

THE IMPORTANCE OF FINDING THE REAL PERSON
AT WORK

DR. DAVID M. WARBURTON, C.Psychol.

Department of Psychology
University of Reading
Whiteknights
Reading
RG6 6AL
United Kingdom

ABOUT THE AUTHOR

Dr. David Warburton has been a research professor at the University of Reading for over 20 years. He is a native of the United Kingdom, he obtained his B.Sc. from the University of London at Birkbeck College. Then, he went to the United States on an English Speaking Union Scholarship with his first Fulbright Grant.

At Indiana University, he obtained an A.M. and Ph.D. (with Highest Distinction). After which, he taught at the University of California at Irvine, before returning to the United Kingdom. Since being at the University of Reading, he has had another Fulbright Scholarship and so worked as a Visiting Professor at Indiana University and at the University of California at La Jolla.

He is currently the Director of the Human Psychopharmacology Group at the University of Reading and is working on psychological states and health most recently. His research interests in psychobiology have resulted in eight books and he has published over one hundred and eighty papers on various aspects of this area. He is an Editor for the "Psychopharmacology" and on the Editorial Boards of Behavioural Pharmacology, Current Psychology, Journal of Psychopharmacology and Neuropsychobiology.

He is a Chartered Psychologist and has been elected a Fellow of the British Psychological Society. He has acted as Technical Advisor to the World Health Organization, Technical Advisor to the British, Cyprus, Japanese, New Zealand, Taiwan and Venezuelan Governments, to the Surgeon-General of the United States, and to the Food and Drug Administration of the United States. He is Membre du Jury, Fondation de France pour la Recherche sur la Maladie d'Alzheimer.

He is the founder of ARISE (Associates for Research Into the Science of Enjoyment), which is devoted to studying the positive contribution of pleasure to health and to everyday life, in general. For this program, he has been elected as a Fellow of the Royal Society of Arts (FRSA) for "His Integration of Arts and Science".

In 1999, he was awarded the biennial "Dr. Rainer Wild-Preis" by the Executive Committee of Dr. Rainer Wild Foundation in Heidelberg. The citation was for "his contributions to the Science of Psychology, especially for his studies of wellness."

INTRODUCTION

It has long been recognized that the precision (reliability) and the accuracy (validity) of verbal instruments are determined to a large degree by the design and construction of the scales (Thurstone, 1928). In particular, the precision and accuracy of training instruments are determined by the extent to which measurement errors can be avoided. An instrument with low precision can cause Type II Errors by failing to detect real differences, while those with low accuracy may detect spurious differences that do not exist (Type I Errors). Type I Errors would cause problems for training programs, as an educational tool for personal understanding, while the avoidance of Type II Errors (failing to detect real differences) is essential for job selection.

In a broad sense, the quality of a scale describes its ability to faithfully reflect the trait being measured. The questionnaire designer must provide the person with the most suitable instrument for its expression and produce data in a form that can be readily collected and easily interpreted. A significant part of this construction involves the semantics of the scale - the adjectives, words, phrases, and sentence construction that produce response items which will elicit an accurate, internal response in the respondent. A second aspect is the response scale and the number of scale points for each item. If a response scale closely matches the internal response of the person, then there is a reduction in the measurement error, which might have been caused by forcing people to express their reactions through an inadequate scale. A final aspect is the minimization of error due to response biases¹.

These aspects will be considered with respect to the DISC Instrument. A version of this Instrument has been described in *The Universal Language DISC: A Reference Manual*. (Bonnstetter, Suiter and Widrick, 1993).

THE DISC INSTRUMENT

All of us have developed behavioral styles, which are represented as distinct cognitions, emotions and actions. The central core of our behavioral style tends to remain stable because it reflects our individual identities. However, the demands of the work setting often require different responses that evolve into a work behavioral style. In order to help a person understand themselves, we need to evaluate this disparity.

The DISC model of behavioral style helps people understand themselves and others in their work environment and other situations. By increasing their understanding of their behavioral style, they can identify the conditions, which are most conducive to their personal fulfillment. At the same time, people learn about the different, behavioral styles of colleagues and their ideal work conditions. This is the information which they require for maximum productivity and to build multiform, harmonious relations with others.

A major aim of the language of DISC is communication between people at work and in their home life (Bonnstetter et al, 1993). Contrary to statements in recent papers

¹ This is a whole topic in itself (Block, 1965), but only social desirability will be considered here.

which have been distributed (e.g. Lothian, 1998), DISC self-assessment is not designed as the exclusive instrument for personnel selection or for rejection of employees for promotion, nor should it be. Selecting employees is a complex process, which must assess attitudes, behaviors and skills. The DISC assessment compliments conventional, job selection methods and should not be used as the sole criterion, an issue to which I will return later with respect to psychometrics.

WHAT ARE THE IDEAL NUMBER OF SCALING CHOICES

When assessing behavioral style for whatever purpose, the clear aim is to match the instrument to the internal response of the people being assessed. Consequently, an important question for scale designers is how is the information being internalized by the person?

An important assumption in the field of attitude and trait measurement is the Principle of Reciprocal Evaluative Action and it is this assumption which allows measurement to be simplified (Cacioppo & Berntson, 1994). The Principle says that positively and negatively valent 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.

By assuming the Principle, the endpoints of the continuum are placed on a single (bipolar) scale, thus allowing traits to be measured. It was on the basis of this assumption that Likert (1932) proposed a simple bipolar scaling procedure, known as the method of summated ratings. People are presented with a set of alternative responses on an agree-disagree type of continuum. Because the final score is obtained by summing the appropriate scale value for each item, the scales are referred to as "summative scales".

Given a bipolar scale, we would like to know at a more specific level, whether the degree of agreement is continuous or discrete. If it is discrete, then we would like to know the number of scale points (discrete levels of response) that are being used by the individual. Thus, one objective of this paper is to discuss the number of scale points, which should be included in the response scale for each item.

Initially, Likert (1932) did not consider the number of choices to be an important issue and was more interested in the response scale properties, as Thurstone (1928) had been. Thus, Likert experimented with different weightings for the choices and concluded that "If five alternatives are used, it is only necessary to assign values from one to five with three assigned to the undecided position." Many researchers have adopted this sort of five choice scale where participants are asked to tick a box to state whether they "strongly agree", "agree", "undecided", "disagree", or "strongly disagree", in response to items. However, some researchers just choose the number of scale choices for the response items arbitrarily.

In the traditional DISC instrument (e.g. the Personal Insights Profile of Bonnstetter et al, 1993), the questionnaire contains 96 adjective items, which are framed in 24 blocks of four response items. Correlational analysis has grouped the adjectives into five sets

- those which load onto **D** for Dominance, **I** for Influence, **S** for Steadiness, **C** for Compliance and the remainder who do not correlate above 0.30 and are termed “blanks”. (I will return to the issue of “blanks” later).

The person is instructed to choose one MOST item (agree) and one LEAST item (disagree) selection from each of the 24 boxes, in order to describe themselves. Effectively, the response scale has only two points or three if neither MOST nor LEAST is counted as a neutral point. A major consideration is the extent to which this response choice scaling can represent the internal structure of the four behavioral styles for a person.

There are at least three aspects to the issue of response scaling. First, there is the balance between the number of response choices and number of scale points. Second, there is the relative importance of direction of response (polarity) and strength of agreement or disagreement (intensity). A third consideration is the question of response cost, the influence on a person’s accuracy of responding of the length of the questionnaire.

AN IDEAL NUMBER OF RESPONSE ITEMS AND SCALE POINTS?

Likert addressed the implications of the number of response items for scale quality, by comparing the balance between the number of items and the number of response choices in an overall measure of instrument quality, split-half reliability². (Murphy and Likert, 1938). Their hypothesis was that the optimum number of response choices on a scale (scale points) does not exist independently of the number of response items.

Murphy and Likert found that when the total number of response items was reduced from 44 to 36, there was a decrease in the split-half reliability. This decrease in split-half reliability could be reversed by increasing the number of scale points from three to five, so that the split-half reliability rose from 0.88 to 0.94 as a consequence. They concluded that this increase was due to the fact that the total number of response choices for the participant were increased in reality from 132 (44 times 3) to 180 (36 times 5). In other words, reliability was a function of the *total* number of response choices in the instrument. The total number of response choices was computed as the intervals per scale (scale points) multiplied by the number of response items.

This hypothesis, that the number of scale points can be balanced against the number of response items to achieve high reliability, was supported by Bendig (1954), Komorita (1963) and Komorita and Graham (1965). However, this balance was only important with a small number of response items, which gave a total number of response choices of less than 50. Over this number, the number of scale points per item was not important for reliability.

As the DISC Instrument has 96 items which are divided into 24 blocks of four items, there is a MOST, a LEAST and a neither MOST nor LEAST response choice for 24

² An estimate of the reliability of an instrument, which is obtained from the correlation between a pair of scores which are obtained from a division of the test into equivalent “halves”, for example random assignment of items to the two data sets.

blocks, i.e. a total of 72 response choices. This leads to the conclusion that the precision of the DISC Instrument does not depend on the number of scale points, which are used for evaluation.

Summary

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. The DISC Instrument exceeds this minimum.

RELATIVE IMPORTANCE OF POLARITY AND INTENSITY

In this Section, I will consider the relative importance of the direction of response (polarity) and the strength of agreement (intensity).

One of the first studies to look at the relative importance of the polarity of scoring versus the number of scale points was Peabody (1962). He separated the Likert scale into a direction or polarity component (agree or disagree) and extremeness or intensity component (degree of agreement or disagreement), in order to determine the part played by intensity in determining scale quality. He reported the dominance of the directional component of the Likert scale over the intensity component for determining the composite score measures.

His conclusion was:

“The results indicate 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, Pp 73)

This conclusion was later supported by other researchers (Komorita; 1963; Komorita and Graham, 1965). These authors also separated the Likert scale into a polarity component (agree or disagree) and an intensity component (degree of agreement or disagreement). Using split-half correlation and Cronbach Alpha coefficients³ as measures of consistency they found that when the number of items is large, the intensity component was unimportant. These findings imply that internal consistency is not affected by the number of choices beyond the minimum of two needed to establish polarity.

The findings of Komorita and Peabody were generalized by Jacoby and Matell (1971) and Matell and Jacoby (1972) to include other measures of scale quality. They confirmed the dominance of dichotomy in determining reliability and the relative lack of importance of additional scale choices to assess intensity. They showed that this unimportance also applied to test-retest reliability, concurrent validity, and predictive validity. The conclusion of Jacoby and Matell was that a three choice scale was sufficient, with agree, disagree and uncertain.

³ Cronbach's Alpha Coefficient can be thought of as the mean of all possible split-half correlations (Cronbach, 1951).

This conclusion supports the use of MOST, LEAST and neither MOST nor LEAST for the DISC Instrument, with MOST and LEAST to indicate directionality.

This MOST/LEAST approach views behavioral style not as points on a continuum, but as categories, a conception which stems from work on knowledge representation in cognitive psychology⁴. Personal knowledge is thought of as being organized into a hierarchy of cognitive elements and the relations between them. For example, a trait like Dominance, will comprise beliefs, evaluations and behaviours, and the relations between those three cognitive elements. The content of these elements, as well as the relations between them, will be different categorically from those relevant to another behavioral style.

Defining traits as categories means there is a prototypical set of beliefs, evaluations, and behaviours that identify a specific behavioral style. Of course, there are no “pure” Dominant individual, a **D** 24⁵ as in the example by Lothian (1998), we are all a combination (Bonnstetter et al, 1993).

Summary

Altogether these findings justify the use of a two-point or dichotomous agreement scale of either agree (MOST) or disagree (LEAST) for the DISC Instrument, indicating the polarity of responding with respect to a behavioral style⁶.

RESPONSE COST

An additional consideration is response cost; fewer choices are more economical and easier to use. In general, an increase in quality beyond two or three choices is not justified by the complexity of the instrument (with increased possibilities for error) and the additional cost of the data gathering process. Thus, the addition of additional response scale points must be justified by an increase in reliability⁷.

A PROPOSED ALTERNATIVE SCORING SYSTEM

It is on the grounds of increased reliability that Lothian (1998) has argued against the MOST/LEAST Scoring method. One major criticism is that two items in each block of four are being ignored in the traditional scoring method. Instead, he proposed an Alternative Scoring System (ASS) in his pamphlet “DISC and MOST/LEAST ONLY Questionnaires: Does An Alternative Scoring System Generate More Accurate

⁴ See discussion for attitudes in Latane and Nowak (1994)

⁵ In fact, this example is spurious, because the maximum number of D items is 20, because four are D blanks, which will be discussed in the next Section.

⁶ Three other notable examples of dichotomous trait measures are the Minnesota Multiphasic Personality Inventory (Dahlstrom, Welsh and Dahlstrom, 1972), the Sixteen Personality Factor Questionnaire (16PF) of Cattell, Eber and Tatsuoka (1980) and the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975)

⁷ However, Cronbach (1950) stated from his own studies that “The argument that the finer scale gives more reliability is not a sound one, since this is precisely what we would expect if all of the added reliable variance were response-set variance and had no relation to beliefs about the attitude-object in question. (Pp 22). In other words, intensity scores largely represent response sets.

Outputs?" (Lothian, 1998). With the ASS, participants are required to select firstly MOST in each frame (weighting one); then LEAST in each frame (weighting zero); and also to grade the remaining two items from each frame into second place (weighting 2/3) and third place (weighting 1/3). Not surprisingly, different hypothetical graphs are produced.

However, there are no data, which indicate their importance for personal evaluation. For example, did participants "see" themselves as more like the ASS-rated person or the traditionally-evaluated person? Certainly, it is an important to know how well the four point scale represented the internalized structure of behavioral style of the person. MOST would represent "strongest agreement" and LEAST "strongest disagreement". Second place could be interpreted as "strong agreement, but not strongest agreement" or "uncertain or weak disagreement". Similarly, third place could mean "uncertain or weak disagreement" or "strong disagreement, but not strongest disagreement".

In psychometric terms, we cannot legitimately consider these different ratings as necessarily interval in nature. A rating of one on the scale may not be viewed by respondents as three times greater than a rating of one third. Similarly, a rating of two thirds is not necessarily double one third. Therefore, a summed score of a MOST item (one) and a Least item (zero) is not equal to a second place item (2/3) and a third place item (1/3). The scale is an ordinal one, and all we can say about the Lothian scoring system is that MOST is greater than the second place and LEAST is less than third place⁸.

Of course, the MOST/ LEAST Scoring System does not have this interval scaling problem, because the respondents only make polarity judgements. These polarity judgements are summed and plotted to produce the graphs for the pencil and paper version. The computer reports are based on item analyses rather than summed data. The person's responses are condensed for the computer report, because of the enormous number of different combinations of "most" and "least" descriptor items. Nevertheless, the outcome of the condensed item analyses by the computer is 384 different reports (Bonnstetter et al, 1993).

Summary

In summary, the research in the earlier Section, would indicate that an Alternative Scoring System would not have any significant advantage over a MOST/LEAST, forced choice. The disadvantages of the four point scale are in terms of the interval scaling problem. The MOST/LEAST Method does not depend on interval scaling and the reports for the computer version are based on item analyses not simple, summed scores.

In terms of the increased complexity of the responding with its opportunities for errors and the extra time taken to complete the instrument (response cost), the disadvantages of ASS are considerable. The response cost problem is magnified as the number of

⁸ It is worth mentioning in passing that for the "Insights *Discovery* Preference Evaluator" © Andrew Lothian, Insights, uses an even more complex system of a seven point, rating scale (See Appendix).

response decisions is increased as in the Insights *Discovery* Preference Evaluator (See Appendix).

SOCIAL DESIRABILITY

A pervasive issue of the measurement errors in the assessment of traits, like behavioral styles, has been to what extent do people endorse favorable items on questionnaires (and reject unfavorable items), i.e. social-desirability responding (e.g. Block, 1965;; McCrae and Costa, 1983), bluntly to “fake good” (Dahlstrom, Welsh and Dahlstrom 1972; Eysenck and Eysenck, 1975). This can be in the form of impression management or self-deceptive enhancement and self-deceptive denial (Paulhus, 1991; Paulhus and Reid, 1991).

Although positivity biases may not be a tremendous problem of measurement error for a group, there is the possibility for individuals to distort (a very important factor for a personal training instrument or selection instrument). An extreme example would be individuals with a DSM IV definition of Narcissistic Personality, which includes a grandiose sense of self-importance and a tendency to exaggerate talents (American Psychiatric Association, 1994; Pp 661). However, this disposition is present to some extent in most people, except for some evangelical missionaries.

The development of DISC by John Geier at the University of Minnesota followed the usual format for measuring traits using a questionnaire containing a set of many descriptor items⁹, from which a smaller number were chosen for scoring on the final instrument on the basis of their correlation with the four behavioral styles.

Correlational analysis of every item with the pure factors is required to demonstrate that the scales correspond to uniquely definable, functionally unitary, and psychologically significant source traits. A correlation cut-off of 0.3 (accounting for nine percent of the variance) has been used for DISC. Thus DISC Instrument is not a questionnaire, which is composed of arbitrary items, but have been carefully oriented and groomed with respect to the primary traits of **D**, **I**, **S** and **C**.

The summed responses to the selected descriptor items serve as the trait measure. However, the poorly correlated items have been left in the scale. This was because Geier was concerned about the “Fakeability Factor”, i.e. sociability desirability bias. This is not surprising, because the Minnesota Multiphasic Personality Inventory (Dahlstrom, Welsh and Dahlstrom, 1972) was one of the first instruments to include a Lie Scale to enhance its validity.

Later, the “Fakeability Factor was labelled in a less defamatory way as the “N Factor” - for neutral factor (O’Connor, 1987) and the items are termed “blanks” in the Bonnstetter et al (1993) Reference Manual. Inspection of the blank items do give some evidence of their usefulness for indicating sociability desirability bias. From six of what Lothian (1998) has called “C blanks” are “Easily led, follower”, “Fussy, hard to please, “Fearful, afraid”, “Tolerant¹⁰”, “Cultured, educated, knowledgeable”, and “Accurate, correct”, it is easy to see how a sociability desirability bias could lead to

⁹ Allport and Odbert (1936) found that there are at least 17,953 personality trait names and terms in the English language. However, only about a quarter were “real traits of personality, i.e. terms and words which were generalized and personalized determining tendencies.

¹⁰ For example, “I am tolerant” is a Lie Scale item on the MMPI.

the endorsement or rejection of these statements. The Bonnstetter et al (1993) Reference Manual points out that the use of the blanks to indicate faking has not been validated for DISC, but it is a legitimate concern.

As a rule of thumb, O'Connor (1987) recommended that "Zero to 1 or 2, even 3 N's are not so bad. With 4 or 5, get nervous. With 6 N's - definitely forget it." At the very least 25 percent of the data is missing and at the worst 25 percent of the answers are biased. In either case, the profile will be inaccurate. In training, social desirability may not be an important bias, but it can be a significant factor when the instrument is used as part of a selection procedure (Eysenck and Eysenck, 1975). Certainly, it indicates a need for caution in interpreting the pattern of data

The DISC instrument (Bonnstetter et al, 1993) has also left the poorly correlated items ($r < 0.30$) in the questionnaire, but they are not used in the trait scoring. The result is discarding seven D items, 12 I items, ten S items and 17 C items from the MOST and LEAST categories.

Lothian (1998) comments on this procedure

"There is a concerning feature about these blanks. Since they are not good items they should not really be there. Their presence and selection may have stopped the subject scoring on one of the other good dimensions." (Pp 10).

The reason that they are included in the DISC Instrument, is because eliminating them would alter its psychometric structure; it would not be the same instrument, which has been validated in many studies world-wide.

In the ASS, Lothian scores all items and plots graphs using 'blanks' as if they were valid responses to style items and shows that "The incidence of blank questions influences the score." (Pp 10). It may well do. However, he seems not to understand that "blanks" have not been assigned to any behavioral style, because they did not correlate highly enough to load onto any of the styles (less than 0.3).

Using scores from these items will result in measurement errors. Using a scoring system of this sort may result in Type I Errors (detection of differences that do not exist) which would cause problems for training. In job assessment, the elimination of Social Desirability Bias from measurement error is crucial for the avoidance of Type II Errors (failing to detect real differences) for selection. High scores on the blank items may indicate impression management, and self-deceptive enhancement and self-deceptive denial.

Summary

In summary, the user of the DISC Instrument cannot score all 24 items in the D, I, S and C categories. This over-inclusiveness would score items ("blanks"), which did not load highly enough on any behavioral style. Inclusion of the "blanks" will result in significant measurement errors.

INTERPRETATION OF SUMMED SCORES

A consequence of the elimination of blanks is that there are different numbers of items which are scored on the MOST Graph (Graph I) and LEAST Graph (Graph II). Therefore, a summed score of six on each style are not equivalent for each style on the same graph or for the same style on Graphs I and II. Consequently, the raw scores are converted to percentages or plotted on different scales (See Bonnstetter et al, 1993). When the items are summed and converted, the outcome is a rank order of the four behavioral styles for each respondent.

Technically, the DISC scales from the Personal Insights Profile are ipsative¹¹ and are subject to the limitations of ipsative measurement (Horst, 1965). An ipsative score does not represent strength of a trait in absolute terms, but rather the strength of that trait *in relation* to the strength of other traits. Thus, ipsative measures are intrinsically within-subject measures and the scale legitimately only produces a rank order of the four behavioral styles for that individual.

It is important that ipsative judgements correspond to the internal thinking of a person about themselves. Lamiell (1980; 1987; 1997). When formulating and expressing judgements about their own and one another's personality characteristics, lay people do not rely on normative reasoning processes, i.e. their judgements are not framed by contrasting themselves with others. Instead, they operate with a conception of what they are not, but might otherwise be, i.e. a conception of the *negation* of themselves (Lamiell, 1997). Related research on the implicit standards that individuals use when they rate themselves supports this view. Ipsative self-assessment ratings were more similar to their own (implicit) standards than normative ratings (Chaplin and Buckner, 1988). In other words, ipsative judgements corresponded more closely to person's view of themselves.

An ipsative scale cannot be used to compare individuals, except for the following.

- (1) The scores for a single behavioral style can be used to make valid, between-person comparisons, e.g. for the primary behavioral style.
- (2) The scores on a single behavioral style, e.g. one which was ranked as most (or least) important, can be used to categorize respondents.

However, a major use of DISC assessment is the evaluation of the behavioral style pattern of a person in terms of the Point Spread, the difference in points between each of the six relationships on the Graphs, starting with the differences from the primary behavioral style, i.e. the highest plotting point above the center line (Bonnstetter et al, 1993; Pp 110-112). Lothian (1998) makes a big issue of the fact that the center line is not the same percentage for each style on the graph. However, the scales are ipsative and must be considered in relation to one another for an individual. Hence, identification of the largest Point Spread to indicate the strongest tendencies, which give the most accurate descriptors of behavior (Bonnstetter et al, 1993; Pp 115).

¹¹ The ASS also results in ipsative scales

Later, Bonnstetter et al (1993) discuss unusual graphs, where there are no large strong points spread, i.e. tight graphs. A tight graph would be obtained by the raw scores for **D 6, I 6, S 6** and **C 6**, which was given as a hypothetical example by Lothian (1998). After ruling out obvious factors, like erroneous scoring and failure to understand the words, unusual graphs can indicate an individual “attempting to outsmart the instrument” (Bonnstetter et al 1995; Pp 134)!!!.

A final important step in a profile analysis is examining the disparity between Graph II (Core Behavioral Style) and Graph I (Responses to the Work Environment). Again this is carried out at an ipsative level. As an educational tool, disparities are important for the organization and for the individual. A study of 150 middle managers was conducted using the DISC measure (percentage points) of Primary Behavioral Style Disparity only (Warburton and Suiter, 1993). It was found that DISC disparity was predictor of problems at work and at home. It was related to the amount of alcohol use, physical health, mental health, job satisfaction and absenteeism. The immense physical and financial costs of job dissatisfaction have been documented for the individual and the organization (Warburton and Suiter, 1996).

Summary

The outcome of the DISC assessment is a rank order of the four behavioral styles on Graph I and Graph II for that individual, ipsative scaling.

Studies have shown that ipsative judgements correspond to the way people view themselves.

The DISC scores are used to make within-person assessments in terms of the, which is permissible with ipsative scaling. The interpretation of these scores can then be used for personal training, group understanding and the detection of job incompatibility, which would predict physical and mental health problems

CONCLUSIONS

(1) The precision (reliability) and the accuracy (validity) of verbal instruments are determined to a large degree by the design and construction of the scales, which will avoid measurement errors. An instrument with low precision can cause Type II Errors by failing to detect real differences, while those with low accuracy may detect spurious differences that do not exist (Type I Errors). Type I Errors would cause problems for training programs, as an educational tool for personal understanding, while the avoidance of Type II Errors is essential for job selection.

(2) A significant part of this construction involves the semantics of the scale - the descriptor items which will elicit an accurate, internal response in the respondent. Valid scales, like DISC, have been through a process of refinement to ensure that only items which meet a correlational criterion (typically greater than 0.03 as in DISC). While the DISC Instrument includes the poorly correlated descriptors, they are not scored, because it would result in significant Type I and Type II Errors. High scoring

on these items indicates a need for caution in interpretation, because the individual may have a conscious or unconscious social desirability bias.

(3) A second aspect is the response scale and the number of scale points for each item.

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. The DISC Instrument exceeds this minimum. Further, research findings have shown that direction of responding (polarity) is more important than intensity of agreement or disagreement (the number of scale points). Consequently justify the use of a two-point or dichotomous agreement scale of either agree (MOST) or disagree (LEAST) for the DISC Instrument.

Suggestions that an Alternative Scoring System (ASS) with more scale points would have any significant advantage over a MOST/LEAST, forced choice are unfounded. Increasing the number of scale points introduces the interval scaling problem. In terms of the increased complexity of the responding with its opportunities for errors and the extra time taken to complete the instrument (response cost), the disadvantages of ASS are considerable.

(4) If a response scale closely matches the internal response of the person, then there is a reduction in the measurement error, which might have been caused by forcing a people to express their reactions through an inadequate scale. The outcome of the DISC assessment is a rank order of the four behavioral styles for that individual in terms of ipsative scaling.

Studies have shown that ipsative judgements correspond to the way people view themselves. Consequently, the DISC scores can be used with confidence to make within-person assessments for personal training, mutual group understanding and the detection of job incompatibility, which would predict physical and mental health problems.

ACKNOWLEDGEMENT AND DISCLAIMER

Preparation of this Report has been aided by a donation from Target Training International Ltd. Nevertheless, the views which are expressed here are solely those of the Author and are not necessarily those of Target Training International Ltd. or of any of its Companies world-wide.

REFERENCES

Allport, G.W. and Odbert, H.S. (1936) Trait-names: a psycholexical study. Psychological Monographs: General and Applied, 47, whole of No. 211.

American Psychiatric Association (1994) Diagnostic and Statistical Manual (DSM IV) Washington, DC: American Psychiatric Association.

Bendig, A.W., (1954) Reliability and the number of rating scale categories. Journal of Applied Psychology, 38, 8-10.

Block, J (1965) *The Challenge of Response Sets*. Springfield, IL: Charles C. Thomas.
Bonnstetter, W., Suiter, J.I., and Widrick, R. (1993) *The Universal Language DISC: A Reference Manual*. Scottsdale, AZ: Target Training International.

Cacioppo, J. T. and Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115, 401-423.

Cattell, R.B., Eber, H.W., and Tatsuoka, M.M. (1990) *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing, Inc.

Chaplin, W.F. and Buckner, K.E. (1988) Self-ratings of personality: a naturalistic comparison of normative, ipsative, and idiothetic standards. *Journal of Personality*, 56, 509-530.

Cronbach, L.J. (1950) Further evidence on response sets and test design. *Educational Psychology and Measurement*, 10, 3-31.

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

Dahlstrom, W.G., Welsh, G.S., and Dahlstrom, L.E. (1972) *An MMPI Handbook Volume 1 - Clinical Interpretation*. Minneapolis, MN: University of Minnesota.

Eysenck, H.J. and Eysenck, S.B.G. (1975) *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.

Horst, P. (1965) *The Factor Analysis of Data Matrices*. New York: John Wiley.

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

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

Lamiell, J.T. (1980) On the utility of looking in the "wrong" direction. *Journal of Personality*, 48, 82-88.

Lamiell, J.T. (1987) *The Psychology of Personality: An Epistemological Inquiry*. New York: Columbia University Press.

Lamiell, J.T. (1997) Individuals and the differences between them. In Hogan, R., Johnson, J., and Briggs, S. (Eds.) *Handbook of Personality Psychology*. San Diego: Academic Press.

- Latane, B. and Nowak, A. (1994). Attitudes as catastrophes: From dimensions to categories with increasing involvement. In R. R. Vallacher & A. Nowak (Eds). *Dynamically Systems in Social Psychology*. San Diego, CA: Academic Press.
- Likert, R. (1932) A technique for the measurement of attitudes. *Archives of Psychology*. 140, whole issue.
- Lothian, A (1998) DISC and MOST/LEAST ONLY Questionnaires: Does An Alternative Scoring System Generate More Accurate Outputs. Presentation at Business Psychology Conference, London, UK
- Matell, M. and Jacoby, J. (1971) Is there an optimal number of alternatives for Likert items? Study 1: Reliability and validity. *Journal of Educational and Psychological Measurement*, 31, 657-674.
- Matell, M. and Jacoby, J. (1972) Is there an optimal number of alternatives for Likert items? *Journal of Applied Psychology*. 56, 506-509.
- McCrae, RJR, and Costa, PT (1983) Social desirability scales: more substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882-888.
- Murphy, G., and Likert, R. (1938) *Public Opinion and the Individual: A Psychological Study of Student Attitudes on Public Questions With a Re-test Five Years Later*. New York, NY: Harper Books.
- O'Connor, M. (1987) *The disc Needs Model: The Professional Trainer and Consultant's Reference Encyclopaedia*. Minneapolis, MN: Perform Systems International.
- Paulhus, DL (1991) Measurement and control of response bias. In Robinson, JS, and Shaver, PR, and Wrightsman, L.S. (Eds.) *Measures of Personality and Social Attitudes*. San Diego, CA: Academic Press.
- Paulhus, DL and Reid, D. (1991) Attribution and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317.
- Peabody, D. (1962) Two components in bipolar scales: direction and extremeness. *Psychological Review*, 69, 65-73 .
- Target Training International (1994) *Personal Insights Profile™*. Scottsdale, AZ: Target Training International Ltd.
- Thurstone, L.L. (1928) Attitudes can be measured. *American Journal of Sociology*, 33, 529-554.
- Warburton, D.M. and Suiter, J.I. (1993) *Discovering the person at work*. In Bonnstetter, W., Suiter, J.I., and Widrick, R. (1993) *The Universal Language DISC: A Reference Manual*. Scottsdale, AZ: Target Training International

Warburton, D.M. and Suiter, J.I (1996) The costs of job dissatisfaction. In Warburton, D.M. and Sherwood, N. (Eds.) *Pleasure and Quality of Life*. New York: John Wiley.

APPENDIX

It is instructive to use the Conclusions on validated DISC to evaluate a new questionnaire, the “Insights *Discovery* Preference Evaluator”¹².

The instructions for the “Preference Evaluator” are (1) “Select the one statement that MOST describes you in your work environment and circle **M** next to it.”; (2) “From the remaining three statements, select the statement that LEAST describes you in your work environment and circle **L** next to it.”; and (3) “From each of the next two statements circle a weighting from the values 1, 2, 3, 4 and 5, where 1 represents ‘not likely to describe me’, and 5 represents ‘very likely to describe me’. *Please do NOT choose the same weighting twice.*”

This is an example from the Preference Evaluator, with the choices struck through.,

Careful and evaluating	L 1 2 3 4 5 M
Constant and attached	L 1 2 3 4 5 M
Influencing and gesturing	L 1 2 3 4 5 M
Strategic and pushing	L 1 2 3 4 5 M

Effectively, a sort of seven choice scale has been produced with the person being forced to choose two extreme ratings, M and L and grading the remaining two items of the block of four into second and third place on a five point scale. As in ASS, the different ratings are not necessarily interval in nature. A rating of three on the scale may not be viewed by respondents as three times greater than a rating of one. Therefore, a summed score of nine (three, three and three) is not the same as nine (five, three and one). The scale is an ordinal one, and a rating of three can legitimately be viewed only as greater than one. Like the ASS, it is a unique hybrid of two scoring systems, with two items rated dichotomously and two items scaled. These types of responses cannot be combined in a single graph.

As in all scales of this kind, it is hard to interpret the middle categories. Often, they are used when the person cannot understand the item, or the connection between adjectives. Many of Lothian’s adjective pairs in the Insights *Discovery* Preference Evaluator fall into this category. In fact, they appear to have been derived arbitrarily and no justification for their source has ever been given. It is instructive to examine the descriptor pairs in this block, the first block of his questionnaire, in order to demonstrate their capricious nature.

(1) “Careful and evaluating”, combines a C item of “Careful” from the Personal Insights Profile of DISC (Bonnstetter et al, 1993) with an adverb implying appraising¹³.

¹² © Andrew Lothian, Insights.

¹³ Roget’s Thesaurus

(2) “Constant and Attached are only related, because both are synonyms of loyal¹¹, but not of each other. Loyal is an **S** item on the Personal Insights Profile of DISC.

(3) The adverbs “Influencing and Gesturing” seem to be designed as **I** descriptors, since **I** stands for the trait Influencing on the Personal Insights Profile of DISC. However, is “Gesturing” a substitute for “Demonstrative”? It bears only a tenuous relationship to it and it certainly does not combine with “influencing” in a meaningful way.

(4) The most arbitrary combination of this block is the fourth descriptor - “Strategic and Pushing”. By a process of elimination, we must suppose that the descriptor combination of “Strategic and Pushing,” is assessing **D**. But what is the association between strategic and pushing? Is “Strategic” assumed to correspond to “Aggressive” on DISC? “Strategic” is associated with warlike and antagonistic in a Thesaurus¹¹. It means “serving the ends of strategy” and “materials for fighting a war”. But no way does it correspond to any classical **D** descriptors. “Pushing” can mean “aggressively ambitious”, but it hardly fits with “Strategic”.

COMMENT

The only conclusion, which can be drawn from an examination of this questionnaire is;

(1) Because of the semantics of the scale, the descriptor items will NOT elicit an accurate, internal response in the respondent. It has not been validated psychometrically by a process of refinement to ensure that only items which meet a relevant, correlational criterion are included. Consequently it will result significant Type I and Type II Errors.

(2) A second problem is that the number of scale points for each item is seven. This number introduces a clear interval scaling problem. In addition, the increased complexity of the responding with its opportunities for errors and the extra time taken to complete the instrument, the **DISADVANTAGES** of this questionnaire are **SUBSTANTIAL**.

(3) The third problem is that the combination of invalidated items and large number of scale points ensures that the response scale cannot claim to match the internal response of the person. As a result, the Insights *Discovery* Preference Evaluator **CANNOT** be used with assurance for personal training, mutual group understanding and the detection of job incompatibility.

(4) The value of the “Insights *Discovery* Preference Evaluator”¹⁴ questionnaire, as a person evaluation instrument, has **NOT** been established.

¹⁴ © Andrew Lothian, Insights.