

DISC Instrument Validation Study

TECHNICAL REPORT

Dr. Larry A. Price

Texas State University

Statistics and Psychometrics



Table of Contents

Section 1	Study Overview
Section 2	Psychometric Development of the Institute for Motivational Living's DISC Instrument
Section 3	Sample Composition and Demographics
Section 4	Evidence of Internal Structure: Score and Scale Reliability
Section 5	Evidence of Validity: Logical, Content, Construct
Section 6	Special Problems Associated with Personality Assessment Inventories
Section 7	Conclusions and recommendations for future studies

Section I: Study Overview

Study Overview

The purpose of this study was to examine the psychometric properties of the *Institute for Motivational Living (IML)* DISC Personality System Analysis survey instrument within the context of the workplace environment. Ideally, the findings reported herein will serve to provide the IML with deeper understanding about the psychometric properties of the DISC thereby enhancing the organization's precision when using the DISC in the workplace environment.

Personality is a vague construct that is often used to explain behavioral consistency within persons and behavioral distinctiveness between persons. In fact, there are many definitions of personality in the psychology literature with some focusing on particular aspects of an individual (Goldstein, 1963a) while others view the individual within the context of society or particular social setting (Sullivan, 1953). Historically personality assessment has focused on the measurement of interpersonal emotions, attitudes, and personality styles and /or motivational attributes rather than cognitive abilities such as intelligence or achievement. The earliest personality theory attempted to sort individuals into discrete categories or types. For example, the early Greek physician Hippocrates proposed a humoral theory with four personality types (sanguine, choleric, melancholic, and phlegmatic). In consideration of the historical research contributions pertaining to the notion of personality (e.g., Traits, Types, and States) and personality assessment, a contemporary definition that is offered by Cohen & Swerdlik (2005) and is a flexible yet parsimonious definition of *personality* - "an individual's unique constellation of psychological traits and states" (p. 336). Further, Cohen & Swerdlik (2005) define *personality assessment* as "the measurement and evaluation of psychological traits, states, values, interests,, attitudes, worldview, acculturation, personal identity, sense of humor, cognitive behavior styles, and/or related individual characteristics" (p. 336). Cohen and Swerdlik's definition of personality merges with previous research to offer the notion that personality is multidimensional in nature. With this multidimensional definition in mind, this study will use both classical psychometric analytic methods as well as newer approaches that are well-suited for the study of multidimensional constructs.

About the Institute for Motivational Living | PeopleKeys®

The Institute for Motivational Living | PeopleKeys DISC Insights Personality System Analysis is classified as a self-report inventory and is administered either in a paper-pencil or computer aided format. The Institute for Motivational Living is a training and publishing company designed to help people communicate better and work together more effectively. "The Institute" trains and certifies individuals in product usage with their behavior analysis course, Introduction to Behavioral Analysis. This training course provides the expertise to consultants, entrepreneurs, business managers, pastors and counselors in the use of the DISC Personality System and other behavioral assessment profiles for use in team building, career planning, hiring, conflict resolutions, family counseling, personal counseling, marriage counseling and executive coaching. The Institute's online PeopleKeys system, designed for administering DISC in the workplace, was used within this study to collect data as a part of the research.

Steps in item development in relation to logical and content validity

The process of determining and deriving the operational definitions of the item content was based primarily on the original theoretical work of William Marston, M. D. as published in *Emotions of Normal People* (1928). Based on Marston's theory, the items on the IML DISC were developed to tap four latent dimensions of personality as espoused by Marston. Additionally, item content and wording was refined through an iterative procedure so as to allow for the maximum potential for actually tapping a particular latent construct or dimension. Item quantity (e.g., the number of items comprising the scale) and content redundancy are important issues to be considered during the development for reasons of ensuring an adequate level of scale reliability and validity. In order to ensure that the latent construct is being measured in a variety of ways, the instrument development team balanced the number of items on the final scale as well as the amount and degree of redundancy the items exhibited. Although some of the items on the DISC appear to be redundant, the inclusion of the items was deliberate and based on content experts at IML in order to assure adequate assessment of the latent dimensions or constructs according to Marston's theoretical framework. In summary, during the item development and refinement process, the IML development team displayed evidence of adhering to the Standards for Educational and Psychological Testing published by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME, 1999).

Brief Review of Research of the Forced-Choice Item Format

The forced-choice rating technique is used on the IML DISC where the subject responds to four sets of words or phrases and is required to select one of the four sets that MOST describes his/her behavior in the work environment and one of the four sets that LEAST describes his/her behavior in the work environment. Highland & Berkshire (1951) conducted an extensive study of forced-choice rating forms in connection with rating instructors. Highland-Berkshire investigated six different configurations of forced-choice item formats in an effort to determine which format yielded the highest level of internal consistency reliability; validity against a criterion of rank ordering of the instructors; susceptibility of the scores to biasing (with raters told to assure ratees with high scores); and popularity of the rating form from the point of view of the raters. Highland & Berkshire's findings include the following points:

- 1) The highest reliability coefficients were attained when the item formats were composed of four words or statements, two being favorable and two unfavorable; the rater selecting the most and the least descriptive. This finding regarding item format provides psychometric support for the approach used in the IML DISC.
- 2) The highest criterion validity evidence was observed when the item formats were composed of four words or statements, all being favorable; the rater selecting the most descriptive and the least descriptive. This finding regarding item format provides psychometric support for the approach used in the IML DISC.
- 3) Finally, the item formats that were composed of four words or statements, all being favorable; the rater selecting the most descriptive and the least descriptive revealed the smallest evidence of response bias. Overall, this item format, the same as used in the IML DISC instrument, was regarded as displaying "optimum" psychometric characteristics.

Important Considerations for the Development and use of Forced-Choice Ipsative Scales

Criticisms of the forced-choice item based instrument include a substantial amount of time and energy to construct this type of scale, and that fact that this type of scale is constructed in a way so as to display evidence of construct validity for a particular population or group of interest. In consideration of this issue, the sample was purposefully selected to represent personality style in the workplace environment. A second issue that is important for developers and users of the forced-choice type of scale is the fact that the type of measurement is “ipsative”, meaning that the subject’s responses and the presumed strength of a measured trait are interpreted relative to the measured strength of other traits of that same subject, or alternatively, the measurement is “*within a person*”. Users of this type of scale must always be mindful of the underlying theoretical nature of the ipsative measurement and scoring approach so as to use the scores or classification information in an appropriate manner.

Section III: Sample Composition and Demographics

This primary goal of this study was to investigate the psychometric aspects of the IML DISC within the context of the workplace environment. Therefore, a purposive sampling strategy was employed for data collection process. Respondents were employed in the social service and counseling field, and ranged in job position from entry level administrative to National Executive Director.

The sample obtained was cross-sectional representing seven different regions of the country in 26 separate locations. Specifically, represented in the sample are the northeast, Midwest, South, and Southwest regions of the United States. The ethnic composition of the sample was 76% Caucasian, 12% African-American, 11% Hispanic, and 1% Other (Asian/Pacific Rim). Eighty-seven percent (87%) of the subjects were female, 13% male with an age range of between 21 – 85 years. All of the subjects had attained a minimum of 12 years of formal education with most having 16 years. Three hundred thirty one (331) survey instruments were completed and returned for analysis.

Section IV: Evidence of Internal Structure – Score and Scale Reliability

Within the context of psychological measurement the term “reliability” broadly refers to the accuracy, consistency, and stability of scores obtained on a measurement instrument. This consistency, accuracy, or stability does not represent a qualitative description of how “good” or “bad” the scores obtained by the scale or instrument are in terms of the construct that it purports to measure, rather only the degree to which the scale produces consistent and/or stable scores. An important aspect also related to the estimation of reliability is known as “measurement error”. Plainly stated, no psychological scale or instrument produces “error free” measurement. Therefore, according to classical true score theory (Allen & Yen, 1979; Nunnally & Bernstein, 1978), score reliability estimates are represented by two components: 1) the examinee’s true score on the underlying psychological construct; 2) errors of measurement that have nothing to do with measuring the examinee’s underlying trait. The error or errors of measurement is/are mathematically related to the reliability coefficient and represent the inaccuracy of measurement on a particular testing occasion or over time as in a test-retest situation.

The methods of estimation of score reliability in the present study are based on classical measurement theory also known as the classical true score model (Allen & Yen, 1979; Nunnally & Bernstein, 1978). There are several procedures for estimating the reliability of scores within the context of the classical true score model. These include internal consistency, stability, alternate forms, parallel forms, inter-scorer/interrater agreement, and decision consistency. The decision regarding which method of reliability estimation to use is primarily driven by the nature of the scale or instrument under investigation and the specific purpose for which the instrument is being used. In the present study, evidence of reliability is presented based on indices of internal structure and decision consistency. Finally, a rigorous structural equation modeling approach was used to examine the congruence between the estimated reliability of the IML DISC subscales and its construct validity.

Evidence of Internal Consistency

Internal consistency estimates of reliability represent the homogeneity of items that comprise the instrument or scale. Ideally, the individual items on a scale will be moderately to highly correlated among themselves thereby lending support to the notion that the items are likely tapping the same underlying psychological construct or share the same cause. Although the underlying construct is unobservable or “latent” and therefore not directly measurable, the items are measurable and subsequently the intercorrelations among them can be assessed. Within the context of the present study, the underlying factor structure of the DISC is theoretically represented by four factors or latent constructs first proposed by Marston (1928). These four factors (e.g., dimensions) are: 1) Dominance, 2) Influence, 3) Steadiness (submission originally), and 4) Compliance. Therefore, the latent factor structure underlying the 24-item DISC instrument is multidimensional consisting of certain items within the total twenty-four that reflect a particular piece of the multidimensional structure of the DISC theory.

Given the multidimensional nature of the IML DISC, internal consistency estimates of reliability and associated standard errors of measurement are reported for each factor or dimension and also for the twenty-four item scale. Table 1 below provides the intercorrelations among each of the DISC subscales. Table 2 provides the estimates of internal consistency reliability (coefficient alpha) and associated standard errors of measurement derived from a random subsample of 200 subjects from the total sample (N=331).

Inspection of correlation coefficients allows one to compare the strength and direction of associations between different pairs of variables. The results provided in Table 1. illustrate that the patterns of associations among the four subscales covary in magnitude and direction as expected according to Marston’s original

theory. Interpretation of the practical effect of a correlation coefficient include the following guidelines: 1) coefficients below .35 (in absolute value) are considered as reflecting a “low” degree of association; 2) coefficients between .36 and .65 (in absolute value) are considered as reflecting a “moderate” degree of association; 3) coefficients above .65 (in absolute value) are considered as reflecting a “high” degree of association. Finally, whether the coefficient is preceded by a “+” or “-” sign, dictates whether the observed relationship between the two variables increases in a positive manner (e.g., as a value on one variable increases, so does the value on the other variable being examined), or is “inverse” (e.g., as a value on one variable increases, the value on the other variable being examined decreases).

With the aforementioned guidelines in mind, if the dominance dimension is considered as the “baseline” dimension, then the following descriptions are offered based on the results observed herein:

- 1) Dominance exhibits a “low, positive” relationship with the Influence dimension indicating intensity of extroversion.
- 2) Dominance exhibits a “moderate, negative” relationship with the Compliance dimension in terms of task-oriented style.
- 3) Dominance exhibits a “strong, negative” relationship with the Steadiness dimension indicating intensity of extroversion.

Table 1. Spearman correlation coefficients for four DISC subscales (N = 331)

Subscale	1	2	3	4
1. Dominance	--			
2. Influence	.10	--		
3. Steadiness	-.80**	-.29**	--	
4. Compliance	-.35**	-.69**	.23**	--

Note. Spearman correlation coefficients are derived based on standard scoring procedure.

**Significant at $p < .01$

Table 2. Subscale Reliability (N = 200)

Subscale	<i>r</i>	SEM
Dominance	.84	3.16
Influence	.70	4.29
Steadiness	.73	4.00
Compliance	.84	1.89
Overall	.87	7.39

r = coefficient alpha, SEM = standard error of measurement

Note. Sample is composed of 200 subjects randomly selected from the total (N = 331) sample.

Evidence of Agreement between DISC Theory and Chance

Another form of reliability evidence presented herein is percent agreement between the responses obtained from the examinees and the classification of personality style according to the underlying DISC theory. The percent agreement between the examinee's responses within a "primary" style or a close secondary "primary" style according to DISC theoretical expectations was high (91%). In order to further examine the proportion of agreement between DISC theory and the observed examinee responses, the McNemar nonparametric test for two related (matched) dichotomous outcome variables was used. McNemar's test is particularly useful in detecting changes in patterns of examinee responses in "before-and-after" or "related subjects" designs. The results of the McNemar were observed to be statistically significant at the $p < .05$ level of significance. Interpretation of this finding reveals that the percent agreement was not only "good", but may also be stated to be "statistically significant". Finally, the Kappa coefficient, an agreement statistic that statistically adjusts for capitalization due to chance was also calculated. In this study, coefficient Kappa was observed as $r = .40$. This level of congruence between the theory and actual observed responses is classified as "good" agreement beyond the level of chance (Landis & Koch, 1977a).

Section V: Evidence of Validity

Establishing evidence for the validity of a particular psychological or educational instrument, scale or test is likely to be the most important aspect of instrument development. According to the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (1999), the term “validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests”, (p. 9). Moreover, in psychological measurement and assessment, the term “validity” represents a judgment or expression of how well an instrument or test measures what it purports to measure. It is a judgment based on evidence about the appropriateness of inferences drawn from test scores. Therefore, validation, involves the process of developing a sound, scientifically-based argument that will support the intended use of the scores or classifications derived from using the instrument. The process of test or instrument validation involves conducting investigations and accumulating evidence to provide a sound framework for the proposed score or classification interpretations. The score interpretations are theoretically linked to the theoretical underpinnings or construct that the instrument purports to be measuring. In the present study, the theoretical framework that the IML DISC instrument incorporates was first proposed by William Marston in his text “Emotions of Normal People” published in 1928. Within the context of the present study, the underlying factor structure of the DISC is theoretically represented by four factors or latent constructs first proposed by William Marston (1928). These four factors (dimensions) are: 1) Dominance, 2) Influence, 3) Steadiness (submission originally), and 4) Compliance. Therefore, the latent factor structure underlying the DISC instrument is multidimensional and consists of certain items within the total twenty-four items that reflect a particular piece of the multidimensional structure of the DISC theory.

Finally, the validity of an intended interpretation of test scores or classification outcomes relies collectively on adequate score reliability, appropriate test administration, accurate scoring and scaling, and cultural fairness to examinees.

Evidence Based on Instrument Item Content – Content Validity

Content validity evidence refers specifically to themes, wording with current cultural issues in mind, item formats, tasks, and guidelines for administration and scoring. In the present study, evidence for appropriate content validity was established by qualified experts at the Institute for Motivational Living by compiling a series of statements and traits that are universally accepted as the basis for the four DISC styles. This exercise was based on Marston’s original theoretical work and descriptors as published in *Emotions of Normal People* (1928). Furthermore, the composition of the descriptors and statements for the four styles was also compared with other leading DISC publications of modern times to establish universal evidence of agreement for the “core” observable behaviors or traits. This exercise also served to establish evidence of “face” or “logical” validity for the IML DISC instrument.

To extend the content validation process, a group of 12 certified behavior analysts were asked to rate themselves using this survey instrument as well as have their colleague review their choices to ensure congruence between how the analysts rated themselves using the DISC related to how others observed their behavior. Finally, since this validation study was conducted within the context of a workplace setting or environment, the Bair Foundation National Office began the study by taking 25 people from the National Office and had them take the 24 item DISC. This step served as baseline or pilot information for the study as well as provided an orientation to the instrument to the Bair Foundation who provided access to the overall sample.

Evidence Based on Internal Structure – Construct Validity

The primary purpose for investigating the internal structure of an instrument is to gain evidence that the relationships among items conform to the theoretical construct on which proposed test score interpretations are based. As previously discussed, the underlying structure of the DISC is multidimensional in nature and is composed of four subscales each representing a separate but correlated dimension of personality style. Therefore, an important goal of the present study was to rigorously investigate the factor structure of the IML DISC instrument in relation to Martson's theoretical framework using two specific types of factor analytic approaches: 1) the *Q – Technique*, and 2) the *R – Technique* within a structural equation modeling framework.

The Q – Technique factor analysis method is particularly appropriate for the study of classifying individuals based on preferences or systems of typology – such as in personality assessment. Typically, Q – technique studies employ a small number of subjects because the issue of interest is the notion of “intra-individual differences” (within subjects) rather than “inter-individual” (between subjects) differences. The inherent nature of the forced-choice measurement approach make traditional factor analytic techniques (*R – technique*) often inappropriate due to the “correlated” nature of the item formats and also because of the distributional shapes of the data obtained from respondents. The Q – Technique analyses proceeded as follows. First, ten separate random samples of size 20 were selected for analysis using Q – Technique factor analyses. Next, each random sample was factor analyzed using both Principle Components Analysis (PCA) with Varimax rotation and then using Principle Axis Factor Analysis (PAF) with Promax Oblique rotation. In all 10 analyses, the results provided evidence of a four factor dimensional structure underlying the IML DISC instrument. Across all analyses, an average of 70% of the total variance was explained by the four factor DISC model. As a general rule, the guideline for the percentage of variance accounted for in a factor analytic model to be evaluated as “acceptable” in the psychometric literature is 70%.

The inherent nature of the forced-choice measurement approach make traditional factor analytic techniques (*R – technique*) inappropriate due to the “correlated” nature of the item formats and also because of mathematical scaling problems. To account for this artifact, a modified scaling approach to the items was employed by the author thereby allowing for the ***exploratory*** (preliminary) investigation of the theoretical factor structure using structural equation modeling. A detail examination of the factor loadings and dimension correlation coefficients is instructive. Specifically of interest is how several items load on more than a single item, indicating clear evidence of item overlap and multidimensionality of the scale. Since the IML DISC (or any DISC) instrument has not previously been examined in this way, this is new information regarding the construct validity evidence of this scale.

Figure 1. The IML DISC multidimensional four-factor oblique structural equation model.

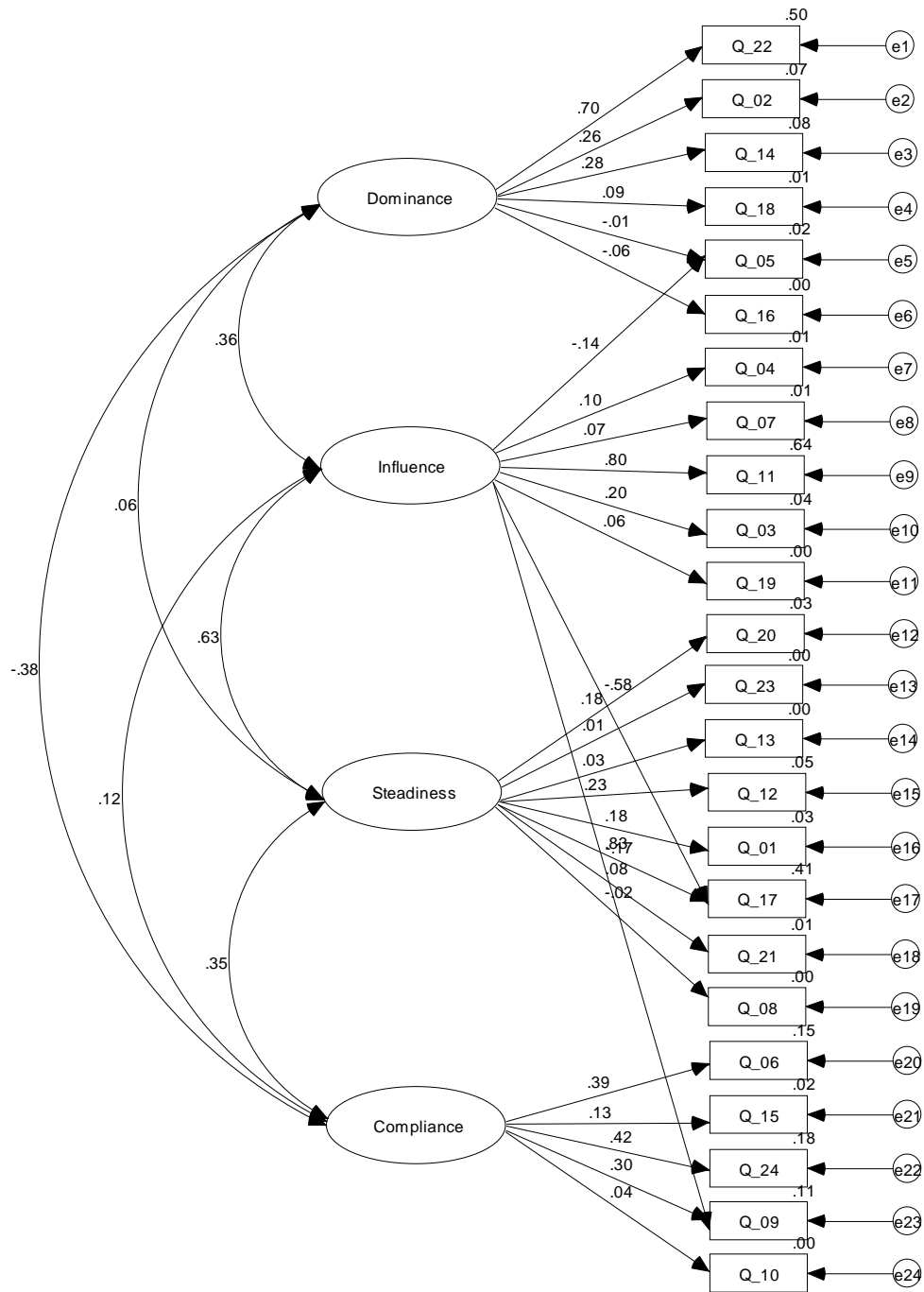


Figure 1 Notes. Numbers of the one-way arrows moving from the ovals (latent constructs) to the rectangles (items on the DISC) are standardized regression weights (correlation coefficients). Numbers on the curved two-headed arrows are correlation coefficients between latent constructs. Numbers at the upper right corner of each item represent the squared multiple correlation for each item in the model respectively. Small circles on the far right represent measurement error although the actual errors of measurement are not represented in this diagram.

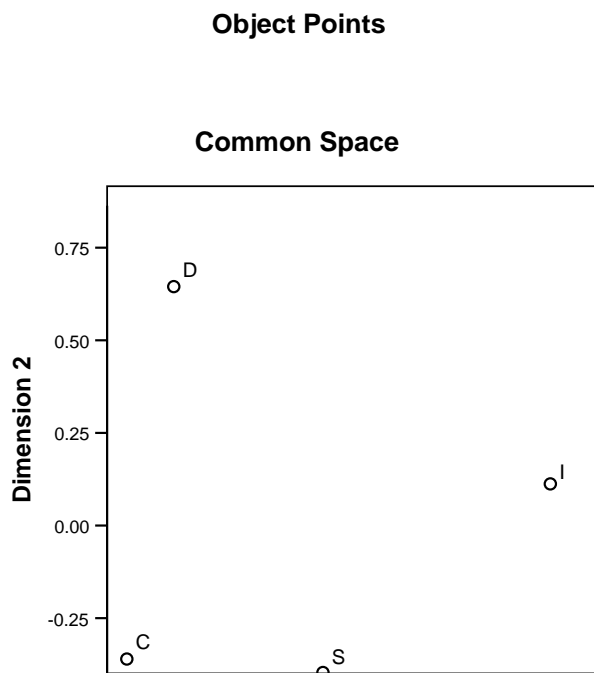
Model Estimation. Data on the four-factor structure of the DISC instrument were assessed using confirmatory factor analysis, conducted with the Analysis of Moment Structures program (AMOS; Arbuckle & Wothke, 1999). After screening the data for multivariate normality and finding that this assumption was tenable, the method of maximum likelihood was used to derive parameter estimates with all subsequent analyses performed on the covariance matrices. In order for models to be identified, scales on each latent variable were established at unity.

Assessment of Measurement Models. Several fit indices were used as indicators for the goodness-of-fit of the DISC structural equation model. Fit indices used were: (a) the overall chi-square statistic; (b) the root mean square residual (RMR), (c) the Akaike Information Criterion (AIC), and (d) the root mean square error of approximation (RMSEA; Steiger, 1990). The GFI and TLI have values ranging from 0 to 1 with values above .90 indicating a good fit of the empirical data to the implied model. The RMSEA provides values that represent the goodness-of-fit of the model if it were estimated in the population. RMSEA values between .05 and .08 are viewed as acceptable with values closer to zero indicating a more closely approximated model fit in relation to the population. The TLI typically has values between 0 and 1, however TLI indices are not limited to that range. For the four-factor DISC model evaluated herein, the following fit statistics were observed: $\chi^2(244, n=200)=272.65, p > .05, GFI = .90, AIC = 348.67, RMR = .08, RMSEA = .02.$

Multidimensional Scaling

Multidimensional scaling (MDS) is a multivariate statistical technique that helps researchers to identify key dimensions underlying examinee's responses. To accomplish this, MDS offers a scaling approach that allows for inferences about the underlying dimensions (in geometric space such as was posited in Marston's original work) from a series of similarity or preference judgments or statements. The results of MDS provide a perceptual map representing respondent's perceived personality style.

Figure 2. below provides the MDS perceptual map obtained with this study. The perceptual map demonstrates that the four DISC dimensions relate to one another as posited in Marston's original theory.



Section VI: Special Problems Associated with Personality Assessment Inventories

Three common problems associated with personality assessment (a noncognitive form of assessment) as compared to the assessment of ability or achievement (cognitive) include: 1) a greater propensity toward malingering, 2) submitting false (untrue or inaccurate) responses, and 3) situational specificity. Also, the behavior being measured fluctuates more frequently than does ability or achievement. This potential for frequent or sporadic fluctuation of the behavior or personality type increases the need for the test developer to frequently assess the various forms of score reliability such as stability and internal consistency (Coefficient Alpha or Split-Half).

Malingering and Submission of False Responses. The standard DISC administration procedure that the IML uses incorporates a timed approach thereby reducing the propensity for examinees to malingering and/or submit false responses on individual items. The IML DISC instrument also makes use of “false” positive responses that are not included as part of the final individual tally scores, thus reducing the propensity for examinees to intentionally skew results.

Situational Specificity. The frequent fluctuation of behavior related to personality type dictates that when the DISC instrument is administered and scores or classifications are obtained, the interpretation of the scores or classifications must be interpreted with the context of the environment and sample. This is essential in order for the interpretation of the scores or classifications to be psychometrically valid and reliable. The IML DISC instrument and related training materials make clear reference and instruction as to this consideration, both in the administration and interpretation of the results.

Section VII: Conclusions

This study was designed to investigate the psychometric properties of the IML DISC instrument. The study incorporated a multiple analytic approach in demonstrating the multidimensional factorial validity and reliability of the IML DISC instrument. Results from the present study indicate that the psychometric properties of the IML DISC are internally consistent and reliable. Classical true score model estimates of internal consistency reliability were found to be consistent with previous reliability estimates for the IML DISC instrument. Furthermore, the development process that the IML organization adhered to for the item selection and refinement are closely aligned with the Standards for Educational and Psychological Testing (1999) extending support for the collective validation process. Results of the application of structural equation modeling methodology revealed theoretical support for a four-factor oblique multidimensional model accurately explains the broad construct of personality type within the context of Marston's original theory. This support is evidenced through an extremely close approximation of the hypothesized IML DISC instrument factor structure to the empirical data.

Collectively, the results of the present study provide evidence that the IML DISC provides a robust measure of personality style. The reliability and validity evidence indicates that the IML may confidently use the DISC in their effort to better understand the personality style of individuals in the workplace environment.

References

- American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Arbuckle, J. & Wothke, W. (1999). *Analysis of Moment Structures (AMOS) 4 user's reference guide*. Chicago: Smallwaters Corporation.
- Allen, M., and Yen, W. (1979) *Introduction to Measurement Theory*. Belmont, CA: Wadsworth.
- Cohen, R. J., and Swerlik, M. E. (2005) *Psychological Testing and Assessment, 6th ed.*. New York, NY: McGraw-Hill.
- Comrey, A. L., and Lee, H. B. (1992). *A First Course in Factor Analysis, 2nd ed.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Geier, J. G. (1977). *The Personal Profile System*. Minneapolis, MN: Performax Systems, International.
- Goldstein, K. (1963a). *The Organism*. Boston: Beacon.
- Guilford, J. P. (1954). *Psychometric Methods, 2nd ed.* New York, N.Y: McGraw-Hill.
- Gulliksen, H. (1987). *Theory of Mental Tests*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Highland, R. W. & Berkshire, J. R. (1951). *A methodological study of forced-choice performance rating*. Research Bulletin 51-9. San Antonio, TX: Human Resources Research Center.
- Landis, J. R., and Koch, G. G. (1977a) The measurement of observer agreement for categorical data. *Biometrics*, 33, 159 – 174.
- Martson, W. M. (1928). *The Emotions of Normal People*. London: Harcourt, Brace & Co.
- Nunnally, J. C., and Bernstein, I. H. (1978). *Psychometric Theory, 3rd ed.* New York, NY: McGraw-Hill
- Sullivan, H. S. (1953). *The Interpersonal Theory of Psychiatry*. New York: Norton.
- Traub, R. E. (1994) *Reliability for the Social Sciences*. Thousand Oaks, CA: Sage.