Target Training International – Performance Systems, Ltd.

16020 N. 77th St.
Scottsdale, AZ 85260
1.800.869.6908

Success Insights, Inc.

4336 E. Karen Drive
Phoenix, AZ 85032, U.S.A.
1.602.788.8114

Style Insights – DISC

Instrument Validation Manual

Published: 15 March 2004

Copyright © 2004 retained by Target Training International – Performance Systems, Ltd., Success Insights Inc., Russell J. Watson, Ed.D. and Peter T. Klassen, Ph.D. All rights reserved. Except as permitted under national and international laws, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Table of Contents

1. Introduction and background

This manual is developed for the purpose of demonstrating the construct reliability and the validity of the *Style Insights* instrument published by Success Insights, Ltd., Scottsdale, Arizona, USA. It is intended to be used by both distributors of the Success Insights instrumentation, by potential clients and customers, and by researchers and academics interested in more information about the technical specifications of the instrument and reports. The intent herein is to provide a verifiable pattern of evidence that establishes the *Style Insights* instrument as a sound, reliable, valid, and usable instrument for a variety of purposes in personal and organizational development, and for organizational and corporate use in a number of venues.

The intent of this manual is to provide the essential information needed, while remaining somewhat brief in text length. This increases the usability of the manual and also provides the information in a clear-cut manner. There will be some parts of this manual that will address specific statistical procedures and methods. That discussion will be specific, and will treat the information in an empirical and statistical manner. It is assumed that some readers will have a background in statistics and others will not. The intent will be to provide the information in both statistical tabular form, and also to provide a discussion of the meaning of the tables in clear narrative form.

The research and statistics provided in this manual have been written and conducted to the specifications published in *Standards for Educational and Psychological Testing* (1999) cooperatively by the American Educational Research Association, American Psychological Association and the National Council on Measurement in Education. The guidelines provide the standards against which many US-based and international assessments are designed and validated. It is the purpose of this manual to respect those specifications and to encourage the reader to explore the standards in more detail. The reader is also encouraged to ask active questions about other assessments in the marketplace and to discover the extent to which those assessments followed similar guidelines to the *Style Insights* instrument and reports.

Measurement of one's 'style' – A brief history

The *Style Insights* instrument is generically loaded into a category of assessments sometimes called 'personality tests.' The authors of this manual and the principals of Success Insights and Target Training International Performance Systems (TTI) prefer the use of the term 'style' instead of 'personality' for a variety of reasons. First, the term 'personality' is a very complex and global term indicating a wide bandwidth of behavior and applications of the entire individual. Second, the term 'style' as originally suggested by Fritz Perls, relates more to the specifics of *how* someone does something, and is therefore more applicable to the purposes and goals of the *Style Insights* instrument and reports.

Historically, there are a variety of ways by which one's personality and 'style' has been measured. Early work by Kraepelin (1892) with the free association test involved the subject being given a list of stimulus words to which the subject was asked to provide the first word that came to mind. The free association

methodology has been used for a variety of assessment purposes and it remains in use today. Some criticism of the method remains with issues of scoring, inter-rater reliability, and malingering by the subject.

In answer to the critical issues of scoring and inter-rater reliability came the self-report inventory. A very early form of this assessment technique was developed by Woodworth during World War I (DuBois, 1970; Goldberg, 1971; Symonds, 1931). The original purpose was that of a screening test for identifying those unfit for military service. The War ended before the model was deployed; however, civilian forms were developed for both adults and children. The Woodworth Personal Data Sheet served as a prototype and early model for many inventories to follow. Some designs explored specific areas such as vocational adjustment, school adjustment, home, etc. Other assessments explored interpersonal responses in social settings, and later came assessments focused on interests and attitudes. It is in the self-report genre that the ***Style Insights*** instrument and reports are based.

The *'performance'* or situational test is another commonly used assessment method. With this model, the subject is asked to perform a task, and is measured based on their performance. The specific purpose for some of these tests is concealed from the subject. An early application of this model was developed by Hartshorne and May, et al., (1928, 1929, 1930), and standardized on schoolchildren. Situational tests for adults were developed during World War II by the Assessment Program of the Office of Strategic Services. These tests were high in complexity for the time, and needed some detailed staging and skilled administration. Even so, issues of inter-rater reliability and interpretation of responses were rather subjective.

Another methodology is that of the *projective test* design. In this method, the subject is presented with an ambiguous or open-ended task or description to provide of a stimulus card or process. Again, the purposes of these tests are somewhat disguised from the subject to reduce the potential of the subject creating a preferred response, or malingering. As with free association and some situational tests, there is room for inter-rater reliability errors and variability in scoring due to the subjective nature of the instrumentation.

The ***Style Insights*** instrument and reports use the self-report methodology that eliminates inter-rater reliability issues because of the objective scoring method of the instrument. Using the self-report method, the instrument captures one's own self-perception and records responses. While inter-rater reliability is eliminated, an inherent issue with all self-report instruments is the accuracy of one's responses and the focus of their self-perception. Therefore, the respondent is always encouraged to be honest in their response and clear in their situational focus when they respond. This methodology has been widely used and adopted in many academic and commercial applications.

Connection of DISC to Target Training International's and Success Insight's published instruments

In 1983-84 Target Training International, Ltd. (TTI) acquired a DISC-based instrument under a license agreement. Since that time TTI and Success Insights have invested substantial amounts of attention, energy, and resources into the continued statistical validation of the instrument and the reports. Changes have

been made to the newer versions of the instrument to keep pace with current terms and descriptors in use, and to up-date those terms and descriptors that were useful decades ago, but are less valid in the 21st Century. TTI and Success Insights are rare among DISC providers in that their statistical validation work features current scores from the 21st century that are based in the language / cultural groups using an instrument. This allows for increased reliability and validity of the report printouts by comparing one's scores against a large, well-defined, contemporary culturally relevant database.

2. Foundations

Theory

The *Style Insights* model is based on the four-dimensional DISC model which is widely used throughout the world, and has received broad acceptance by commercial and organizational enterprises. Four-dimensional models have a long history in philosophy and psychology. Empodocles (444 B.C.) was the founder of the school of medicine in Sicily. He based his teaching on the notion that everything was made up of four elements: air, earth, fire, and water. These elements can be combined in an infinite number of ways to form all other materials. Hippocrates (400 B.C.) offered the notion of four different types of climate and terrain having an impact on the behavior and appearance of people living in those environments. From this notion, he developed the concept of the four temperaments (sanguine, melancholic, choleric, and phlegmatic) and associated these temperaments with four bodily fluids (blood, black bile, bile, and mucous).

In 130 – 200 A.D., Galen in Rome offered the notion of four bodily fluids (blood, black bile, yellow bile, and phlegm) and their effect on behavior. He also offered the idea that our bodies are acted upon by four external conditions of warm, cold, dry, and moist. Major empirical contributions were made in 1921 by Carl Jung in Germany through his work *Psychological Types*. He identified four 'types' of psychological functions: thinking, feeling, sensing, and intuiting. These four types are further divided by two energy forces of introversion and extraversion which were categories in addition to the four types of psychological functions.

The primary developer of the DISC model is Dr. William Moulton Marston, of Harvard University. He is well-known for his 1938 book, *The Lie Detector*, and made major contributions in that area. The *Style Insights* instrument as well as most other DISC instruments is based on Marston's original design. Marston was born in 1893, and received three degrees from Harvard University: A.B. in 1915, L.L.B. in 1918, and Ph.D. in 1921. Marston served as a teacher and consulting psychologist, and was a member of the faculty at The American University, Tufts, Columbia, and New York University. He contributed articles to the American Journal of Psychology, the *Encyclopedia Britannica*, the *Encyclopedia of Psychology*, and authored five books.

In 1928 Marston published his book, *The Emotions of Normal People*, in which he submitted the DISC theory that is used today. He viewed people as behaving along two axis with their actions tending to be active or passive, depending upon the individual's perception of the environment as either favorable or antagonistic. By placing the axes at right angles, four quadrants form with each describing a behavioral pattern.

1. **Dominance produces activity in an antagonistic environment.**
2. **Inducement produces activity in a favorable environment.**
3. **Steadiness produces passivity in a favorable environment.**
4. **Compliance produces passivity in an antagonistic environment.**

Marston believed that people tend to learn a self-concept which is basically in accord with one of the four factors. It is possible, using Marston's theory, to apply the powers of scientific observation to behavior and to increase objectivity and description while reducing subjectivity and judgment. Walter Clark in the 1950s was the first person to build a psychological instrument based on the Marston theory. The form was called the 'Activity Vector Analyses." Some of Clark's original associates subsequently left his company, and each refined the format as they created their own self-report adjective checklist forms of the instrument. There are many companies today using the Marston theory as the basis for exploring behavior via a self-report descriptive behavioral instrument. The *Style Insights* and its ancillary forms enable us to identify patterns of behavior in such a way as to make practical applications of the Marston theory.

Historical validity

A wide variety of research studies have confirmed the theoretical constructs. Those major studies will be briefly mentioned herein. Since several commercial organizations use the model, full disclosure of their research is somewhat difficult to obtain because of certain proprietary rights claimed by each of the organizations. Success Insights and TTI Performance Systems have been leaders in disclosure of certain research findings, while protecting those proprietary results such as specific scoring keys and item analyses. A brief summary of significant research contributing to the advancement of the DISC theory and design will be presented herein.

- **1967 – Dr. John G. Geier, University of Minnesota. "A Trait Approach to the Study of Leadership in Small Groups." *The Journal of Communication*, December, 1967.**
  - **1977 – Dr. John G. Geier, University of Minnesota. *The Personal Profile System*. Minneapolis, MN: Performax Systems, Int'l.**
- **1983 – Dr. Sylvan J. Kaplan completed a study of the Personal Profile System, and compared it to the following psychological instruments:**
  - **Wechsler Adult Intelligence Scale**
  - **Myers-Briggs Type Inventory**
  - **Cattell 16 Personality Factor Questionnaire**
  - **Minnesota Multiphasic Personality Inventory**
  - **Strong Interest Inventory**
- **1989 – Dr. Russell J. Watson, Wheaton College, "A Statistical Comparison of the TTI Style Analysis and the Performax Personal Profile System," Wheaton, IL. This study at the time indicated that there were no statistically significant differences between the scores of the Style Analysis and the Personal Profile System.**
- **2002-2004 – Success Insights and TTI Performance Systems have engaged in robust statistical analyses of their instrumentation and have made subtle modifications to the instrument and scoring under the direction of Dr. Peter T. Klassen, Professor Emeritus, College of DuPage, Glen Ellyn, IL. This statistical analysis has resulted in an increase in all construct reliability variables measured in the initial sample, and has resulted in an increase in the face-validity of the reports as well.**

Over the past several decades there have been numerous studies using the Style Analysis model as a base for behavioral research across a variety of topics. A selected list of those higher level research studies granted dissertation status is listed below.

A selected list of additional academic study using the DISC / Style Analysis as a base:

- **"A case study which utilizes type indicators to analyze 360-degree performance assessments." Doctoral dissertation in Educational Psychology by George Landon Anderson, University of Louisville, Louisville, KY, USA, 1995,**
- **"Investigating the effects of behavior constructs on academic persistence in engineering, creativity, and risk-taking." Doctoral dissertation in Psychology by Viveca K. Deanes, Texas A & M University, USA, 2003.**
- **"Behavioral Style as a predictor of hearing aid return for credit." Doctoral dissertation in Psychology by Steven A. Huart, Central Michigan Univesity, USA. 2002**
- **"Market segmentation: Exploring the need for further consumer behavior analysis. (behavioral profiling)." Doctoral dissertation in Personality Psychology by James Joseph Kolacek, III, Nova Southeastern University, USA, 1999.**
- **"Jury deliberation style and just world belief." Doctoral dissertation by Harry Naifach, Kent State University, USA, 2002.**

Utilization confirmation

In addition to the formal academic research about the DISC model, or using the DISC model as a basis for descriptive or experimental research, there have been numerous informal studies conducted at the team or organizational levels. While many of those studies are not published, others remain proprietary to the consultant or organization, and there are also studies made public through trade press articles and consultants who publish their results to the greater community. Some of those results will be mentioned herein. The reader is encouraged to explore the topic through a variety of academic web-based search engines using the descriptors: DISC Style Analysis; Style Analysis; Behavioral Style; and Behavioral Profiling to explore a variety of sources of additional information.

The DISC model has experienced a very wide usage in business, industry, and organizational environments over the past thirty years. Rather than tracing another history of such usage herein, we will provide the reader with several current published resources, and again suggest individual searches using the descriptor protocols above. The sources indicated below are from commonly available trade press and journals in the US and internationally. These sources, while not academic in sense of a juried journal article or doctoral dissertation, provide a pattern of evidence and multiple measures of the use of the DISC model in a variety of business environments. Some recent publications include:

- **"What is Your Communication Style?" Lisa Aldisert in *Bank Marketing*; Oct. 2000, Vol. 32 Issue 10, p. 46. Presents communication tips for bank**

marketing. Advantage of adapting to a behavioral style in communication; Information on the different types of behavioral styles in communication.

- "D.I.S.C. Drives." Evan Cooper in *Financial Planning*; Sept. 2000, Vol. 30 Issue 9, p. 107. Explores the four behavioral styles of brokers identified by Dr. Joseph Marshall. Behavioral type of financial planners; How brokers can benefit from their behavioral styles.
- "What's Your Client's Style?" Susan Foster in *Selling*; Dec. 1998, Vol. 6 Issue 5 p. 8. Discusses individual behavioral preference classification as a guide to salespersons in better understanding their clients. Three key steps of applying knowledge of behavioral styles to a sales situation.
- "Cracking the Communication Ice. What's Below the Surface?" Kenneth R. Kramm and Deborah A. Kramm in *Training & Development Journal*; April 1989, Vol. 43 Issue 4, p. 56. Presents warm-up exercises to initiate discussion and audience participation during seminars or workshops. Behavioral styles which help to decide how to use the word-selection ice breaker, and evaluation of the exercise.
- "Preferences for Behavioral Style of minority and majority members who anticipate group interaction." Alain Van Hiel and Ivan Mervielde in *Social Behavior & Personality: An International Journal*; 2001, Vol. 29 Issue 7, p. 701. The research investigates whether prospective minority and majority members ascribe high effectiveness to particular behavioral styles in order to exert influence in a forthcoming group interaction.

While this list is brief, it is a selection of a variety of uses of the DISC model across a spectrum of business topics.

3. Instrumentation

Instrument overview

Style Insights **– DISC is anchored in design and development of TTI's prior DISC instrument as well as based on prior DISC instruments. Respondents select from within frames of four descriptive items one item as "like-me" and one item as "not-like-me." These selections are used to build the four DISC scales. The four scales are labeled as "Dominance-Challenge," "Influence-Contact," "Steadiness-Consistency," "Compliance-Constraints."**

These scale scores are communicated through two graphs, and text output. The graphs, one for the adaptive scales and one for the natural scales, graphically display the relative values of the four scale scores. Graphing is based on equating respondent's frequencies of selected descriptions with the percentile ranking of similar item selections in the norm-population. Further discussion of this issue is reported on page 22.

Scale structure

Style Insights **instrument contains ninety-six (96) phrases organized in twenty-four (24) frames of four items each. Each frame contains descriptive items associated with each of the four scale constructs. Respondents select a forced choice of "most-like" and "least-like" themselves or rank the four items from most like me to least like me, depending on the delivery format. Two dimensions of four scales are constructed from these responses. The two dimensions are adaptive and natural, each of which has the four scales D, I, S, and C. Items selected as "most-like-me" that are associated with each of the four scales are summed up for the adaptive scales. Items selected as "least-like-me" that are associated with each of the four scales are summed up for the natural scales, in which not identifying descriptions yields a higher ranking on the scale. The importance and utility of the two dimensions of scales is discussed later (see page 13).**

Construction of these scales using this pattern of association and rejection is well established. "Studies have shown that the number of scale points does determine the reliability of an instrument only when the number of descriptor items is below 50" (Warburton, n.d.). The *Style Insights* DISC instrument exceeds this minimum. Peabody examined the relative importance of the polarity of scoring versus the number of scale points. He concluded that: "…composite scores reflect primarily the direction of responses, and only to a minor extent their extremeness. The practical implication is that there is justification for scoring the items dichotomously according to the direction of response" (Peabody, 1962, p. 73). This conclusion is supported by other researchers (Komorita, 1963; Komorita & Graham, 1965; Jacoby & Matell, 1971; Matell & Jacoby, 1972) and is the reason the *Style Insights* DISC is scored using a three choice responses pattern (most-like-me, unselected, and least-like-me).

Primary DISC scales

Each of the four DISC scales is considered a constellation of behaviors that are coherently related to a psychological style for an individual. As discussed in both the theory and the setting sections (see page 17) of this manual, the specific expression of descriptive behaviors may be impacted by setting, culture, and other environmental variations. For the purpose of communication each of the four scales is referred to by the first letter of a descriptive label intended to summarize the constellation of related behaviors. An individual's behavior will tend to demonstrate higher or lower levels of the described behaviors based on the dominance or recessiveness of the underlying psychological characteristic. Therefore, some of the described behaviors may be expressed or not depending on the setting and cultural expectations or constraints. The presence or absence of any specific described behavior is not an indication of the DISC style.

Dominant, Driver, Choleric	Influencer, Expressive, Sanguine	Steadiness, Relater, Amiable, Phlegmatic	Compliance, Analytical, Melancholic
Adventuresome	Affable	Accommodating	Accurate
Aggressive	Assuring	Amiable	Agreeable
Ambitious	Captivating	Caring	Analytical
Argumentative	Charming	Compassionate	Careful
Authoritative	Cheerful	Complacent	Cautious
Blunt	Companionable	Considerate	Conscientious
Bold	Confident	Contented	Conservative
Brazen	Convincing	Conventional	Contemplative
Challenging	Cordial	Deliberate	Conventional
Competitive	Delightful	Devoted	Courteous
Confrontational	Demonstrative	Dutiful	Deliberative
Courageous	Effusive	Even-tempered	Diplomatic
Daring	Emotional Generous	Faithful	Disciplined
Decisive	Enthusiastic	Friendly	Evasive
Demanding	Expressive	Good Listener	Exacting
Determined	Friendly	Inactive	Fact-finder
Direct	Good mixer	Loyal	Factual
Domineering	Gregarious	Mild	Follower
Energetic	Impulsive	Modest	High standards
Enterprising	Influential	Nonchalant	Logical
Exploring	Inspiring	Non-demonstrative	Mature
Forceful	Open-minded	Obedient	Methodical
Goal-oriented	Optimistic	Obliging	Orderly
Impatient	Outgoing	Passive	Organized
Independent	Persuasive	Patient	Patient
Innovative	Playful	Possessive	Peaceful
Inquisitive	Poised	Predictable	Perfectionist
Persistent	Popular	Relaxed	Precise
Pioneering	Self-promoting	Reliable	Quiet
Powerful	Sociable	Satisfied	Receptive
Responsible	Spontaneous	Serene	Reserved
Results-oriented	Stimulating	Sincere	Respectful
Risk-taker	Talkative	Stable	Restrained
Self-starter	Trusting	Steady	Self-controlled
Strong Ego strength		Sympathetic	Sensitive
Strong-willed		Team player	Structured
Stubborn		Thoughtful	Systematic
Unwavering		Understanding	Tactful
Unyielding		Willing	Tranquil

Adaptive and natural dimensions

The *Style Insights* instrument utilizes two related approaches to measuring this DISC behavioral style (characteristics). The Principle of Reciprocal Evaluative Action posits "that positively and negatively valiant activation functions are reciprocally determined. In other words, on a bipolar scale of agreement, maximum agreement is the reciprocal of minimum disagreement, and vice versa" (Warburton, n.d.). These dichotomous responses result in two sets of DISC scales referred to as the adaptive and natural dimensions.

The adaptive and natural sets of DISC scores are constructed by summing up descriptions that characterize each of the four scale concepts. Each of the twenty-four (24) frames of the instrument contains four words or phrases. Each of those words or phrases is associated with one of the scales. Respondents select one of these four items as "most-like-me" and one as "least-like-me" or rank the four items from most to least like me. Thus, the each of the dimensions' four scales is constructed from the association of the item with self-perception characteristics.

Because the adaptive dimension DISC scales are constructed from the selections of "most-like-me," these scales are an indication of the public self as presented in a setting. It is logical to conclude that these responses may be most susceptible to an intentional individual bias.

Because the natural dimension DISC scales are constructed from the selections of "least-like me," these scales are reversed in valiance. That is, the more items ranked as "least-like-me" the lower the score, and the less the associated characteristic is perceived as expressed by the individual. Highest scores are the result of NOT selecting a descriptive item associated with a scale. It is both logical, and demonstrated in comparisons among various groups of respondents, that the natural scale is less susceptible to setting and/or an intentional individual bias.

Validity and reliability

Reliability based on response processes and internal structure

The issue of instrument reliability is the initial question asked when exploring how 'good' an instrument is, or if it is actually useful. The word 'reliability' always means 'consistency' when applied to instruments and tests. There are several procedures that are commonly used for this routine statistical treatment. Test-retest reliability is the consistency of scores obtained by the same persons when re-tested with the identical instrument. Alternate-form reliability provides the subject with two similar forms of the instrument. Both test-retest and alternate-form reliability documentation should express both the reliability coefficient and the length of time passed between the first and second testing events. Both of these procedures focus on the consistency of measurement. Such consistency and the "learning the test" advantage is a major concern with ability and knowledge measurements. The *Style Insights* is not subject to an advantage from repeated administration because it asks for self-reports. The instrument's scales are as stable as the individual's perception of situational demands and self-concept is constant.

Split-half reliability involves a single administration of the instrument, and uses the technique of 'splitting' the instrument in half, e.g., odd and even question items, and determining a correlation between the two sets of scores. This technique reduces some of the concerns of test-retest and alternate-form reliability by eliminating the passage of time between testing events. Kuder-Richardson reliability is also based on a single form and single administration of the instrument, and measures the consistency of responses to all items on the test. The Kuder-Richardson formula is actually the mean of all split-half coefficients based on different splittings of the test. The

Spearman-Brown reliability formula is another statistical treatment that provides a reliability coefficient, and is frequently used with the split-half procedures. Spearman-Brown differs by including a method for doubling the number of items on an instrument as a part of its formula. By doubling the number of items on the instrument, reliability usually increases. Some critics of the Spearman-Brown formula say that it may artificially raise the reliability coefficient of a test. Each of the reliability coefficients discussed so far are ones that can be calculated by hand, or using a simple calculator.

The alpha coefficient is the expression of an instrument's reliability and ranges from –1.00 through zero to +1.00. An instrument with a perfect reliability would have an alpha coefficient of +1.00, and no instrument has yielded that score to date. Additionally, there is no standard, agreed-upon 'levels' of what makes a good or bad correlation for testing purposes. However, there is general agreement on a minimum standard for alpha equal to .6 or greater, with some experts advocating use of a .7 or higher standard. Obviously, the higher the alpha coefficient the stronger is the coherence of items. Cronbach's alpha (α) (Cronbach, 1951) is considered by many to be the most robust reliability alpha to date (Anastazi, 1976; Reynolds, 1994). "Coefficient α is the maximum likelihood estimate of the reliability coefficient if the parallel model is assumed to be true" (SPSS, p.873). For dichotomous data, "Cronbach's alpha is equivalent to the Kuder-Richardson formula 20 (KR20) coefficient" (SPSS, p.873). Cronbach's alpha is used to determine all of the reliability coefficients for the ***Style Insights*** instruments. The reader is encouraged to compare the reliability coefficients presented in this manual to the reliabilities of other instruments, and also to ask how other vendors compute their alpha numbers.

Validity based on context and relationships to other variables

Validity helps answer the question, "Does the instrument measure what it is supposed to measure?" It also asks a deeper quality-related question: "How well does the instrument make these measures?" These questions are obviously more difficult to answer and may leave room for subjectivity. With regard to any questions of validity, the critical issue is the relationship between performance on the instrument and other observable facts about the behavior being studied. When someone says, "The test wasn't fair," the comment is usually directed to the test's validity, not reliability. A more accurate way to state the same expression is, "The test wasn't valid." There are three primary forms of validity: Content, criterion-related, and construct validity.

Content validity examines the instrument's content to determine if it covers the behavioral topic being measured. Simple examination of items in a biology or chemistry test should indicate questions related to the topic or subject being studied. When used in the development of the DISC themes, it is important that all four descriptor categories are represented in rather equal proportion for selection of D, I, S, or C descriptors. Additionally, it is important to explore social desirability as an element of content validity. If there is an imbalance between words that are socially desirable versus descriptors that are less desirable, then content validity is affected. The ***Style Insights*** instrument is screened for content validity and since its initial

printing some descriptors have been replaced to boost both the content validity and the reliability of the instrument.

Criterion-related validity refers to the ability of an instrument to predict a participant's behavior in certain future situations. One's scores on an instrument are compared with any variety of external 'criterions.' In the use of the ***Style Insights*** instrument and reports, there are a variety of studies available from Success Insights and TTI Performance Systems that have clearly linked specific scores and patterns of scores to job success in specific, well-defined areas (Bonnstetter, et al., 1993). Criterion-related validity has two forms: concurrent validity and predictive validity. Concurrent validity examines one's scores and compares them to external criterion at the same time as taking the instrument. Predictive validity explores one's instrument scores against criterion after a specified time interval. Both methods provide robust support for the ***Style Insights*** instrument and reports (Bonnstetter, et al., 1993).

Construct validity examines the ability of an instrument to measure a theoretical construct or trait. Construct validity is built from a pattern of evidence and multiple measures across a variety of sources. Some constructs explored in behavioral trait analysis include: Developmental changes of participants responding to the instrument at different ages and stages of their lives, or under different response focus points. Correlation with other tests is a form of construct validation. There have been a variety of comparisons of the ***Style Insights*** instrument with other behavioral instruments such as MBTI (Myers Briggs Type Indicator), MMPI (Minnesota Multiphase Personality Inventory), 16-PF (16 Personality Factor), and other instruments (Bonnstetter, et al., 1993). All of these studies assist in establishing the overall constructs of these instruments.

One very important technique within construct validity activity is the factor analysis. This is a technique that 'refines' an instrument by comparing and analyzing the interrelationships of data. In this process the interrelationships are examined and 'distilled' from all initial combinations, to a smaller number of factors or common traits. Through factor analytic work using other instruments, it has been discovered that instruments from some other vendors have specific descriptors that actually factor-load into different categories than the ones in which they are scored on the instrument (Golden, Sawicki, & Franzen, 1990). The ***Style Insights*** instrument has been refined through the factor analysis process and has made subtle scoring changes that increase both the overall validity and reliability of the instrument and reports (see "Examination of theoretical coherence" beginning on page 21).

A comment about face validity: Face validity is not to be confused with any of the above mentioned forms of validity. Face validity is not really validity in the pure technical definition of the word. Face validity considers whether the instrument 'looks valid' to participants, and if the report 'sounds valid' to the reader. Fundamentally, while face validity is not technically a form of validity, it is nevertheless a very important consideration in both instrument construction and report writing. One of the authors (Watson) has used the ***Style Insights*** instrument and reports for research with thousands of participants. When in audiences of any size, the participants are asked, "For how many of you is your report 85% to 90% accurate for you?" This is a

question of face validity. Invariably, when asked that question, nearly 90% of an audience will raise their hand, or indicate the same statistics (85% to 90% accuracy) in written statements (to avoid hand-response majority bias). While this method is not using a statistical formula, it nevertheless addresses the concept of face validity.

Convergent and discriminate evidence

Two additional issues are part of examining validity. These issues basically ask the question of whether classification using an instrument appropriately identifies common individuals (convergent) and differentiates among individuals belonging to a different classifications (discriminate). Once again most of the evidence to these powers lies with the successful application experiences of consultants using the instrument. During the 2002 initial review of the ***Style Insights*** instrument, an opportunity arose to compare samples from respondents from different employment classifications. Four data sets came from respondents classified as employee-managers, executives, sales, and customer services. From a total of 120,898 responses a sample of 1076 cases were randomly selected. These cases were submitted to a canonical discriminate analysis. Based on using the eight scales as independent variables, a discriminate classification model was calculated. Applying this classification model resulted in over 42% of the cases being classified convergent with their data source. At the high end of correct classification 75% of the customer service respondents were accurately classified. On the low end 30% of the executives were accurately classified. Given the diversity within each classification of respondents, and the cross-over of individuals of different employment classification these results demonstrate a substantial level of capacity to identify and discriminate among individuals based on their ***Style Insights*** DISC scales.

Setting and environment

All DISC instruments focus on behavior, and that behavior takes place in a public setting. Thus, the psychological characteristics being measured are expressed through interactions. Unlike measurement of purely internal beliefs and values, behavior is encouraged and discouraged through these interactions. When considering these effects three issues arise that impact the measurement of such an expressed characteristic. One issue focuses on the situational demands and individual perceptions of the setting in which the measurement is taking place. A second issue focuses on the social desirability perceived by the respondents with reference to each of the descriptions of behaviors used as indicators of the four DISC scales. A third issue is introduced when descriptions are translated and the instrument is used in a different language/cultural environment. Each of these issues is discussed below.

Situational demands

Situational demands arise from the setting and conditions as perceived by an individual. Since the instrument is based on self-reports of behaviors, an individual may enhance or censor such self-reports based on conscious or sub-conscious perceptions of the setting in which the instrument is being

used. If these perceptions are that the setting is low-risk and trusting, the individual may be more candid than in high-risk settings. Conversely, if these perceptions are that the setting is high-risk or judgmental, then the individual may be less candid in order to present a positive self-image. This is an issue shared by all self-report (ipsitive) instruments and it is discussed on page 19.

The effect of these perceptions is more easily expressed in choosing descriptions that are "most-like-me." With the choice of "least-like-me" or low rankings of descriptions, it is harder for an individual to express a bias. Comparison of scales based on "least" and "most" "–like-me" are discussed on page 13. However, these "least-like-me" statements are subject to social desirability influences as discussed next.

Social desirability

Not all descriptions or characteristics of a scale can be used as scaleable items in an instrument. There are some descriptions that are socially loaded as either attractive or undesirable. That is, some descriptions of a scale may be socially desirable. Persons displaying a higher level of a scale characteristic may act more frequently in that socially desirable manner. But, everyone would like to see themselves in the positive light. So, if that description is included among the choices and many respondents choose it, it does not differentiate between someone high, moderate, or low in the scale behavior. In a similar way there may be descriptions of behaviors at many people would choose as "least-like-me" even when it is a good description of a scale behavior. In both cases the descriptions do not work to differentiate among respondents, and they are therefore not used in scoring an instrument. These values of desirable and undesirable behaviors are socially established and shared among most members of a culture. Culture further impacts selection of descriptions as discussed next.

Cultural impacts

Although there may be many cultures and sub-cultures present in a population, the effects of language groups are the level of differentiation implemented in the ***Style Insights*** instrument's versions. Cultures differ in how specific behaviors are defined and judged. Anyone visiting another culture may notice such differences immediately. Loud simultaneous talking may be the norm of a good friendship in one culture, and signs of a fight about to erupt in another. A description of a behavior utilizing similar words in two different languages may have very different connotations. For example solidarity and compassion may carry different connotations with reference to the role of equality and sympathy in different cultures. It is important to consider these differences when using an instrument in different cultures. In response to these differences, specific versions of ***Style Insights*** are developed, evaluated and tested for different language groups. The descriptions used as items in the instrument are tested for reliability and coherence with the scale concepts for each language version. If usage of the instrument is sufficient and clients conclude that it is important, specific distributions and norms can be calculated for any specific sub-population that can be defined.

Measurement limits

The process of self-report is referred to as ipsitive measurement. All efforts at this type of measurement are limited. Three issues are of specific interest related to this instrument. These issues consider the nature of the score measurement, a bias of self-reporting, and the effects of the situational demands and perceptions on scores.

Frequency counts and a score

First, the process of summing up the frequency of responses produces a score that is a comparative measure, not a quantity measure. A score is a count of descriptions selected by the respondent. The count is compared with other people's counts among a norming population. These raw counts across several scales cannot be compared directly. That is, selecting 10 x items and 5 y items does not mean one is more x. However, if in the norming population the average is selecting 5 x items and 7 y items, then an individual selecting 10 x items can be reasonably evaluated as seeing themselves as behaving in a more x manner than generally expected in the population. In this instrument the comparison is made by reporting individual scores as the percentile ranking found in the norm population. Remember, it is important to note that the scales are not quantities of the characteristics.

These comparisons are clearly based on grounding the norming population as representative of people like those who look to an instrument for feedback. In this instrument the norms for comparison are representative of current instrument users. Wherever possible, specific norms are developed for unique language/cultural groups. Each norm-distribution used as reference for a version of the instrument is clearly identified. Further discussion of the issue of the norming population is reported on page 22 and in each of the technical summaries in the Appendix beginning on page 28.

Ipsitive measures

The process of self-report as a source for behavioral scaling has a long tradition in psychological measurement. Such measurement is accepted as a method to gain insight based on self-perception; however, it does have its limitations. Ipsitive measurement is subject to bias, which may be either sub-conscious or conscious.

An example of this sub-conscious bias may occur when we do not see ourselves as others do. Our self-perceptions, while founded in feedback from others, may not be congruent with the way others would describe us. Awareness of and addressing this incongruence may be an outcome worth the effort in team-building and human resource development.

Effects of situational expectations

An example of conscious bias may occur when an individual believes certain behavior characteristics are valued by others in a specific setting and therefore tries to present their "best foot forward." This issue was presented in more detail on page 18.

4. Technical information

Review and revision

Target Training International (TTI) and Success Insights (SI) initiated a review of their *Style Analysis - DISC* instrument during the spring of 2002. The core issue being addressed with this review was scale and item reliability for the ninety-six (96) descriptions used as indicators when constructing the two dimensions of the four DISC scales.

Scale reliabilities and item cohesion with its assigned scales were examined for samples from each of five English language version sub-populations, German language version and Hungarian speaking response-populations. The review process examined each of the five English speaking sub-populations separately, and, when appropriate, combined. The same process was applied to the German version. Revisions and editing of each language version were conduct independently. The following description of the review and revision process outlines the steps taken to examine the reliability of items, and scale constructions. Specific technical information about each language version can be found in the Appendix beginning on page 28.

All of the cases reviewed and examined were from respondents completing the *Style Analysis* during the prior year (2001-2002). The five English language sub-populations yielded over one-hundred twenty thousand cases (120,896). Because these sub-populations came from different applications of the instrument with respondents from different corporate statuses, it was possible to identify both similarities and differences in scale scores and reliability patterns. An initial review of the data from each sub-population involved confirming the computed scale scores agreed with the scoring matrix, and computing descriptive statistics and item/scale reliabilities. The sub-populations were compared with each other.

Most statistical procedures do not require use of the large numbers of cases available for examination. Therefore, for most statistical evaluations random samples were drawn from the sub-populations. The use of samples allowed for development of hypotheses that could then be tested against another sample that was independent of the first. This testing process was frequently applied to confirm recommendations for editing and revision.

Two approaches were taken in examining the coherence of the DISC scales. One examination took a "naive" approach of looking for patterns of common variance. This addressed the question of whether responses presented a pattern of coherence that justified the theoretical construction of the scales. Discussion of this approach follows in the section "Examination of theoretical coherence." A second examination applied the matrix of scale construction looking at the coherence of each item to its assigned scale, and the overall reliability of that scale construction. This examination utilized Cronbach's alpha and is discussed beginning of page 21.

Examination of theoretical coherence

Construction of a scale starts with implementation of theoretical constructs into operational measurement. In the case of DISC instruments, this construct validity occurred in the publication of early instruments. In order to confirm the coherence of the descriptions assigned to each scale a sample of responses was examined using a Principle Component Factor Analysis. In this statistical procedure the ninety-six (96) items were examined to find patterns of similar variation. Each factor is a latent construct, an unmeasured characteristic. The procedure results in a listing of factors with a measure of covariance for each of the variables. These coefficients may be positive or negative or neutral. By selecting the items with significant positive or negative coefficients to a factor one identifies a constellation of items that describe a latent factor. Frequently a factor will reflect two contrasting sets of items. One characteristic can be found among the items sharing positive coefficients, and a second among the items sharing negative coefficients. If the listing of items agrees with the listing of items theoretically assigned to a scale, then one may conclude that the implementation of the theory as a scale is well founded. When an item has a strong positive coefficient with other items assigned to a scale to which it is not assigned, then the theory and/or item needs to be examined. Most items aligned with their assigned scales. However, the most common anomaly is that an item is does not have a strong positive coefficient with any scale. In this case the item is not a usable indicator of a characteristic for measurement, even if it may be a good description. This paradox is explained in the section of social desirability on page 18.

A principal component factor-analysis of the items from the different sub-populations examined the continuity of the scales as constructed. Based on analysis of all of these findings, a limited number of items were revised, edited, and field-tested. Item revisions were based on finding new descriptions that were theoretically based combined with linguistic considerations that focused on current usage and minimization of social desirability bias.

Each revised ***Style Insights*** version was, prior to release, subjected to several rounds of field-testing, further editing and confirmation of revisions. Once again several different responding-populations were utilized. The current release confirms increased reliability in each of the scales, and improves independence between the S and C scales.

The process of scale editing mandated a revision of population distributions. This process changed the reference point for comparison of style from its historic point of development up to the 21st century with recognition of changing behaviors and social expectations.

Item and scale reliabilities

Scale reliabilities were calculated using Cronbach's alpha (α). Cronbach's α is considered the most appropriate statistical test for reliability, given the dichotomous responses used to construct the scales. For dichotomous data, this is equivalent to the Kuder-Richardson formula 20 (KR20) coefficient. These statistics model internal consistency, based on the average inter-item correlation. These evaluations are a more rigorous approach than a traditional

split-half statistic.[1] Cronbach's α is a statistic bounded by 0 to 1. In general an α equal to or greater than .6 is considered a minimum acceptable level, although some authorities argue for a stronger standard of at least .7.

Since this manual is released in association with several language versions of *Style Insights* – DISC, *specific* reliability statistics are summarized in each of the language version summaries included in the Appendix beginning on page 28. Most of the ninety-six items are assigned to scale construction for both dimensions. A few items are associated with only one or the other of the dimensions. Specific coefficients for individual items with their assigned scale are not released since they constitute proprietary information that guards the specific scoring and item analysis of the instrument.

Norms and population parameters

The pedigree of the current versions of *Style Insights* is based on the culmination of multiple evaluations involving a diversity of data sources and samples. Examination of prior versions which began in 2002 involved over one-hundred thousand respondents. Current item and scale reliability is the culmination of these repeated evaluations using different samples. The instrument's pedigree is strengthened by these repeated independent evaluations. Samples have come from current users of the instrument. These users represent a full range of individuals utilizing the instrument. The specific characteristics of users are included with each of the technical information sheets for each version published in the Appendix starting on page 28.

Sex/gender

One concern for any instrument designed to serve business and individual users in the 21st century is the effect of sex/gender on response patterns. One issue examined in this instrument review has been differences in response patterns between males and females. As one might expect, there are some differences in the average scale scores for males and females. However, these differences indicate relatively minor shifts of dominance of specific expression of behaviors. Whether these differences arise from biology, socialization, or both is not important to the effectiveness of the instrument. What is important is that the instrument measurements reflect measurement and feedback that does not induce a sex/gender bias. In response to this challenge the samples used to establish distribution norms are evaluated. When a sample contains a representative proportional sampling of females and males, no adjustment is required. However, when the proportion of males and females is disproportional an adjustment is applied to these data to equalize the effects of patterns of males and females. Such adjustment is noted in the technical information sheets for each version published in the appendix starting on page 28.

---

[1] When variance is spread throughout a series of items and sufficient cases examined, Cronbach's α equals a traditional split-half statistic. One issue that arises in comparison of Cronbach's α with split-half calculations is the adjustments for an increased number of items that inflates a split-half statistic. While such an adjustment is justified in measurement of ability and knowledge, it is not appropriate when used in ipsitive measurement. Cronbach's α can be thought of as an average of all the results possible from all split-half permutations. We warn those wishing to compare Cronbach's α with split-half statistics that comparison of adjusted split-half statistics with Cronbach's is statistically inappropriate.

Language versions

Style Insights **is available in several language versions. With the release of the current revisions each of those versions may be developed as a separately developed and evaluated instrument. When such development takes place the item descriptions that are initial translations from the English version are analyzed for their coherence with their assigned scale, and those scales' reliabilities appraised. This process results in further editing of items, and when necessary, revision of scales in order to develop an instrument that is reliable and appropriate to the targeted language/cultural group. Distribution norms specific to a language version are developed based on responses to that language version in order to provide clients with clear feedback that is relevant to the language/cultural group that uses the instrument. Technical information sheets are then released for each specific version. These are published in the Appendix starting on page 28.**

5. Instrument protocols and utilization

Administration, utilization

Training for utilization and interpretation may be arranged through TTI.

Scaling & graphing

The most direct report of scale outcomes are reported in two graphs. One graph portrays the adaptive DISC scores, and the other the natural DISC scores. The DISC graphing of scale scores is constructed using a percentile rank distribution of the scale scores observed in a norming population. Two important issues need to be kept in mind with this type of self-report and comparison. First, the four scales are relative to each other. So higher self-reports of descriptions associated with a scale lower the possible associations with the other three scales. Thus, the scales feed back the relative levels of association with the four styles of behavior, not a quantity or frequency of that style. Second, the percent comparisons are of the respondents' frequency of association with the described behavior and the cumulative frequency of such self-reports in the norm sample. Thus the comparison of individual to norm population is with reference to self-reports of relative levels of association, not a quantity of actions.

While the current graphs have been updated for the current samples, the procedure is the same as that used in the original instrument. The percentile rank distribution is a "universal" distribution, not a "normal" distribution. That is, the cases are not distributed in a "normal" bell shaped curve; but instead are a flatter distribution with approximately equal numbers at each scale point. That is why raw scores that infrequently occur are grouped together at the top or bottom of a graph.

The graphs are percentile ranks found in the reference population. In a group of 100 people containing a full range of individuals (not a concentration of similar people), individuals would cover the full range of each scale point from low on the graph to high on the graph. The graph of distributions would be relatively flat from low to high. However, given that the number of items associated with each scale is less than 100, a number of respondents share each possible scaled point.

These graphs are comparisons across the four scales. The scales are not independent measures. Therefore, as the self-reports (selection of descriptive items) increase on one scale, some other scale or scales is reduced. With the revisions in the current version, these comparisons are updated from comparisons of individuals to some English speaking, probably white-male dominated norm population of the 1960's or 1970's (maybe as early as the late 50's), to comparisons representative of current users.

Each of the scales is not a quantity of that type of behavior, but rather a percent of people in the norm group whose answers produce a score lower than the individual responder. So, a graphed score of 90% D means that around 90% of the norming population responded by reporting a lower level of D type behavior relative to the other descriptions of behavior. It does not mean that 90% of the

individuals' behavior is D. Using that same logic then, a scale score of 10% S means that the person's association with S type behavior descriptions (or not with a description indicated by the ranking of 4 or "not-like-me") was higher than about 10% of the norm sample. Given individuals' similarities in scores, it also means that about 78% of the norm sample reported higher association with S behavior, and 12% of the norm sample chose similar levels of S descriptions as the individual. These last two numbers are not easily read from the graph, but it is important to note that there are significant percents of people with similar scores.

Report

This technical manual covers DISC scale construction, recent revisions and the scale output through two graphs. Other reports produced by various distribution systems include text and graphic based interpretations. Those wishing more information should contact Target Training International, Success Insights, or their authorized representatives for specific information about reports and feedback formats.

Training and interpretation

Training for utilization and interpretation may be arranged through TTI. Interpretative information can be found in *The Universal Language DISC: A Reference Manual* by "Bill" J. Bonnstetter, Judy I. Suiter, and Randy J. Widrick published by Target Training International, Ltd. Training and certification programs are available for those wishing to develop expertise in using the *Style Insights* and other instruments.

6. References

Anastazi, A. (1976). *Psychological Testing*. New York: Macmillan Publishing Co.

Bonnstetter, W. J., Suiter, J. I., & Widrick, R. J. (1993). *The Universal Language DISC: A Reference Manual.* Scottsdale, AZ: Target Training International, Ltd.

Clark, W. (1954). *Activity Vector Analyses.* Self-published.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297-334.

DuBois, P. H. (1970). *A history of psychological testing.* Boston: Allyn and Bacon.

Goldberg, L. R. (1971). A historical survey of personality scales and inventories. In P. McReynlods (Ed.), *Advances in psychological assessment.* Vol. 2, 293-336. Palo Alto, CA: Science and Behavior Books.

Golden, C.J., Sawicki, R.F., & Franzen, M.D. (1990). Test construction. In G. Goldstein & M. Hersen (Eds.) *Handbook of Psychological Assessment.* New York: Pergamon Press.

Geier, J. G. (1967, December). "A trait approach to the study of leadership in small Groups." *The Journal of Communication*, 17, (4), 316-323.

Geier, J. G. (1977). *The Personal Profile System*. Minneapolis, MN: Performax Systems, Int'l.

Hartshorne, H. & May, M. A. (1928) *Studies in Deceit.* New York: Macmillan.

Hartshorne, H., May, M. A., & Maller, J. B. (1929) *Studies in Service and Self-control.* New York: Macmillan.

Hartshorne, H., May, M. A., & Shuttleworth, F. K. (1930) *Studies in Organization of Character.* New York: Macmillan.

Jacoby, J. & Matell, M. (1971). Three point Likert scales are good enough. *Journal of Marketing Research*, 8,495-500.

Kaplan, S. J. (1983). *A Study of the Validity of the Personal Profile System. Minneapolis, MN: Performax Systems Int'l.*

Komorita, S. S. (1963). Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology*, 61, 327-334.

Komorita, S.S. & Graham, K. (1965). Number of scale points and the reliability of scales. *Journal of Educational and Psychological Measurements*, 25,987-995.

Kraepelin, E. (1892). *Uber die Beeinflussung einfacher psychischer Vorgange durch einige Arzneimittel.* Jenna: Fischer.

Matell, M. & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert items? *Journal of Applied Psychology,* 56, 506-509.

Marston, W. M. (1928). *The Emotions of Normal People.* London: Harcourt, Brace & Co.

Peabody, D. (1962). Two components in bipolar scales: direction and extremeness. *Psychological Review,* 69, 65-73.

Reynolds, C.R. (1994). Reliability. In R.J. Sternberg (Ed.), *Encyclopedia of Human Intelligence.* New York: Macmillan Publishing Co.

SPSS. (1988). *SPSS-X User's Guide, 3rd Ed.* Chicago: SPSS Inc.

Standards for Educational and Psychological Testing, Committee to Develop (c/o American Psychological Association). (1999). *Standards for educational and psychological testing.* Washington D.C.: American Educational Research Association.

Symonds, P. M. (1931). *Diagnosing personality and conduct.* New York, NY: Century.

Warburton, D. M. (n.d.). *The Importance of Finding the Real Person at Work.* Reprints available *at* http://enlightenedpartners.com/profiles/ValidityStudies.pdf or http://www.cgroupinc.com/files/Validity_Study.pdf (Retrieved: 3 March 2004).

Watson, R. J. (1989). A statistical comparison of the TTI Style Analysis and the Performax Personal Profile System. Unpublished research report, Department of Education, Wheaton College, Wheaton, IL.

7. Appendix

Technical Information: *Style Insights* – DISC, English version 2003d

Summary

Based on a series of examinations of scale and item reliabilities across multiple populations of respondents, revisions were developed and tested for the ***Style Insights*** *that culminated in development of a revised instrument. The following assessment of* ***Style Insights*** *is based on 1130 responses received 1 October 2003. These data are from respondents actively using this instrument through the TriMetrix distribution system. The results of assessment of this revised edition indicate improved reliability for the two dimensions (adaptive, natural) of four parallel scales (D, I, S, C) ranging from .69 to .85. Each of the ninety-six items used to construct the scales contributes at a significant level to one or both the scale dimensions. Correlations between adaptive and natural scales indicate that these two dimensions of parallel scales are highly related, as one would expect, but also that the scales are sufficiently independent measures to justify separate interpretations and comparisons. Scores on the scales are distributed across a wide range of scale points, which supports making comparison between individuals and the self-reported behaviors in a population. Revision of the instrument included utilization of new population distributions that anchored comparisons in a population distribution representative of the 21st century. Overall, the* ***Style Insights*** *is a strong, reliable instrument applicable across a variety of populations.*

Background

Style Insights **- DISC English version 2003d is anchored in design and development of their prior DISC instrument. During the late summer of 2002, Target Training International, Ltd. initiated a review of the reliability of the eight scales and the associated items. That assessment utilized samples from five different response-populations of respondents containing at total of 120,898 responses submitted within a three month period. These original data contained 43% female and 57% male respondents. The following assessment of the *Style Insights* - DISC English 2003d is based on 1130 responses received 1 October 2003. These data contain 49.5% female and 50.5% male responses that are representative of individuals actively using this instrument through the TriMetrix distribution system.**

The Style Insights instrument contains ninety-six (96) phrases organized in twenty-four frames of four items each. Each frame contains descriptive items associated with each of the four scale constructs. Respondents select a forced choice of "most-like" and "least-like" themselves. Two dimensions of four scales are constructed from these responses. The two dimensions are adaptive and natural. The importance and utility of the two dimensions of scales are discussed later. The four scales are labeled as "Dominance-Challenge," "Influence-Contact," "Steadiness-Consistency," "Compliance-Constraints."

Scale reliabilities and item cohesion to its assigned scales were examined for samples from each of the five response-populations. A factor-analysis on the items was used to further confirm the continuity of the scales as constructed.

Based on analysis of all of these indicators, a limited number of items were revised, edited, and field-tested. Item revisions were based on theoretical construction of items combined with linguistic considerations that focused on current usage and minimization of social desirability bias.

The revised *Style Insights* instrument was, prior to release, subjected to several rounds of field-testing, further editing and confirmation of revisions. Once again several different responding-populations were utilized. The current release confirms increased reliability in each of the scales, and improves independence between the S and C scales.

The process of scale editing mandated a revision of population distributions. This process changed the reference point for comparison of style from its historic point of development up to the 21st century with recognition of changing behaviors and social expectations.

Norming sample

The pedigree of the current version has involved a diversity of data sources and samples. Current item and scale reliability is the culmination of these repeated evaluations using different samples. Thus, the instrument's pedigree is strengthened by repeated independent evaluations. The norms used in the current iteration, *Style Insights* – DISC English 2003d version, utilize a sample of 1130 respondents compiled from users of the instrument during 2002 and 2003. The sample is compiled from one delivery system – TriMetrix – in active use among businesses in the United States of America. The sample contained 49.5% females and 50.5% males. Respondents' age ranged from 20s through over 50 years old. Occupations include managers, service workers, trade personnel, clerks, accountants, entrepreneurs, engineers, teachers, consultants and trainers. Thus, the sample represents a full range of individuals making use of the instrument in a variety of settings.

Evaluation of reliability and calculation of the distribution norms was conducted using this sample of 1130 to confirm selection and editing conducted with multiple prior assessments. The table below summarizes this population.

Style Insights – English 2003d		Males = 50.5%, Females = 49.5%							9-Oct-03
		Adaptive D	Adaptive I	Adaptive S	Adaptive C	Natural D	Natural I	Natural S	Natural C
Valid responses		1130	1130	1130	1130	1130	1130	1130	1130
Missing		0	0	0	0	0	0	0	0
Mean		6.66	6.15	5.79	4.94	6.89	4.31	4.43	7.63
Std. Error of Mean		.144	.123	.115	.100	.139	.110	.090	.117
Median		6.00	5.00	5.00	5.00	7.00	3.00	4.00	7.50
Std. Deviation		4.852	4.123	3.868	3.351	4.668	3.688	3.017	3.947
Minimum		0	0	0	0	0	0	0	0
Maximum		24	21	19	18	23	21	14	20
Percentiles	10	1	1	1	1	1		1	2
	20	2	2	2	2	2	1	2	4
	30	4	3.30	3	3	4	2	2	5
	40	5	4	4	4	5	3	3	7
	50	6	5	5	5	7	3	4	7.5
	60	7	7	7	5	8	4	5	9
	70	8	8	8	6	9	6	6	10
	80	10	10	9	7	11	7	7	11
	90	14	12	11	10	13	10	9	13

Revised scale reliability

Scale reliabilities were calculated using Cronbach's alpha (α). Cronbach's α is considered the most appropriate statistical test for reliability, given the dichotomous responses used to construct the scales. For dichotomous data, this is equivalent to the Kuder-Richardson formula 20 (KR20) coefficient. These statistics model internal consistency, based on the average inter-item correlation. These evaluations are a more rigorous approach than a traditional split-half statistic. Cronbach's α is a statistic bounded by 0 to 1. In general an α equal to or greater than .6 is considered a minimum acceptable level, although some authorities argue for a stronger standard of at least .7.

The following table compares the original SA1 reliabilities and Cronbach's α from the TriMetrix 2003 data that were utilized to set the distribution profile for the revised SA2 scales. All scale reliabilities have improved substantially, with the largest increases in the S and C scale. These findings document the revised SA2 as an instrument with solid scale construction and reliability.

Cronbach's α reliabilities				
	Original SA1 (Samples from N= 120,898, M=57%, F=43%)		SA2.d TriMetrix (N= 1130, F=49.5%, M=50.5%)	
Scale	Adaptive	Natural	Adaptive	Natural
Dominance-Challenge	.77	.81	.85	.84
Influence-Contact	.62	.69	.78	.79
Steadiness-Consistency	.65	.62	.78	.69
Compliance-Constraints	.54	.58	.74	.77

Scale relationships -- Correlations

Examination of the relationship among the scales focuses on two issues. First, the relationship of the "adaptive" scales, based on respondents' selection of "most like" phrases and the "natural" scales based on selection of "least like" phrases, has a theoretical foundation. While some may argue that the DISC scales are strengthened by simply combining these two dimensions, examination of the following correlation table supports a conclusion that these two dimensions measure subtle, but significant differences. Correlations between same scale adaptive and natural values range from .681 to .797. Based on observations made across each of response-populations, I judge that there is strong support for concluding that the natural scales are less prone to social desirability biases and variation due to the setting, environment, and responders' expectations.

Spearman rank order correlations among scales									
		Adap. D	Adap. I	Adap. S	Adap. C	Nat. D	Nat. I	Nat. S	Nat. C
AD	Correlation Coefficient	1							
	Sig. (2-tailed)	.							
AI	Correlation Coefficient	-0.089	1						
	Sig. (2-tailed)	0.003	.						
AS	Correlation Coefficient	-0.722	-0.320	1					
	Sig. (2-tailed)	0.000	0.000	.					
AC	Correlation Coefficient	-0.334	-0.622	0.337	1				
	Sig. (2-tailed)	0.000	0.000	0.000	.				
ND	Correlation Coefficient	0.797	0.090	-0.731	-0.390	1			
	Sig. (2-tailed)	0.000	0.002	0.000	0.000	.			
NI	Correlation Coefficient	0.027	0.734	-0.282	-0.640	0.077	1		
	Sig. (2-tailed)	0.369	0.000	0.000	0.000	0.010	.		
NS	Correlation Coefficient	-0.674	-0.257	0.711	0.368	-0.737	-0.330	1	
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	.	
NC	Correlation Coefficient	-0.461	-0.562	0.538	0.681	-0.600	-0.661	0.4492	1
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.0000	.
		Non-significant		S & C convergences			Adaptive & Natural agreement		

A second issue worth noting is that of the increased independence between the S and C scales as a result of editing items. The common variance in between the natural S and natural C is reduced to 20%, and the common variance between the adaptive S and adaptive C reduced to 11%.

Conclusions

I've reached the following conclusions with reference to TTI's *Style Insights* based on an analysis of response patterns from a diverse population of respondents.

- **Scores on the scales – while not a "statistically normal" distribution – are distributed with enough variance across all scale points to make interpretations and comparisons between respondents meaningful when interpreted as comparisons of individuals to distributions of these self-reported behaviors in a population.**
- **The items tend to co-vary around consistent latent-construct indicators of the theoretical concepts represented by instrument descriptions.**
- **The eight scales are constructed from coherent items with a solid reliability as indicated by Cronbach's α ranging from .7 to .8.**
- **The revisions presented in SA2 improve scale and item reliabilities significantly.**
- **Each of the ninety-six items used to construct the scales contributes in a significant way to one or both of the scale dimensions.**
- **The two dimensions of "adaptive" and "natural" contain parallel scales that are consistent with each other, but that also represent substantial potential for meaningful complimentary interpretation.**
- **The instrument is referenced in current populations, thus anchoring comparisons in the 21st century.**

With continued assessment and review of TTI's *Style Insights*, this revision initiates a process of continual quality improvement.

Submitted by:

Peter T. Klassen, Ph.D.
Principal – DocumentingExcellence.com
Professor Emeritus – College of DuPage

9 October 2003

Technical Information: *Style Insights* – DISC German version 2004d

Summary

Based on a series of examinations of scale reliabilities and item correlations across multiple populations of respondents, revisions were developed and tested for a German language version of the ***Style Insights****. The 2004d version is based on a sample of 881 responses from individuals using the instrument in business settings. This norming sample contained 237 females (27%) and 644 males (73%). The findings from assessments of this revised edition include improve reliability for the two dimensions (adaptive, natural) of four parallel scales (D,I,S,C) . Three of the eight scales' reliabilities have attained a Cronbach's α exceeding .8 and four additional scales are between .7 and .8 with one scale in the high .6 range. The scale reliabilities improved, and the instrument demonstrates a strong, reliable measure of consistent theoretical constructs. Correlations between adaptive and natural scales indicate that these two dimensions of parallel scales are highly related, as one would expect. Also, the parallel scales are sufficiently independent measurements to justify separate interpretations and comparisons. Scores on the scales are distributed across scale dimensions, which support making comparison between individuals and the self-reported behaviors in a population. Revision of the instrument included utilization of new population distributions that anchored comparisons in a population representative of the 21st century. Overall, the German version of the* ***Style Insights*** *is a strong, reliable instrument applicable across a variety of populations.*

Background

Style Insights **- German version 2004d is anchored in design and development of prior DISC instruments. During the late summer of 2002, TTI and Success Insights initiated a review of the reliability of the eight scales and the associated items for both their English and German language versions DISC instrument.**

The *Style Insights* instrument contains ninety-six (96) phrases organized in twenty-four frames of four items each. Each frame contains descriptive items associated with each of the four scale constructs. Respondents select a forced choice of "most-like" and "least-like" themselves. Two dimensions of four scales each are constructed from these responses. The two dimensions are adaptive and natural. The importance and utility of the two dimensions of scales is discussed later. The four scales are labeled as "Dominance-Challenge," "Influence-Contact," "Steadiness-Consistency," "Compliance-Constraints."

Scale reliabilities and item cohesion to its assigned scales were examined utilizing German language responses. Based on analysis of these findings, a limited number of items were revised, edited, and field-tested. Item revisions were based on theoretical construction of items combined with linguistic considerations that focused on current usage and minimization of social desirability bias. Thus, the German language version is based on independent assessment of response patterns and issues based in cultural and linguistic characteristics. The German version is more than a translation, it is a version tested and confirmed as applicable to those utilizing it.

The revised *Style Analysis* instrument was, prior to release, subjected to several rounds of field-testing, further editing and confirmation of revisions. The current analysis is based on a sample of 691 responses in which 43% were female and

57% were male that you provided as representative of the population using this instrument. The current release confirms increased reliability in each of the scales, and improves independence between the S and C scales.

The process of scale editing mandated a revision of population distributions. This process changed the reference point for comparison of style from its historic point of development up to the 21st century with recognition of changing behaviors and social expectations.

Norming sample

The pedigree of the current version has involved a diversity of data sources and samples. Current item and scale reliability is the culmination of these repeated evaluations using different samples. Thus, the instrument's pedigree is strengthened by repeated independent evaluations. The norms used in the current iteration, the *Style Insights* German language 2004d version, utilize a sample of 881 respondents compiled from users of the German language instrument during 2003 and 2004. The sample is compiled from two delivery systems – both of which are in active use among businesses in Germany, Switzerland, and Austria. The sample contained 237 females (26.9%) and 644 males (73.1%). Respondents' ages range from 20s through over 50. Occupations include managers, service workers, trade personnel, clerks, accountants, entrepreneurs, engineers, teachers, consultants and trainers. Thus, the sample represents a full range of individuals making use of the instrument in a variety of settings.

Evaluation of reliability is conducted using this sample of 881. For the purpose of calculating the distributions used as the current norms for comparison, an adjustment was applied which equalized the affects of patterns of responses among males and females. The two tables below summarize a comparison of the unadjusted and adjusted distributions.

Statistics Unadjusted for SEX		**Males = 73%, Females = 27%**							**29-Feb-04**
		Adaptive D	Adaptive I	Adaptive S	Adaptive C	Natural D	Natural I	Natural S	Natural C
Valid responses		881	881	881	881	881	881	881	881
Missing		0	0	0	0	0	0	0	0
Mean		7.369	6.664	5.005	4.173	-5.982	-3.555	-4.529	-9.037
Std. Error of Mean		0.170	0.140	0.136	0.118	0.146	0.109	0.097	0.133
Median		7	6	4	3	-5	-3	-4	-9
Std. Deviation		5.037	4.163	4.034	3.497	4.330	3.232	2.886	3.955
Minimum observed		0	0	0	0	-24	-18	-13	-20
Maximum observed		23	20	20	17	0	0	0	0
Percentiles	10	1	2	1	0	-12	-8	-9	-14
	20	3	3	1	1	-9	-6	-7	-12
	30	4	4	2	2	-8	-5	-6	-11
	40	5	5	3	2	-6	-3	-5	-10
	50	7	6	4	3	-5	-3	-4	-9
	60	8	7	5	4	-4	-2	-3	-8
	70	10	9	7	6	-3	-1	-3	-7
	80	12	10	8	7	-2	-1	-2	-6
	90	15	13	11	9	-1	0	-1	-4

Statistics adjusted for equal SEX distribution									29-Feb-04
		Adaptive D	Adaptive I	Adaptive S	Adaptive C	Natural D	Natural I	Natural S	Natural C
Valid responses		881	881	881	881	881	881	881	881
Missing		0	0	0	0	0	0	0	0
Mean		7.034	6.794	5.314	4.066	-6.413	-3.396	-4.350	-8.96
Std. Error of Mean		0.169	0.142	0.143	0.114	0.154	0.106	0.097	0.136
Median		6	6	4	3	-6	-3	-4	-9
Std. Deviation		5.007	4.226	4.240	3.376	4.560	3.138	2.888	4.023
Minimum observed		0	0	0	0	-24	-18	-13	-20
Maximum observed		23	20	20	17	0	0	0	0
Percentiles	10	1	2	1	0	-13	-8	-8	-14
	20	2	3	1	1	-10	-6	-7	-12
	30	4	4	2	2	-8	-4	-6	-11
	40	5	5	3	2	-7	-3	-5	-10
	50	6	6	4	3	-6	-3	-4	-9
	60	8	7	5	4	-4	-2	-3	-8
	70	9	9	7	5	-3.8	-1	-2	-7
	80	12	11	9	7	-3	-1	-2	-5
	90	14	13	12	9	-1	0	-1	-3

Reliability

Scale reliabilities were calculated using Cronbach's alpha (α). Cronbach's α is considered the most appropriate statistical test for reliability, given the dichotomous responses used to construct the scales. For dichotomous data, this is equivalent to the Kuder-Richardson formula 20 (KR20) coefficient. These statistics model internal consistency, based on the average inter-item correlation. These evaluations are a more rigorous approach than a traditional split-half statistic. Cronbach's α ranges in value from 0 to 1. In general an α equal to or greater than .6 is considered a minimum acceptable level, although some authorities argue for a stronger standard of at least .7.

Improvement in the reliability of the eight scales constructed in the *Style Analysis* as measured by Cronbach's α is clearly documented in the following table. The average German translation reliability in SA1 of .64 increased to .77 in the German version SA2. At this point, three of the eight scales' reliabilities have attained a Cronbach's α exceeding .8 and four of the scales are between .7 and .8 with one scale in the mid .6 range. Overall, the scale reliabilities improved, and the instrument is a strong, reliable measure of consistent theoretical constructs.

	SA2004.d German (N=881)	SA2003.c German (N=691)	SA1 German (N=6,318)		SA2004.d German (N=881)	SA2003.c German (N=691)	SA1 German (N=6,318)
Adaptive D	0.852	0.827	0.780	**Natural D**	0.828	0.819	0.782
Adaptive I	0.767	0.735	0.522	**Natural I**	0.755	0.774	0.613
Adaptive S	0.807	0.779	0.685	**Natural S**	0.687	0.663	0.600
Adaptive C	0.793	0.762	0.531	**Natural C**	0.772	0.803	0.608

Correlations among scales

The following table lists the correlations among the eight scales. This round of data confirms improved independence resulting in differentiation between the S and C scales. The S and C adaptive scales share 3% common variance and the S/C natural scales share 12%.

	Dominance-Challenge	Influence-Contact	Steadiness-Consistency	Compliance-Constraints	Dominance-Challenge	Influence-Contact	Steadiness-Consistency	Compliance-Constraints
AD	1							
AI	-0.014	1						
AS	-0.767	-0.266	1					
AC	-0.257	-0.672	0.146	1				
ND	0.768	0.116	-0.699	-0.296	1			
NI	0.073	0.685	-0.226	-0.599	0.075	1		
NS	-0.690	-0.171	0.705	0.176	-0.701	-0.212	1	
NC	-0.452	-0.586	0.437	0.668	-0.588	-0.684	0.353	1

Conclusions

I've reached the following conclusions with reference to the *Style Insights* German version 2004d based on an analysis of response patterns from a diverse population of respondents.

- **Scores on the scales – while not a "statistically normal" distribution – are distributed with enough variance across scale dimensions to make interpretations and comparisons among respondents meaningful when interpreted as comparisons of individuals to distributions of these self-reported behaviors in a population.**
- **The instrument is referenced in current populations, thus anchoring comparisons in the 21st century.**
- **The items tend to co-vary around consistent latent-construct indicators of the theoretical concepts represented by instrument descriptions.**
- **Three of the eight scales' reliabilities have attained a Cronbach's α exceeding .8 and four of the scales are between .7 and .8 with one scale in the mid .6 range.**
- **The revisions presented in *Style Insights improve* scale and item reliabilities significantly.**
- **The two dimensions of "adaptive" and "natural" contain parallel scales that are consistent with each other, but that also represent substantial potential for meaningful complimentary interpretation.**

This assessment and review of Success Insights' German version of the *Style Insights* initiates a process of continual quality improvement.

Submitted by:

Peter T. Klassen, Ph.D.
Principal - DocumentingExcellence.com
Professor Emeritus – College of DuPage

29 February 2004