most important types recognized in educational research are construct, content, and criterion-related validity (Kerlinger, 1973, p. 457). Nevertheless, it has come to be recognized that validity is a unitary concept (Gay & Airasian, 2000, p. 162; Wiersma & Jurs, 2005, p. 327), that “there are different types of evidence of validity” (Wiersma & Jurs, 2005, p. 327), and that “there are multiple ways to establish the various forms of test validity” (Gay & Airasian, 2000, p. 169).

Construct validity assesses the underlying theory of the test. It is the extent to which the test can be shown to measure hypothetical constructs which explain some aspect of human behavior (Gay & Airasian, 2000, pp. 162-163; Wiersma & Jurs, 2005, pp. 328-329). For PIE, these are the hypothetical constructs of personality type as conceptualized by Jung. Evidence for construct validity can be both logical and empirical analyses (Wiersma

& Jurs, 2005, p. 329).

The process of establishing construct validity for Personality Identity Estimator used logical evidence. The constructs that were used for constructing the Personality Identity Estimator were from the Myers-Briggs Type Indicator.

[Since the construct validity of these items had already been established (Myers & McCaulley, 1985), their validity did not have to be re-established and, therefore, was inferred to the Personality Identity Estimator. Thus, the Personality Identity Estimator has construct validity as a result of using the Myers-Briggs Type Indicator to identify concepts for its items and for its development in the process of establishing content validity..

Content Validity

Content validity refers to the sampling adequacy of the content of the instrument (Wiersma & Jurs, 2005, p. 328). To establish content validity, it is necessary to "clearly identify and examine for completeness the bounds or the content area to be tested before constructing or selecting a test or measuring instrument" (Gay & Airasian, 2000, p. 163). For the Personality Identity Estimator, content validity is concerned with the degree to which the items are representative of the Jungian personality types as they are depicted in the Myers-Briggs Type Indicator (MBTI). The MBTI is a summated-rating scale in which the Jungian concept of personality types is groups people along four dimensions: Extraversion (E) and Introversion (I), Sensing (S) and iNtuition (N), Thinking (T) and Feeling (F), and Judging (J) and Perceiving (P). MBTI scoring is conducted by the individual completing the questionnaire and then being categorized into a personality type along each of the dimensions. This results in each respondent being given a 4-letter personality type such as ENTP with a letter from each dimension. In this process of

assigning a letter to each dimension, the continuous values from the scores from the individual dimensions are converted to a categorical score. Both the individual items in the MBTI and the categories created by summing these items into the dimensional classifications were used in the content validity analyses.

Discriminant Analysis

The concepts for the items for PIE were identified by the multivariate statistical procedure of discriminant analysis using the MBTI to represent the universe of concepts for Jungian personality type. Discriminant analysis is a powerful multivariate statistical procedure for examining the differences between groups using several discriminating variables simultaneously (Kachigan, 1991, p. 216; Klecka, 1980, p. 5). This procedure produces a structure matrix which shows the interactions within the analysis and which can be used for naming the process that separates the groups (Klecka, 1980, pp. 31-34). Consequently, the structure matrixes from the discriminant analyses were used to determine the concepts for the item for inclusion in PIE.

A series of discriminant analyses were conducted to simultaneously examine all items in each of the four dimensions to determine the differences between each of the two groups in that dimension. In these analyses, the individual MBTI items that make up the scales for the dimension were used as the discriminating variables, and the participants were grouped according to their personality type on each dimension. Discriminant analysis was used because this statistical procedure produces a structure matrix that shows the correlation between the individual discriminating variables and the overall discriminant function (Klecka, 1980, p. 31). The structure matrix contains the coefficients which show the similarity between each individual variable and the total discriminate function, which is

the equation that expresses the statistical relationship of the significant variables in the analysis and which is used for placing people in groups. The variables with the highest coefficients show how closely the variable and the overall discriminant function are related. Interpreting the structure matrix distinguishes the groups from each other (p. 31). Consequently, by using the groups for each dimension as the grouping variable and by using the individual MBTI items for that dimension as the set of discriminating variables, an analysis can be generated which produces a structure matrix which describes the *process* that separates the two groups in each dimension (Conti, 1996, p. 71). Therefore, a separate discriminant analysis was conducted for each dimension on the MBTI, and the nine items with the highest correlations from the structure matrix for each of these discriminant analyses were used as the constructs to determine the wording of the items in PIE.

Complete MBTI data were available on 553 participants. For each separate discriminant analyses, the participants were grouped on one of the personality type dimensions. There were 94 discriminating variables from the MBTI. These items were distributed as follows: Extraversion (E) and Introversion (I)–21, Sensing (S) and iNtuition (N)–26, Thinking (T) and Feeling (F)–23, and Judging (J) and Perceiving (P)–24. Each analysis was run using the Wilks' lambda method for selecting the variables for inclusion in the discriminant function. Wilks' lambda is a stepwise procedure (Klecka, 1980, p. 54) that systematically adds variables to the discriminant function through a series of steps in which variables that account for the most variance are added to the equation "continuing until the inclusion of another variable would account for only an insignificant amount of variance in the criterion variable" (Kachigan, 1991, p. 153).

The criterion used for judging the usefulness of the discriminant function produced by the analysis was that it had to be at least 75% accurate in correctly classifying the participants. Although 75% is more than 1.5 times greater than the chance placement rate of 50% for two groups, the judgement criterion was set at this level because any formula that could not correctly place at least three-fourths of the participants would have little value in identifying the key items for separating the groups.

For the first analysis, the 553 participants were grouped as Extraversion or Introversion. The Extraversion group consisted of 308, and the Introversion group contained 245 participants. The nine most relevant items in the structure matrix of the discriminant analysis for these two groups were items 1, 65, 21, 61, 69, 9, 57, 37, and 45. This process was 94% accurate in discriminating between the Extraversion group and the Introversion group. The constructs from these items that describes this process made up the Extraversion-Introversion scale in PIE.

For the second analysis, the 553 participants were grouped as Sensing or iNtuition. The Sensing group consisted of 359, and the iNtuition group contained 194 participants. The nine most relevant items in the structure matrix of the discriminant analysis for these two groups were items 10, 78, 6, 46, 54, 62, 22, 82, and 34. This process was 95% accurate in discriminating between the Sensing group and the iNtuition group. The constructs from these items that describes this process made up the Sensing-iNtuition scale in PIE.

For the third analysis, the 553 participants were grouped as Thinking or Feeling. The Thinking group consisted of 298, and the Feeling group contained 255 participants. The nine most relevant items in the structure matrix of the discriminant analysis for these two

groups were items 35, 43, 11, 39, 63, 3, 51, 31, and 71. This process was 94% accurate in discriminating between the Thinking group and the Feeling group. The constructs from these items that describes this process made up the Thinking-Feeling scale in PIE.

For the fourth analysis, the 553 participants were grouped as Judging or Perceiving. The Judging group consisted of 298, and the Perceiving group contained 255 participants. The nine most relevant items in the structure matrix of the discriminant analysis for these two groups were items 68, 4, 16, 72, 8, 28, 80, 52, and 32. This process was 95% accurate in discriminating between the Judging group and the Perceiving group. The constructs from these items that describes this process made up the Judging-Perceiving scale in PIE.

Thus, a series of discriminant analyses were used to produce a reduced set of items from the MBTI that could be used as the conceptual bases for forming the items of PIE. Nine items were selected from each analysis because these items correlated with the overall discriminating function, and this provided an odd number of items for each section of PIE. An odd number of items was desired because each dimension presents a binary choice for personality type, and having an odd number of items avoids a respondent having an equal preference toward each option in the dimension.

The discriminant analysis also provided additional information about each concept. An examination of the group mean for each item in the selection reveals the *degree* to which each group supported the concept in the item. That is, the mean for each group on the item indicated a range of support from extremely low to extremely high for the concept. Consequently, descriptors were used in each item of PIE such as almost always, always, extremely, fairly, moderately, nearly always, neutral, not especially, often, slightly, sometimes, somewhat, strongly, tend, and usually. As a result, the wording of the items in

PIE more precisely describes the general feeling of each personality type toward the concept in the item. Thus, as well as identifying the constructs for PIE, the discriminant analyses also provided additional information about each concept.

Factor Analysis

Factor analysis was used to provide an additional check on the content validity of the items in PIE. Factor analysis "is a way to take a large number of variables and group them into a smaller number of clusters called *factors*" (Gay, Mills, & Airasian, 2006, pp. 203-204). The factor represents the variables in it as their abstract underlying dimension (Kachigan, 1991, p. 237). Consequently, factor analysis can identify the basic constructs in an instrument.

An important task in factor analysis is determining how many factors best represent the data (Kachigan, 1991, p. 246). Principal components analysis is often used as a preliminary step to help in this process (p. 246). Principal components factor analysis initially extracts as many factors as there are variables in the analysis (p. 245). The degree to which each variable correlates with a factor is referred to as the factor loading (p. 243). Following the principal components factor analysis, the factors can be rotated (i.e., redefined) so that the loadings can "make sharper distinctions in the meaning of the factors" (p. 248). For this rotation, "the most commonly used method is the varimax method, which attempts to minimize the number of variables that have high loadings on a factor" (Norusis, 1988, p. B-54).

In order to check the validity of the items in the Myers-Briggs Type Indicator, the 94 items from the instrument were factor analyzed using a principal components analysis with a varimax rotation. In the initial principal components factor analysis, the scree plot, which

is a graph of the strength of the various factors, indicated the presence of four major factors just as predicted by the structure of the MBTI. Therefore, a second factor analysis was calculated using a principal components analysis with a varimax rotation in which the factors were limited to four. In this analysis, all of the variables loaded into four factors that explained 25.4% of the variance in the analysis. Each of the items loaded on the factors as predicted by the MBTI with the following accuracy: J-P (Factor 1)--83.3% (20/24), S-N (Factor 2)--65.4% (17/26), T-F (Factor 3)--91.3% (21/23), E-I (Factor 4)--100% (21/21), and Total--84% (79/94). Thus, factor anslysis confirmed that the item responses for the sample used for constructing PIE were nearly identical to the generally expected responses to the items.

Additional analyses were conducted to assess the similarity between the items identified for PIE by the discriminant analyses and the items with high loadings in the factor analyses. First, the nine items with the highest factor loadings in each of the four factors were compared to the nine items selected for each PIE scale from the discriminant analyses. For the Thinking-Feeling dimension, the highest loading nine items in the factor matched the items in the discriminant analysis perfectly; for the Extraversion-Introversion dimension almost all matched; for the Judging-Perceiving dimension two-thirds matched; only the Sensing-iNtuition dimension had a poor match. The exact matches were as follows: T-F--100% (9/9), E-I--88.9% (8/9), J-P--66.7% (6/9), and S-N--22.2% (2/9). Overall, there was a high degree of agreement in results between the factor analysis and discriminant analyses concerning the most important items in each dimension.

Additional analyses were conducted by examining each dimension separately. A separate factor analysis was conducted for each dimension with only the items that made

up the dimension and with the factors limited to one. The nine highest factor loadings for this one factor were compared to those of the discriminant analysis for each dimension. Once again, the factor loadings were congruent with the correlations in the structure matrix of the discriminant analyses. For the three dimensions of Extraversion-Introversion, Thinking-Feeling, and Judging-Perceiving, eight of the nine items were identical in each type of analysis; for the Sensing-iNtuition dimension, five of the nine items were the same. For the Extraversion-Introversion, the five highest loading items were the same in both analyses; for the Judging-Perceiving dimension, the three highest loading items were the same; for the Thinking-Feeling dimension, two of the three highest loading items were the same; and for the Sensing-iNtuition dimension, the two highest loading items were the same. Thus, both the factor analyses and the discriminant analyses had a high degree of consistency in the items selected from the total universe of the Myers-Briggs Type Indicator items for the conceptual basis for item construction for the Personality Identity Estimator.

While the results from the discriminant analyses and factor analyses identified similar items for the foundation of PIE, there is one consequential difference in the results pertaining to instrument construction. The factor loadings for the factor analysis only produce correlations, and these correlations reflect the degree to which the individual item in the analysis relates to the overall factor. While this correlation provides information about the strength of the relationship, it does not provide insights concerning the degree to which respondents support the item. In contrast, the items in the structure matrix of the discriminant analysis can be interpreted for the intensity to which each group supports the item by examining the means for each group on the item. This analysis reveals the extent to which each group supports the concept in the item, and this in turn allows for the new

item that is constructed to include descriptors to reflect this magnitude of support. Thus, by using the results from the discriminant analyses, the new items were constructed with a great deal of precision. This in turn increased the content validity of PIE by allowing it to more accurately represent the actual universe of possible items for the constructs being measured in it.

Correlation of Items

The analyses of the 94 items in the MBTI produced a set of 36 concepts to serve as the conceptual bases for forming the items of PIE. There were nine items in each of the four dimensions of personality type. In order to determine which of these concepts should be included in PIE, odd-numbered scales of 9, 7, 5, and 3 items for each of the 4 dimensions were constructed and analyzed to determine the strongest concepts to be included in the final PIE instrument. Each scale contained the items with the highest correlations in the structure matrix from the discriminant analyses. An odd number of items was selected to eliminate the possibility of a tie in a respondent's number of preferences for any personality dimension. Each of these scales was a summated scale in which the total number of preferences for each option in a dimension was added together to produce a total score for that aspect of the dimension. To access the strength of the items, the items were analyzed by examining the correlation between individual responses to items for each scale and the total score for that scale. This procedure was used because each item is part of the overall concept, and the item must contribute to the total score in order for the it to be useful. In order to do this, the item must have a moderate to strong positive correlation with the overall scale.

The standard that was used for judging if an item should be retained for the final

form of PIE was that at item should explain at least its own weight in variance to the total score. That is, for each dimensional scale each individual items should explain at least the following amount of variance in the total scale: 9 items--11%; 7 items--14%, 5 items--20%, and 3 items--33%. This was measured by Pearson correlations which indicates the proportion of variation in one variable that is explained by the other, and therefore the correlation coefficient indicates how closely the variables are related. Squaring the correlation coefficient for each item indicated the proportion of variation in the total score explained by the item. Consequently, the required correlation coefficients required for each item for each scale were as follows: 9 items (.33); 7 items (.38), 5 items (.45), and 3 items (.58).

Each item produced a correlation coefficient for each of the two options of the dimension; thus, the total number of correlations for each scale were as follows: 9 items–72, 7 items–56, 5 items–40, and 3 items–24. The correlation coefficients for all items on all four scales were above the minimum required for retention in the scale. The correlation coefficients for each scale were as follows: 9 items–15 (.7), 34 (.6), 20 (.5), and 3 (.4); 7 items–3 (.8), 17 (.7), 22 (.6), 12 (.5), and 2 (.4); 5 items–6 (.8), 19 (.7), 7 (.6), and 8 (.5); and 3 items–15 (.8) and 9 (.7). While all of the items for each scale were strong enough to meet the minimum standard, three of the items in the 9-item scale and 2 in the 7-item scale barely met the standard. Likewise, eight of the items in the 5-item scale were at the minimum. In sharp contrast, all of the items in the 3-item scale were markedly above the minimum with nearly two-thirds of the items having a coefficient of .8. Because of the strength of these correlation coefficients, the format of three items for each of the four dimensions of personality type was judged as the best for the final version of PIE.

Therefore, the Personality Identity Estimator consists of 12 items with 3-paired items for each of the 4 personality type dimensions of Extraversion-Introversion, Sensing-iNtuition, Thinking-Feeling, and Judging-Perceiving (see Figure 1).

Figure 1: Personality Identity Estimator Items

Extraversion-Introversion Scale	
1a. I am almost always sociable	1b. I am usually restrained
2a. I am inclined to talk a great deal	2b. I am usually quiet
3a. I can usually talk freely with others	3b. I tend to talk mostly in intimate situations
Sensing-iNtuition Scale	
4a. I want to be viewed as very down-to-earth	4b. I want to be viewed as very resourceful
5a. The word “actual” often appeals to me	5b. The word “theoretical” usually appeals to me
6a. I prefer to be with practical people	6b. I generally get along best with creative people
Thinking-Feeling Scale	
7a. Reasoning moderately appeals to me	7b. Being responsive to others strongly appeals to me
8a. Analyzing things moderately appeals to me	8b. Being compassionate strongly appeals to me
9a. Sound judgment is more important than enthusiasm	9b. Enthusiasm is much more important than sound judgment
Judging-Perceiving Scale	
10a. Planning is extremely appealing to me	10b. I like to improvise
11a. I tend to plan the timing and activities for a trip	11b. When going someplace, I prefer to be flexible
12a. I almost always like to plan things	12b. Following a schedule usually restricts me

Scoring PIE

PIE is easy to score. The score or results for PIE consists of four letters with each of the letters representing one of the four dimensions for the Jungian concept of personality types. Consequently, the four letters in the identified personality type are E (Extraversion) or I (Introversion), S (Sensing) or N (iNtuition), T (Thinking) or F (Feeling), and J (Judging) or P (Perceiving). This combination of four letters indicates the personality type identified by PIE.

In PIE each dimension has three items. The highest number of options selected for each of the binary choices in the dimension indicates the respondent’s primary preference in that dimension. Thus, a score of 2 or 3 indicates a preference in that dimension, and a

respondent's overall personality label is based on the person's dominate preference for each of the four dimensions.

When analyzing the scores in each dimension, it is important to keep in mind the logic used in discriminant analysis for judging the accuracy of a discriminant function. Here, the general rule is to ask how much improvement the function is over chance. When there are two groups, there is a 50% chance that a person will be placed in either group simply by random chance; therefore, the usefulness of the function is determined by how much an improvement its placement is over this 50% chance. Since there are three items in each dimension in PIE, selecting two of the three choices for a preference in this dimension indicates a two-thirds or 66.6% preference; this is a 16.6% improvement over chance placement. Selecting all three options for the dimension indicates a 100% preference which is a 50% improvement over chance placement. Thus, while there is only a small number of items for each dimension, each option selected indicates a large improvement over chance placement. As a result, not only can PIE be scored quickly, but also its results are accurate.

Summary

Content validity addresses the sampling adequacy of an instrument, and for the Personality Identity Estimator, content validity also dealt with how its items would be constructed. Several statistical procedures were employed to establish the content validity of the Personality Identity Estimator and thereby to assure this sampling adequacy and to provide the constructs for wording of each item. Since the universe for Jungian personality types was defined as that included in the Myers-Briggs Type Indicator, discriminant analysis and factor analysis were used to identify the constructs. Discriminant analyses

were conducted in each of the four personality type dimensions to isolate these constructs, and the results of these discriminant analyses were confirmed by factor analyses. The results from these analyses produced a pool of 36 constructs with 9 constructs in each of the 4 personality type dimensions. Scales were constructed with 9, 7, 5, and 3 items in each of the 4 dimensions and with each scale containing the items with the highest correlations from the discriminant analyses structure matrixes. Each scale was analyzed to determine the degree to which each item contributed to the total score of the scale. While all items met at least the minimum standard for inclusion in the scale, the scale which contained 12 items with 3 items for each of the 4 personality type dimensions was judged as the best format for the final version of the Personality Identity Estimator (see Figure 1).

Criterion-Related Validity

Criterion-related validity compares an instruments scores with an external relevant criterion variable (Huck, 2004, p. 90). Often establishing criterion-related validity for most instruments is usually the direct procedure of comparing the new instrument to an established concurrent measure such as an instrument or behavior (Wiersma & Jurs, 2005, p. 328). Criterion-related validity for the Personality Identity Estimator was established by comparing responses on it to those on the Myers-Briggs Type Indicator.

Both PIE and the MBTI were completed by 178 respondents. Participants responded to both instruments. Responses on PIE were scored and compared to the dimension preferences on the MBTI. The group, which was 64% female and 36% male, had an average age of 20.9 years. Its racial makeup was as follows: African American–42.4%, White–34.7%, Hispanic–12.9%, Asian–2.9%, Native American–2.4%,

and Other–4.7%. The educational level of the respondents varied as follows: High school level–87%, bachelors degree–3%, and graduate degree–10%.

The chi-square test of independence was used to assess the association between the responses on the Personality Identity Estimator and the Myers-Briggs Type Indicator. Chi-square assesses the statistical independence or association between two or more categorical variables by comparing how the pattern of observed frequencies differs from the pattern of expected frequencies. Both the dimensional results for PIE and MBTI are categorical; therefore, chi square allowed testing for the relationship between PIE and the external criterion of the MBTI.

Differences between the responses on PIE and the MBTI were compared using the chi-square test for independence. A separate chi square was calculated for each of the four dimensions: Extraversion-Introversion (E-I), Sensing-iNtuition (S-N), Thinking-Feeling (T-F), and Judging-Perceiving (J-P). Based upon a significance level of α=0.05, significant differences were found for each of the four dimensions: E-I (χ^2 = 76.9, *df* = 1, *p* = .001) , S-N (χ^2 = 22, *df* = 1, *p* = .001) , T-F (χ^2 = 32, *df* = 1, *p* = .001), and J-P (χ^2 = 49.1, *df* = 1, *p* = .001). These differences confirm that PIE and the MBTI are associated with (i.e., dependent upon) each other and thereby confirm the criterion-validity of PIE.

The strength of the association between PIE and the MBTI can be further verified by examining the residuals from the analyses. “A residual is the difference between the observed and expected values for a cell. The larger the residual, the greater the contribution of the cell to the magnitude of the resulting chi-square value” (Sharpe, 2015, p. 2). The standardized residuals are a measure of the strength of the difference between

observed and expected values, and they measure how significant the cells are to the chi-square value. As standardized values, any residual greater than 2 is outside of 2 standard deviations from the mean. Consequently, standardized residuals as high as 3 or 4 indicate that extreme differences exist. The standardized residuals for the chi-square analyses were all extreme: E-I (8.8), S-N (4.7); T-F (5.7), and J-P (7.1). Thus, the association between PIE and its criterion-related MBTI is extremely strong. Consequently, the significant chi-square values supported by extremely high standardized residuals establish the criterion-related validity of the Personality Identity Estimator.

Reliability

"Reliability is the degree to which a test consistently measures whatever it is measuring" (Gay & Airasian, 2000, p. 169). The reliability of PIE was established by the test-retest method which addresses "the degree to which scores on the same test are consistent over time" (p. 171). PIE was administered to a group of 89 with a 1-week interval. The group was 59.8% female and 40.2% male with an average age of 16.1 years. Its racial makeup was as follows: African American–51.7%, White–25.8%, Hispanic–11.3%, Asian–2.2%, Native American–1.1%, and Other–7.9%. All respondents were at the high school educational level.

The test-retest method has been used to examine the reliability of the Myers-Briggs Type Indicator. Three studies that examined test-retest reliability and met the criteria for inclusion in a meta-analysis all reported Pearson product-moment correlation coefficients (Randall, Isaacson, & Ciro, 2017, p. 15). Based upon this analysis, it can be “cautiously conclude[d] that the MBTI performs reliably over time” (p. 17).

Although reliability information is often reported as a correlation coefficient and

although past MBTI reliability studies have used the instrument's continuous scores to calculate reliability coefficients, chi square was used to establish the reliability for the Personality Identity Estimator because its scores are categorical for each of the four dimensions. The chi-square test of independence was used to assess the association between the first responses on the Personality Identity Estimator and the re-test responses. A separate chi square was calculated for each of the four dimensions to compare the responses for each testing: Extraversion-Introversion (E-I), Sensing-iNtuition (S-N), Thinking-Feeling (T-F), and Judging-Perceiving (J-P). Based upon a significance level of α=0.05, significant differences were found for each of the four dimensions: E-I: (χ^2 = 28.7, *df* = 1, *p* = .001) , S-N (χ^2 = 16, *df* = 1, *p* = .001) , T-F (χ^2 = 23.4, *df* = 1, *p* = .001), and J-P (χ^2 = 8.4, *df* = 1, *p* = .004). All of the standardized residuals for the chi-square analyses indicated a strong significant association: E-I (5.4), S-N (4); T-F (4.8), and J-P (2.9). Thus, the chi-square values and supporting standardized residuals establish the reliability of the Personality Identity Estimator.

References

Gay, L. R., & Airasian, P. (2000). *Educational research* (6th Ed.). Columbus, OH: Merrill.

Gay, L. R., Mills, G., & Airasian, P. (2006). *Educational research: Competencies for analysis and application* (7th Ed.). Columbus, OH: Merrill.

Huck, S. W. (2004). *Reading statistics and research* (4th Ed.). Boston: Allyn & Bacon.

Kachigan, S. K. (1991). *Multivariate statistical analysis*. New York: Radius Press.

Kerlinger, F. N. (1973). *Foundations of behavioral research*. New York: Holt, Rinehart, & Winston.

Klecka, W. A. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage Publications.

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

Norusis, M. J. (1988). *SPSS/PC+ advanced statistics V2.0.* Chicago: SPSS Inc.

Randall, K., Isaacson, M., & Ciro, C. (Spring, 2017). Validity and reliability of the Myers-Briggs Personality Type Indicator: A systematic review and meta-analysis. *Journal of Best Practices in Health Professions Diversity*, 10(1), 1-27.

Sharpe, D. (2015). Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation, 20*(Article 8), 1-6.

Wiersma, W., & Jurs, S. G. (2005). *Research methods in education* (8th Ed.). Boston: Allyn & Bacon.