

AD-R103 025

A FREE-RESPONSE METHOD OF COMPUTER ASSISTED PERSONALITY 1/1  
ASSESSMENT: A RESEARCH UPDATE(U) OFFICE OF NAVAL  
RESEARCH LONDON (ENGLAND) W D CRANO 00 JUL 87

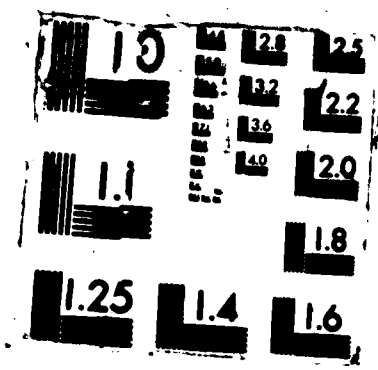
UNCLASSIFIED

ONRL-7-016-R

F/G 5/8

NL







DTIC FILE COPY

①

# ONRL Report

7-016-R

AD-A183 825

A Free-Response Method of Computer Assisted  
Personality Assessment: A Research Update

William D. Crano

8 July 1987

DTIC  
ELECTE  
AUG 04 1987  
S D  
E

Approved for public release; distribution unlimited

U.S. Office of Naval Research, London

87

~~7 81 094~~  
8 4 062

RE

AD-A183 825

1a. REPORT SECURITY CLASSIFICATION  
UNCLASSIFIED

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)  
7-016-R

6a. NAME OF PERFORMING ORGANIZATION  
US Office of Naval Research  
Branch Office, London

6b. OFFICE SYMBOL  
(If applicable)  
ONRL

6c. ADDRESS (City, State, and ZIP Code)  
Box 39  
FPO, NY 09510

8a. NAME OF FUNDING/SPONSORING ORGANIZATION

8b. OFFICE SYMBOL  
(If applicable)

8c. ADDRESS (City, State, and ZIP Code)

3. DISTRIBUTION/AVAILABILITY OF REPORT  
Approved for public release;  
distribution unlimited

5. MONITORING ORGANIZATION REPORT NUMBER(S)

7a. NAME OF MONITORING ORGANIZATION

7b. ADDRESS (City, State, and ZIP Code)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

A Free-Response Method of Computer Assisted Personality Assessment: A Research Update

12. PERSONAL AUTHOR(S)  
William D. Crano

13a. TYPE OF REPORT  
Technical

13b. TIME COVERED  
FROM \_\_\_\_\_ TO \_\_\_\_\_

14. DATE OF REPORT (Year, Month, Day)  
8 July 1987

15. PAGE COUNT  
6

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	10	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Reliability  
Validity  
Personality  
Assessment  
Selection

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The free-response methodology, first reported by ESN in early 1984, has been tested and extended, and the results appear to support the high initial expectations. Recent developments in the techniques are reported, and the author discusses means that could be employed to further develop this important measurement method.

15

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT  
 UNCLASSIFIED/UNLIMITED  SAME AS RPT  DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION  
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL  
C.J. Fox

22b. TELEPHONE (Include Area Code)  
(44-1) 409-4340

22c. OFFICE SYMBOL  
11

## A FREE-RESPONSE METHOD OF COMPUTER ASSISTED PERSONALITY ASSESSMENT: A RESEARCH UPDATE

Not long ago in these pages, Richard Snow described some intriguing research being conducted in Belgium on the development of a free-format response measure of personality (ESN 38-4:163-165 [1984]). In a collaborative effort involving psychologists from the University of Leuven (Belgium) and the psychological research center of the Belgian Army, civilians and army officers served as research subjects in a series of investigations designed to develop and to explore the validity of the free-format approach. The work was in an initial stage when Snow described it, but even then it appeared promising. Over the past few years, the methodology has been tested and extended, and the results appear to support the high initial expectations of the developers of the approach. Indeed, the research results derived from the free-format measurement system now play a role in the officer selection process of the Belgian Army.

The test constructors were motivated by a dissatisfaction with traditional personality inventories that forced respondents to use pre-set categories to describe themselves. Some of the descriptors to which subjects responded seemed to the researchers to be obviously irrelevant. Such irrelevancies, of course, would adversely influence both the reliability and validity of the assessment device. To remedy the problem of "descriptor irrelevancy," the researchers decided to allow the respondents to describe themselves, in their own words. They assumed that these self-descriptions would more precisely reflect the true state of the respondent, i.e., that people had sufficient self-insight to be able to describe themselves in a meaningful and coherent fashion. A recent validity study at the University of Leuven appears to have supported this assumption (Claeys et al., 1985). In the Claeys et al. study, a free-response method was employed in conjunction with established forced-choice personality measures, and the validity coefficients obtained in the

resulting comparison were of acceptable magnitude.

In a recent meeting with Dr. Arnold Bohrer (Psychology Section, Belgian Army), I learned more about recent developments in the free-format personality measurement technique. I will describe these developments on the following pages. But first, let me provide a brief description of the general measurement approach.

In using the free-format method, the respondent is told: "Describe your personality, as completely as possible, listing 10 adjectives. Do not say how you want to be, but say how you really are. Try to use words of common usage."

Although the demands on the subject are relatively minimal, the procedure used to assign scores to responses is the result of considerable effort. In developing the scoring system, Bohrer and his colleagues reasoned that a personality-descriptive term might be conceptualized along any number of personality-relevant dimensions. For example, the self-description "friendly" carries some information regarding the respondent's degree of introversion-extroversion, his or her level of neuroticism, conscientiousness, etc. This "incidental" information is very useful, since it allows for more fine-grained differentiations among respondents.

To develop the response norms to be used in scoring self-descriptions, the researchers assembled a large number of trait-descriptive adjectives (3000 in the initial Dutch version), and asked judges to rate these descriptors along seven dimensions, which include extroversion, agreeableness, conscientiousness, neuroticism, open-mindedness, leadership, and creativity. In addition, each term was also judged in terms of its social desirability. Five categories were available for each judgment. So, for example, if evaluating the self-descriptive term *friendly* on the extroversion dimension, the judge could decide that the response indicated that the respondent was "very extroverted," "extroverted to some degree," "uncertain, or nothing can be said about extroversion on the basis of the

term," "introverted," or "very introverted." Overall, 3000 personality-descriptive terms were rated by a total of 40 different judges (10 judges per trait, with appropriate tests for commonality of judgments over the entire group) along each of the critical dimensions.

A very useful aspect of the test is that its scoring is computer-assisted. The researcher need only input the particular self-descriptions used by his or her subjects (or collect them directly from the computer terminal), and the computerized dictionary assigns the appropriate scores on each dimension, to each term, and produces a file of each subject's self-placement along each of the eight critical dimensions.

The psychometric qualities of the instrument attest to the expertise and motivation of the researchers. For example, interjudge reliability in evaluating the stimulus terms was greater than 0.90. Test-retest reliability ( $r_{tt}$ ) for subjects administered the test with a 1-day testing separation was quite high ( $r_{tt}=0.75$ ), and this stability remained at a high level when the test administrations were separated by 1 year ( $r_{tt}=0.50$ ).

In addition to the validity tests noted in Claeys et al., the researchers have found the concurrent validity of the various trait dimensions to be very high. Scores on the first four dimensions were correlated with established tests of extroversion, agreeableness, conscientiousness, and neuroticism, respectively, and strong and significant relationships were obtained. In addition, the correlations between subjects' trait dimension scores and behavioral judgment checklists completed by their friends and parents also resulted in high validity coefficients. Predictive validity of the leadership, conscientiousness, and social desirability dimensions have been investigated in a military selection context, and this research, too, has demonstrated the quality of the free-format measurement approach.

Versions of the dictionary have been developed for Dutch, French, and English speakers. The Dutch and French editions

are currently in use in the Belgian Army's officer selection center. The English edition is just now available in a commercial version. Further information about the availability of these dictionaries can be obtained from the Center for Basic Interactive Research, Vangramberenstraat 28, B-3071 Erps-Kwerps, Belgium.

#### Observations and Future Directions

Free-format self-description techniques are desirable for a number of reasons. There is reason to believe, for example, that they might tap aspects of importance to the individual respondents with greater sensitivity and fidelity than do forced-choice measures. The scale development costs of such instruments are miniscule, although, of course, the resources that must be expended in developing a valid scoring system more than offset the initial cost advantage. However, once developed, the scoring dictionary should prove useful at least over the intermediate term. Much remains before this work can be considered established, but secondary analysis of data already in hand could contribute greatly in further establishing the methodology.

What follows, therefore, is a detailed discussion of the means that could be employed to further develop the free format approach. On one hand, this discussion can be viewed as an extended evaluation of the work of the Belgian researchers. The importance of their approach merits such an appraisal, for if this work fulfills its promise, it will have ramifications throughout the field of personality psychology--from the applied problems of selection to the treatment of personality disorders to the development of theory. At the same time, this critique may be read as a set of general methodological rules of thumb that could be used in the development of any free-format technique, and as such, it is relevant not only to the Belgian work, but to personality research in general. Hopefully, these suggestions will prove useful to the reader in evaluating the Belgian work, to the Belgians in the further development of their methodology,

and to personality researchers in general, who might be tempted to embark on a free-format venture of their own.

Internal Consistency. If one were to attempt to refine the Belgian approach through secondary analysis of the results of Claeys et al. (1985), three supplementary analyses immediately suggest themselves. First, although considerable evidence points to the test-retest reliability of the approach, it would be useful to know the pattern of item-whole correlation coefficients within dimensions, and the internal consistency reliability of the various dimensions that are tapped. Further, it would be very useful if mean (within dimension) item-whole values could be enhanced, thereby improving internal consistency reliability. Since different subjects employ different self-descriptive terms ("items," in the language of classical test theory), application of the standard methods of internal consistency analysis is not possible. However, a simple (but very useful) internal consistency analogue can be developed, as follows. Suppose that instead of determining within-dimension item-whole correlation across all subjects (impossible here, since items differ from subject to subject), we instead calculated the mean and standard deviation of each subject's scores on each of the critical dimensions. Then, for each subject, within each dimension, we could determine which scores, if any, deviated substantially from the others on that dimension. For example, suppose that in scoring a subject's extroversion scores over 10 self-descriptive adjectives, we observed the following pattern of scores:

1, 1, 1, 1, 1, 1, 1, 1, 1, 5

If we adopt a classical test theory perspective, it is clear that the last score does not belong--it is three standard deviations beyond the mean of its distribution. It does not "hang together" with the other item scores that define the attribute in question, and accordingly should be discarded (Nunnally, 1967). If this approach were taken over the entire sample, with some arbitrarily chosen

cutoff for deletion (say, 1.5 or 2 standard deviations from the mean of each subject's distribution), it is clear that within-dimension item-total correlations would be enhanced. This enhancement, in turn, would of necessity enhance validity correlations between the free-format device and established instruments, perhaps substantially. Of course, for some subjects no items would be discarded, but for others, some of the extreme and probably erroneously scored (for that particular subject) traits would be coded as missing. This process would almost inevitably enhance the reliability, and the consequent validity, of the instrument. Carried out across all dimensions of the test, such a procedure could have a powerfully beneficial impact on the measure.

There is a psychological, as well as psychometric, rationale for this suggested exclusionary approach. The acknowledged advantage of the free-format technique is that it allows the subject to define himself, in his own words. However, the hidden disadvantage in all free-format measures is that a particular term employed by a subject might have a very idiosyncratic meaning, which is incorrectly reflected in the dictionary values obtained over a set of judges who do not share the respondent's particular or peculiar definition of the term in question. Under such circumstances, discarding the idiosyncratically defined term (since no value can be assigned to it with any confidence) represents a means of maximizing the positive aspects of the free-format approach, and minimizing perhaps its greatest liability.

Order. A second consideration of interest concerns the response hierarchy of trait descriptors, as might be inferred from the order in which the terms are produced by the subject. Are the most important self-descriptors at the top of the subject's list of self-descriptions, at the end, or is order of no consequence? We might also wish to know whether this question is answered in the affirmative for some dimensions, but not for others. These issues could be investigated by calculating item-total correlations  $r_{IT}$  across all subjects (vs.

within a given subject, as in the previous recommendation), on each of the eight dimensions. In this type of analysis, the particular trait that appears in a given ordinal position is irrelevant, as it is the order of appearance itself that is crucial. (It is expected, of course, that the self-descriptive traits produced initially will vary from subject to subject.)

If the mean  $r_{iT}$  of the initial items is substantially greater than that of the later ones, this would have both substantive and methodological significance. At the theoretical level, it would suggest a hierarchical structure of self-relevant terms, which could have important implications for research on self-schemas, implicit personality, etc. Of course, a recency effect also is possible, where later-mentioned self-descriptors weigh more heavily than earlier ones in determining the total score. Either primacy or recency effects would prove interesting. For research purposes, it is helpful that data on the first five critical dimensions are available in the Claeys et al. archive, since analyses of later dimensions can be used as replications of the results of earlier ones.

At the methodological level, high initial  $r_{iT}$ 's might foster the practice of limiting the number of self-descriptors elicited--if the important data are produced early, then why go to the bother of collecting probably misleading, and certainly extraneous, extra information. This option appears to fly in the face of the common psychometric truism, "Longer tests are better tests," but high initial  $r_{iT}$ 's and attenuated later ones would provide some grounds for this unorthodox advice. On the other hand, if the later traits exhibited the greater  $r_{iT}$ 's, then we might argue for a longer list of self-descriptors, with the initial ones discarded in the scoring process. These possibilities are empirical questions which can be solved by the data currently at hand.

The Ambiguous Midpoint. As noted, in this free-format approach each self-descriptive trait in the dictionary has

been scored on eight different dimensions by judges using five-point rating scales. A potential problem concerns the interpretation of traits that have received intermediate values. Theoretically, mid-range values should mean that the trait has nothing to do with the dimension in question, or that it represents a "mid-dling" value on the dimension. A self-description of "blue eyed," for example, would not appear to have much to do with conscientiousness. If such a description were encountered by a judge, he or she would assign a midpoint value of 3 (on a 5-point scale) to the response. This seems to me to be a mistake. If a trait has nothing to do with a dimension, it probably should be scored as missing. If a trait in fact is seen to fall midway on a given dimension, then its score is meaningful, and it should be included. Judges can, and should, distinguish between "irrelevancies" and "intermediates."

Unfortunately, such a distinction would not completely answer the question of the proper handling of midrange values, since another instance in which a trait could be assigned an intermediate value on a dimension would occur when half the judges assigned it a value of 1, while the remainder assigned it a value of 5. Depending upon context and idiosyncracies of judges' definitional sets, it is conceivable that one judge might see the self-descriptor "aggressive" as indicating high extroversion (an aggressive used car salesman comes to mind), while another might see the term as indicating extreme introversion (the psychopath who sits quietly in a corner until he explodes).

Differentiation between "true" intermediate values and intermediate scores that arise as a consequence of the combination of extreme high and low scores is a problem with which attitude researchers have had to contend since the time of Thurstone and Chave (1929). A useful means of determining the underlying cause of an intermediate score is to inspect its variance. The variability of a true intermediate score should not



differ from that of the other traits in the distribution, with the exception of traits having the most extreme values (which must, of necessity, exhibit little variability--a trait with a mean score of 1 must have a variance of 0). A "pseudo-intermediate" trait will have a very high variance relative (1) to those of other traits in the dictionary, and (2) to other intermediate values of the "true" intermediate variety.

It is important that the pseudo-intermediates be flagged, and either counted as missing or rescored so as to be brought into congruity with the other self-ratings of a given subject on a given dimension. And, this procedure should be employed before either of the earlier mentioned analyses are undertaken.

Social Desirability. It is not likely that an individual asked for a set of self-descriptions would list "hateful," "child abuser," "deranged," "liar," "spiteful," "envious," "nauseous," "disgusting," "loathsome," and "diseased." Much more likely are positive self-descriptors. In part, this reflects reality. Although some people probably do fit the ugly description presented here, it is clear that many more do not. In addition to the distribution of positive traits in the population, however, we must also deal with the tendency of people to present themselves in the most favorable light possible, to answer personality inventories in positive ways. This tendency has been termed the social desirability bias, and there is a long research history in psychology devoted to its study (Crano and Brewer, 1986). Put most simply in the present context, it would be useful to know the extent to which people gild the lily when describing themselves on a free-response instrument. We already have available judged ratings of the social desirability of each trait. However, additional information might be useful in considering the extent to which social desirability influences responses in the free-format context.

A useful approach would involve coupling the free-response measure with standard forced-choice personality measurement devices (as in Claeys et al.) and measures of social desirability. Then, in addition to the obvious comparisons (including the concurrent validity comparison of the social desirability dimension score with the established forced-choice instrument), all of the secondary analyses that have been suggested here could be employed to discover deeper interrelationships between self-descriptions and the social desirability bias. We then could compare the extent of bias in the established scales with that found in the various dimensions of the free-format device. This would be the most fair comparison, since a simple assessment of bias in the free-response instrument without reference to the forced-choice measures would appear to imply that no bias existed in the more established instruments, and this decidedly is not the case. If the free-format measure proved no more susceptible to the social desirability bias than the more standard tests, then it would appear to be a useful addition to the methodological tool box of the personality psychologist.

Finally, it would be very useful to know whether some of the seven substantive dimensions are more susceptible to social desirability bias than others, and the contexts in which such bias is most likely. (In an officer selection situation, it seems likely that candidates would be intent on demonstrating leadership ability, for example.)

#### Conclusions

Bohrer, DeBoeck, and their colleagues have developed a technique for measuring various aspects of personality that has enormous potential. Coupled with the power of computer assisted scoring, their approach could become a model for personality research in the future. Researchers need not be bound by the particular dimensions that the authors have chosen to incorporate into their model, and which represents their

own implicit theory of personality. While the dimensions chosen are obviously important, others might have been used, depending upon the particular theoretical or substantive orientation of the investigator. The free-format approach is general and easily expandable, and can accommodate enhancements. ~~I am hopeful~~ that further research will be undertaken on this approach which, while presenting some difficult problems, nonetheless offers personality psychology with unique opportunities. Some possible routes to progress in this area have been described here. In my opinion, adoption of these recommendations could foster a major advance in this technique.

References

Claeys, W., P. DeBoeck, W. Van Den Bosch, R. Biesmans, and A. Bohrer, "A Comparison of One Free-Format and Two Fixed-Format Self-Report Personality Assessment Methods," *Journal of Personality and Social Psychology*, 49 (1985), 1028-1039.

Crano, W.D., and M.B Brewer., *Principles and Methods of Social Research* (Boston: Allyn and Bacon, 1986).

Nunnally, J.C., *Psychometric Theory* (New York: McGraw-Hill, 1967).

Thurstone, L.L., and E.L. Chave, *The Measurement of Attitudes* (Chicago: University of Chicago Press, 1929).

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input checked="" type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



END

9-87

Dtic