

## Trait Relevance and Trait Assessment

**Thomas W. Britt**

*Walter Reed Army Institute of Research*

**James A. Shepperd**

*Department of Psychology  
University of Florida*

*Growing agreement exists among psychologists that all traits are not equally relevant to all people. The assumption that traits differ in their relevance across individuals has implications for psychometric theory and the assessment of reliability and validity. Specifically, the less relevant a trait is to an individual, the more error that person is contributing to the measurement process. This article discusses the construct of trait relevance, explicates why differences in trait relevance need to be incorporated into models of trait assessment, and suggests new possibilities for the measurement of trait relevance.*

In attempting to understand human thought, emotion, and behavior, psychologists have found it essential to develop instruments designed to measure important intrapsychic constructs. Without the development of valid and reliable instruments to measure theoretically relevant variables, the field of psychology would be hard pressed to make inferences and test hypotheses about intrapsychic phenomena. Given the importance of construct assessment in psychology, it is not surprising that researchers are continually examining the validity and reliability of their instruments. The state of a given area within psychology is often defined in terms of the reliability and validity of the indicators of variables central to the area (Goldberg, 1993).

Researchers developing instruments to assess constructs and the psychologists who use the instruments make the implicit assumption that the latent variable that is being measured is relevant to all individuals and that the only task for the researcher is to determine where individuals fall on the continuum representing the construct. However, numerous researchers in personality and social psychology have argued that not all personality traits, attitudes, and values are equally relevant to all individuals (Allport, 1937; Baumeister, 1991; Baumeister & Tice, 1988; Bem & Allen, 1974;

Britt, 1993; Converse, 1970; Fazio, 1989; Hershberger, Plomin, & Pedersen, 1995; Higgins, 1997; Kelly, 1955; Kenrick & Stringfield, 1980; Lamiell, 1987; Lanning, 1988; Pelham, 1993; Pelham & Swann, 1994; Petty & Krosnick, 1995; Tice, 1989). Often termed *idiographic* (Allport, 1937; Pelham, 1993),<sup>1</sup> these approaches argue that because of unique social experience and genetic endowment, people will vary in the extent to which a given construct influences their experience and determines their behavior.

The idea that all psychological constructs are not equally relevant to all individuals has important implications for how researchers assess the validity of constructs. Specifically, individuals for whom a trait is less relevant can be argued to contribute more error to the measurement process than individuals for whom a trait is more relevant. By not taking into account differential trait relevance, researchers are ignoring a systematic source of error that may reduce the reliability and validity of the constructs they are examining. To explicate these claims, we first discuss the nature of trait relevance, distinguishing it from related personality moderators. We then discuss the implications of trait relevance for measurement theory and for assess-

---

Both authors contributed equally to this article. We thank Bill Graziano, Eric Knowles, and Mark Leary for their helpful comments.

Requests for reprints should be sent to either Thomas W. Britt, who is now at the Department of Psychology, King College, 1350 King College Road, Bristol, TN 37620-2699, E-mail: twbritt@king.edu; or James A. Shepperd, Department of Psychology, University of Florida, P.O. Box 112250, Gainesville, FL 32611-2250, E-mail: shepperd@psych.ufl.edu.

---

<sup>1</sup>We distinguish between the idiographic assumption of differential trait relevance and the idiographic methodology of using a within-subjects design to test hypotheses (Pelham, 1993). The assumption refers to the idea that individuals differ in the extent a given trait is relevant to their behavior and experience. The idiographic method assumes that the best way to study psychological phenomena is to use the individual as the unit of analysis, rather than combining individuals to obtain a group mean. Idiographic assumptions can be tested using both nomothetic and idiographic methodologies (see Jaccard & Dittus, 1990; Kelly, 1955).

ing the reliability and validity of traits that are hypothesized to be differentially relevant to a sample of individuals. Finally, we address the various ways trait relevance has been measured and offer new approaches to measuring the construct.

### The Nature of Trait Relevance

We use the term *trait* broadly to refer to such constructs as personality traits, attitudes, and values, intrapsychic structures that develop through the complex interplay of social experience and genetic constitution. These constructs influence the fundamental processes of perception, interpretation, and action. Several researchers have noted commonalities between attitudes and traits (Azjen, 1987; Blass, 1984; Sherman & Fazio, 1983). Traits and attitudes are similar in that both function to explain consistencies in behavior and guide relevant perceptions through the processes of attention and encoding (Sherman & Fazio, 1983).

We define *trait relevance* as the degree to which a trait is consequential in influencing an individual's thought, affect, and behavior. The more relevant a trait is to an individual, the more the individual can be said to be "traited" (Baumeister & Tice, 1988; Reise & Waller, 1993) on the underlying construct and therefore the more likely the trait will influence behavior and experience. The relevance of a trait to an individual is to some extent independent of the individual's score on the underlying trait dimension. For example, an individual might score very low or very high on the trait of aggressiveness, but the underlying dimension of aggressiveness may not be relevant to the individual in terms of determining his or her behavior or experience. Biesanz, West, and Graziano (1998) noted that an assumption of differential trait relevance implies that the researcher must first determine where the individual falls on the trait dimension (trait score or standing), and then determine the extent to which the trait influences the individual's experience and behavior (trait relevance).

In understanding the distinction between trait relevance and trait score in the context of personality traits, values, and attitudes, it is helpful to consider dimensions relevant to all people, such as height and weight. All people "have" height and weight (height and weight are by definition relevant to all people), so the task of measurement involves simply determining where people fall along these dimensions. However, as we noted earlier, numerous personality and social psychologists have argued that people differ in the extent to which personality traits, values, and attitudes are relevant to them. For these dimensions, the researcher must determine not only where an individual falls on

the dimension, but also the extent to which the dimension is relevant to the individual in the first place.

### Trait Relevance Versus Trait Awareness (Metaknowledge of the Trait)

Our definition of trait relevance makes no assumptions regarding awareness of a trait; a trait can be highly relevant to an individual without his or her awareness. Wicklund and Eckert (1992) argued for a major distinction between possessing self-knowledge about an attribute and being consciously aware of that knowledge. This distinction is similar to James's (1990) distinction between self as object and self as knower (see also Wicklund & Eckert, 1992). In this context, a trait may influence an individual's experience and determine his or her behavior without the individual being consciously aware of the trait's relevance.

This point provides a crucial distinction between trait relevance and similar personality moderators proposed in the past. For example, researchers who examine such moderators as trait certainty, uniqueness, and importance assess these constructs by having people rate the certainty, uniqueness, and importance of different traits (Cheek, 1982; Markus, 1977; Pelham, 1993; Zuckerman, Bernieri, Koestner, & Rosenthal, 1989; Zuckerman et al., 1988; Zuckerman, Miyake, Koestner, Baldwin, & Osbourne, 1991). Although research in these areas shares the idiographic assumption underlying research on trait relevance, these constructs represent inferential processes on the part of individual, where the individual calls a given trait to consciousness and then reflects on different aspects of the trait. In short, these constructs differ from our conception of trait relevance in that they assume that people possess an awareness of underlying trait qualities (i.e., they possess a metaknowledge of the trait). Our distinction between trait relevance and trait metaknowledge has been echoed recently within the attitude literature, where researchers have distinguished between the relevance of an attitude and subjective perceptions of an attitude's importance or certainty. Specifically, Bassili (1996) proposed two major classes of attitude strength measures: meta-attitudinal (e.g., ratings of an attitude's importance, certainty, etc.) and operative (e.g., assessment of attitude strength through such operations as response latency or extremity).<sup>2</sup>

<sup>2</sup>The distinction between trait relevance and trait metaknowledge also has important implications for assessing the reliability of peer ratings of personality. We argue that if trait relevance is high, but trait metaknowledge is low, it should still be the case that peer ratings of an individual's standing on the trait are fairly reliable. However, if trait metaknowledge is high, but trait relevance is low, peer-ratings of an individual's standing on the personality trait will not be highly reliable. We thank Bill Ickes for bringing this point to our attention.

## Theoretical Perspectives on Trait Relevance

The nomothetic approach assumes that all traits are relevant to all people, and that people differ only with regard to where they fall on a given trait continuum. For example, according to the nomothetic approach, the trait of friendliness is relevant to all people. People differ from one another, however, in the extent to which they are friendly—some individuals are more (or less) friendly than others.

In contrast to the nomothetic approach is the idiographic approach, an approach that has its basis in antiquity (see Jaccard & Dittus, 1991), yet is most often identified with Gordon Allport (1937, 1962). According to Allport, not all traits apply equally well to all people. Allport (1937) argued that cardinal traits consist of those traits most influential to an individual and that distinguish the individual from others (see Zuckerman et al., 1991). Central traits, although not as influential as cardinal traits, still have a sizable influence on the experience and behavior of individuals. Allport (1937) argued that many traits are peripheral to an individual's personality, playing much less of a role in determining the individual's experience and behavior. For example, although some people may be more (or less) friendly than others, for some people friendliness may be an inconsequential aspect of their personality. For these individuals, the construct of friendliness plays a minor role in the organization of their experience and in the mobilization of their behavior. Allport (1937) suggested not only that different individuals will come to incorporate different traits into their personality, but also that researchers must take into account the interactions among traits to fully explain the role of personality in an individual's experience and behavior (see also Funder, 1991). In line with Allport, and in sharp contrast to other theories of personality during his time, George Kelly (1955) proposed that individuals come to possess unique ways of knowing and experiencing their interpersonal worlds. Kelly argued that as a result of experiencing the world, and operating as "naive scientists," individuals come to develop personal constructs, which he viewed as interpretational frameworks that allow individuals to anticipate and make sense of experience. Each individual's unique social history will influence the constructs that he or she habitually uses to deal with the external world (see also Higgins, 1990).

These perspectives, as well as modern derivatives of these perspectives (Baumeister & Tice, 1988; Biesanz et al., 1998; Bem & Allen, 1974; Fazio, 1990; Higgins, 1990; Lanning, 1988; Markus, 1977; Pelham, 1993; Tellegen, 1988; Zuckerman, et al., 1989), all share in common the belief that traits are

differentially relevant to individuals. For example, Bem and Allen (1974) proposed that individuals differ on the traits that guide cross-situation consistency in behavior. Likewise, Baumeister and Tice (1988) introduced the concepts of "metatrait" and "traitedness" to refer to the extent to which a trait structure determines an individual's experience and behavior (see also Biesanz et al., 1998; Lanning, 1988; Reise & Waller, 1993). Finally, Markus's (1977) notion of self-schemata refers to the extent to which individuals possess internal frameworks for a given trait or attribute (see also Fazio, 1990; and Higgins, 1990).<sup>3</sup> These perspectives are discussed in more detail in a later section of the article. Our point in this section of the article is to emphasize the multitude of perspectives endorsing the belief that traits are differentially relevant to individuals.

In summary, our conception of trait relevance suggests that the nomothetic assertion that all traits are equally relevant to all people is incorrect; all traits are not universally relevant to all people. The idiographic assumption raises questions about the construction and use of instruments designed to tap latent psychological constructs. For example, in the current practice of constructing measures of attitudes, traits, and values, researchers administer an instrument to a group of individuals and assume that the scores obtained are equally meaningful for all individuals. The very fact that individuals supply responses to items comprising a personality instrument is presumed to verify that a latent trait is contributing to the trait score. However, supplying responses does not by definition indicate that a latent construct is relevant to a given respondent. The responses supplied, although meaningful and predictive of behavior for some people, are less meaningful or predictive for others. Our discussion of trait relevance therefore has important implications for psychometric theory and construct validation. It is to these issues that we now turn.

<sup>3</sup>Of note, Markus (1977) was interested in self-conceptions—how people perceive or conceptualize themselves—which, although perhaps fitting under our broad definition of trait, is nevertheless distinct. Moreover, a trait may be highly relevant to a person yet not be part of his or her self-concept. An example might be self-monitoring, which may not figure prominently in a person's self-concept, but which nevertheless may be highly relevant to the person's behavior. That is, although it may be obvious to others, the person may be unaware that his or her behavior is guided by the underlying trait. Markus (1977) unfortunately confounded trait relevance with trait extremity and trait awareness (Baumeister & Tice, 1988; Britt, 1993). Although people with extreme trait scores are probably more likely to be high in trait relevance than are people with less extreme scores, trait extremity is not a necessary precondition of trait relevance. It is possible for a person to score in the middle of a trait continuum yet be high in trait relevance (see Baumeister, & Tice, 1988, for additional discussion).

## Trait Relevance and Psychometric Theory

### Classical Test Theory

The measurement of psychological attributes has a long and venerable history. Over 4,000 years ago the Chinese measured various abilities for purposes of matching civil servants to particular jobs (DuBois, 1966). Likewise, the ancient Greeks employed tests to assess mastery of both mental and physical abilities, and European universities from their inception in the middle ages have used formal examinations for purposes of awarding degrees and honors (Anastasi, 1982). Today countless psychological tests and measurement instruments are used to predict behavior and performance, to select personnel, and to identify individuals' attitudes, values, traits, abilities, skills, aptitudes, and pathologies.

Early psychometric theorists developed formulas for assessing the validity of these measurement tools. The assumption underlying these early and later assessment efforts was that the attribute being measured was equally relevant to all individuals. Indeed, according to classical test theory, an early theoretical framework guiding test development and validation procedures, an individual's score on a measurement instrument has two parts: the "true score" and "error" (Spearman, 1907, 1927). The *true score* represents the individual's score on the latent variable measured by the instrument, whereas *error* represents deviation from the true score due to factors such as imprecision or unreliability in the measurement instrument and transient fluctuations in an individual's responding. According to Spearman, the presence of error serves to diminish the correlation between theoretically related variables. That is, were it possible to measure a person's true score without error, the correlation between theoretically related variables would be considerably higher.

Although Spearman's (1907) classical true score model has seen revisions over the years (see Crocker & Algina, 1986; Guilford, 1954; Gulliksen, 1950; Magnusson, 1967; Lord & Novick, 1968), the central assumption that the score a person receives on a test is to some degree an index of the person's true score on the construct assessed by the test has not been altered. All forms of classical test theory assume that there is an underlying or latent attribute within the person that is measured by the assessment instrument and is reflected in the individual's test score, and that no individual differences exist in the relevance of the latent attribute. In addition, the assumption of a latent trait being partly responsible for observed scores is the cornerstone of item response theory (Crocker & Algina, 1986; Lord, 1980). That is, a latent trait is assumed to influence the probability of people respond-

ing correctly to any given test item (Crocker & Algina, 1986).

To illustrate the issues involved in incorporating the concept of differential trait relevance into the components of an individual's observed score on a trait dimension, we start with the classical true score model:

$$X = T + E$$

That is, a person's observed score ( $X$ ) equals the true score ( $T$ ) plus error ( $E$ ). The more reliable (and presumably valid) the measurement instrument, the smaller the error. This model assumes that a given attribute is relevant to all individuals.

Although the assumption of universal relevance may hold for attributes such as ability and intelligence, as we noted earlier, there is growing sentiment among personality and social psychologists that not all traits, attitudes, and values are equally relevant to all individuals. The concept of differential trait relevance suggests that all individuals' true scores should not be weighted equally. The true scores of individuals for whom a given trait is highly relevant should receive more weight than the true scores of individuals for whom the trait is less relevant. Furthermore, we argue that differences in trait relevance result in an additional source of error being present in observed scores, a source stemming from error due to variation in trait relevance. The following formula illustrates the adjustments that are necessary when accounting for observed scores on traits that are presumed to differ in relevance across a sample of individuals:

$$X = W_{tr}T + E_{tr}(1 - W_{tr}) + E_{nr}$$

That is, an individual's observed score equals the true score weighted by the relevance of the trait for the individual ( $W_{tr}$ , which can range from 0 to 1), plus the error created by differential trait relevance ( $E_{tr}$ ), plus the error not due to differential trait relevance ( $E_{nr}$ ). As trait relevance decreases, the contribution of the true score to the person's observed score decreases and the contribution of error increases. In extreme cases, when a trait is completely irrelevant to an individual, the observed score consists solely of error. Moreover, given the distinction between error attributable to trait relevance versus error attributable to other aspects of the test and situation, an individual's observed score could consist primarily of error even were the instrument itself perfectly reliable.

The extreme case vividly illustrates that differential trait relevance is not captured by traditional psychometric models. Traditional models assume that traits are relevant to all people, and that if we could just eliminate random error from the measurement process (e.g., error due to imperfect items, participant

mood, etc.), we could obtain the person's true score on the trait and the true score could be used to predict the person's experience and behavior in relevant domains. However, in the extreme case, when the trait is completely irrelevant to the individual, the individual has no true score. As such, improving the reliability of the instrument would neither provide a better assessment of this individual's true score nor would it increase the predictive power of the instrument for this person. The person's trait score represents nothing more than error. Moreover, this error, which has been regarded as random by traditional psychometric models, actually represents an important psychological construct in its own right that cannot be incorporated under the rubric of error variance under traditional models.

Importantly, the formula we offer is for illustration purposes only. It is intended principally to convey the complexities encountered in accounting for differential trait relevance. Clearly, considerable psychometric work would be necessary to verify the specific algebraic relations among the components in the formula. The formula simply illustrates that when a trait is not equally relevant to all members of a sample, there is an additional source of error present in observed scores, namely, error due to variation in trait relevance.

Recent empirical evidence supports the hypothesis that low trait relevance is related to greater error in trait measurement. Building on research using response latency as an indicator of attitude strength (Fazio, 1990), Britt and Kraus (1994) used response latency to items from several personality, attitude, and value scales to assess trait relevance. Presumably, the faster an individual responds to a set of items measuring a given trait, the more relevant the trait is to the individual. Britt and Kraus examined the correlation between the speed of responding to scale items and the amount of variability exhibited in responses to items comprising the scale. Consistent with the hypothesis that greater trait relevance would be associated with less error in measurement, response latency correlated with interitem variability. Specifically, across 13 different measures of personality traits, attitudes, and values, the faster individuals responded to items comprising a scale, the less variability they displayed in their responses across the scale items. These relations were obtained after controlling for participants' trait score, speed of responding across all measures, and variability of responding across all measures. In short, the results support the idea that an additional source of error exists in the measurement of traits that are not relevant to all people.

Other evidence for an extra source of error in the scores of people for whom a trait is less relevant comes from a study by Hershberger et al. (1995). These authors found that interitem variability, a proposed mea-

sure of trait relevance, possessed a genetic component independent of trait level. Therefore, a genetic influence appears to exist that affects variability in responding to trait scales. This finding again suggests that there is a source of error in the observed scores of individuals apparently attributable to differential trait relevance.

### Internal Consistency

The idea that lower trait relevance is related to greater variability in responding across scale items has implications for the assessment of a scale's internal consistency. Spearman (1904) was among the first to use the internal consistency of a scale as a measure of the scale's reliability. Spearman reasoned that responses to scale items designed to tap the same latent construct should evidence sizable covariation. Accordingly, if responses to the items do not covary, then the items are not capturing a unidimensional construct. Spearman and others developed formulas to assess the extent to which items within a scale covary (see also, Cronbach, 1951; Kaiser & Michael, 1975; Novick & Lewis, 1967; Rulon, 1939).

The computation of the most common method of assessing internal consistency, coefficient alpha, is determined in part by the degree to which the responses to the items comprising the scale covary for each individual (see Ebel, 1972). If an individual responds consistently across items, then the covariation is high. If an individual responds variably across items, then the covariation is low. High item covariation within individuals leads to greater internal consistency. Yet, as noted earlier, Britt and Kraus (1994) found that lower trait relevance is related to greater variability in responding to scale items. In addition, genetic influences exist that cause some individuals to respond more variably to a given trait scale (Hershberger et al., 1995). Clearly, the presence in a sample of people for whom the trait is less relevant and who display low covariation in their item responses diminishes the internal consistency of the scale.

The consequences can be quite serious. For example, consider two scales, A and B, being completed by a sample of people. If the trait measured by scale A is more relevant to the sample than the trait measured by scale B, the internal consistency of scale B may be lower than scale A not because of poor item quality, but because of greater measurement error as a function of lower trait relevance. Indeed, a scale may be perfectly reliable, measuring a trait without error, yet appear unreliable because it is highly relevant only to handful of people in the sample. The presence in the sample of people for whom the trait is not relevant would serve to obscure the assessment of a scale's in-

ternal consistency. The more of these individuals present in the sample, the greater attenuation of the scale's internal consistency.

This discussion has implications for how we think about threats to internal consistency. Measures of internal consistency traditionally have been viewed as possessing two sources of error variance: error variance attributable to content sampling and error variance attributable to content heterogeneity (Anastasi, 1982). Both sources of error arise from imperfections in the items selected to represent the construct. Our discussion of trait relevance suggests that a third source of error variance exists for traits that are not universally relevant to all individuals. The error variance arises not from the items comprising the measurement instrument, but from the individuals who complete the instrument. We call this third source error attributable to differential trait relevance and propose that the magnitude of this error variance is a direct function of the extent to which a given trait is relevant to a sample of individuals.

Identifying a third threat to internal consistency raises questions about the usefulness of formulas developed to correct for attenuation. Spearman (1904) developed a procedure to correct a correlation for attenuation resulting from unreliability in the measurement of the constructs being correlated. The correction formula provides an estimate of the relation between two constructs were it possible to measure one or both constructs perfectly. However, Spearman's formula for correcting for attenuation does not address the problem of error variance attributable to differential trait relevance. Indeed, using the formula to correct for attenuation without regard for the source of attenuation could provide misleading results. Such would be the case were a scale's modest coefficient of internal consistency attributable to the fact that the trait measured by the scale is irrelevant to much of the sample rather than to error in the measurement of the construct.

### Trait Relevance and Validity

Our discussion of differential trait relevance has important implications for the validity of instruments assessing traits that differ in their relevance to a group of individuals. Typically, during the development of an instrument, it is administered along with numerous other measures to a large sample of individuals, and various procedures are used to examine the validity of the instrument. Rarely is any thought given to the fact that individuals differ in trait relevance. Yet as we have noted, there is a systematic source of error that is created as a function of differential trait relevance. As such, those individuals for whom a trait is less relevant contribute more error to the validation process. The net

effect is that ignoring differences in trait relevance during the validation process can produce a systematic underestimation of the validity of the instrument. As evident in our discussion of internal consistency, which directly affects the size of the validity and reliability coefficients that can be obtained (Crocker & Algina, 1986), it is possible that a perfectly reasonable measure of a particular trait will be labeled as less valid not because of any inherent flaw in the instrument, but because the trait being measured is less relevant in the sample completing the instrument. In the following section we discuss trait relevance in regard to two types of validity: construct validity and criterion-related validity. We also propose an additional type of validity that taps the extent to which a trait is relevant to a given sample.

### Construct Validity

*Construct validity* has two components: convergent and discriminant validity (Cronbach & Meehl, 1955). *Convergent validity* for a construct is obtained when responses on the instrument measuring the construct correlate with responses on instruments measuring other theoretically related constructs (Anastasi, 1982; Campbell & Fiske, 1959; Cronbach & Meehl, 1955). The correlation between instruments assessing two theoretically related constructs should be strong enough to evidence the presumed relation, but not so strong as to raise questions about whether the two constructs are distinct. *Discriminant validity* is obtained when responses on instruments measuring theoretically distinct constructs do not correlate strongly.

Two areas of research on trait relevance bear on the issue of convergent validity. The first addresses the moderating effects of trait relevance on the correlation between self and observer ratings of a person's standing on a given trait. Specifically, several studies have examined whether measures of trait relevance moderate the extent to which self and peer reports of personality correlate with one another (Biesanz et al., 1998; Chaplin, 1991; Lanning, 1988). Many of these studies defined *trait relevance* in terms of some form of variability across scale items (Chaplin, 1991; Lanning, 1988), with the logic being that people who respond more variably across scale items, or who conform less to a normative response pattern, are less likely to be traited on the underlying trait dimension (or in the terminology of this article, the trait is less relevant to the individual). The results of these studies supported the idea that trait relevance moderates the size of self-peer correlations, but that the size of the moderating effect is small, and somewhat inconsistent across moderators (see Chaplin, 1991).

The recent research by Biesanz et al. (1998) used a different measure of trait relevance, and obtained more convincing evidence that trait relevance moderates the extent of self-peer correlations. These authors proposed temporal response pattern stability (the stability of an individual's score over repeated administrations of a test) as a measure of what we are referring to as trait relevance. Their findings revealed replicable moderating effects on self-peer correlations, such that the correlations between self and peer reports of personality were larger when the individual's self-ratings on the personality traits were consistent across time than when the self-ratings were inconsistent across time.

The second area of research on trait relevance that pertains to convergent validity concerns the moderating effects of trait relevance on the correlations among theoretically related constructs. Britt (1993), using moderated multiple regression, found that correlations among theoretically related personality traits were higher when the traits were highly relevant to the individual (as evidenced by low interitem variability to items tapping the same trait) than when one or both traits were less relevant to the individual (as evidenced by higher interitem variability to items tapping the same trait). In conceptually related research, Judd and Krosnick (1989) found that between-attitude consistency was greater among experts than among novices, suggesting that conceptually related attitudes are more likely to be related when the attitudes are relevant to the individual by virtue of the individual possessing knowledge about the attitudes. The moderating effects of trait-relevance on self-peer correlations and on the correlations between theoretically related traits speak directly to the implications of trait relevance for the assessment of construct validity. In particular, these findings have implications for research and thinking on nomological networks. Cronbach and Meehl (1955) argued that constructs within a given domain should be placed within a nomological network so that relations among constructs are clearly identified. By indicating the relations among constructs, researchers could prepare the theoretical domain for the entry of new constructs into the network, insuring that the variables are not redundant with existing variables in predicting important criteria. These findings suggest that low trait relevance on a construct within a network can obscure the relation between that construct and other theoretically related constructs. For example, consider five theoretically related constructs within a nomological network. To the extent that trait relevance for some people is low on one of the constructs, then the corresponding correlations between that construct and the remaining four constructs may be diminished. In short, trait relevance can influence

the empirical interrelations of theoretically related constructs, an effect researchers should consider when examining new (and old) constructs within a existing theoretical system.

Although research has focused primarily on the implications of trait relevance for the assessment of convergent validity, the findings also have implications for the assessment of discriminant validity. For example, a researcher may find no correlation between two constructs that are presumed to be unrelated. However, the absence of a correlation may be attributable to no theoretical relation between the two constructs or to the presence in the sample of people for whom one or both constructs are low in relevance. These people could weaken the relation between two constructs that are actually related. This problem may be especially severe for constructs that are highly relevant to a relatively small proportion of the population. In such cases, nonsignificant correlations with other variables may be more a function of the large number of people who are low in trait relevance for one or both of the constructs than a function of the lack of a presumed theoretical relation. Of course, if the personality traits are "actually" unrelated, then differential trait relevance should have no effect on the correlation among the traits.

Consider, for example, the trait of objectivism, which taps the tendency to rely on observable, empirical information in making judgments, rather than on feelings or intuitions (Leary, Shepperd, McNeil, Jenkins, & Barnes, 1986). This trait is likely highly relevant to some people but not others. Indeed, it is possible that the trait of objectivism is consequential in the experience and behavior of only a few people and is inconsequential in the experience and behavior of most people. Presumably, the few people for whom the trait of objectivism is highly relevant have trait scores that span the entire continuum of objectivism. However, people for whom the trait of objectivism is low in relevance likely also have trait scores that span the entire continuum of objectivism. The presence of people in a validation sample for whom the trait is low in relevance are adding error to the measurement process, and the error is likely to obscure the relation (i.e., diminish the correlation) between the measure of the trait and measures of other traits. As discussed in greater detail shortly, we argue that traits that deal with more circumscribed areas of behavior will likely evidence high relevance for a limited proportion of the population.

Finally, our discussion of trait relevance suggests a new way to examine convergent validity within a nomological network. Specifically, for convergent validity to be obtained, individuals for whom a given trait is highly relevant should also display high trait relevance for theoretically related traits. Thus, for ex-

ample, a person for whom the trait of shyness is highly relevant should also be high in relevance for other traits that fall within the same nomological network as shyness, such as audience anxiety and interaction anxiety. Similarly, a person for whom shyness is relatively low in relevance should show low relevance on measures of audience anxiety and interaction anxiety. Support for this novel use of information about trait relevance is exemplified in research on political ideologies (McGuire, 1989), in which the possession of one attitude is presumed to be related to the possession of other attitudes (see Judd & Krosnick, 1989). For example, individuals who possess an attitude on the abortion issue should also possess attitudes on capital punishment, gun control, and government spending.

### Criterion-Related Validity

*Criterion validity* is obtained when the construct being assessed predicts theoretically relevant behaviors (Anastasi, 1982; Crocker & Algina, 1986; Cronbach & Meehl, 1955). For example, a measure of social anxiety should predict various verbal and nonverbal behaviors when an individual is placed in a social encounter. To the extent that scores on an assessment instrument meaningfully correlate with trait relevant behaviors, increased confidence is placed in the validity of the instrument as a measure of the construct.

Criterion validity coefficients should be greater for traits that are highly relevant to an individual. For traits that are not relevant, the trait should be relatively unpredictable of the individual's behavior. Research from several studies is consistent with this reasoning (Baumeister & Tice, 1988; Bem & Allen, 1974; Fazio & Williams, 1986). For example, Bem and Allen (1974), in a highly influential article, examined whether some people are more situationally consistent than are other people. Bem and Allen argued that individuals differ in their consistency of behavior with regard to a given personality trait and that this inconsistency would affect the size of trait-behavior correlations. In line with this reasoning, Bem and Allen showed that individuals who reported greater situational variability with regard to a particular trait,<sup>4</sup> or who responded more variably to the items measuring the trait, evidenced lower trait-behavior correlations than did individuals who reported less cross-situational variability or responded less variably to items measuring the trait.

<sup>4</sup>Reporting of situational variability relevant to a given trait also requires some metaknowledge of the trait in question. However, individuals are asked specifically to comment on their behavior, rather than the aspects of the trait in question.

In terms of this article's terminology, Bem and Allen showed (1974) that a given trait may be more relevant and thus more predictive of behavior for some people than for others. In the attitude domain, Fazio (1990) have also shown that "strong" attitudes (attitudes to which people exhibit fast response latencies when queried) are more likely to be accessible in memory and therefore are more likely to guide behavior. These findings suggest that failing to consider a trait's relevance may lead researchers to underestimate the criterion validity of an instrument, perhaps leading to the erroneous conclusion that an instrument enjoys poor criterion validity when in fact it predicts quite well for individuals for whom the trait is highly relevant, but poorly for individuals for whom the trait is not relevant. Importantly, other researchers have failed to replicate the findings of Bem and Allen (Chaplin & Goldberg, 1984; Cheek, 1982; Mischel & Peake, 1982). We discuss possible reasons for this failure to replicate in the section on measuring trait relevance.

### Trait Relevance Validity

We have characterized trait relevance as a potential source of error that can influence estimates of validity. However, we could also characterize trait relevance as a distinct type of validity in itself. Measurement theorists have traditionally identified three types of validity: Content validity, criterion validity, and construct validity (Anastasi, 1982). *Content validity* addresses whether the items comprising the instrument adequately represent the domain of interest. *Criterion validity* addresses whether the instrument effectively predicts behavior in a specified situation. *Construct validity* addresses whether the instrument measures the theoretical construct or trait it was designed to measure. Our discussion suggests a fourth type of validity—trait relevance validity. *Trait relevance validity* addresses the extent to which the psychological construct being measured is relevant to the target population. If a trait is highly relevant for the target population, then trait relevance validity is high. If the trait is largely irrelevant to the target population, then trait relevance validity is low. For example, the argument could be made that because of their pervasiveness and emergence in natural language, the Big Five personality dimensions are relevant to all individuals (John, Hampson, & Goldberg, 1991; McCrae, 1993). If this is the case, then the Big Five personality dimensions possess a high degree of trait relevance validity. However, even proponents of the Five-Factor Model of personality do not argue for universal trait relevance across cultures. For example, McCrae and Costa (1997) presented evidence in support of the generaliza-



tion of the Five-Factor Model across different cultures, but also noted that

The fact that the five factors can be found in different cultures does not mean that they play the same role everywhere. . . . Individual differences in Agreeableness versus Antagonism may be of little consequence in societies in which interpersonal relationships are rigidly dictated by social roles, and variations in Conscientiousness may be unimportant in cultures that devalue personal ambition. (p. 514)

In the context of the terminology of this article, McCrae and Costa (1997) argued that traits may differ in their relevance to different cultures.

Prior researchers have examined differences among traits that speak to the issue of trait relevance validity. For example, Funder and his colleagues (Funder & Colvin, 1988; Funder & Dobroth, 1987) demonstrated that traits that are more observable, in the sense of having more visible behaviors that reflect the trait, are more likely to show higher self-peer correlations. In addition, Hayes and Dunning (1997) recently argued that traits differ in their *ambiguity*, which they define as the potential number of behaviors to which a trait is applicable. Hayes and Dunning (1997) showed that ambiguous traits were more likely to manifest low self-peer correlations than unambiguous traits. In addition, trait observability and trait ambiguity independently added to the strength of self-peer correlations.

Although trait observability and trait ambiguity may contribute to greater trait relevance validity, we would argue that the main determinant of trait relevance validity is the prominence of behaviors relevant to the trait in the population of individuals being sampled. In fact, it is most likely the prominence and breadth of behaviors relevant to the Big Five personality traits that supports the argument for these traits possessing a high degree of trait relevance validity (John et al., 1991; McCrae, 1993). Traits that are less pervasive and refer to more restricted types of behavior could be argued to be lower in trait relevance validity. For example, traits such as physique anxiety and private self-consciousness refer to circumscribed domains of behavior and feelings that many people may not have experienced or thought about. When these individuals are then asked to provide self-ratings on the trait dimension, the responses of many may not reflect the presence of a trait that is highly relevant to their personality. Some researchers have argued that personality variables at this mid-level of analysis are actually at the optimal level for predicting behavior and experience (Briggs, 1989; Paunonen, 1998). If this is true, and trait relevance is most important at a mid-level analysis of personality dimensions, then variations in trait relevance might make the most dif-

ference at the level of trait organization most relevant to psychological functioning.

## Measuring Trait Relevance

### Traditional Measures of Trait Relevance

The preliminary work on trait relevance by Allport and others was largely theoretical. However, Bem and Allen (1974) proposed that it was possible to measure individual differences in what we are calling trait relevance. Bem and Allen measured this construct in two ways. The first involved a single 7-step item asking participants how much they varied from one situation to another in behavior indicative of the trait. In their case, the trait was friendliness. According to this article's terminology, this method is predicated on the assumption that people low in trait relevance will evidence greater inconsistency in behavior relevant to the trait than will people high in trait relevance. The second method, called the *ipsatized variance index*, involved computing for each participant the ratio of the participant's interitem variance on the trait measure of interest (in their case, conscientiousness) to the participant's interitem variance on a number questionnaires she or he completed (including the conscientiousness scale). According to this article's terminology, this second method is predicated on the assumption that people low in trait relevance will respond more variably across items tapping the trait than will people high in trait relevance. For both methods, participants identified as high in trait relevance displayed higher trait-behavior correlations than did participants identified as low in trait relevance.

Bem and Allen's (1974) purpose in developing a measure of perceived situational variability was to improve the ability of trait instruments to predict behavior (see also Kenrick & Stringfield, 1980). To this end, their efforts were met with mixed results, largely because of problems in their two measures of consistency (Paunonen & Jackson, 1985; Rushton, Jackson, & Paunonen, 1982), problems that undermined the ability to replicate Bem and Allen's general findings (Chaplin & Goldberg, 1984; Cheek, 1982; Mischel & Peake, 1982). Specifically, the ipsatized variance index introduced an element of arbitrariness in the assessment process (Chaplin & Goldberg, 1984; Paunonen & Jackson, 1985; see Baumeister & Tice, 1988, for a review). Furthermore, single-item measures of a construct tend to be unreliable.

It is noteworthy that single-item self-report measures have been used in most of the studies examining moderators assessing trait metaknowledge, both in the attitude and personality domains (Basilli, 1996; Judd

& Krosnick, 1989; Pelham, 1993; Zuckerman et al., 1988, 1989). However, the problem of low reliability of single-item measures (i.e., single-item moderators) has been virtually ignored. Although some may argue that finding an effect with a single-item moderator is especially convincing given its inherent low reliability, we are skeptical. There is no substitute for demonstrated reliability. Because of their low reliability, the effects of single-item moderators may be difficult to replicate. Moreover, if the reliability of a moderator is unknown, as is the case with single-item measures, then there is no way to determine whether an unsuccessful moderator failed because of low reliability or because the moderator being tapped by the single-item measure is, in fact, unrelated to the behavior or outcome of interest. Indeed, the modest effect of some moderators may be due to their low reliability (Chaplin, 1991).

Baumeister and Tice (1988) offered two other methods for assessing trait relevance, both of which they quickly dismissed with good reason. The first, trait extremity, was dismissed because it systematically confounds trait score with trait relevance. The second, developing a detailed measure of behavioral consistency for each trait, was dismissed because it requires a seemingly infinite number of new scales.

One of the more popular methods for assessing trait relevance is computing for each individual the interitem variance of the items comprising the instrument (Baumeister & Tice, 1988). This procedure assumes that individuals high in trait relevance will respond more consistently (either at one of the extremes or in the middle) to items assessing the trait than will individuals low in trait relevance. Take, for example, the trait of friendliness. In a given sample of people, friendliness may be highly relevant to some people but not others. Among the people high in trait relevance, those who are high in friendliness will consistently supply responses reflecting high friendliness, those who are low in friendliness will consistently supply responses reflecting low friendliness, and those who are moderate in friendliness will consistently supply responses reflecting moderate friendliness. By contrast, people for whom friendliness is less relevant will respond inconsistently across items comprising the friendliness measure. The result will be lower interitem variability on the friendliness scale for the high relevance people than for the low relevance people.

Although interitem variance may appear a suitable measure of trait relevance, it can be misleading if calculated on raw item responses. Specifically, Tellegen (1988) argued that an apparently consistent pattern of responding may be quite inconsistent once the mean and standard deviation of each item are taken into account. For researchers using interitem variability as a

measure of trait relevance, Tellegen (1988) recommended standardizing the item responses before calculating individuals' interitem variability.

Interitem variance, especially when calculated on standardized item scores, is an appealing method for assessing trait relevance; it is relatively easy to calculate, and seems intuitively correct. However, interitem variance has at least three drawbacks: (a) It is often confounded with trait extremity, (b) it is ill suited to some response formats, and (c) other factors in addition to low trait relevance can produce interitem variability.

The first drawback to using interitem variance to assess trait relevance is that trait relevance and extremity are often naturally confounded. That is, individuals who supply responses that consistently fall at one or the other end of a trait continuum are more likely to have lower interitem variability than are individuals who supply responses toward the middle of the continuum (Paunonen & Jackson, 1985; Rushton et al., 1982; Tellegen, 1988). Baumeister and Tice (1988) argued that a moderate correlation between trait relevance and attribute extremity is to be expected; individuals receiving extreme scores, after all, are probably more likely to be high in trait relevance than are individuals with less extreme scores (see also Markus, 1977). Nevertheless, Baumeister and Tice advised either discarding from the sample individuals who are low in trait relevance or statistically checking the extent of the confound. Of note, hierarchical multiple regression can control for the correlation between trait relevance and trait extremity (Bissonnette, Ickes, Bernstein, & Knowles 1990a, 1990b; Britt, 1993; Paunonen & Jackson, 1985; Tellegen, 1988). The researcher need only enter into a regression model (a) the score on the predictor and (b) the interitem variabilities of the predictor and criterion before entering the two- and three-way interactions of trait score and interitem variabilities (see Britt, 1993, for an illustration). Thus, researchers can address the first drawback of interitem variability through multiple regression.

The second drawback of using interitem variability as a means of assessing trait relevance is that some instruments are not suited to this method. Specifically, instruments with a dichotomous response format, by definition, provide little or no interitem variance. Baumeister and Tice (1988) suggested fashioning instruments so as to maximize interitem variance. Although this is fine for the development of new instruments or for instruments that already rely on a Likert-type format, it is problematic for established instruments that use a true-false (e.g., the original self-monitoring scale, Snyder, 1974), forced-choice (e.g., the Myers-Briggs Type Indicator, Myers, 1975), Guttman (e.g., the Conflict Tactic Scale, Straus, 1979), or other formats that produce no interitem variance or

for which interitem variance is confounded with the scale score.<sup>5</sup>

A third drawback to interitem variability is that other factors besides low trait relevance can produce variability in item responses. For example, interitem variability may be attributable to error in the content or homogeneity of the items, deliberate random responding, poor reading ability, low intelligence, and limited command of the language in which the instrument is written (Nichols & Greene, 1986, as cited in Tellegen, 1988). These alternative sources of variability in interitem responding threaten the validity of interitem variance as a measure of trait relevance. They can be particularly problematic if they lead researchers to underestimate the proportion of the population high in trait relevance. For example, a researcher may conclude that a high level of interitem variability in responding indicates that a given trait is relatively rare in the population when in fact the validation sample includes many individuals with poor reading skills.

In sum, interitem variability (calculated across standardized item scores), represents a reasonable, albeit limited, measure of trait relevance. Given the potential importance of trait relevance for the field of personality, we encourage the use of multiple methods to converge on the construct. In the following, we offer several other methods for measuring trait relevance. Although data on the utility of these methods is limited, these methods seem reasonable directions for future research efforts. Our selection of these methods was based on two considerations: (a) The approach in some way assesses the extent to which a trait is consequential for an individual, and (b) the approach does not require the individual to consciously comment on aspects of the trait being assessed.

### Alternative Methods for Measuring Trait Relevance

One potentially fertile approach to assessing trait relevance comes from assessing people's response patterns on personality measures. Biesanz et al. (1998) determined an individual's degree of "temporally stable response patterns" by calculating the variability of an individual's score over repeated (three) administrations of a personality scale. The authors reasoned that individuals who show greater consistency in their trait

scores over repeated administrations of the trait instrument should also evidence greater consistency in trait-relevant behavior, which would result in higher correlations between self and peer ratings of personality. As we noted earlier, Biesanz et al. (1998) found higher self-peer correlations among individuals who were consistent in their scores on the personality dimensions over repeated administrations of the test than among individuals who were less consistent in their scores. Furthermore, in comparing different measures of trait relevance (scalability, interitem variance, temporally stable response patterns), the authors found that temporally stable response pattern was the only assessment to significantly moderate the relation between self and peer reports of personality in two different studies. Therefore, this measure of trait relevance may prove to be very useful in assessing the implications of trait relevance in different areas.

A second approach to assessing trait relevance is to measure the response latency to scale items. Fazio (1990) used response latency to index having an attitude along an attitude-nonattitude continuum. Markus (1977) also used response latency to examine the effect of being *schematic* (possess a cognitive representation) on a given self-descriptor. As a measure of trait relevance, individuals high in trait relevance should respond more quickly to items assessing the trait than individuals low in trait relevance.

Markus's (1977) conceptualization of schemata as organized bodies of knowledge about the self suggests a third method for assessing trait relevance. Individuals' autobiographical memories relevant to a specific trait may be *prima facie* evidence for high trait relevance. Kihlstrom and his colleagues (Kihlstrom et al., 1988) provided evidence that autobiographical memories are linked to a vast body of knowledge about the self. We propose that the more autobiographical memories an individual possesses relevant to a particular trait, the greater the relevance of the trait to the individual. In assessing trait relevance through the number of autobiographical memories relevant to a given trait, researchers could use a method similar to McAdams's (1989) use of self-narratives. McAdams (1990) referred to *self-narratives* as personal accounts about peoples' lives, in which they define who they are, who they have been in the past, and who they may become. These types of narratives could be examined for the frequency with which certain traits are mentioned, or the extent to which behavior relevant to a given trait is recalled. Presumably, the more times the person mentions the trait in the context of his or her life story, the more relevant the trait is to the person.

Note that although this approach assesses the presence of trait constructs in narrative, it does not require asking participants to reflect on aspects of the trait in question. The approach would simply require the anal-

<sup>5</sup> Another circumstance in which interitem variability may provide little useful information regarding trait relevance is when a researcher knows that the trait is not relevant to a significant proportion of the population, yet the instrument assessing the trait has high internal consistency when administered to a representative sample (i.e., Cronbach's  $\alpha > .90$ ). Such may be the case when the items comprising the scale are so similar as to be little more than restatements of the same single idea.

ysis of the natural flow of the narrative in determining which traits appear dominant in the individual's experience. This approach is reminiscent of Allport's (1965) idea of personal structure analysis, and is epitomized in Allport's (1965) study of "letters to Jenny," in which the key elements of Jenny's personality were extracted through an analysis of letters she had written to another couple.

Another method for assessing trait relevance involves borrowing procedures developed in the cognitive and achievement testing literatures to detect aberrancy in individual responses to an assessment instrument. Sometimes termed *appropriateness indices* (Drasgow, Levine, & Williams, 1985; Levine & Rubin, 1979), *caution indices* (Harnisch & Linn, 1981), or *norm conformity indices* (Tatsuoka & Tatsuoka, 1982), these procedures examine the extent to which a particular individual's pattern of responding corresponds to a normative pattern. If an individual's response pattern is nonnormative (e.g., if the individual scores low on the test, but answers questions that only the top portion of the sample answered correctly), then the test is regarded as not valid for that individual.

Appropriateness indexes were developed by educational researchers to identify individuals for whom a test was not valid, either because they were responding haphazardly, cheating, or for some other reason supplying aberrant responses. We believe these measures may be useful in assessing trait relevance, a belief shared by other researchers (Lanning, 1988; Reise & Waller, 1993). That is, for traits that are not universally relevant, nonnormative responding would indicate the trait was less relevant to the individual.

Bem and Allen (1974) alluded to this approach in their discussion of the scalability of behaviors. Specifically, trait-relevant behaviors could be arranged along a Guttman-like continuum from *easy* to *difficult*. All people for whom the trait was at least moderately relevant would endorse easy items. However, only people high in trait relevant would endorse the difficult items. Generally speaking, a high score on the scale would be analogous to a high trait score, whereas a low score on the scale would be analogous to a low trait score. *Low relevance* would be defined as an aberrant pattern of responding to the items. That is, individuals would not be regarded as high in trait relevance if they endorsed difficult items while failing to endorse easier items. This final approach to measuring trait relevance is intriguing and may be suitable for instruments with response formats that do not lend themselves to an interitem variability approach. A limitation of this approach, however, is that it assumes that the items comprising an instrument can be sorted (preferably a priori) on a continuum from *easy* to *difficult*. For many existing psychological instruments, this is not possible.

Finally, it seems reasonable that researchers could assess a person's trait relevance by surveying the perceptions of peers. Funder (1991) argued that an individual's standing on a personality trait is best assessed by peer ratings. The same may be the case for ratings of trait relevance. For example, close others should be able to rate the extent to which a given trait is consequential for an individual, as in whether a trait is useful in understanding a person's behavior across different situations. Once such peer assessment measures were developed, researchers could examine the congruence between peer-ratings of trait relevance with other indexes of trait relevance.

## Conclusions

Our central argument is that the concept of differential trait relevance has important consequences for the procedures used to evaluate instruments assessing psychological constructs. First, differential trait relevance may influence a scale's internal consistency by affecting the variability with which individuals respond to items tapping the same construct. Second, to the extent that a trait is not universally relevant, attempts to correct for error due low trait relevance by using procedures more appropriate for error attributable to imperfect items can produce misleading findings. Finally, differential trait relevance can affect estimations of construct and criterion validity.

The premise that people for whom a trait is not relevant produce an underestimation of validity may seem overstated. Specifically, one could argue that a validity estimate calculated on a sample that includes such people is an accurate portrayal of how valid the instrument is for the sample as a whole. Accordingly, the validity estimate should be thought of as an aggregate value that may be inaccurate for any particular individual but is quite accurate for the sample as a whole. We believe this thinking is misleading. To illustrate, imagine a trait that is highly relevant to a small proportion of the population. If administered without regard to relevance, the instrument assessing the attribute, even if highly valid, would likely appear to have unacceptably low validity. However, were the validation sample separated into two groups (those for whom the trait was highly relevant and those for whom the trait was low in relevance), the instrument would likely predict well for the segment of the population for whom the trait was relevant, yet not at all for the segment of the population for whom the trait was relatively irrelevant. In short, the validity estimate obtained from administering an instrument without regard to trait relevance, although perhaps accurate for the aggregate, would be misleadingly high for people for whom the trait is low

in relevance and misleadingly low for people for whom the trait is highly relevant.

Clearly, psychologists need to go beyond merely determining where a person falls on a given trait continuum by summing the responses to scale items measuring the trait. Just because a person supplies responses to items comprising a scale does not mean that the underlying trait is relevant to that person. Indeed, a person could intentionally respond randomly to items comprising a trait measure, and these random responses could then be summed by an unknowing researcher to produce a trait score for that person. However, the trait score would not be meaningful because the responses were random. They would not reflect the person's true standing on the trait nor would they predict the person's trait relevant behaviors. In a similar vein, a person for whom the trait is less relevant could supply responses to a trait instrument, and these responses could be summed to produce a trait score. However, as in the case of the intentional random responder, the trait score is largely meaningless, not because the person is intentionally responding randomly, but because the underlying latent trait is not relevant to the individual.

It is important to note that our discussion of trait relevance has one encouraging implication: It is likely that the assessment instruments used by psychologists and other social scientists are more accurate and valid than previously believed. With perhaps the exception of instruments designed to assess some forms of mental illness, we can think of no case in which the evaluation of an assessment instrument has been limited to individuals for whom the trait is known to be relevant. Thus, we believe that the precision of the instruments measuring constructs that are differentially relevant to individuals is far better than psychologists realize.

## References

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G. W. (1962). The general and the unique in psychological science. *Journal of Personality*, 30, 405-422.
- Allport, G. W. (1965). *Letters from Jenny*. New York: Harcourt Brace Jovanovich.
- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.
- Azjen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 1-63). New York: Academic.
- Bassili, J. N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. *Journal of Personality and Social Psychology*, 71, 637-653.
- Baumeister, R. F. (1991). On the stability of variability: Retest reliability of metatraits. *Personality and Social Psychology Bulletin*, 17, 633-639.
- Baumeister, R. F., & Tice, D. M. (1988). Metatraits. *Journal of Personality*, 56, 571-598.
- Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 81, 506-520.
- Biesanz, J. C., West, S. G., & Graziano, W. G. (1998). Moderators of self-other agreement: Reconsidering temporal stability in personality. *Journal of Personality and Social Psychology*, 75, 467-477.
- Bissonnette, V., Ickes, W., Bernstein, I., & Knowles, E. (1990a). Item variances and median splits: Some discouraging and disquieting findings. *Journal of Personality*, 58, 595-601.
- Bissonnette, V., Ickes, W., Bernstein, I., & Knowles, E. (1990b). Personality moderating variables: A warning about statistical artifact and a comparison of analytic techniques. *Journal of Personality*, 58, 568-587.
- Blass, T. (1984). Social psychology and personality: Towards a convergence. *Journal of Personality and Social Psychology*, 47, 1013-1027.
- Briggs, S. R. (1989). The optimal level of measurement for personality constructs. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 246-260). New York: Springer-Verlag.
- Britt, T. W. (1993). Metatraits: Evidence relevant to the validity of the construct and its implications. *Journal of Personality and Social Psychology*, 65, 554-562.
- Britt, T. W., & Kraus, S. J. (1994, August). *Trait strength and response variability*. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality*, 59, 143-178.
- Chaplin, W. F., & Goldberg, L. R. (1984). A failure to replicate the Bem and Allen study of individual differences in cross-situational consistency. *Journal of Personality and Social Psychology*, 47, 1074-1090.
- Check, J. M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer rating study. *Journal of Personality and Social Psychology*, 43, 1254-1269.
- Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue. In E. R. Tufté (Ed.), *The quantitative analysis of social problems* (pp. 168-189). Reading, MA: Addison-Wesley.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart & Winston.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Drasgow, F., Levine, M. V., & Williams, E. (1985). Appropriateness measurement with polychotomous item response models and standard indices. *British Journal of Mathematical and Statistical Psychology*, 38, 67-86.
- DuBois, P. H. (1966). A test-dominated society: China 1115 B.C.-1905 A.D. In A. Anastasi (Ed.), *Testing problems in perspective* (pp. 29-36). Washington, DC: American Council on Education.
- Ebel, R. L. (1972). *Essentials of educational measurements yearbook*. Highland Park, NJ: Gryphon.
- Fazio, R. H. (1989). On the power and functionality of attitudes: The role of attitude accessibility. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 153-179). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark

- (Eds.), *Review of personality and social psychology: (11). Research methods in personality and social psychology* (pp. 74–97). Newbury Park, CA: Sage.
- Fazio, R. H., & Williams, C. S. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology, 51*, 505–514.
- Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science, 2*, 31–39.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149–158.
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52*, 409–418.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement, 18*, 133–146.
- Hayes, A., & Dunning, D. (1997). Construal processes and trait ambiguity: Implications for self-peer agreement in personality judgment. *Journal of Personality and Social Psychology, 72*, 664–677.
- Hershberger, S. L., Plomin, R., & Pedersen, N. L. (1995). Traits and metraits: Their reliability, stability, and shared genetic influence. *Journal of Personality and Social Psychology, 69*, 673–685.
- Higgins, E. T. (1990). Personality, social psychology, and person-situation relations: Standards and knowledge activation as common language. In L. A. Pervin (Ed.), *Handbook of personality* (pp. 301–338). New York: Guilford.
- Higgins, E. T. (1997). The “self-digest”: Self-knowledge serving self-regulatory functions. *Journal of Personality and Social Psychology, 71*, 1062–1083.
- Jaccard, J., & Dittus, P. (1990). Idiographic and nomothetic perspectives on research methods and data analysis. In C. Hendrick & M. S. Clark (Eds.), *Review of personality and social psychology: (11). Research methods in personality and social psychology* (pp. 312–351). Newbury Park, CA: Sage.
- James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.
- John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology, 60*, 348–361.
- Judd, C. M., & Krosnick, J. A. (1989). The structural bases of consistency among political attitudes: Effects of political expertise and attitude importance. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 99–128). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kaiser, H. F., & Michael, W. B. (1975). Domain validity and generalizability. *Educational and Psychological Measurement, 35*, 31–35.
- Kelly, G. A. (1955). *The psychology of personal constructs* (Vol. 1 & 2). New York: Norton.
- Kenrick, D. T., & Stringfield, D. O. (1980). 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, 87*, 88–104.
- Kihlstrom, J. F., Cantor, N., Albright, J. S., Chew, B. R., Klein, S. B., & Niedenthal, P. M. (1988). Information processing and the study of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 145–178). San Diego, CA: Academic.
- Lamiell, J. T. (1987). *The psychology of personality: An epistemological inquiry*. New York: Columbia University Press.
- Lanning, K. (1988). Individual differences in scalability: An alternative conception of consistency for personality theory and measurement. *Journal of Personality and Social Psychology, 55*, 142–148.
- Leary, M. R., Shepperd, J. A., McNeil, M. S., Jenkins, T. B., & Barnes, B. D. (1986). Objectivism in information utilization: Theory and measurement. *Journal of Personality Assessment, 50*, 32–43.
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics, 4*, 269–290.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley.
- Magnusson, D. (1967). *Test theory*. Boston: Addison-Wesley.
- Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35*, 63–78.
- McAdams, D. P. (1989). The development of a narrative identity. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 160–174). New York: Springer-Verlag.
- McAdams, D. P. (1990). Unity and purpose in human lives: The emergence of identity as a life story. In A. L. Rabin, R. A. Zucker, R. A. Emmons, & S. Frank (Eds.), *Studying persons and their lives* (pp. 148–200). New York: Springer.
- McCrae, R. R. (1993). Moderated analyses of longitudinal personality stability. *Journal of Personality and Social Psychology, 65*, 577–585.
- McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509–516.
- McGuire, W. (1989). The structure of individual attitudes and attitude systems. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 37–69). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Mischel, W., & Peake, P. K. (1982). Beyond *deja vu* in the search for cross-situational consistency. *Psychological Review, 89*, 730–755.
- Myers, I. B. (1975). *Manual: The Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1–13.
- Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology, 74*, 538–556.
- Paunonen, S. V., & Jackson, D. N. (1985). Idiographic measurement strategies for personality and prediction: Some unredeemed promissory notes. *Psychological Review, 92*, 486–511.
- Pelham, B. (1993). The idiographic nature of human personality: Examples of the idiographic self-concept. *Journal of Personality and Social Psychology, 64*, 665–677.
- Pelham, B. W., & Swann, W. B., Jr. (1994). The juncture of intrapersonal and interpersonal knowledge: Self-certainty and interpersonal congruence. *Personality and Social Psychology Bulletin, 20*, 349–357.
- Petty, R. E., & Krosnick, J. A. (1995). *Attitude strength: Antecedents and consequences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

- Reise, S. P., & Waller, N. G. (1993). Traitendness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology, 65*, 143–151.
- Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test of split-halves. *Harvard Educational Review, 9*, 99–103.
- Rushton, J. P., Jackson, D. N., & Paunonen, S. V. (1982). Personality: Nomothetic or idiographic? A response to Kenrick and Stringfield. *Psychological Review, 88*, 582–589.
- Sherman, S. J., & Fazio, R. H. (1983). Parallels between attitudes and traits as predictors of behavior. *Journal of Personality, 51*, 308–345.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*, 526–537.
- Spearmen, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72–101.
- Spearmen, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18*, 161–169.
- Spearmen, C. (1927). *The abilities of man*. New York: Macmillan.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The conflict tactics (CT) scale. *Journal of Marriage and the Family, 41*, 75–88.
- Tatsuoka, K. K., & Tatsuoka, M. M. (1982). Detection of aberrant response patterns. *Journal of Educational Statistics, 7*, 215–232.
- Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality, 56*, 621–663.
- Tice, D. M. (1989). Metatraits: Interitem variance as personality assessment. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions*. New York: Springer-Verlag.
- Wicklund, R. A., & Eckert, M. (1992). *The self-knower: A hero under control*. New York: Plenum.
- Zuckerman, M., Bernieri, F., Koestner, R., & Rosenthal, R. (1989). To predict some of the people some of the time: In search of moderators. *Journal of Personality and Social Psychology, 57*, 279–293.
- Zuckerman, M., Koestner, R., DeBoy, T., Garcia, T., Maresca, B. C., & Sartoris, J. M. (1988). To predict some of the people some of the time: A reexamination of the moderator variable approach in personality theory. *Journal of Personality and Social Psychology, 54*, 1006–1019.
- Zuckerman, M., Miyake, K., Koestner, R., Baldwin, C. H., & Osbourne, J. W. (1991). Uniqueness as a moderator of self-peer agreement. *Personality and Social Psychology Bulletin, 17*, 385–391.