11.2: Personality Assessment

By David Watson

University of Notre Dame

This module provides a basic overview to the assessment of personality. It discusses objective personality tests (based on both self-report and informant ratings), projective and implicit tests, and behavioral/performance measures. It describes the basic features of each method, as well as reviewing the strengths, weaknesses, and overall validity of each approach.

learning objectives

• Appreciate the diversity of methods that are used to measure personality characteristics.
• Understand the logic, strengths and weaknesses of each approach.
• Gain a better sense of the overall validity and range of applications of personality tests.

Introduction

Personality is the field within psychology that studies the thoughts, feelings, behaviors, goals, and interests of normal individuals. It therefore covers a very wide range of important psychological characteristics. Moreover, different theoretical models have generated very different strategies for measuring these characteristics. For example, humanistically oriented models argue that people have clear, well-defined goals and are actively striving to achieve them (McGregor, McAdams, & Little, 2006). It, therefore, makes sense to ask them directly about themselves and their goals. In contrast, psychodynamically oriented theories propose that people lack insight into their feelings and motives, such that their behavior is influenced by processes that operate outside of their awareness (e.g., McClelland, Koestner, &
Weinberger, 1989; Meyer & Kurtz, 2006). Given that people are unaware of these processes, it does not make sense to ask directly about them. One, therefore, needs to adopt an entirely different approach to identify these nonconscious factors. Not surprisingly, researchers have adopted a wide range of approaches to measure important personality characteristics. The most widely used strategies will be summarized in the following sections.

Do people possess the necessary awareness to see themselves as they are and provide accurate insights into their own personalities? [Image: fotEK10, https://goo.gl/GCBDJL, CC BY-NC-SA 2.0, goo.gl/Toc0ZF]

### Objective Tests

#### Definition

Objective tests (Loevinger, 1957; Meyer & Kurtz, 2006) represent the most familiar and widely used approach to assessing personality. Objective tests involve administering a standard set of items, each of which is answered using a limited set of response options (e.g., true or false; strongly disagree, slightly disagree, slightly agree, strongly agree). Responses to these items then are scored in a standardized, predetermined way. For example, self-ratings on items assessing talkativeness, assertiveness, sociability, adventurousness, and energy can be summed up to create an overall score on the personality trait of extraversion.

It must be emphasized that the term “objective” refers to the method that is used to score a person’s responses, rather than to the responses themselves. As noted by Meyer and Kurtz (2006, p. 233), “What is objective about such a procedure is that the psychologist administering the test does not need to rely on judgment to classify or interpret the test-taker’s response; the intended response is clearly indicated and scored according to a pre-existing key.” In fact, as we will see, a person’s test responses may be highly subjective and can be influenced by a number of different rating biases.
Basic Types of Objective Tests

Self-report measures

Objective personality tests can be further subdivided into two basic types. The first type—which easily is the most widely used in modern personality research—asks people to describe themselves. This approach offers two key advantages. First, self-raters have access to an unparalleled wealth of information: After all, who knows more about you than you yourself? In particular, self-raters have direct access to their own thoughts, feelings, and motives, which may not be readily available to others (Oh, Wang, & Mount, 2011; Watson, Hubbard, & Weise, 2000). Second, asking people to describe themselves is the simplest, easiest, and most cost-effective approach to assessing personality. Countless studies, for instance, have involved administering self-report measures to college students, who are provided some relatively simple incentive (e.g., extra course credit) to participate.

The items included in self-report measures may consist of single words (e.g., assertive), short phrases (e.g., am full of energy), or complete sentences (e.g., I like to spend time with others). Table 1 presents a sample self-report measure assessing the general traits comprising the influential five-factor model (FFM) of personality: neuroticism, extraversion, openness, agreeableness, and conscientiousness (John & Srivastava, 1999; McCrae, Costa, & Martin, 2005). The sentences shown in Table 1 are modified versions of items included in the International Personality Item Pool (IPIP) (Goldberg et al., 2006), which is a rich source of personality-related content in the public domain (for more information about IPIP, go to: http://ipip.ori.org/).

![Sample self-report measure assessing the general traits comprising the influential five-factor model (FFM) of personality: neuroticism, extraversion, openness, agreeableness, and conscientiousness.](https://socialsci.libretexts.org/Bookshelves/Psychology/Map%3A_Discover_Psychology_-_A_Brief_Introductory_Text_(Noba)/...)
Table1: Sample Self-Report Personality Measure

Self-report personality tests show impressive validity in relation to a wide range of important outcomes. For example, self-ratings of conscientiousness are significant predictors of both overall academic performance (e.g., cumulative grade point average; Poropat, 2009) and job performance (Oh, Wang, and Mount, 2011). Roberts, Kuncel, Shiner, Caspi, and Goldberg (2007) reported that self-rated personality predicted occupational attainment, divorce, and mortality. Similarly, Friedman, Kern, and Reynolds (2010) showed that personality ratings collected early in life were related to happiness/well-being, physical health, and mortality risk assessed several decades later. Finally, self-reported personality has important and pervasive links to psychopathology. Most notably, self-ratings of neuroticism are associated with a wide array of clinical syndromes, including anxiety disorders, depressive disorders, substance use disorders, somatoform disorders, eating disorders, personality and conduct disorders, and schizophrenia/schizotypy (Kotov, Gamez, Schmidt, & Watson, 2010; Mineka, Watson, & Clark, 1998).

At the same time, however, it is clear that this method is limited in a number of ways. First, raters may be motivated to present themselves in an overly favorable, socially desirable way (Paunonen & LeBel, 2012). This is a particular concern in "high-stakes testing," that is, situations in which test scores are used to make important decisions about individuals (e.g., when applying for a job). Second, personality ratings reflect a self-enhancement bias (Vazire & Carlson, 2011); in other words, people are motivated to ignore (or at least downplay) some of their less desirable characteristics and to focus instead on their more positive attributes. Third, self-ratings are subject to the reference group effect (Heine, Buchtel, & Norenzayan, 2008); that is, we base our self-perceptions, in part, on how we compare to others in our sociocultural reference group. For instance, if you tend to work harder than most of your friends, you will see yourself as someone who is relatively conscientious, even if you are not particularly conscientious in any absolute sense.

Informant ratings

Another approach is to ask someone who knows a person well to describe his or her personality characteristics. In the case of children or adolescents, the informant is most likely to be a parent or teacher. In studies of older participants, informants may be friends, roommates, dating partners, spouses, children, or bosses (Oh et al., 2011; Vazire & Carlson, 2011; Watson et al., 2000).

Generally speaking, informant ratings are similar in format to self-ratings. As was the case with self-report, items may consist of single words, short phrases, or complete sentences. Indeed, many popular instruments include parallel self- and informant-rating versions, and it often is relatively easy to convert a self-report measure so that it can be used to obtain informant ratings. Table 2 illustrates how the self-report instrument shown in Table 1 can be converted to obtain spouse-ratings (in this case, having a husband describe the personality characteristics of his wife).
Table 2: Sample Spouse-Report Personality Measure

Informant ratings are particularly valuable when self-ratings are impossible to collect (e.g., when studying young children or cognitively impaired adults) or when their validity is suspect (e.g., as noted earlier, people may not be entirely honest in high-stakes testing situations). They also may be combined with self-ratings of the same characteristics to produce more reliable and valid measures of these attributes (McCrae, 1994).

Informant ratings offer several advantages in comparison to other approaches to assessing personality. A well-acquainted informant presumably has had the opportunity to observe large samples of behavior in the person he or she is rating. Moreover, these judgments presumably are not subject to the types of defensiveness that potentially can distort self-ratings (Vazire & Carlson, 2011). Indeed, informants typically have strong incentives for being accurate in their judgments. As Funder and Dobroth (1987, p. 409), put it, “Evaluations of the people in our social environment are central to our decisions about who to befriend and avoid, trust and distrust, hire and fire, and so on.”

Informant personality ratings have demonstrated a level of validity in relation to important life outcomes that is comparable to that discussed earlier for self-ratings. Indeed, they outperform self-ratings in certain circumstances, particularly when the assessed traits are highly evaluative in nature (e.g., intelligence, charm, creativity; see Vazire & Carlson, 2011). For example, Oh et al. (2011) found that informant ratings were more strongly related to job performance than were self-ratings. Similarly, Oltmanns and Turkheimer (2009) summarized evidence indicating that informant ratings of Air Force cadets predicted early, involuntary discharge from the military better than self-ratings.

Nevertheless, informant ratings also are subject to certain problems and limitations. One general issue is the level of relevant information that is available to the rater (Funder, 2012). For instance, even under the best of circumstances, informants lack full access to the thoughts, feelings, and motives of the person they are rating. This problem is magnified when the informant does not know the person particularly well and/or only sees him or her in a limited range of situations (Funder, 2012; Beer & Watson, 2010).
Informant personality ratings are generally a reliable and valid assessment instrument, however in certain cases the informant may have some significant biases that make the rating less reliable. Newly married individuals for example are likely to rate their partners in an unrealistically positive way. [Image: Sociales El Heraldo de Saltillo, https://goo.gl/3g3Qhh, CC BY-NC-SA 2.0, goo.gl/Toc0ZF]

Informant ratings also are subject to some of the same response biases noted earlier for self-ratings. For instance, they are not immune to the reference group effect. Indeed, it is well-established that parent ratings often are subject to a sibling contrast effect, such that parents exaggerate the true magnitude of differences between their children (Pinto, Rijsdijk, Frazier-Wood, Asherson, & Kuntsi, 2012). Furthermore, in many studies, individuals are allowed to nominate (or even recruit) the informants who will rate them. Because of this, it most often is the case that informants (who, as noted earlier, may be friends, relatives, or romantic partners) like the people they are rating. This, in turn, means that informants may produce overly favorable personality ratings. Indeed, their ratings actually can be more favorable than the corresponding self-ratings (Watson & Humrichouse, 2006). This tendency for informants to produce unrealistically positive ratings has been termed the letter of recommendation effect (Leising, Erbs, & Fritz, 2010) and the honeymoon effect when applied to newlyweds (Watson & Humrichouse, 2006).

Other Ways of Classifying Objective Tests

Comprehensiveness

In addition to the source of the scores, there are at least two other important dimensions on which personality tests differ. The first such dimension concerns the extent to which an instrument seeks to assess personality in a reasonably comprehensive manner. At one extreme, many widely used measures are designed to assess a single core attribute. Examples of these types of measures include the Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994), the Rosenberg Self-Esteem Scale (Rosenberg, 1965), and the Multidimensional Experiential Avoidance Questionnaire (Gamez, Chmielewski, Kotov, Ruggero, & Watson, 2011). At the other extreme, a number of omnibus inventories contain a large number of specific scales and purport to measure personality in a reasonably comprehensive manner.
These instruments include the California Psychological Inventory (Gough, 1987), the Revised HEXACO Personality Inventory (HEXACO-PI-R) (Lee & Ashton, 2006), the Multidimensional Personality Questionnaire (Patrick, Curtin, & Tellegen, 2002), the NEO Personality Inventory-3 (NEO-PI-3) (McCrae et al., 2005), the Personality Research Form (Jackson, 1984), and the Sixteen Personality Factor Questionnaire (Cattell, Eber, & Tatsuoka, 1980).

**Breadth of the target characteristics**

Second, personality characteristics can be classified at different levels of breadth or generality. For example, many models emphasize broad, “big” traits such as neuroticism and extraversion. These general dimensions can be divided up into several distinct yet empirically correlated component traits. For example, the broad dimension of extraversion contains such specific component traits as dominance (extraverts are assertive, persuasive, and exhibitionistic), sociability (extraverts seek out and enjoy the company of others), positive emotionality (extraverts are active, energetic, cheerful, and enthusiastic), and adventurousness (extraverts enjoy intense, exciting experiences).

Some popular personality instruments are designed to assess only the broad, general traits. For example, similar to the sample instrument displayed in Table 1, the **Big Five** Inventory (John & Srivastava, 1999) contains brief scales assessing the broad traits of neuroticism, extraversion, openness, agreeableness, and conscientiousness. In contrast, many instruments—including several of the omnibus inventories mentioned earlier—were designed primarily to assess a large number of more specific characteristics. Finally, some inventories—including the HEXACO-PI-R and the NEO-PI-3—were explicitly designed to provide coverage of both general and specific trait characteristics. For instance, the NEO-PI-3 contains six specific facet scales (e.g., Gregariousness, Assertiveness, Positive Emotions, Excitement Seeking) that then can be combined to assess the broad trait of extraversion.

---

**Projective and Implicit Tests**

**Projective Tests**

Projective tests, such as the famous Rorschach inkblot test require a person to give spontaneous answers that "project" their unique personality onto an ambiguous stimulus. [Image: CC0 Public Domain, goo.gl/m25gce]

As noted earlier, some approaches to personality assessment are based on the belief that important thoughts, feelings,
and motives operate outside of conscious awareness. Projective tests represent influential early examples of this approach. Projective tests originally were based on the projective hypothesis (Frank, 1939; Lilienfeld, Wood, & Garb, 2000): If a person is asked to describe or interpret ambiguous stimuli—that is, things that can be understood in a number of different ways—their responses will be influenced by nonconscious needs, feelings, and experiences (note, however, that the theoretical rationale underlying these measures has evolved over time) (see, for example, Spangler, 1992). Two prominent examples of projective tests are the Rorschach Inkblot Test (Rorschach, 1921) and the Thematic Apperception Test (TAT) (Morgan & Murray, 1935). The former asks respondents to interpret symmetrical blots of ink, whereas the latter asks them to generate stories about a series of pictures.

For instance, one TAT picture depicts an elderly woman with her back turned to a young man; the latter looks downward with a somewhat perplexed expression. Another picture displays a man clutched from behind by three mysterious hands. What stories could you generate in response to these pictures?

In comparison to objective tests, projective tests tend to be somewhat cumbersome and labor intensive to administer. The biggest challenge, however, has been to develop a reliable and valid scheme to score the extensive set of responses generated by each respondent. The most widely used Rorschach scoring scheme is the Comprehensive System developed by Exner (2003). The most influential TAT scoring system was developed by McClelland, Atkinson and colleagues between 1947 and 1953 (McClelland et al., 1989; see also Winter, 1998), which can be used to assess motives such as the need for achievement.

The validity of the Rorschach has been a matter of considerable controversy (Lilienfeld et al., 2000; Mihura, Meyer, Dumitrascu, & Bombel, 2012; Society for Personality Assessment, 2005). Most reviews acknowledge that Rorschach scores do show some ability to predict important outcomes. Its critics, however, argue that it fails to provide important incremental information beyond other, more easily acquired information, such as that obtained from standard self-report measures (Lilienfeld et al., 2000).

Validity evidence is more impressive for the TAT. In particular, reviews have concluded that TAT-based measures of the need for achievement (a) show significant validity to predict important criteria and (b) provide important information beyond that obtained from objective measures of this motive (McClelland et al., 1989; Spangler, 1992). Furthermore, given the relatively weak associations between objective and projective measures of motives, McClelland et al. (1989) argue that they tap somewhat different processes, with the latter assessing implicit motives (Schultheiss, 2008).

Implicit Tests

In recent years, researchers have begun to use implicit measures of personality (Back, Schmuckle, & Egloff, 2009; Vazire & Carlson, 2011). These tests are based on the assumption that people form automatic or implicit associations between certain concepts based on their previous experience and behavior. If two concepts (e.g., meand assertive) are strongly associated with each other, then they should be sorted together more quickly and easily than two concepts (e.g., me and shy) that are less strongly associated. Although validity evidence for these measures still is relatively sparse, the results to date are encouraging: Back et al. (2009), for example, showed that implicit measures of the FFM personality traits predicted behavior even after controlling for scores on objective measures of these same characteristics.
Behavioral and Performance Measures

Observing real world behavior is one way to assess personality. Tendencies such as messiness and neatness are clues to personality. [Image: Crumley Roberts, https://goo.gl/6Ahn8q, CC BY 2.0, goo.gl/BRvSA7]

A final approach is to infer important personality characteristics from direct samples of behavior. For example, Funder and Colvin (1988) brought opposite-sex pairs of participants into the laboratory and had them engage in a five-minute “getting acquainted” conversation; raters watched videotapes of these interactions and then scored the participants on various personality characteristics. Mehl, Gosling, and Pennebaker (2006) used the electronically activated recorder (EAR) to obtain samples of ambient sounds in participants’ natural environments over a period of two days; EAR-based scores then were related to self- and observer-rated measures of personality. For instance, more frequent talking over this two-day period was significantly related to both self- and observer-ratings of extraversion. As a final example, Gosling, Ko, Mannarelli, and Morris (2002) sent observers into college students’ bedrooms and then had them rate the students’ personality characteristics on the Big Five traits. The averaged observer ratings correlated significantly with participants’ self-ratings on all five traits. Follow-up analyses indicated that conscientious students had neater rooms, whereas those who were high in openness to experience had a wider variety of books and magazines.

Behavioral measures offer several advantages over other approaches to assessing personality. First, because behavior is sampled directly, this approach is not subject to the types of response biases (e.g., self-enhancement bias, reference group effect) that can distort scores on objective tests. Second, as is illustrated by the Mehl et al. (2006) and Gosling et al. (2002) studies, this approach allows people to be studied in their daily lives and in their natural environments, thereby avoiding the artificiality of other methods (Mehl et al., 2006). Finally, this is the only approach that actually assesses what people do, as opposed to what they think or feel (see Baumeister, Vohs, & Funder, 2007).

At the same time, however, this approach also has some disadvantages. This assessment strategy clearly is much more cumbersome and labor intensive than using objective tests, particularly self-report. Moreover, similar to projective tests, behavioral measures generate a rich set of data that then need to be scored in a reliable and valid way. Finally, even the
most ambitious study only obtains relatively small samples of behavior that may provide a somewhat distorted view of a person’s true characteristics. For example, your behavior during a “getting acquainted” conversation on a single given day inevitably will reflect a number of transient influences (e.g., level of stress, quality of sleep the previous night) that are idiosyncratic to that day.

**Conclusion**

No single method of assessing personality is perfect or infallible; each of the major methods has both strengths and limitations. By using a diversity of approaches, researchers can overcome the limitations of any single method and develop a more complete and integrative view of personality.

**Discussion Questions**

1. Under what conditions would you expect self-ratings to be most similar to informant ratings? When would you expect these two sets of ratings to be most different from each other?
2. The findings of Gosling, et al. (2002) demonstrate that we can obtain important clues about students’ personalities from their dorm rooms. What other aspects of people’s lives might give us important information about their personalities?
3. Suppose that you were planning to conduct a study examining the personality trait of honesty. What method or methods might you use to measure it?

**Vocabulary**

**Big Five**

Five, broad general traits that are included in many prominent models of personality. The five traits are neuroticism (those high on this trait are prone to feeling sad, worried, anxious, and dissatisfied with themselves), extraversion (high scorers are friendly, assertive, outgoing, cheerful, and energetic), openness to experience (those high on this trait are tolerant, intellectually curious, imaginative, and artistic), agreeableness (high scorers are polite, considerate, cooperative, honest, and trusting), and conscientiousness (those high on this trait are responsible, cautious, organized, disciplined, and achievement-oriented).

**High-stakes testing**

Settings in which test scores are used to make important decisions about individuals. For example, test scores may be used to determine which individuals are admitted into a college or graduate school, or who should be hired for a job. Tests also are used in forensic settings to help determine whether a person is competent to stand trial or fits the legal definition of sanity.

**Honeymoon effect**

The tendency for newly married individuals to rate their spouses in an unrealistically positive manner. This represents a specific manifestation of the letter of recommendation effect when applied to ratings made by current romantic partners. Moreover, it illustrates the very important role played by relationship satisfaction in ratings made by romantic partners: As marital satisfaction declines (i.e., when the “honeymoon is over”), this effect disappears.

**Implicit motives**

These are goals that are important to a person, but that he/she cannot consciously express. Because the individual cannot verbalize these goals directly, they cannot be easily assessed via self-report. However, they can be measured using projective devices such as the Thematic Apperception Test (TAT).
Letter of recommendation effect
The general tendency for informants in personality studies to rate others in an unrealistically positive manner. This tendency is due a pervasive bias in personality assessment: In the large majority of published studies, informants are individuals who like the person they are rating (e.g., they often are friends or family members) and, therefore, are motivated to depict them in a socially desirable way. The term reflects a similar tendency for academic letters of recommendation to be overly positive and to present the referent in an unrealistically desirable manner.

Projective hypothesis
The theory that when people are confronted with ambiguous stimuli (that is, stimuli that can be interpreted in more than one way), their responses will be influenced by their unconscious thoughts, needs, wishes, and impulses. This, in turn, is based on the Freudian notion of projection, which is the idea that people attribute their own undesirable/unacceptable characteristics to other people or objects.

Reference group effect
The tendency of people to base their self-concept on comparisons with others. For example, if your friends tend to be very smart and successful, you may come to see yourself as less intelligent and successful than you actually are. Informants also are prone to these types of effects. For instance, the sibling contrast effect refers to the tendency of parents to exaggerate the true extent of differences between their children.

Reliability
The consistency of test scores across repeated assessments. For example, test-retest reliability examines the extent to which scores change over time.

Self-enhancement bias
The tendency for people to see and/or present themselves in an overly favorable way. This tendency can take two basic forms: defensiveness (when individuals actually believe they are better than they really are) and impression management (when people intentionally distort their responses to try to convince others that they are better than they really are). Informants also can show enhancement biases. The general form of this bias has been called the letter-of-recommendation effect, which is the tendency of informants who like the person they are rating (e.g., friends, relatives, romantic partners) to describe them in an overly favorable way. In the case of newlyweds, this tendency has been termed the honeymoon effect.

Sibling contrast effect
The tendency of parents to use their perceptions of all of their children as a frame of reference for rating the characteristics of each of them. For example, suppose that a mother has three children; two of these children are very sociable and outgoing, whereas the third is relatively average in sociability. Because of operation of this effect, the mother will rate this third child as less sociable and outgoing than he/she actually is. More generally, this effect causes parents to exaggerate the true extent of differences between their children. This effect represents a specific manifestation of the more general reference group effect when applied to ratings made by parents.

Validity
Evidence related to the interpretation and use of test scores. A particularly important type of evidence is criterion validity, which involves the ability of a test to predict theoretically relevant outcomes. For example, a presumed measure of conscientiousness should be related to academic achievement (such as overall grade point average).

References


