

Personality: Nomothetic or Idiographic? A Response to Kenrick and Stringfield

J. Philippe Rushton, Douglas N. Jackson, and Sampo V. Paunonen
University of Western Ontario, London, Ontario, Canada

It is suggested, contrary to the interpretation presented by Kenrick and Stringfield, that substantial consistency can be observed in personality as reflected in both behavior and judges' ratings when the principle of aggregation is applied to traditional nomothetic assessment procedures and results are interpreted in terms of classical reliability theory. Moreover, it is demonstrated that conclusions by Kenrick and Stringfield about the supposed improvement in predictive power stemming from an idiographic analysis do not follow from their data because they confound trait consistency and trait extremity and fail to take account of restriction and inflation of range effects.

Over the past 15 years, the traditional wisdom of the classical nomothetic approach to personality has come under repeated attack. The nomothetic view is that the science of personality consists of a search for general laws having wide applicability to people in which consistent patterns of individual differences in behavior, sometimes called traits, play a central role. Basic assumptions of this approach include substantial consistencies of people's behavior when reliably assessed and considerable predictive power of measures of traits in accounting for behavior. Critics have challenged assumptions about the predictive power of broad traits, arguing that when personality measures are used to predict a specific behavioral event, validity coefficients often fail to exceed .30.

A proposed solution to the alleged problem of the .30 validity coefficient is that of the idiographic approach. Idiographic conceptualizations, although usually not defined explicitly, involve a tacit assumption that personality is idiosyncratically organized within individuals, thus precluding the existence of general traits thought to apply universally. Kenrick and Stringfield (1980), building on earlier work by Bem and Allen (1974) and Bem and Funder (1978), ad-

vocated a "more fully idiographic" (p. 92) approach; they allowed subjects to select their own characteristic dimensions of personality by permitting them to specify trait dimensions on which they were the most consistent across situations. Kenrick and Stringfield administered one trait adjective rating scale based on each of Cattell's 16 personality factors to subjects, the subjects' parents, and subjects' peers. Judges completed 7-point scale ratings of the subjects for each of the 16 dimensions as well as nominations of the traits thought to represent the most and least consistent dimensions for each subject. The ratings purportedly showed substantially higher convergences across raters when judgments for the most consistent target traits were compared with those for the least consistent.

We take strong exception to Kenrick and Stringfield's claim that their idiographic approach is useful in helping to understand the data of personality. Our contention is that (a) the Kenrick and Stringfield interpretation of previous literature, which they see as not supporting the validity of traditional personality measures, fails to consider other data related to reliability theory and the principle of aggregation; and (b) their conclusions about the improvement in predictive power stemming from an idiographic analysis do not follow from their data because of an erroneous and idiosyncratic interpretation of correlational values and a confounding of behavioral consistency with trait extremity. We address each of these issues in turn.

We would like to thank William A. Fisher and Henry L. Roediger III for their comments on an early draft of this paper, and Natalie J. Allen and Roland D. Chrisjohn for their help with the data.

This research was supported in part by Research Grants 410-78-1061 and 410-80-0576 from the Social Sciences and Humanities Research Council of Canada, J. Philippe Rushton and Douglas N. Jackson, principal investigators, respectively.

Requests for reprints should be sent to J. Philippe Rushton, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.

The Validity of Nomothetic Personality Measures

Kenrick and Stringfield (1980) challenge the validity of indexes of nomothetic personality traits:

Studies of nomothetically applied personality tests, clinical inference, and person perception have been interpreted as supporting the view that the naive "trait" based personality conceptions of the layman (and psychologist) are largely erroneous constructions of the perceiver. Recent work has suggested that the assumption of nomothetic applicability of traits may have been incorrect and that only some people may be consistent on any given trait. (p. 88).

This is a pervasive view in current personality theorizing, having led to a general disenchantment with the nomothetic approach. It is based on a belief that a correlation of .20 to .30 is representative of the degree of cross-situational generality of behavior. One can, however, point to a major error of interpretation. The error is to believe that correlations between two items or behavioral events are stable and representative. They very rarely are. Predictive validity usually occurs only when dimensions of behavior are represented as the sum or mean of multiple measurements. This greater predictability occurs because the inevitable randomness found in any one measure is usually large compared with the non-random component of interest. By combining and summing over situations, the randomness ("error variance") tends to average out because it is not cumulative, whereas behavioral consistencies ("true score variance") do accumulate. These expectations are made explicit in psychometric theory in one of its oldest and most frequently confirmed formulations, namely, that reliability of measurement increases as a function of the number of observations (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Gulliksen, 1950; Lord & Novick, 1968). Spearman's (1910) words bear repeating. Referring to errors of measurement as "accidents" he writes:

It is the superposed accident that the present paper attempts to eliminate, herein following the custom of all sciences, one that appears to be an indispensable preliminary to getting at nature's laws. This elimination of the accidents is quite analogous to, and serves just the same purpose as, the ordinary processes of "taking means" or "smoothing curves."

The method is as follows. Let each individual be measured several times with regard to any characteristic to be compared with another. (pp. 273-274)

Because the principle of aggregation is such a fundamentally important one and is so often overlooked in the area of personality, it bears repeating here. We examine the issue separately in the context of (a) the cross-situational stability of behavior and (b) the utility of judges' ratings.

The Cross-Situational Stability of Behavior

The classic study used frequently to shed light on the "specificity versus generality" controversy is the enormous "character education inquiry" carried out by Hartshorne and May in the 1920s and published in three books: *Studies in Deceit* (Hartshorne & May, 1928), *Studies in Self-Control* (Hartshorne, May, & Maller, 1929), and *Studies in the Organization of Character* (Hartshorne, May, & Shuttleworth, 1930). These investigators gave 11,000 elementary and high school students some 33 different behavioral tests of altruism (referred to as the "service" tests), self-control, and honesty in home, classroom, church, play, and athletic contexts. Concurrently, extensive ratings of the children's reputations with teachers and their classmates were made in all these areas. By intercorrelating the children's scores on all tests it was possible to discover whether their behavior was specific to situations or generalizable across them. To the extent that behavior was generalizable, cross-situational correlations should be high. This is the largest study of the question ever undertaken, raising most of the major points of interest, and has been seriously misinterpreted by many investigators, as has been noted by Rushton (1980).

Let us consider their measures of altruism. The behavioral indexes intercorrelated .20 on the average, suggesting support for the specificity viewpoint. If the five measures were combined into a battery, however, they had a much higher correlation (.61) with the measures of the child's altruistic reputation among his or her teachers and classmates. In this regard Hartshorne et al. (1929) wrote:

The correlation between the total service score and the total reputation score is .61. . . . Although this seems low, it should be borne in mind that the correlations between test scores and ratings for intelligence seldom run higher than .50. (p. 107)

On the "guess who" test of reputation (e.g., "Guess who is kind to younger children?") the teachers' perceptions of the students' altruism agreed extremely highly with that of the students' peers ($r = .81$; Hartshorne, May, & Maller, 1929, p. 91). Taken together these latter results indicate a considerable degree of generality and consistency in altruistic behavior.

Virtually identical results were found for the measures of honesty and self-control. Any one behavioral test correlated, on the average, only a modest .23 with any one other behavioral test. When, however, the measures were combined into more reliable batteries, much higher relationships were found either with teachers' ratings of the children or with any single measure taken alone.

Typically these correlations were on the order of .50 and .60. (See Hartshorne & May, 1928, p. 130, Table 97; Hartshorne et al., 1929, p. 104, Table 20 and pp. 351-352).

Hartshorne and May (1928; Hartshorne et al. 1929, 1930), however, focused on the smaller correlations as supporting the notion of behavioral specificity: "Neither deceit nor its opposite, honesty, are unified character traits, but rather specific functions of life situations. Most children will deceive in certain situations and not in others" (Hartshorne & May, 1928, p. 411). Their conclusions, and their data, have often been referred to in defending such a position. For example, Mischel (1968), in a highly influential review, again extolled the notion of specificity in behavior, pointing out that the average correlation between two behavioral instances of a trait is .20 to .30. Persons, therefore, are said to exhibit "discriminative facility" between situations.

This specificity doctrine has provided a service by emphasizing that contexts are important and that people acquire different ways of dealing with different situations. Unfortunately, some have interpreted this as meaning that consistency does not exist. By focusing on correlations of .20 and .30 between any two behaviors or items, rather than the higher ones based on a battery of items, a very misleading impression can be created. Although Hartshorne and May employed the Spearman-Brown formula to estimate reliabilities based on increasing the number of items in their measures, they failed to realize the implications of doing so in interpreting their data.

Not only did total scores within the battery of altruism tests and measures yield evidence of consistency, but so too did measures of control, persistence, honesty, and moral knowledge. There is, indeed, evidence for a pervasive general factor of moral character (e.g., Hartshorne, May, & Shuttlesworth, 1930, p. 230, Table 32). One of the first to note this was Maller (1934). He dissented from the conclusion that there exists little commonality across moral behaviors. Using Spearman's tetrad difference technique, Maller (1934, p. 100) demonstrated a common factor throughout the inter-correlations of the character tests of honesty, altruism, self-control, and persistence. Subsequently, Burton (1963), in a reanalysis of the original Hartshorne and May data, found a large general factor that accounted for between 35% and 40% of the total variance.

On the one hand, as has been reviewed elsewhere (Jackson & Paunonen, 1980; Rushton, 1980), the typical correlation between any two behavioral exemplars is about .30. Combining measures, on the other hand, typically leads to considerably greater predictability. This principle

is not new and has often been argued for in the past (Fishbein & Ajzen, 1974; Green, 1978; Jaccard, 1974; Tryon, 1973a, 1973b). More recently, Epstein (1979, 1980) demonstrated that when measures of behavior are averaged over an increasing number of events, stability coefficients increase from around .20 from one day to the next to .70 and higher from one week to the next. These figures were found for even such supposedly transient mood states as feeling happy, kind, or calm.

In their advocacy of an idiographic approach to personality assessment, Kenrick and Stringfield have largely ignored the published research supporting traditional nomothetic conceptions and procedures. In addition, they themselves have employed one-item measures to scale the personality characteristics of respondents, thereby exposing their study to the same criticisms that have been directed at poorly formulated traditional assessment approaches.

The Utility of Judges' Ratings

Just as the validity of behavioral assessments increases as the number of behavioral items becomes larger, so does the validity of judgments as the number of judges becomes larger. The principle of aggregation is perfectly general. One early demonstration of this was provided by Eysenck (1939). He showed (Figure 1) that the function relating judgmental validity to number of judges was essentially the same for highly subjective estimates of the aesthetic value of pictures using an independent criterion (curve A) and judgments of the weight of objects, for which there was an objective criterion (curve B).

Another study demonstrating the incremental validity of using aggregated judges was carried out by Moskowitz and Schwarz (in press). These authors systematically varied the number of raters

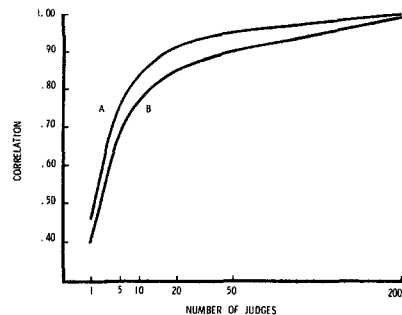


Figure 1. Relation between number of judges (square root) and correlation of their pooled judgments with independent criterion. (A = aesthetic judgments; B = weight judgments. After Eysenck, 1939).

judging a target and the number of observation periods allowed prior to the prediction of actual behavior counts of dominant behaviors (Figure 2). The findings are clear; validity increases linearly as the number of raters increases and as the number of groups of observations increases. Under these conditions validity coefficients are considerably in excess of .30 and operate in accord with predictions from classical test theory.

Kenrick and Stringfield used two sets of single judges (parents and peers) as predictors of target behavior. The reliabilities of these ratings were quite low (Kenrick & Stringfield, 1980, p. 94, Table 1), imposing an upper bound on validity and suggesting that potential improvements in judgmental validity would have occurred had peer and parent ratings been combined.

Judges' ratings of various kinds, when assessed reliably, come to have more than trivial predictive utility. For example, Eron (1980) found that average peer ratings of aggressiveness at age 8 correlated .43 with the average of a different set of peer ratings of aggressiveness at age 19. Further, those who had been rated as aggressive at age 8 were three times as likely to have been entered on police records by the time they were 19 than those not so rated. This suggests that perceptions of personality can be both stable over time and predictive of behavior. Other researchers have found equally important results using raters' perceptions of personality. Block (1971), for exam-

ple, has shown that independent ratings of personality are stable over decades when measured reliably; Jackson, Peacock, and Smith (1980) have found them to be important for the employment selection interview; Hirschberg and Itkin (1978) and Rushton, Murray, and Paunonen (Note 1) were able to predict such important aspects of university functioning as time to complete a doctoral program and success at publishing and teaching using personality information.

In summary, the two sections just covered support the view that consistent patterns of individual differences in behavior are highly stable over time, that these individual differences are readily perceived, and that the more measurements are replicated, the more the aggregate score is predictive of behavior. Kenrick and Stringfield can be faulted both for misguided conclusions concerning the utility of traditional assessment procedures, and for committing the same errors that have perpetuated dissatisfaction with those procedures.

Problems With the Kenrick and Stringfield Correlation Coefficients

Kenrick and Stringfield (1980) sought to demonstrate that modest correlations between self-ratings and ratings by peers and parents could be enhanced by separately considering consistent and inconsistent traits as nominated by respondents. Each subject in the study completed self-ratings of personality and subjective estimates of consistency on 16 one-item bipolar trait scales. Ratings of the subjects were also obtained from peers and parents using the same materials. In addition, each rater selected out of the 16 trait scales the one most and one least consistent personality dimension for the target. Self-report trait scores on the "most consistent" target dimensions were intercorrelated with peer and parent ratings on these same dimensions. These correlations were compared with the corresponding values computed across the "least consistent" target traits. The average rating congruence was quite substantial for the former sets of ratings and quite meager for the latter (e.g., .61 vs. .23; p. 95, Table 5). Unfortunately, these correlations are (a) based on sets of data not appropriately combined and (b) influenced by a confounding of consistency ratings and trait ratings. We consider each of these problems below and provide new data to evaluate the latter interpretation.

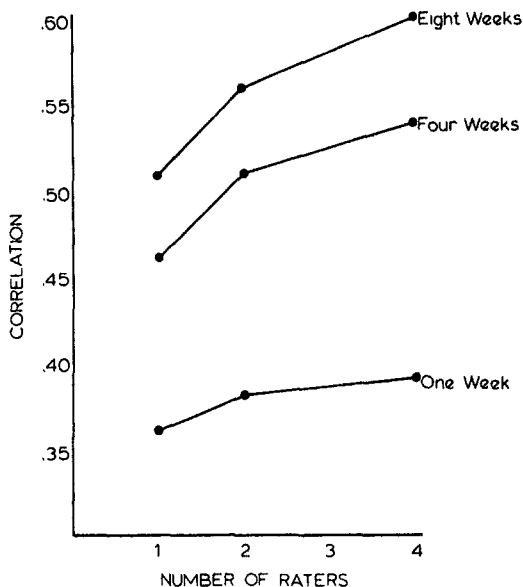


Figure 2. Validity of dominance ratings as a function of number of raters and weeks of observation. (After Moskowitz and Schwarz, in press.)

Heterogeneous Traits and Computing Correlations

Ever since the development of correlational procedures, accepted practice has been to consider

two observations, x_i and y_i , representing two measurements of attributes for entity i . Normally the same variable x and the same variable y are employed for each person, and the correlation represents the degree to which variable x is associated with variable y for the sample of persons under consideration. An estimate is ordinarily made regarding the degree of covariation between x and y , for example, the validity of trait measure x with respect to criterion y . This is not what Kenrick and Stringfield report as their correlation coefficients. Their coefficients refer not to the same trait, x , measured across different individuals, but to different traits for different individuals linked to parallel criterion ratings, y , which again refer to different traits for different individuals. These correlational values cannot be said to refer to the validities of particular trait measures, nor can they be said to reflect the validity of a battery of such measures. In fact, their interpretation is obscure.

Confounding Trait Consistency and Extremity

The second problem with the correlations computed across ratings for consistent and inconsistent traits is the failure to account for an expected relationship between the consistency of trait-related behavior and the strength of the trait (Stones & Burt, 1978). Consider three persons, one extremely high on the trait of dominance, one in the middle range, and one in the extremely low range. Suppose each was asked for a self-rating of consistency. The person high on the trait probably decides that he/she often behaves dominantly and thus rates himself/herself as "consistent." Similarly, the person low on dominance decides that he/she rarely behaves dominantly and thus may also provide a self-rating of "consistent." Those in the mid-range of dominance could reasonably conclude that dominant behavior for them depended on the situation and would rate themselves as "inconsistent." Thus, consistency becomes correlated with extremity of scoring. What then would be the effect of comparing only consistent (and extreme) individuals on a trait with inconsistent (and moderate) individuals? The data reported below bear on this question.

Empirical Analysis of Trait Consistency and Extremity

Method

To examine the association between consistency ratings and personality ratings, data were collected following the procedure of Kenrick and Stringfield. A total of 52 undergraduate students volunteered to complete a set of trait adjective ratings. Self-ratings, using 7-point bipolar scales,

were collected on 34 dimensions of behavior. These included the 16 adjective pairs employed by Kenrick and Stringfield (1980), such as intelligent-unintelligent followed by 16 parallel adjective pairs using different but synonymous anchors (e.g., clever-dull), and 2 adjectives related to social desirability.¹ Subjects then estimated the consistency/variability of their behaviors with respect to each of the 16 primary dimensions. They were instructed to circle a number on a 7-point scale indicating the extent to which their "behavior varies from situation to situation for each of the dimensions." Following these self-description tasks, subjects were asked to rate the public observability of their behaviors related to the 16 dimensions using the verbatim instructions of Kenrick and Stringfield (1980). Finally, each rater nominated his/her most and least consistent behavioral dimensions from the list of 16.

Results and Discussion

The first column of Table 1 reports the correlations between the two measures of the same trait, which range from .06 to .58, with a mean of .38. Reliabilities such as these impose an upper bound on validity. Strictly speaking, these reliabilities refer to the sampling of adjectives within a trait domain but are relevant in illustrating the somewhat limited reliability of single items.

The question of the degree to which ratings of consistency are a function of the rater's extremity on the trait may be addressed by correlating the judgments of consistency with an index of extremity. Column 2 of Table 1 refers to these correlations between consistency judgments and extremity as measured by the absolute deviation of one's trait rating from the center of the 7-point scale (4). Thus, trait adjective ratings of 1 and 7 were recoded as 3, ratings of 2 and 6 recoded as 2, and so on. A total of 10 out of 16 of these correlation coefficients are statistically significant

¹ The 16 bipolar adjective pairs and following synonyms and social desirability items, counterbalanced to control for direction effects, were: emotional-calm, intelligent-unintelligent, reserved-outgoing, assertive-not assertive, sober-happy-go-lucky, conscientious-expedient, shy-venturesome, tender-minded-tough-minded, suspicious-trusting, practical-imaginative, shrewd-forthright, self-assured-apprehensive, conservative-experimenting, group oriented-self-sufficient, undisciplined-self-disciplined, relaxed-tense, stable-easily upset, dull-clever, friendly-unfriendly, humble-dominant, playful-serious, irresponsible-responsible, thrill seeking-timid, unsentimental-sentimental, faith in people-cynical, creative-matter-of-fact, straightforward-calculating, unconfident-confident, try new things-prefer old ways, independent-oriented to others, follow schedule-follow impulses, driven-easygoing, cruel-kind, good-bad.

Table 1
Self-Rating Reliabilities and Correlations With Consistency and Desirability Ratings for 16 Personality Traits

Trait	Reliability	Consistency ^a	Desirability
Emotional	.53	.27	.11
Intelligent	.45	.53	.25
Friendly	.33	.24	.21
Assertive	.37	.41	-.22
Sober	.30	-.03	-.09
Conscientious	.06	.45	.42
Shy	.41	.29	.07
Tender-minded	.27	.41	.54
Suspicious	.55	.19	-.21
Practical	.32	-.30	-.16
Shrewd	.32	-.06	-.24
Self-assured	.58	.35	-.03
Conservative	.42	-.05	-.03
Group oriented	.55	.28	.14
Undisciplined	.32	-.04	-.16
Relaxed	.33	.40	.04
Mean	.38	.21	.04

Note. $N = 52$; $r = .24$, $p < .05$.

^a Trait ratings for this column were transformed into trait extremity scores, calculated as the numerical distance between a subject's mean self-rating and the scale midpoint of 4.

in the predicted direction for our data, indicating that the majority of personality adjective ratings are associated with ratings of consistency. This analysis suggests that what Kenrick and Stringfield term consistency may well be subject to the alternative interpretation of extremity on the trait dimension. Indeed, Jack Block (Note 2) also observed the possibility of a confound between self-rated consistency and the extremity of self-assigned trait ratings in the Kenrick and Stringfield study. He suggested to Kenrick and Stringfield that their data be reanalyzed, recoding the self-assessed trait ratings following the procedure outlined above. Kenrick (Note 3) has reported, in a letter to Block, the correlations between self-rated consistency and the recoded trait ratings for the 16 traits studied to be all positive, ranging in value from .14 to .57 with a mean of .33. These correlations are lower bounds, attenuated by the unreliabilities involved, and provide direct evidence for the presence of the hypothesized alternative explanation.

One can select, for each trait dimension, the subjects with the highest variability ratings and compare the dispersion of their trait scores with the trait score variance of subjects with the lowest variability ratings. This will reflect increases or decreases in criterion variance as a function of the sampling procedure. The variances of the trait scores for the 26 subjects with the highest consistency ratings and the 26 with the lowest ratings for each trait are contrasted in Table 2. (Each trait score was the sum of two items. Subjects

with the same consistency ratings were randomly divided into the two groups.) Ten of the 16 dimensions show larger trait score variances for the consistent group relative to the inconsistent group. One would, in general, obtain larger correlations

Table 2
Variances of Trait Scores for the 26 Most Consistent and 26 Least Consistent Respondents Along Each of 16 Personality Dimensions

Trait	Most consistent	Least consistent
Emotional	7.28	4.07
Intelligent	3.28	1.92
Friendly	4.30	3.23
Assertive	6.83	2.53
Sober	4.65	3.48
Conscientious	2.50	1.87
Shy	7.10	3.10
Tender-minded	4.33	2.16
Suspicious	9.13	5.33
Practical	3.50	7.23
Shrewd	3.10	3.83
Self-assured	3.39	5.38
Conservative	4.92	4.98
Group oriented	4.83	5.02
Undisciplined	3.36	4.21
Relaxed	8.12	2.24
Mean	5.04	3.79

Note. Subjects with the same consistency ratings were randomly divided into the two subgroups.

between these 10 trait measures and a criterion, such as parent or peer ratings, for the consistent versus the inconsistent subgroups. Kenrick and Stringfield (1980) found such differential correlations for 11 of the 16 dimensions in their "trait-by-trait analysis" (p. 97, Table 6), noting that for the remaining variables "just the reverse seems to be the case" (p. 97). However, as we have illustrated above, disparities in trait score variances can account for all disparities in the pairs of correlation coefficients presented by Kenrick and Stringfield (1980, Table 6, column 1).

As a last analysis, subjects' trait scores were divided into two groups based on their self-nominated "most consistent" and "least consistent" dimensions. (Thirty-five subjects completed this judgment.) For each subject the raw total score on his/her most consistent trait (regardless of which trait) was put into one group and his/her score on the dimension selected as least consistent was put into the other group. This procedure replicates that on which Kenrick and Stringfield have based their central conclusions. The mean trait score for the 35 most consistent dimensions is 7.9 with an inflated variance of 12.7. For the least consistent traits, the mean total score is 8.0 with an attenuated variance of 3.5. These two distributions are illustrated in Figure 3. (Across all subjects and all traits, the mean total trait score is 7.97 and the mean of the trait score variances is 4.78.) It is apparent that high consistency can occur at all points along the trait score continuum, but a high degree of behavioral variability is precluded at the extremes. For statistical reasons alone one would expect the correlation of trait scores and other ratings for the consistent group to be much higher than the corresponding coefficient for the inconsistent group, particularly when predictor ratings (by parents and peers) are accurate and, hence, similarly affected through selective sampling. The difference in predictive validity for the two groups would be artificially exaggerated. There are, thus, good psychometric bases for Kenrick and Stringfield's findings. Invoking an idiographic interpretation as the primary reason for such findings is gratuitous.

There is an additional problem with the Kenrick and Stringfield paper that should be mentioned briefly, namely, the possibility that self-ratings may be confounded with desirability bias. The third column of Table 1 reports the correlation between each set of self-ratings and the sum of the two reference variables designed to identify desirability bias in ratings. Whereas most of these values are low or moderate, they do range as high as .54, suggesting that this source of contamination should also be considered in reviewing validity data. Self-ratings reflecting socially desirable

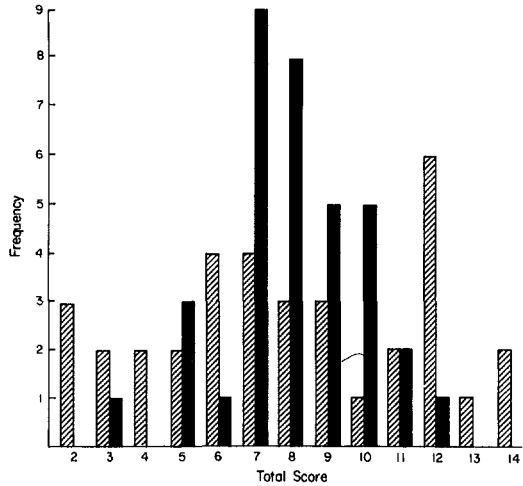


Figure 3. Frequency distributions of total scores for traits identified by subjects ($N = 35$) as their *most* consistent dimensions (hatched bars) and those identified as their *least* consistent (black bars).

response tendencies will show poor convergence with other ratings (e.g., peer ratings) predicated on the nonevaluative or descriptive attributes of trait terms (Jackson & Paunonen, 1980). Consistency ratings might also be mediated by social desirability. For example, few respondents will report inconsistency for the trait of honesty because to do so will imply socially undesirable behavior.

Conclusions

Published empirical findings have been reviewed that demonstrate that when a sufficient number of measurements of personality are taken, reliable personality profiles of individuals result. If these measurements are based on observer perceptions, they can allow for the reliable prediction of behavior, particularly if the number of judges and observations is sufficient to yield reliable criterion scores. Kenrick and Stringfield (1980) have largely ignored these findings in presenting their objections to nomothetically based assessment procedures.

Attempts to approach the data of personality with techniques thought to be idiographic are fraught with problems of interpretation, most of which yield to nomothetic analyses. Thus, in attempting to isolate subgroups of predictable individuals through the identification of consistent behavioral dimensions, ratings of trait consistency frequently yield measures confounded with trait extremity. This confounding may have selectively inflated and deflated the validity coefficients re-

ported by Kenrick and Stringfield, resulting in spurious conclusions concerning the utility of their procedure. Our analysis fails to support the view that traditional nomothetic assessment techniques can be improved on by idiographic approaches. A more fruitful course, in our view, is the substantive improvement of predictor and criterion measures and the adherence to classical requirements for reliable measurement.

Reference Notes

1. Rushton, J. P., Murray, H. G., & Paunonen, S. V. *Personality characteristics of university professors*. Paper presented at the 22nd International Congress of Psychology, Leipzig, German Democratic Republic, July 1980.
2. Block, J. Personal communication, March 1981.
3. Kenrick, D. T. Personal communication to J. Block, June 1980.

References

- Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 1974, *81*, 506-520.
- Bem, D. J., & Funder, D. C. Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, 1978, *85*, 485-501.
- Block, J. *Lives through time*. Berkeley, Calif.: Bancroft Books, 1971.
- Burton, R. V. Generality of honesty reconsidered. *Psychological Review*, 1963, *70*, 481-499.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. *The dependability of behavioral measurements*. New York: Wiley, 1972.
- Epstein, S. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 1979, *37*, 1097-1126.
- Epstein, S. The stability of behavior: II. Implications for psychological research. *American Psychologist*, 1980, *35*, 790-806.
- Eron, L. D. Prescription for reductions of aggression. *American Psychologist*, 1980, *35*, 244-252.
- Eysenck, H. J. The validity of judgments as a function of the number of judges. *Journal of Experimental Psychology*, 1939, *25*, 650-654.
- Fishbein, M., & Ajzen, I. Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 1974, *81*, 59-74.
- Green, B. F., Jr. In defense of measurement. *American Psychologist*, 1978, *33*, 664-670.
- Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.
- Hartshorne, H., & May, M. A. *Studies in the nature of character: Vol. 1. Studies in deceit*. New York: Macmillan, 1928.
- Hartshorne, H., May, M. A., & Maller, J. B. *Studies in the nature of character: Vol. 2. Studies in self-control*. New York: Macmillan, 1929.
- Hartshorne, H., May, M. A., & Shuttleworth, F. K. *Studies in the nature of character: Vol. 3. Studies in the organization of character*. New York: Macmillan, 1930.
- Hirschberg, N., & Itkin, S. Graduate student success in psychology. *American Psychologist*, 1978, *33*, 1083-1093.
- Jaccard, J. J. Predicting social behavior from personality traits. *Journal of Research in Personality*, 1974, *7*, 358-367.
- Jackson, D. N., & Paunonen, S. V. Personality structure and assessment. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology* (Vol. 31). Palo Alto, Calif.: Annual Reviews, 1980.
- Jackson, D. N., Peacock, A. C., & Smith, J. C. Impressions of personality in the employment interview. *Journal of Personality and Social Psychology*, 1980, *39*, 294-307.
- Kenrick, D. T., & Stringfield, D. O. Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review*, 1980, *87*, 88-104.
- Lord, R. M., & Novick, M. R. *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.
- Maller, J. B. General and specific factors in character. *Journal of Social Psychology*, 1934, *5*, 97-102.
- Mischel, W. *Personality and assessment*. New York: Wiley, 1968.
- Moskowitz, D. S., & Schwarz, J. C. A validity comparison of behavior counts and ratings by knowledgeable informants. *Journal of Personality and Social Psychology*, in press.
- Rushton, J. P. *Altruism, socialization, and society*. Englewood Cliffs, N.J.: Prentice-Hall, 1980.
- Spearman, C. Correlation calculated from faulty data. *British Journal of Psychology*, 1910, *3*, 271-295.
- Stones, M. J., & Burt, G. Quasi-statistical inference in rating behavior: I. Preliminary investigations. *Journal of Research in Personality*, 1978, *12*, 381-389.
- Tryon, R. C. Basic unpredictability of individual response to discrete stimulus presentations. *Multivariate Behavioral Research*, 1973, *8*, 275-295. (a)
- Tryon, R. C. Generality and reliability of individual differences in discrete stimulus-item situations. *Multivariate Behavioral Research*, 1973, *8*, 297-311. (b)

Received March 11, 1981

Revision received June 8, 1981 ■