

In Defense of Personality Measurement: New Wine for Old Whiners

Robert Hogan
Hogan Assessment Systems
Tulsa, Oklahoma

The key to success in business is money and people. Personality psychology is about people—it's about the nature of human nature. Some understanding of human nature—and the ability to measure its key components—would seem to offer a huge advantage to applied psychologists. Despite its practical significance, personality has lived a troubled existence in academic psychology. The topic was popular after World War II. But the 1960s brought the response set (or faking) controversy, which challenged the foundation of personality assessment. Then the 1970s and 1980s brought Mischel's revolution, which taught us that we can't measure personality because it doesn't exist—people's actions are not determined by their personalities but by "situational factors." Situational factors are like dark matter in Physics—they are strange, undefined, invisible forces that exist "out there," that capture us, and then make us do their bidding.

In the 1990s, personality made a comeback in industrial psychology. The comeback was fueled by the news that well-constructed measures of personality predict job performance almost as well as measures of cognitive ability, but with no adverse impact. The critics went silent for about 10 years, but now they are back (cf. Schmitt, 2004). They argue that claims for the validity of personality measures have been vastly overstated, that the data reveal only trivial relations between these measures and occupational performance. The point of this article is to try to put the various issues in perspective and show that well-constructed measures of personality are an indispensable tool for applied psychologists. The article is organized in terms of four sections: (a) What is wrong with personality psychology? (b) What is

This article is based on articles presented by Robert Hogan at the 32nd International Congress on Assessment Center Methods, October, 2004, Las Vegas, NV.

Correspondence should be sent to Robert Hogan, Hogan Assessment Systems, 2622 East 21st Street, Tulsa, OK 74144. E-mail: rhogan@hoganassessments.com

wrong with the critics of personality psychology? (c) What is personality psychology? (d) What do the data actually say?

WHAT IS WRONG WITH PERSONALITY PSYCHOLOGY?

No one is more aware of the problems of personality psychology than I am. I can summarize these problems in terms of three points. First, there are only about 200 personality psychologists in the United States, and not many of them are active researchers. Among the active researchers, there is virtually no agreement regarding an agenda for the discipline. One group studies the sources of psychopathology. A second group tries to identify the basic structure of personality, with no concern for practical applications. A third group evaluates the neuropsychological foundations of personality and the heritability of the major dimensions. A fourth and very tiny group, most of whom live in Oklahoma, is concerned with predicting important practical outcomes—competence, effectiveness, leadership, creativity, integrity. Overall, however, there is no consensus about an intellectual agenda for the field, which means that the cumulative impact of the few existing researchers is minimized. In my view, the field only stays alive because it is intrinsically appealing to the general public.

A second problem with personality psychology is a generalized lack of concern for measurement validity. There are perhaps 2,500 test publishers in the United States, and only a very few pay attention to validity. The furor caused by the recent volume, *The Cult of Personality*, by Annie Murphy Paul (2004), is due entirely to the willingness of publishers to sell tests with no demonstrated validity. Not surprisingly, the reaction of the profession to this volume is a big yawn.

Moreover, the pioneers of personality measurement—J. P. Guilford, Raymond Cattell, and Hans Eysenck—regarded correlations between test scores and real-world criteria as “peripheral validity”; real validity is the degree to which factor structures replicate across samples. Ironically, the empirical tradition from Minnesota, which focuses on validity, is typically derided as “dust bowl empiricism.” The “pioneers” were interested in replicating factor structures, whereas the empirical tradition is interested in predicting outcomes. The “pioneers” had an academic agenda that is of little use to consumers of assessment services; the empirical tradition has an applied agenda that interests consumers but not academics.

The last problem concerns the quality of the research in personality psychology. Like the ability to play the piano, the ability to do research is normally distributed. A few people, like my wife, have a real talent for research, but most people don't. The popularity of meta-analysis has exacerbated the problem because it allows people to do research who have little talent for it. Many of the meta-analyses that evaluate the validity of personality have significant limitations. For example, researchers often include in the same analysis measures that are not commensurable.

Thus, they combine measures of normal personality with measures of psychopathology and values and interests. In addition, they include well validated measures (e.g., the California Psychological Inventory, or CPI) with poorly validated measures (e.g., the Myers–Briggs and the self-monitoring scale).

Moreover, high scores on the Agreeableness scale of the NEO Personality Inventory indicate people who try not to give offense, whereas high scores on the Likeability scale of the Hogan Personality Inventory (HPI) indicate people who are actively charming. These scales predict different things and don't belong in the same analysis—but they are routinely combined. Furthermore, researchers often fail to align predictors with criteria; this results in using measures of conscientiousness to predict service orientation, or measures of extraversion to predict training performance. The resulting correlations are low and critics then use them to indict personality research rather than the personality researchers. And then many researchers ignore the problem of bidirectionality—sometimes measures of conscientiousness are positively correlated with outcomes and sometimes negatively, but for sound theoretical reasons in both cases. Imagine that conscientiousness is negatively correlated with one kind of performance (rated creativity), but positively correlated with another kind of performance (rated compliance with rules). This suggests that conscientiousness is a robust predictor of performance. However, if you simply add the two sets of correlations together, they cancel each other, and lead to the conclusion that personality is a weak predictor of performance. Finally, the best known meta-analyses regarding the validity of cognitive ability measures used one test, the General Aptitude Test Battery, because this avoids the problems of classification and measurement equivalence. In the same way, meta-analyses of personality research should only use one inventory, rather than combining a bunch of them, all of which have different measurement goals.

We judge piano playing by the performance of the best players, not by the performance of the average players. Similarly, we should judge personality research by the performance of the best researchers, not by the performance of the average researchers. All researchers are not created equal.

WHAT IS WRONG WITH THE CRITICS OF PERSONALITY PSYCHOLOGY?

My argument here is frankly *ad hominem*—many of the critics of personality psychology are behaviorist ideologues and this is the source of their hostility to the field. Most behaviorists will no more be persuaded by data supporting the validity of personality measurement than creationists will be persuaded by data supporting evolutionary theory. As Nils Bohr, the father of atomic theory, remarked, in science we never persuade our critics, we have to wait for them to die.

The most important claim of personality psychology is that there are structures inside people (hopes, dreams, fears, and aspirations) that determine their behavior. This claim is anathema to behaviorists. B. F. Skinner, the king of behaviorism, was notoriously insensitive to human feelings (when his brother died, Skinner insisted on helping the county medical examiner with the autopsy, which he found quite interesting). Skinner went to his grave denying the truth of evolutionary theory and the existence of instincts—that is, denying that internal structures influence social behavior in important ways. Although the modern cognitive behaviorists have now discovered internal structures, they still don't know how to spell *Darwin*—that is, they invented structures with no concern for evolutionary theory. A good bit of the antipersonality sentiment in our profession is sheer ideology, promoted by critics who won't be persuaded by data.

Note also that academic psychologists don't compete in the real marketplace of ideas; rather, they play to captive audiences. They sell antipersonality arguments to students and like-minded academics, but people in the business community understand that individual differences in attitudes and values affect job performance, and they want to use assessment to make better hiring decisions. The problem is that business people have trouble getting good advice from academic psychology. This, in turn, explains the widespread interest in bogus measures of personality such as the Myers–Briggs Type Indicator and Goleman's Emotional Competence Inventory.

WHAT IS PERSONALITY?

Everyone has a theory of personality; we can't go to work without one. The problem is that these theories are informal, implicit, and unspecified. The same is true for industrial psychologists; although they use personality measures, their theory of personality is rudimentary. That is, they define personality as traits, and that is a mistake because trait theory has significant flaws. The problems can be quickly outlined. On the one hand, it confuses description with explanation, and is, therefore, completely tautological. For example, Mike Tyson is usually described as aggressive. Trait theorists want to explain Tyson's aggressive behavior in terms of a trait for aggression, and that's just dumb.

Sophisticated trait theorists try to escape from the tautology by arguing that regularities in behavior are caused (and explained) by underlying "neuropsychic structures." This is psychological reductionism, an effort to explain phenomena at one level in terms of phenomena at the next lower level of analysis. Thus, biology should be reduced to the laws of chemistry, chemistry should be reduced to the laws of nuclear physics, and overt behavior should be reduced to neuropsychic structures.

There are two problems with reductionism. On the one hand, the rest of science has simply moved on. For example, the Nobel Prize–winning physicist P. W. Anderson (1972) argued 30 years ago that biologists can study biology without worrying about chemistry, that chemists can study chemistry without worrying about physics, and so on. More recently, string theory—the theory of everything—is intended to make Einstein’s relativity theory (it’s about gravity) consistent with quantum mechanics (particles that have no gravitational properties). But some physicists think that relativity and quantum mechanics are separate disciplines that can be studied fruitfully on their own. Similarly, we can study occupational performance without resorting to physiology.

The second problem with reductionism in personality psychology is that, after 70 years, we still haven’t found any underlying neuropsychic structures—and don’t hold your breath. Obviously people are biological animals, and our actions reflect our genetic makeup, but that is all we need to say. Neuroscientists can study neuropsychology, and applied psychologists can study social behavior on its own terms. The bottom line is that trait theory is not a competent theory of personality.

It is a mistake to confuse the way we use trait words with trait theory. Trait words are indispensable for describing other people. However, other people don’t have traits; rather, we assign trait terms to them as a way of summarizing recurring themes in their behavior. There is a difference between description and explanation, and trait theorists ignore the distinction. We describe other peoples’ behavior with trait words, but we explain their behavior in terms of what they are trying to accomplish.

Personality is two things: (a) generalizations about human nature, and (b) explorations of individual differences. What generalizations can we make about human nature? Sociology, anthropology, and evolutionary psychology suggest three. First, people always live in groups. Second, every group has a status hierarchy. Third, every group has a religion, which is typically used to justify the status hierarchy and the existing moral and legal systems. This suggests that there are three overriding themes in individual lives: (a) efforts to get along with other people (because we live with them); (b) efforts to attain some power, status, and control of resources (more is always better); and (c) efforts to make some sense out of our lives (by interpreting them in terms of a quasi-philosophical system).

Personality psychology is also about individual differences. People differ from one another in many, many ways. These three generalizations—that people want acceptance, status, and meaning—suggest what the most important domains of individual differences might be. The first domain will concern individual differences in the desire for, and the ability to obtain, social acceptance and support. The second will concern individual differences in the desire for, and the ability to obtain, status, power, and the control of resources. The third will concern individual differences in the desire for meaning and purpose in life.

I have suggested three important generalizations about human nature. I have suggested three important vectors of individual differences. Together, these point

to a measurement agenda for personality psychology. In addition, please note that leaders are people who excel in their ability to gain acceptance and support, power and status, and to make meaning.

Finally, it is important to note that there is not one definition of personality, there are two. There is personality from the view of the actor, and personality from the view of the observer. Personality from the view of the actor—your view of you—is identity. Personality from the view of the observer—our view of you—is reputation. Identity and reputation are different, although somewhat related, concepts, and they have different implications for assessment. Self-reports—statements about who you think you are—are almost useless as data sources in and of themselves. As Freud might say, the you that you know is hardly worth knowing—because you made it all up. In any case, the study of identity is not very advanced, and has yielded few reliable generalizations.

On the other hand, reputation has several attractive features as a data source. First, we have a well-developed vocabulary for talking about reputations, and that is the vocabulary of trait words. Trait words are what we use to describe other people, and our descriptions of others are, in fact, their reputations. Second, we have a well-developed taxonomy of trait words, and it is the so-called Five-factor Model. Recent research indicates that there are more than five dimensions of normal personality, but the point is that we have a well-defined structure in terms of which trait terms can be organized. Third, we can use trait words reliably to characterize others' reputations. And finally, reputations are immensely useful for predictive purposes. The best predictor of future behavior is past behavior, reputations are a summary of past behavior, therefore reputations are the best information we have about future behavior.

The measurement model that we use involves taking the statements that people make about themselves and then determining, in an empirical way, the links between their self-descriptions and their reputations. Take, for example, such statements as, "I can get this country moving again" or "I can take this company to the next level." Our research indicates that peers and subordinates describe people who say this as egotistical narcissists. The trick, then, is to take what people say about themselves and translate it into what other people say about them. Conversely, it is always risky to take what people say about themselves at face value; the data are quite clear that people are poor judges of how they are seen by others (cf. Harris & Schaubroeck, 1988; Mabe & West, 1982). When a manager tells me, "My staff respects me," I always say, "Can we take a vote?"

WHAT DO THE DATA ACTUALLY SAY?

Before examining the validity data, it might be useful to consider briefly something that the data do not say. The most frequent complaint about personality mea-

tures that we hear is that they can be faked, which means that they are useless for preemployment selection. The empirical literature on this point is vast and extraordinarily tedious. Our view of the faking issue can be summarized in terms of four points. First, some people can improve their scores on well-validated measures of personality if you ask them to. However, there are individual differences in this capability, and many people, when trying to “enhance” their scores, actually “degrade” them. Second, when we evaluate a person for hiring purposes, we look at a profile, not a score on a single scale. Very few people know what the profile for a particular job is, and virtually no one can fake an entire profile. Third, although some people can, in principle, alter their scores, it is not clear that job applicants try to do this, or that it makes any difference. In perhaps the best single study of this topic, Smith and Ellington (2002) showed that response distortion has little impact on the construct validity of personality measures used in selection contexts. But finally, what does it mean to fake on a personality measure, or to fake in everyday life? Does it mean to act in a way that is different from the real you? Recall Freud’s view that you made up the you that you know, so that the real you is very hard, if not impossible, to define. Imagine that you are walking down a city street and suddenly feel an urge to poop. If you don’t poop on the sidewalk, then you are not being yourself. You are faking, and after that, it is a slippery slope—in everyday life it is impossible to say where faking ends and authenticity begins. The entire question of faking founders on the question of authenticity.

What is it that we know about scores on well-constructed measures of personality? One thing we know is that they are remarkably stable over time. Costa, Herbst, McCrae, and Siegler (2000) studied 40-year-olds over a 6- to 9-year interval and reported a median test–retest correlation of .83 over the five factors of the NEO Personality Inventory. Costa and McCrae (2002) provided a nice review of the stability issue. But what about validity?

Please consider Table 1. This table concerns the validity of some common medical measures and procedures, and provides an essential point of comparison regarding the size of correlations in other fields. Now consider Table 2. This contains

TABLE 1
Validities of Common Medical Procedures

<i>Procedure</i>	<i>r</i>	<i>N</i>
Coronary bypass surgery and 5 year survival	.08	2,649
Smoking and lung cancer within 25 years	.08	3,956
Antihistamines and reduced snot and sneezing	.11	1,023
Effect of ibuprofen on pain reduction	.14	8,488
Effects of Viagra on headaches and flushing	.25	861
Viagra and improved sexual functioning	.38	779
Height and weight of U.S. adults	.44	16,948

TABLE 2
Validity of Other Assessments for
Predicting Overall Job Performance

<i>Predictor</i>	<i>r</i>
Conscientiousness tests	.18
Integrity tests	.21
Structured interviews	.18
Unstructured interviews	.11
Situational judgment tests	.26
Biodata	.26
General mental ability	.21

Note. These correlations are uncorrected, observed score correlations.

sample weighted validity coefficients for the seven most common measures used to predict occupational performance. These correlations, which vary between .11 and .26, provide another frame of reference for evaluating the validity of personality measures. Consider next that Judge, Colbert, and Ilies (2004) reported a meta-analytic and fully corrected correlation of .27 between intelligence and leadership. These data suggest that a validity coefficient of .30 is unusual at any time for any measure.

My major claim is that there have always been good data supporting the validity of personality measures, but that the critics of personality measurement won't pay attention. Here are some data from the 1960s. Gough (1965) reported a point-biserial correlation of .73 between a delinquency–nondelinquency criterion and scores on the Socialization scale of the CPI in a sample of 10,296 people. In a related study, Gough reported a cross-validated, point-biserial correlation of .63 between a CPI regression equation (Socialization had the largest beta weight) and the delinquency criterion in an American sample of 2,981 cases, and of .60 in a Japanese sample of 149 cases (Gough, DeVos, & Muzushima, 1968). In the best study of creativity ever published, Hall and MacKinnon (1969) reported a cross-validated multiple *r* of .61 between a three (personality) variable regression equation and a solid criterion of real-world creativity. Also in the 1960s, I put together a large sample of college students from two campuses in the northeast and developed an index of marijuana use ranging from frequent use to principled nonuse. A CPI-based discriminant function correctly classified 81% of all the users and 81% of all the nonusers (cf. Hogan, Mankin, Conway, & Fox, 1970). These studies only bear indirectly on occupational performance, but they indicate the kind of correlations that one can expect from competent personality research.

More recently, Judge, Bono, Ilies, and Gerhardt (2002) conducted a careful meta-analytic study of the links between personality (defined in terms of the Five-factor Model) and leadership (defined in terms of emergence and effective-

ness). They reported the following estimated corrected correlations: (a) Neuroticism, $-.24$ ($N = 8,025$); (b) Extraversion, $.31$ ($N = 11,705$); (c) Openness, $.24$ ($N = 7,221$); (d) Agreeableness, $.08$ ($N = 9,801$); and (e) Conscientiousness, $.28$ ($N = 7,510$); with a multiple R , using the five dimensions of $.48$ (see Table 3).

In the best meta-analytic study of personality and job performance ever published, Hogan and Holland (2003) confined their investigation to one inventory (the HPI; Hogan & Hogan, 1995)—rather than trying to combine scales across inventories—and they carefully aligned predictors with the relevant criteria. That is, they did not try to predict training performance using the HPI Adjustment scale, or sales performance using the HPI Prudence scale. They reported the following estimated true validities: (a) Emotional Stability, $.43$ ($N = 2,573$); (b) Extraversion and Ambition, $.34$ ($N = 3,698$); (c) Agreeableness, $.36$ ($N = 2,500$); (d) Conscientiousness, $.36$ ($N = 3,379$); and (e) Intellect and Openness, $.34$ ($N = 1,190$; see Table 4).

TABLE 3
Meta-Analysis of the Relation Between
Big Five Personality Dimensions and Leadership

Dimension	Average			
	<i>k</i>	<i>N</i>	<i>r</i>	<i>p</i>
Neuroticism	48	8,025	-.17	-.24
Extraversion	60	11,705	.22	.31
Openness	37	7,221	.16	.24
Agreeableness	42	9,801	.06	.08
Conscientiousness	35	7,510	.20	.28

Note. *k* = number of correlations; *p* = corrected correlation.

TABLE 4
Summary of Hogan and Holland Results

HPI Scale	<i>k</i>	<i>N</i>	<i>r</i>	<i>p</i>
Adjustment	24	2,573	.25	.43
Ambition	28	3,698	.20	.34
Sociability	NA	NA	NA	NA
Likeability	17	2,500	.18	.36
Prudence	26	3,379	.22	.36
Intellectance	7	1,190	.20	.34
School success	9	1,366	.15	NA

Note. HPI = Hogan Personality Inventory.

That the article was published at all is a small miracle because the reviewers tried valiantly to suppress it.

PERSONALITY MEASUREMENT AND ASSESSMENT CENTERS

The first assessment centers in the United States contained measures of personality other than simulations, and were used to study the links between personality and occupational performance (cf. Bray, 1982; MacKinnon, 1978; Murray, 1938; OSS Assessment Staff, 1948). Standard practice has been to include the personality data, along with the data from the simulations, to create a final score which represents a composite of the judgment of the assessment center raters. Assessment center ratings are valid predictors of occupational performance (Gaugler, Rosenthal, Thornton, & Bentson, 1987). By definition, the personality measurement data will be correlated with the overall assessment center score. The question becomes, to what extent do personality measures predict overall ratings from assessment centers?

Collins et al. (2003) found 524 articles containing correlations between measures of personality and cognitive ability and performance in assessment centers. From this set, they retained those studies in which the personality data could be coded in terms of the Five-factor Model; this resulted in a database of 65 correlations in a total sample size of 9,738. After doing all the relevant corrections, they reported the following average construct level correlations with Overall Assessment Center Ratings: (a) Cognitive Ability, .67; (b) Extraversion, .50; (c) Emotional Stability, .35; (d) Openness, .25; and (e) Agreeableness, .17. This leads to a multiple R of .84, suggesting that most of the valid variance in Overall Assessment Center Ratings can be captured with good measures of cognitive ability and normal personality.

What can we conclude regarding the use of personality measures to predict occupational performance? We can conclude four things. First, when the research is done correctly, the correlations between the standard dimensions of normal personality and job performance criteria that are relevant to that dimension are reliably above .30, and multiple correlations approach .50. Second, personality predicts occupational performance almost as well as measures of cognitive ability. Third, unlike cognitive ability measures, personality measures do not discriminate: Blacks get the same scores as Whites, women get the same scores as men, and those over 40 get somewhat higher scores than those under 40 (reflecting their greater maturity). Finally, we have a solid conceptual account for these findings. Well-constructed measures of personality are designed to predict reputation; performance appraisal at work is all about evaluating reputation. The bottom line is, personality measures work pretty well, especially when compared with all the other measures. When listening to the critics, I find myself asking, "If you don't like personality measures, what are the alternatives?"

REFERENCES

- Anderson, P. W. (1972). More is different. *Science*, *177*, 393–396.
- Bray, D. W. (1982). The assessment center and the study of lives. *American Psychologist*, *37*, 180–189.
- Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment*, *11*, 17–29.
- Costa, P. T., Jr., Herbst, J. H., McCrae, R. R., & Siegler, I. C. (2000). Personality at midlife. *Assessment*, *7*, 365–378.
- Costa, P. T., Jr., & McCrae, R. R. (2002). Looking backward: Changes in the mean levels of personality traits from 80 to 12. In D. Cervone & W. Mischel (Eds.), *Advances in personality science* (pp. 219–237). New York: Guilford.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, *72*, 493–511.
- Gough, H. G. (1965). Conceptual analysis of psychological test scores and other diagnostic variables. *Journal of Abnormal Psychology*, *70*, 294–302.
- Gough, H. G., DeVos, G., & Muzushima, K. (1968). Japanese validation of the CPI social maturity index. *Psychological Reports*, *22*, 143–146.
- Hall, W. B., & MacKinnon, D. W. (1969). Personality inventories as predictors of creativity among architects. *Journal of Applied Psychology*, *53*, 322–326.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*, 43–61.
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job performance relations. *Journal of Applied Psychology*, *88*, 100–112.
- Hogan, R., & Hogan, J. (1995). *Hogan personality inventory manual*. Tulsa, OK: Hogan Assessment Systems.
- Hogan, R., Mankin, D., Conway, J., & Fox, S. (1970). Personality correlates of undergraduate marijuana use. *Journal of Consulting and Clinical Psychology*, *35*, 58–63.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*, 765–780.
- Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, *89*, 542–555.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*, 280–296.
- MacKinnon, D. W. (1978). *In search of human effectiveness*. Buffalo, NY: Creative Education Foundation.
- Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for research and practice in human resource management. *Research in Personnel and Human Resource Management*, *13*, 153–200.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- OSS Assessment Staff. (1948). *Assessment of men*. New York: Rinehart.
- Paul, A. M. (2004). *The cult of personality*. New York: Free Press.
- Schmitt, N. (2004). Beyond the Big Five: Increases in understanding and practical utility. *Human Performance*, *17*, 347–357.
- Smith, D. B., & Ellington, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, *87*, 211–219.