INDUSTRIAL PSYCHOLOGY

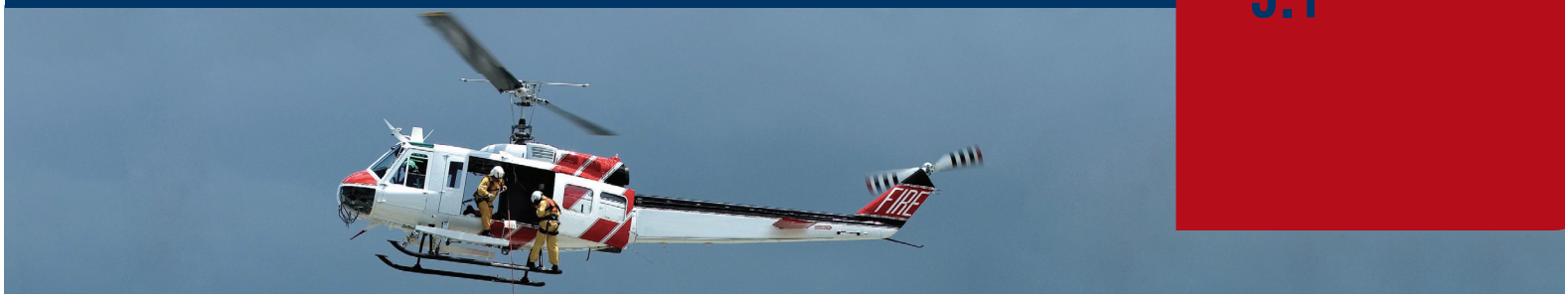
Part 2

3

INDIVIDUAL DIFFERENCES AND ASSESSMENT

An Introduction to Individual Differences

MODULE 3.1



What do Britney Spears, the Pope, Yo-Yo Ma, Stephen King, Carmello Anthony, George W. Bush, your grandmother, and your instructor have in common? Not much. They are different in abilities, interests, experiences, personality, age, gender, race, and backgrounds. Indeed, the only thing we can say with certainty about these individuals is that they are substantially different from one another. We would not expect your grandmother to try out for an NBA team, or Stephen King to officiate at a religious service, or your instructor to meet with heads of state of foreign countries. Many psychologists, including I-O psychologists, believe that the differences among individuals can be used, at least in part, to understand and predict their behavior.

But it isn't good enough to say simply that people are different. You don't have to be a psychologist to recognize that. Some types of differences prove more useful than others in predicting and understanding behavior. The differences among people on various attributes like intelligence, personality, and knowledge are important in understanding a wide variety of socially important outcomes (Lubinski, 2000), including:

- academic achievement;
- intellectual development;
- crime and delinquency;
- vocational choice;
- income and poverty;
- occupational performance.

This chapter will deal first with the concept of **individual differences**, and then with how the assessment of these differences can help to predict occupational performance.

INDIVIDUAL DIFFERENCES
Dissimilarities between or among two or more people.

SOME BACKGROUND

Psychology began in a laboratory in Germany in 1876. The father of the discipline, Wilhelm Wundt, was anxious to show that psychology was different from philosophy and medicine. Since this was a new science and the existing physical sciences like chemistry, biology, and physics had discovered many general principles that enhanced their importance, Wundt set out to uncover general principles of human behavior as well. He developed techniques for studying the sensations and reactions of people, examining the dimmest light that individuals could see, the faintest sound they could hear, and how quickly they could react to a signal. But those who assisted in conducting his experiments quickly discovered that not everyone had the same reaction time, or could see the same dim light, or hear the same soft tone. In other words, they discovered that there were *differences* among individuals.

These differences detracted from the precise results Wundt sought, but to one of his students they represented a fascinating discovery. James McKeen Cattell (1860–1944), an American who received a PhD in psychology under Wundt's direction, soon began measuring and charting the differences among people using "psychological" variables. In 1890 Cattell developed the concept of a **mental test** as a way of charting these differences. Since the subject matter of this research was differences, the study of differences became known as **differential psychology** (Landy, 1997).

MENTAL TEST
Instrument designed to measure a subject's ability to reason, plan, and solve problems; an intelligence test.

DIFFERENTIAL PSYCHOLOGY
Scientific study of differences between or among two or more people.

INTELLIGENCE
The ability to learn and adapt to an environment; often used to refer to general intellectual capacity, as opposed to cognitive ability or mental ability, which often refer to more specific abilities such as memory or reasoning.

MENTAL ABILITY
Capacity to reason, plan, and solve problems; cognitive ability.

METRIC
Standard of measurement; a scale.

PSYCHOMETRICS
Practice of measuring a characteristic such as mental ability, placing it on a scale or metric.

INTELLIGENCE TEST
Instrument designed to measure the ability to reason, learn, and solve problems.

After leaving Wundt's laboratory at the University of Leipzig, Cattell went to England and worked with another researcher very interested in individual differences, Francis Galton. Galton was gathering information that would support his cousin Charles Darwin's radical theory of evolution. In earlier years, Galton had measured inherited characteristics like height, weight, reach, and hair color. With his new mental test, Cattell was able to expand the number of inherited characteristics that he could examine. After working with Galton for several years in developing a comprehensive mental test, Cattell returned to America and used this test to measure the **intelligence** of incoming college students. He believed that he could use the resulting scores to help students choose curricula and to predict who would successfully complete college. Cattell had developed methods of measuring **mental ability**, placing it on a scale or **metric**. As a result, the actual measurement of abilities became known as **psychometrics**.

While other early psychologists began to focus on *pathological* aspects of mental function, the pioneers of differential psychology were primarily interested in the mental abilities of "normal" people. Several were aware of Cattell's work in measuring intelligence. In France, Alfred Binet was measuring mental abilities of French school children. Lewis Terman was conducting similar studies in California with a translation of Binet's test. Hugo Munsterberg was measuring the abilities of trolley drivers in order to predict the likelihood of accidents. When the United States entered the First World War in 1917, the leading industrial psychologists of the time persuaded the Army to use an **intelligence test** to screen recruits and determine who should attend officer's candidate school. Two years after the war's end, Walter Dill Scott, one of the founding fathers of I-O psychology, proclaimed that "possibly the single greatest achievement of the American Psychological Association is the establishment of individual differences" (Lubinski, 2000).

In the postwar years, intelligence tests were adapted for use in selecting individuals for jobs with government and industry. By 1932 measuring the differences in intelligence among individuals in order to predict things like accidents and productivity was a well-established practice (Landy, 1997; Viteles, 1932). As we will see later in the chapter, intelligence is still one of the most generally assessed characteristics of job applicants.

DIFFERENTIAL PSYCHOLOGY, PSYCHOMETRICS, AND I-O PSYCHOLOGY

Nearly a century later, measuring the differences among individuals to predict later behavior ("psychometrics") remains one of the most common frameworks applied by I-O psychologists. It is different from the framework used by an experimental psychologist. The experimental psychologist usually designs an experiment that will show how all people are alike in their response to a stimulus, and looks outside the individual to the stimulus as a way to explain behavior. In contrast, the differential psychologist is person-centered, looking for qualities or characteristics within the person that will help us understand that person's behavior. In the past, I-O psychology—particularly the applied aspect of it—depended on these differences to predict things like job success, job satisfaction, and counterproductive work behavior. I-O psychology still makes great use of the individual differences approach, but as we will see later in this chapter and in succeeding chapters, there is more to behavior than simply individual differences.

The marriage of psychometrics and differential psychology was a good one. The differential psychologist identified what should be measured, and the **psychometrician** set about measuring it. As we saw from the work of Cattell and his contemporaries, the attribute most commonly measured was some form of intelligence. It was widely believed that **cognitive ability** was the single most important attribute that an individual possessed. We use cognitive abilities to acquire knowledge, solve problems, and apply reason to situations. Consequently, many studies were conducted to show that an individual's general intellectual capacity was closely associated with that individual's occupational and vocational success. The pioneers in theories of intelligence referred to this attribute as "**g**," an abbreviation for **general mental ability** (Hull, 1928; Spearman, 1927). Today's psychologists still use that term, and we will use it in this book.

PSYCHOMETRICIAN
Psychologist trained in measuring characteristics such as mental ability.

COGNITIVE ABILITY
Capacity to reason, plan, and solve problems; mental ability.

"g"
Abbreviation for general mental ability.

GENERAL MENTAL ABILITY
The nonspecific capacity to reason, learn, and solve problems in any of a wide variety of ways and circumstances.

IDENTIFYING INDIVIDUAL DIFFERENCES

As we saw in the earlier section describing the history of individual differences, Francis Galton was one of the early advocates of studying such differences. In 1890, Galton wrote that "One of the most important objects of measurement is . . . to obtain a general knowledge . . . of capacities . . . by sinking shafts at a few critical points" (Lubinski, 2000). By this, Galton meant that we can use psychometric tests to explore individual abilities and other attributes the way miners use drilling to explore minerals in the earth. That is an excellent way to think of what we are doing when we study individual differences: We are sinking shafts to obtain more general knowledge about behavior at work. The concept of sinking shafts also provides a good framework for explaining how I-O psychologists explore individual differences today as opposed to 25 years ago. In the past, we concentrated on only one shaft—intelligence. Today we are sinking many more shafts, as well as deeper ones (e.g., specific aspects of cognitive ability; specific aspects of personality). Before, we were content to stop at a more superficial level ("g"). Today our explorations are broader and deeper, and we can reach more meaningful conclusions because the reliability and validity of our measuring devices are better.

We need to keep in mind that not all individual differences will tell us something important. As in drilling for oil, water, or gold, we don't always "strike it rich." This is one of the reasons we do research: to see which shafts provide encouragement. To continue with the drilling metaphor, we can distinguish among the differential psychologist, the psychometrician, and the applied I-O psychologist. The differential psychologist examines the psychological landscape and identifies some attractive areas for drilling. The psychometrician actually sinks the shaft. The applied I-O psychologist uses what comes out of that shaft, but instead of oil, water, or gold, what come out are valuable predictors of performance. Later in this chapter, we will examine the methods by which these areas can be explored: the actual assessment methods for examining these individual differences.

However, you must continually remind yourself (and we will help remind you) that behavior is complex and people are whole. No single area of individual difference (e.g., intelligence) is likely to completely (or even substantially) explain any important aspect of work behavior. You cannot separate an individual's intelligence from his or her personality, knowledge, or experience. When you look at the behavior of any individual, you need to remember that he or she is a whole, intact entity. To acknowledge a person's individuality, we need to go beyond considering just one or another possible predictor of his or her behavior.

VARIETIES OF INDIVIDUAL DIFFERENCES

G-OCENTRIC MODEL
Tendency to understand and predict the behavior of workers simply by examining "g".

In the past 15 years, there has been a substantial shift in thinking about individual differences. Instead of simply examining "g" to understand and predict the behavior of workers—a tendency that Sternberg and Wagner (1993) called the **g-ocentric model**—

researchers are moving toward broadening the field of examination. In addition to cognitive ability, I-O psychologists now consider individual differences in **physical abilities**, **personality**, **interests**, **knowledge**, and **emotion** in examining the behavior of people in work settings. This is the result of several forces. In the early years of testing, the only available tests were intelligence tests. Since that time, psychologists have developed many reliable methods for measuring personality, knowledge, interests, and emotional reactions to work. In addition, our understanding of the many facets of performance has become more sophisticated. Instead of simply assessing overall performance, like an overall GPA, we now consider specific facets of performance such as organizational citizenship, counterproductive work behavior, and task performance, topics we will address in Chapters 4 and 5. Murphy (1996) proposes that there are many different attributes of people that serve many different demands of the job (see Figure 3.1).

PHYSICAL ABILITIES
Bodily powers such as muscular strength, flexibility, and stamina.

PERSONALITY
An individual's behavioral and emotional characteristics, generally found to be stable over time and in a variety of circumstances; an individual's habitual way of responding.

INTERESTS
Preferences or likings for broad ranges of activities.

KNOWLEDGE
A collection of specific and interrelated facts and information about a particular topical area.

EMOTION
An affect or feeling, often experienced and displayed in reaction to an event or thought and accompanied by physiological changes in various systems of the body.

Let's apply that view to a particular job. Some of the most important things that firefighters do are driving the fire truck to the fire, applying water to the fire, providing medical assistance, rescuing trapped citizens, and learning new procedures and how to use new equipment. To accomplish these tasks, firefighters work in teams. To provide medical assistance and learn new procedures, the firefighter needs cognitive ability. To rescue trapped citizens and apply water to the fire, the firefighter needs both physical ability and courage in addition to problem-solving skills. To accomplish teamwork with fellow firefighters and to deal with victims, the firefighter needs communication skills. To drive the truck to the fire accurately and safely, the firefighter needs good vision, coordination, and the knowledge or memory of how to get to the location of the fire. If we only bothered to examine the differences among individuals in cognitive ability (or "g"), we would only be able to predict and understand a limited portion of the firefighter's job performance. To understand the full

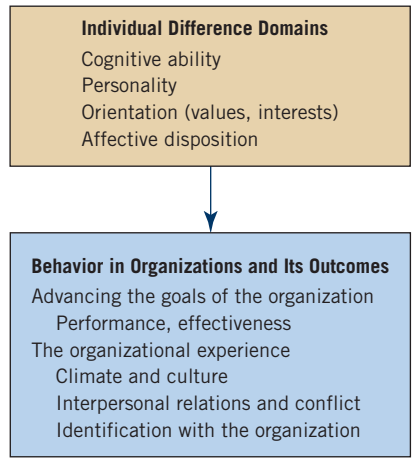


FIGURE 3.1 The Link between Attributes and Behavior in Organizations
SOURCE: K. R. Murphy (1996).

BOX 3.1 EARLY INDIVIDUAL DIFFERENCE CHARACTERISTICS

James McKeen Cattell began testing incoming students, first at the University of Pennsylvania in 1892, then at Columbia in 1900. He wanted to identify the characteristics of "individual differences" of the students so that he could eventually predict which applicants for college admission were likely to get a degree. The following is a list of some of information Cattell gathered on each student:

- Memory
- Reasoning
- Numerical skills
- Reaction time
- Hair color
- Weight
- Height
- Right or left handedness

Questions:

1. Which of the characteristics in the list above are not a part of one of the categories of individual differences in this module?
2. Which of the characteristics in the list would be unlikely to be related to college success?
3. Which characteristics in the list do you think are still routinely gathered in the college admissions process?

range of performance, we need to consider attributes beyond "g."

There is a growing consensus (Guion, 1998; Murphy, 1996) that we can divide the individual differences useful in understanding work behavior into certain categories, including:

- cognitive ability;
- physical ability;
- personality;
- interests.

The performance of most jobs requires multiple abilities. What are some of the abilities called for in the job of firefighter?

In the next section, we will consider these broad categories of attributes, as well as the theories that further define them. Before we do so, we need to consider the fundamental assumptions that I-O psychologists make when they apply the individual differences model. They are listed as follows (adapted from Guion, 1998).

1. Adults have a variety of attributes (e.g., intelligence, personality, interests) and the levels of these attributes are relatively stable over a reasonable time period (several years).
2. People differ with respect to these attributes (i.e., there are "individual differences") and these differences are associated with job success.
3. The relative differences between people on these attributes remain even after training, job experience, or some other intervention. Thus, if individual A has less of an attribute than individual B before training or job experience, and if they both receive the same training or experience to increase that attribute, individual A will still have less of that attribute than individual B after the training or intervention, even though *both* may have higher levels of the attribute after training or experience.
4. Different jobs require different attributes.
5. These attributes can be measured.

With these assumptions in mind, we can now examine these attribute categories in the next modules.

MODULE 3.1 SUMMARY

- The *individual differences* among people on various attributes like intelligence, personality, and knowledge are important in understanding a wide variety of socially important outcomes.
- James McKeen Cattell developed the concept of a *mental test* as a way of charting the differences among people. Since the subject matter of this research was differences, the study of differences became known as *differential psychology*. The actual measurement of abilities became known as *psychometrics*.
- The differential psychologist is person-centered, looking for qualities or characteristics within the person that will help us understand that person's behavior. The differential psychologist identifies what should be measured, and the *psychometrician* sets about measuring it.
- The attribute most commonly measured by early differential psychologists was some form of *intelligence*. It was widely believed that *cognitive ability* was the single most important attribute

possessed by an individual. The pioneers in theories of intelligence referred to this attribute as "*g*," an abbreviation for *general mental ability*.

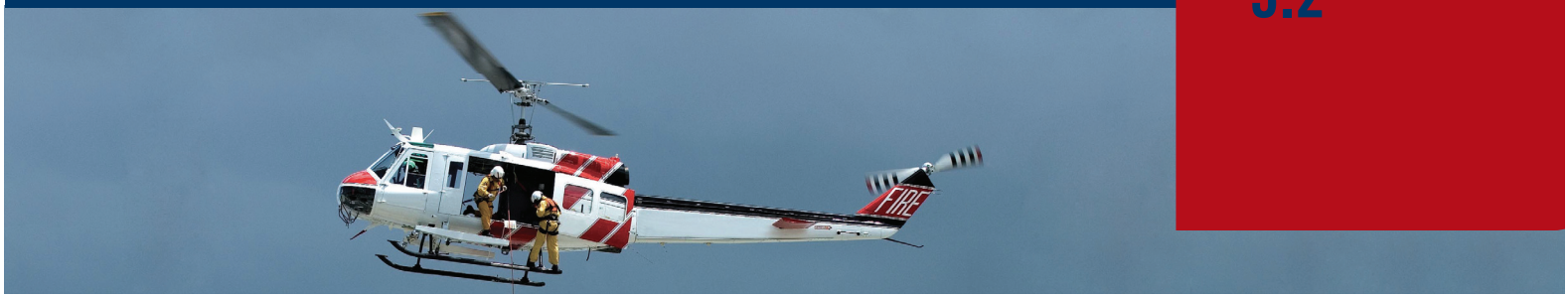
- In addition to cognitive ability, I-O psychologists consider individual differences in *physical abilities*, *personality*, *interests*, *knowledge*, and *emotion* in examining the behavior of people in work settings.

KEY TERMS

individual differences
mental test
differential psychology
intelligence
mental ability
metric
psychometrics
intelligence test
psychometrician
cognitive ability
"g"
general mental ability
g-ocentric model
physical abilities
personality
interests
knowledge
emotion

Human Attributes

MODULE 3.2



ABILITIES

In the 1950s, Edwin Fleishman began a program of research to determine the most common mental and physical abilities associated with human performance, including work performance. Through a combination of field and laboratory research, he and his associates developed a comprehensive list, or **taxonomy**, of 52 abilities (Fleishman & Reilly, 1992). These can be divided into the broad categories of cognitive, physical, and **perceptual-motor abilities**. This taxonomy is more detailed than we need to deal with at this point, so we have presented the full taxonomy on the text website for the interested student. The abilities Fleishman identified cover an impressive variety—and they do not cover personality, **affect**, or interest! Fleishman's work expanded the study of individual differences far beyond his predecessors' focus on differences in intelligence.

Fleishman's list of abilities can be used for many different applied purposes. It is an effective way to analyze the most important abilities in various occupations (Landy, 1989). It can also be used to determine training needs, recruiting needs, and even work design. Once we know the basic abilities that can be brought to the job, it is much easier to identify which of those abilities are truly important. In the next chapter, you will see how the Fleishman ability list has been used to form a comprehensive expert computer system called O*NET that connects human abilities with job demands. O*NET will be described in detail in Chapter 4.

TAXONOMY

An orderly, scientific system of classification.

PERCEPTUAL-MOTOR ABILITIES

Physical attributes that combine the senses (e.g. seeing, hearing, smell) and motion (e.g. coordination, dexterity).

AFFECT

The conscious, subjective aspect of emotion.

COGNITIVE ABILITIES

Intelligence as "g"

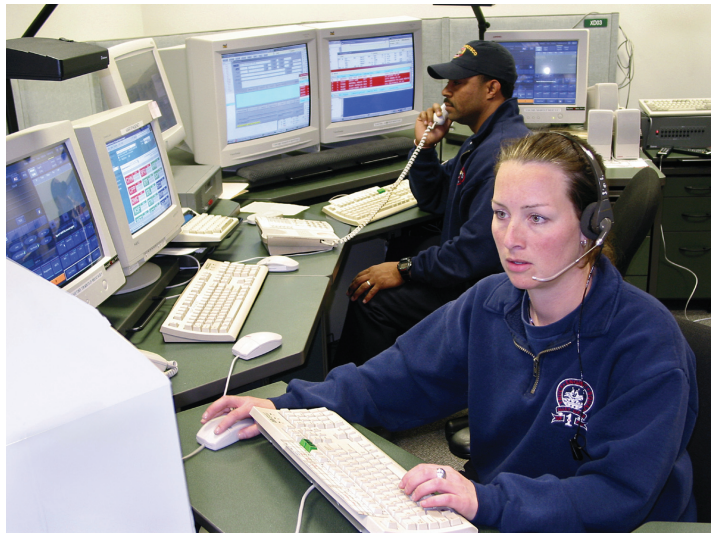
Many people consider the terms intelligence, IQ, cognitive ability, and mental ability to be synonyms for one another. We will make some distinctions. **IQ** is a historical term that stood for **Intelligence Quotient** and refers to the way early intelligence test scores were calculated. The term no longer has scientific meaning, although it is still often used by the general public. Mental ability and cognitive ability are current terms, which scientists often use interchangeably. Cognitive ability and mental ability often refer to specific abilities such as memory or reasoning; intelligence, on the other hand, most often refers to general intellectual capacity (often called "g" for general mental ability). Intelligence can be defined as the ability to learn and adapt to an environment. One or another variation of this definition has been used since at least 1921 (Sternberg & Kaufmann, 1998). A group of leading I-O psychologists defined it as follows: "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly,

IQ

Abbreviation for intelligence quotient.

INTELLIGENCE QUOTIENT

Measure of intelligence obtained by giving a subject a standardized "IQ" test. The score is obtained by multiplying by 100 the ratio of the subject's mental age to chronological age.



Critical abilities for the job of emergency dispatcher include verbal comprehension, reaction time, and problem solving.

comprehend complex ideas, learn quickly, and learn from experience" (Arvey et al., 1995).

Sternberg and Kaufmann (1998) pointed out that no matter how enduring this definition may be for Western cultures, other cultures have different views of who is "an intelligent person." Speed of learning, for example, is not always emphasized in non-Western cultures. In fact, "other cultures may be suspicious of work done quickly" (Sternberg & Kaufmann, 1998), and in some cultures, the word intelligence means "prudence" and "caution." Nevertheless, for our purposes, we will accept the meaning generally assigned by Western psychologists. Intelligence is required whenever people must manipulate information of any type (Murphy, 1996). Measures of "g" assess reasoning ability, knowledge acquisition, and problem-solving ability (Lubinski, 2004).

Is "g" Important at Work?

META-ANALYSIS

Statistical method for combining and analyzing the results from many studies to draw a general conclusion about relationships among variables.

Yes. Almost every job requires some active manipulation of information. This means that your level of general mental ability can affect your performance on any job. The greater the amount of information that needs to be manipulated, the more important "g" becomes. **Meta-analyses** of the relationship between "g" and job performance (Schmidt & Hunter, 2004) have demonstrated very clearly that as the complexity of the job increased, the predictive value (i.e., validity) of tests of general intelligence also increased. In practical terms, this means that if the information-processing demands of a job are high, a person with lower general mental ability is not as likely to be successful as a person of higher general mental ability. That does not mean, however, that high general mental ability guarantees success on that job. If the job also requires interpersonal skills, communication skills, and certain personality traits, even a person with high general mental ability (but lower levels of those noncognitive traits) might fail.

In 1965 Tanner showed that he could accurately predict which Olympic athletes were competing in which sports by looking at their body builds. But *within* each Olympic event, the same individual differences were useless as predictors of who would get a medal (Lubinski, 2000). In this example, think of body build as "g," and all the other attributes of the athletes as specific abilities and attributes; "g" may help a candidate get into the police academy, but it will not ensure that the person will become a successful police officer.

Some, but far from all, of today's psychologists continue to believe that nothing more than measures of "g" are needed to predict training, grades, and job performance. An excellent review of the debate can be seen in an entire issue of the journal *Human Performance* devoted to the issue (Ones & Viswesvaran, 2002). One psychologist framed the issue as follows.

> General mental ability (g) is a substantively significant determinant of individual differences for any job that includes information-processing tasks. . . . The exact size of the relationship will be a function of . . . the degree to which the job requires information processing and verbal cognitive skills. (Campbell, 1990a)

From Campbell's statement we can infer that because "g" represents information-processing ability, then it should logically predict information-processing performance in the workplace. In addition, we can infer that jobs differ in terms of not only how much "information processing" they require, but also how quickly that processing must be completed. A backhoe operator certainly has to process some information, but not as much as a software help-desk operator. The backhoe operator will depend much more heavily on visual/spatial ability than on problem solving or reasoning ability.

Is "g" as Important in Other Countries as It Is in the U.S.?

The simple answer seems to be "yes," at least as far as Europe is concerned. Several meta-analyses have recently been published demonstrating the predictive value of "g" in the European Union (Salgado & Anderson, 2003; Salgado, Anderson, Moscoso, Bertua, & DeFruyt, 2003; Salgado, Anderson, Moscoso, Bertua, DeFruyt, & Rolland, 2003) and specifically in the U.K. (Bertua, Anderson, & Salgado, in press). Salgado and Anderson (2002) also report that the use of tests of general mental ability and other cognitive ability tests is even more prevalent in the E.U. than in the U.S. Much less is known about non-European countries, but it is a fair assumption that as long as globalization is controlled by western nations, "g" will remain important in non-European countries as well. If China emerges as a dominant global player, as it appears will happen, we may see a different and possibly diminished role for "g."

Can Your Level of "g" Change?

Today's researchers observe a fascinating phenomenon: Intelligence continues to rise over time. Individuals appear to be getting smarter and smarter through the lifespan, and new generations appear to be smarter than their parents. The phenomenon is labeled the **Flynn effect** after a political scientist who has done extensive research on the topic (Flynn, 1999). It amounts to a gain of 15 points in average intelligence test scores per generation. This is a substantial increase, considering that the **mean** intelligence on most tests is pegged at 100 with a **standard deviation** of 15. Many psychologists have proposed theories as to why this is occurring, including better health care, better nutrition, increased schooling, and better-educated parents (Sternberg & Kaufmann, 1998). It could also be because we live in an increasingly complex environment both at work and at home (Neisser et al., 1996). This phenomenon is not peculiar to the U.S. Similar effects are found in many different countries (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003). So, at least at a generational level, the answer seems to be that on the average, your generation will be "smarter" than your parents' generation. Within generations, however, it appears that there is a greater stability to "g" (Lubinski, Benbow, Webb, & Bleske-Recheck, in press; Wai, Lubinski, & Benbow, in press) and possibly that, as we get older, the intelligence level we possessed at earlier points in our lives becomes even more prominent and stable (Bouchard, 2004; Plomin & Spinath, 2004). So there is good news and bad news. The good news is that your generation is likely smarter than your parents' generation and that your level of "g" will possibly increase as a result of the complexity of your environment at home and at work. The bad news is that as you age, the amount of change will get smaller – so start working on it now!

FLYNN EFFECT

Phenomenon in which new generations appear to be smarter than their parents by a gain of 15 points in average intelligence test score per generation; named after the political scientist who did extensive research on the topic.

MEAN

The arithmetic average of the scores in a distribution; obtained by summing all of the scores in a distribution and dividing by the sample size.

STANDARD DEVIATION

Measure of the extent of spread in a set of scores.

Cognitive Abilities beyond "g"

The majority of today's psychologists agree that although "g" is important, more specific cognitive abilities also play a role in performance, with some specific abilities important

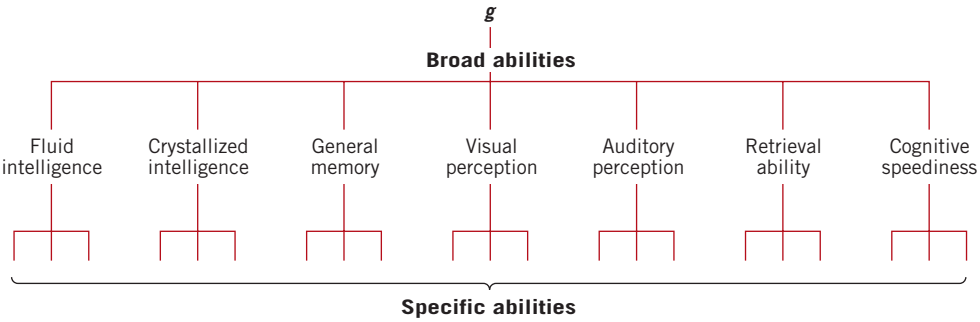


FIGURE 3.2 Carroll's Hierarchical Model
SOURCE: Carroll (1993).

for some jobs and other specific abilities important for other jobs. The example of the backhoe operator and the software help-desk operator points out the importance of specific cognitive abilities.

The question then becomes: How many specific abilities are there? There is no conclusive answer to that question, but we can say with great confidence that there is more than one (i.e., more than just "g"). As we mentioned earlier, Fleishman and his colleagues posited 52 abilities, 21 of which are in the cognitive category, but "g" is not one of them. The reason for this is that Fleishman was more concerned with identifying *specific* abilities than general mental ability. It is now generally accepted that cognitive ability is best conceptualized as having multiple layers of abilities. The specific cognitive abilities identified by Fleishman are available at www.blackwellpublishing.com/landyconte2e.

Carroll (1993) proposed that there are three layers, or strata, to intelligence (see Figure 3.2). The highest layer is "g"; the next layer down consists of seven more specific abilities: fluid intelligence, crystallized intelligence, memory, visual perception, auditory perception, information retrieval, and cognitive speed (Murphy, 1996). The lowest and most specific level includes abilities that are tied to the seven broad abilities in the middle level. For example, information ordering (one of Fleishman's proposed abilities) would be connected to fluid intelligence, and spatial relations would be associated with visual perception.

There are many other theories of cognitive abilities (e.g., Ackerman, Beier, & Boyle, 2002, 2005), but all resemble Carroll's. The important thing to remember is that "g" will only get you so far in understanding work behavior. Different jobs will require additional specific cognitive abilities as well. It is fair to say that a person with a high level of "g" will probably be a successful performer at certain tasks of almost every job (Schmidt & Hunter, 1998), but that other abilities will vary in importance depending on the job in question—the more complex the job, the more important the role of "g." As we will see in the subsequent sections of this module, not only will cognitive abilities play a role in job success and satisfaction, but so also will personality, emotional reactions, and interests.

You might wonder if it is possible to have *too much* intelligence. This reminds us of a story that is often told about a well-known boxer. A radio interviewer was talking with a retired middleweight boxer who had fought for many years and had a relatively undistinguished career, finishing with approximately 60 wins and 30 losses. The interview went something like this, with the interviewer represented by "I" and the boxer by "B."

I: You must have fought many interesting boxers in your career.

B: Yeah, there were plenty of them.

I: I noticed that you fought so-and-so four times and beat him all four times.

B: Yeah, that surprised me because he had a lot better record than me.

I: Why did it surprise you?

B: Because he was so smart. He was always thinking ahead, what combination he would set up, where he wanted to be in the ring, and things like that. He was really smart, always thinking.

I: Then let me ask the obvious question—why do you think you beat him so consistently?

B: I guess it was because when he was thinking, I was punching.

So it does appear that, occasionally, too much "g" can get you hurt!

PHYSICAL, SENSORY, AND PSYCHOMOTOR ABILITIES

Physical Abilities

Hogan (1991a, b) suggested that seven physical abilities are sufficient for analyzing most jobs. Guion compared Hogan's seven abilities with similar abilities identified by Fleishman and Reilly (1992) and found a close match. As you can see in Figure 3.3, several of Hogan's dimensions are combinations of the Fleishman and Reilly (1992) dimensions (e.g., she combines extent flexibility and dynamic flexibility into a single dimension called "flexibility"). In a manner reminiscent of Carroll's theory of intelligence, Hogan then combines her seven measures to form three higher-order physical abilities: muscular strength, cardiovascular endurance, and movement quality. For most jobs, this three-ability taxonomy would likely be sufficient because most physically demanding jobs require **muscular tension**, **muscular power**, and **muscular endurance**, not just one of the three. Similarly, flexibility and balance usually go together in a physically demanding job.

Fairness of Physical Ability Tests Because employers often use physical ability tests to screen applicants for physically demanding jobs, it is important to determine whether such tests are fair to female applicants and older applicants. Because we lose

STAMINA

Physical ability to supply muscles with oxygenated blood through the cardiovascular system; also known as cardiovascular strength or aerobic strength or endurance.

MUSCULAR TENSION

Physical quality of muscular strength.

MUSCULAR POWER

Physical ability to lift, pull, push, or otherwise move an object; unlike endurance, this is a one-time maximum effort.

MUSCULAR ENDURANCE

Physical ability to continue to use a single muscle or muscle group repeatedly over a period of time.

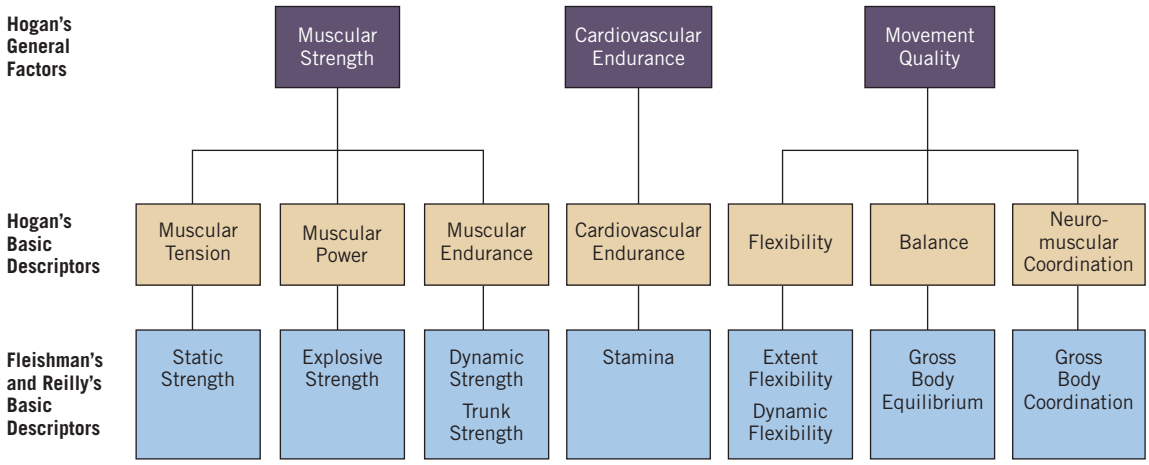


FIGURE 3.3 A Model of Physical Abilities
SOURCE: Guion (1998).

muscle, stamina, and flexibility as we age, the older an applicant is, the less well he or she is likely to perform on physical ability tests. For women the situation has an additional consideration. On average, females have less muscle mass (which means diminished muscular strength) and lower levels of cardiovascular endurance (or stamina) than men (Hogan, 1991a). In contrast, on measures of flexibility (e.g., sit and reach tests) women tend to do better than men. However, most physically demanding jobs require—or are perceived by employers to require—more muscular strength and stamina than flexibility. This has meant that male candidates, who tend to excel on those physical tests, are predominantly hired for such jobs. As a result, women candidates for popular positions such as firefighter have filed employment discrimination suits (*Brunet* v. *City of Columbus*, 1995).

You have probably observed that women and men of all ages can increase their individual physical abilities with exercise and training. In addition, it is clear that many jobs require a fixed level of strength and endurance and that more is not always better. If your job requires you to lift 25-pound boxes from a conveyor belt and place them on a table, the fact that you are strong enough to move 100-pound boxes is irrelevant to the task at hand. In this case, more strength would not lead to higher performance. This means that it is not always necessary for individuals to compete *against* each other on physical ability tests; they merely need to demonstrate sufficient strength and endurance to perform the tasks that comprise the job. By training for several months prior to the administration of physical ability tests, women candidates are able to improve their performance significantly. Thus, one way of helping women to do better on these tests is for employers to encourage them to train ahead of time (McArdle, Katch, & Katch, 2001). We can predict that this same strategy may help older job seekers as well.

In recent years, employers have become more concerned about the cost of medical and disability programs for workers. In essence, the employer would like to predict who is likely to experience an injury and reject those applicants. Physical ability tests have been used as the predictor for future injury. The problem is that while they may be good (but far from perfect) predictors of future injury, such tests may not be particularly relevant for present or future job performance. In a recent case brought against Armour Star Meat Packing facility in Iowa, 52 women successfully sued the company for denying them jobs based on a strength test. A federal judge ruled against the employer and awarded $3.3 million to the women because the test was used to predict injuries, not performance on the job (Business and Legal Reports, 2005a).

Sensory Abilities

SENSORY ABILITIES

Physical functions of vision, hearing, touch, taste, smell, and kinesthetic feedback (e.g. noticing changes in body position).

AMERICANS WITH DISABILITIES ACT

Federal legislation in 1990 requiring employers to give applicants and employees with disabilities the same consideration as other applicants and employees, and to make certain adaptations in the work environment to accommodate disabilities.

Sensory abilities are the physical functions of vision, hearing, touch, taste, smell, and kinesthetic feedback (e.g., noticing changes in body position). Hogan includes kinesthetic feedback in a dimension she calls "movement quality." The sensory abilities of vision and hearing are particularly interesting for applied I-O psychologists because employers often test these abilities in would-be employees.

To prevent employers from using a disability as an excuse to reject an applicant who is capable of performing a job, the **Americans with Disabilities Act** of 1990 forbids them from asking about or testing areas such as sensory or physical abilities that may be considered "disabilities" until after they have made a job offer to the candidate.

Until recently, cognitive psychologists considered sensory abilities to be independent of cognitive abilities, but Carroll's (1993) model of intelligence calls that assumption into question—remember that two of his mid-level abilities are visual perception and auditory perception. In addition, Ackerman's research (e.g., Ackerman et al., 2002, 2005) shows the close association between perceptual speed and other measures of cognitive ability. But in most real-life settings, sensation and perception are inextricably bound together. We usually infer from some kind of report (verbal or behavioral) that a person has sensed some-

thing. Further research will shed light on the extent to which "non-cognitive" abilities are really "non-cognitive."

Psychomotor Abilities

Psychomotor abilities, sometimes called **sensorimotor**, or just **motor abilities**, deal with issues of coordination, dexterity, and reaction time. Once again, Fleishman (Fleishman & Reilly, 1992) has done the most extensive work in identifying these abilities (see Table 3.1). A simple inspection of these abilities immediately brings to mind the jobs for which they may be important (e.g., crane operators, organists, watch repair technicians, surgeons, wait staff, and bartenders). Once again, from this discussion it should be clear that many of these abilities (e.g., rate control and aiming) may very well be associated with visual and/or auditory perception or cognitive speed, facets of Carroll's theory of intelligence.

The work of researchers like Carroll and Ackerman blurs the classical distinctions between cognitive and "noncognitive" abilities. In some senses, this is a good development. Until recently, psychologists tended to treat abilities in isolation when it is clear in real life (and—more importantly for us—in work) that all of these abilities interact within a single person to produce a response or action.

PSYCHOMOTOR ABILITIES

Physical functions of movement, associated with coordination, dexterity, and reaction time; also called motor or sensorimotor abilities.

SENSORIMOTOR ABILITIES

Physical functions of movement, associated with coordination, dexterity, and reaction time; also called psychomotor or motor abilities.

MOTOR ABILITIES

Physical functions of movement, associated with coordination, dexterity, and reaction time; also called psychomotor or sensorimotor abilities.

PERSONALITY AND INTERESTS

PERSONALITY

There is now a broad consensus that personality represents an important area of individual differences for examination by I-O psychologists (Barrick & Ryan, 2003; Schneider & Smith, 2004). There are clear connections between aspects of personality and various work behaviors, both productive (e.g., job performance) and counterproductive (e.g., dishonesty, absenteeism). This consensus is the result of concentrated work on developing a taxonomy of personality factors. This taxonomy is labeled the

An officer directing traffic at a congested intersection is likely to experience information overload.

TABLE 3.1 Psychomotor Abilities

1. Arm-hand steadiness	6. Response orientation
2. Manual dexterity	7. Rate control
3. Finger dexterity	8. Reaction time
4. Control precision	9. Wrist-finger speed
5. Multilimb coordination	10. Speed of limb movement

SOURCE: Adapted from Fleishman et al. (1999); Fleishman & Reilly (1992).

CASE STUDY 3.1 A Level Playing Field

It is common to test for physical abilities before choosing candidates for recruit positions in fire academies. Although physical abilities will be improved in the 16 weeks of the academy training program, you still require a minimum amount of ability to profit from the training. Most fire departments administer physical ability tests that simulate actual tasks performed by firefighters. As examples, candidates may be asked to carry heavy hose bundles up stairs, open fire hydrants with wrenches, or hang heavy exhaust fans in windows. Two tests, in particular, seem to be harder for female applicants than their male counterparts. The first is the "dummy drag" simulation. In this test, the candidate is asked to drag a 150-pound dummy through a 40-foot maze with several left and right turns in it. The second task is pulling 50 feet of a simulated fire hose through a 50-foot maze with two right turns. Since men tend to be larger and stronger, they simply pick up the dummy and carry it through the maze, while women are more likely to drag the dummy along the floor of the maze. Similarly, for the hose pull, men tend to simply loop the hose over their shoulder and pull it through the maze in one single movement. The test is not exactly the same as the actual task, however; in an actual fire situation the firefighter is usually pulling a person or a hose through a burning room and must stay close to the ground because the toxic fumes, smoke, and temperature (often as high as 2000 degrees) are more deadly in the upper part of a room.

If you wanted to make these test components more realistic, how would you redesign the test course? If you did redesign it, do you think that the performance of women would improve? Why or why not?

BIG 5

A taxonomy of five personality factors; the Five Factor model (FFM).

FIVE FACTOR MODEL (FFM)

A taxonomy of five personality factors, comprised of conscientiousness, extraversion, agreeableness, emotional stability, and openness to experience.

Big 5 or the **Five Factor Model** (**FFM**) (Digman, 1990; McCrae & Costa, 1985, 1987). This model is the result of both statistical analyses of personality test information gathered over many decades and a careful conceptual analysis of what most personality tests were trying to assess. Like most innovations, it has its critics, but for our purposes in this book it is a good basic model for describing the potential importance of personality variables in understanding job performance. Further, some recent research has demonstrated that personality measurement based on the FFM produces higher validity coefficients for predicting work performance than non-FFM based measurement devices (Salgado, 2003).

The Five Factor Model

As suggested by its title, the Five Factor Model (FFM) proposes that we can describe someone's "personality" by looking at five relatively independent factors. Personality can be defined in simplest terms as the typical way that an individual has of responding. It is considered a trait because it is fairly stable, even though situations and circumstances might lead a person to behave in a way that is out of character with his or her overall personality. The FFM identifies five different components that, when taken together, give a fair representation of how a person typically responds to events and people. These components and their definitions are presented in Table 3.2. Considerable evidence has been gathered showing that although the five factors might express themselves in slightly different ways in various cultures, the FFM seems to be applicable cross-culturally (Cheung, 2004; McCrae, Terracciano et al., 2005; Tsaousis & Nikolaou, 2001) and that culture and personality may be linked (Hofstede & McCrae, 2004).

CONSCIENTIOUSNESS

Quality of having positive intentions and carrying them out with care.

It is important to keep in mind that the five factors are intended to measure normal personality, not to identify any evidence of psychopathology. We will make that distinction clearer later in this chapter when we discuss how personality is measured. Of the five factors, the first to have attracted most attention from I-O psychologists was **conscientiousness**. More recently, extraversion, openness to experience, and agreeableness are also

TABLE 3.2 The Five Factor Model

FACTOR	CHARACTERISTICS
1. Conscientiousness	Responsible, prudent, self-control, persistent, planful, achievement oriented
2. Extraversion	Sociable, assertive, talkative, ambitious, energetic
3. Agreeableness	Good natured, cooperative, trusting, likeable, friendly
4. Emotional stability	Secure, calm, low anxiety, low emotionality
5. Openness to experience	Curious, intelligent, imaginative, independent

SOURCE: Based on Digman (1990); McCrae & Costa (1985, 1987).

attracting increased attention. In some early research, Barrick and Mount (1991) proposed, on the basis of a meta-analysis, that in all likelihood conscientiousness was positively related to success in all aspects of work for all occupations. That was a strong statement, but it was supported by their analyses. Naturally, there were disagreements with the five factor taxonomy and with the presumed overarching importance of conscientiousness. The first disagreement was that five factors are too few to capture the full range of aspects of personality (Hough, 1992; Tellegen, 1993; Tellegen, Grove, & Waller, 1991; Tellegen & Waller, 2000). The second criticism was that although conscientiousness might be correlated with a wide range of work behaviors, it was not *highly* correlated with them. In addition, extraversion often correlated as highly with behavior as did conscientiousness. A third criticism was that there were combinations of the five factors that led to greater predictive power than any one of the factors by itself (Dunn, 1993; Hogan & Hogan, 1989; Ones, Viswesvaran, & Schmidt, 1993). The first and third criticisms present an interesting dilemma, since one argues for *more* factors, whereas the other seems to be arguing for *fewer* factors.

What seems to be true is that although each of the five factors does predict successful in contrast to unsuccessful performance of certain behaviors, some combinations of the factors may be stronger predictors than any single factor by itself. This introduces the idea of a **functional personality at work** (Barrick, Mount, & Judge, 2001). This means that not just one factor predicts success, but a combination of factors. For example, Ones et al. (1993) found that individuals who were high on the conscientiousness, **agreeableness**, and **emotional stability** factors of the FFM tended to have higher **integrity**. Integrity in this context means honest, reliable, and ethical. Dunn (1993) found that managers believed that a combination of conscientiousness, agreeableness, and emotional stability made applicants more attractive to managers who had hiring responsibilities. Hogan and Hogan (1989) found that the same factors were related to employee reliability. In a review of meta-analyses, Barrick, Mount and Judge (2001) confirm the importance of conscientiousness across a variety of occupations and performance measures. Emotional stability also appeared to predict overall performance across occupations. But extraversion, openness to experience, and agreeableness appeared to be more predictive of success in specific occupations and for specific criteria.

FUNCTIONAL PERSONALITY AT WORK

The way that an individual behaves, handles emotions, and accomplishes tasks in a work setting; a combination of Big Five factors.

AGREEABLENESS

Likable, easy to get along with, friendly.

EMOTIONAL STABILITY

Displaying little emotion; showing the same emotional response in various situations.

INTEGRITY

Quality of being honest, reliable, and ethical.

Other meta-analyses also reveal relationships between the FFM and job performance, both in the United States (Hurtz & Donovan, 2000) and with European data (Salgado, 1997, 1998). The latter series of meta-analyses suggest that, at least for many European countries, culture may not be a moderator variable for the personality/performance

relationship. Nevertheless, remember from Chapter 1 that Hofstede's (1980a, 2001) model of cultural influence showed that the biggest cultural differences seemed to be between Asian and western nations, so the jury is still out on whether the personality/performance relationship holds true in countries like China and Japan. As suggested by the work of McCrae, Terracciano, and colleagues (2005), there is reason to expect that it will be different from its manifestation in Europe or the United States, because the collectivist cultures of China and Japan emphasize group outcomes over individual outcomes.

Implications of the Five Factor Model It appears that as the aspect of work behavior we are trying to predict gets broader (e.g., overall job performance), large FFM factors like conscientiousness do as well as smaller and more discrete factors. There is some debate about whether or not to use broad or narrow personality dimensions (Hogan & Roberts, 1996; Ones & Viswesvaran, 1996; Schneider, Hough, & Dunnette, 1996). It turns out that narrow traits seem to be useful for predicting very specific job behaviors and broader traits for predicting broader behaviors, so it is not necessary to choose between the two approaches (Tett, Steele, & Beauregard, 2003). Each has its own use. In fact, there appears to be a movement to keep the FFM from suffocating research on other possible personality characteristics (Borman, 2004a). Schmitt (2004) has suggested three promising personality characteristics that deserve attention:

1. Core Self-evaluation (Judge & Bono, 2001), a type of inner-directedness and sense of efficacy
2. Tolerance for Contradiction (Chan, 2004), a way of thinking that prefers apparently contradictory information
3. Achievement Motivation and Aggression (James, 1998, 2004), the tendency to aggressively seek out desired ends.

We are not necessarily suggesting that these are "variables to watch"—we are simply echoing the sentiment of Schmitt that it is too early to close the door on variables other than those identified in the FFM.

ACHIEVEMENT

A facet of conscientiousness consisting of hard work, persistence, and the desire to do good work.

DEPENDABILITY

A facet of conscientiousness, consisting of being disciplined, well organized, respectful of laws and regulations, honest, trustworthy, and accepting of authority.

Hough suggested that the FFM factor of conscientiousness should be broken down into two discrete factors called **achievement** and **dependability**. Achievement consists of hard work, persistence, and the desire to do good work. Dependability represents being disciplined, well organized, respectful of laws and regulations, honest, trustworthy, and accepting of authority (Hough, 1992). When we break conscientiousness down into those two facets, it turns out that dependability is a better predictor of employee reliability than conscientiousness, and achievement is a better predictor of effort than conscientiousness. But if we try to predict ratings of overall job performance, then conscientiousness does as well as either achievement or dependability (Mount & Barrick, 1995). Another general finding is that as the behavior we are trying to predict (e.g., effort or reliability) becomes more specific, the correlations with both the FFM factors and the more refined factors go up. The more specific we are about the aspect of performance we are trying to predict, the more accurate the prediction is. A recent and more detailed study of the construct of conscientiousness has suggested as many as six facets (Roberts, Chernyshenko, Stark, & Goldberg, 2005). This study showed that this more refined breakdown of conscientiousness led to higher predictability of behaviors as diverse as work dedication, drug use, and positive health-related behaviors.

Tett (1995) made the point with a few concrete examples. He suggested that the "dependable" (or "rule-bound") aspect of conscientiousness might actually be counterproductive in professions such as musician, sculptor, painter, actor, choreographer, and even management positions in which the manager is expected to "think outside the box" (e.g., marketing manager). He referred to the problem of too much attention to detail and rules as "analysis-paralysis."

Problems can also arise when an individual has plenty of "g" but lacks other attributes. The movie *A Beautiful Mind* tells the story of John Nash, a Nobel Prize-winning mathematician who was brilliant (in a "g" sense") and high on achievement as defined above, but who was severely impaired in social and interpersonal skills. Nash was emotionally unstable, eventually disintegrating into paranoia and delusional states associated with schizophrenia. Early in his career, he was favored because of his sheer brilliance ("g") and tenacity (conscientiousness), but his impairments in other dimensions eventually rendered him useless to the research facility where he was employed. Nash's story illustrates the point that to get a true understanding of behavior, we often need to decompose elements (like conscientiousness) or consider patterns or combinations of elements.

There is a final aspect of the research on the Five Factor Model that deserves discussion. Have you ever had a job in which you were closely supervised and required to follow very detailed work and organizational procedures? In that environment, you would have had little opportunity to let your "habitual way of responding" (i.e., your personality) appear in your behavior. Think of the opposite situation—a job where you had a good deal of control over your work habits. In the latter, you could really be "you," and whether you performed well or poorly probably depended on how well your personality was suited to the job's demands. That is exactly what Barrick and Mount (1993) found with their research on the FFM. In jobs where the employee had a great deal of control (i.e., autonomy), personality was much more predictive of performance than in jobs where the employee had little or no control. In this case, control moderated the relationship between personality and performance. In statistical terms, control would be called a "moderator variable"; that is, a variable that changes the nature of the relationship between two other variables. It has been commonly found that if a situation allows for little discretion on the part of a person (referred to as a "strong" situation), personality will play a minor role in his or her behavior.

How can we summarize what we know about the relationship between personality and work behavior? And what can we say more specifically about the FFM compared to other theories, such as Hough's (1992)? We believe the following conclusions can be drawn with confidence.

1. Personality differences play a role in work behavior independent of the role played by cognitive ability (Mount & Barrick, 1995; Murphy, 1996).
2. Personality is more closely related to motivational aspects of work (e.g., effort expenditure) than to technical aspects of work (e.g., knowledge components). Personality is more likely to predict what a person *will* do and ability measures are more likely to predict what a person *can* do (Campbell, 1990a; Mount & Barrick, 1995).
3. The FFM is a good general framework for thinking about important aspects of personality (Digman, 1990; Guion, 1998; Lubinski, 2000).
4. The more relevant and specific the work behavior we are trying to predict, the stronger the association between personality and behavior (Mount & Barrick, 1995).
5. Conscientiousness is best considered a combination of achievement and dependability. Achievement will predict some behaviors (e.g., effort) and dependability will predict other behaviors (e.g., attendance) (Hough, 1992; Moon, 2001; Mount & Barrick, 1995; Stewart, 1999).
6. Conscientiousness (along with its constituent factors achievement and dependability) has widespread applicability in work settings. It is possibly the most important personality variable in the workplace and it may be the equivalent of "g" in the noncognitive domain (Schmidt & Hunter, 1992).

7. Conscientiousness and its constituent factors (achievement and dependability) have a greater impact on behavior in situations where the worker has substantial autonomy (Barrick & Mount, 1993).
8. Conscientiousness, achievement, and dependability are only a small collection of a number of interesting facets of personality. The single-minded pursuit of "g" slowed down advances in understanding intelligence for almost 80 years. We should not let the same thing happen with the single-minded focus on conscientiousness (Collins, 1998).
9. There is evidence that factors other than conscientiousness have applicability for specific job families and occupations. Extraversion appears related to sales performance; openness to experience predicts training and expatriate success; agreeableness is associated with performance in customer-service and team-oriented jobs; emotional stability contributes to a broad range of jobs including management positions as well as jobs in the safety/security sector (Barrick, Mount, & Judge, 2001; Mount, Barrick, & Stewart, 1998; Vinchur, Schippmann, Switzer, & Roth, 1998).

Psychologists will continue to debate the number of elements of personality, the names of those elements, and the content of those elements, but we think that it is safe to say that personality is divided into no fewer than 5 basic elements and no more than 10 or 11. For the time being, we can use the Five Factor Model (McCrae & Costa, 1987) and the Nine Factor Model (Hough, 1992) as examples of the upper and lower limits. Neither of those models is "right" in any scientific sense. They are both plausible and they both have their applications. It is not uncommon to see some personality tests that measure more than the 10 or 11 elements we have proposed (e.g., the Saville Consulting Wave or the 16PF tests), but these tests are addressing much more distinct facets of personality, rather than its basic dimensions.

Tellegen (1993; Tellegen & Waller, 2000) has proposed an intriguing seven-factor model. His model includes the five dimensions of the FFM plus two other dimensions: positive and negative valence. **Positive valence** is represented by descriptions such as remarkable, extraordinary, excellent, and outstanding, and appears to be a continuum running from normal to exceptional. **Negative valence**, on the other hand, is the dark side of personality and is represented by descriptions such as cruel, evil, wicked, and sickening. It represents a continuum from decent to awful (Lubinski, 2000). Although there needs to be much more research on Tellegen's view of personality, these two additional dimensions strike a chord. We have all known people who were extraordinary in every respect. We want to be like them and be around them. And we have probably known a despicable person whom we avoid like the plague. Furthermore, both types of people might be high or low on conscientiousness, openness to experience, or even agreeableness!

POSITIVE VALENCE

Continuum of favorable personality characteristics running from normal to exceptional.

NEGATIVE VALENCE

Continuum of unfavorable personality characteristics running from normal to abominable.

A Caution One final caution needs to be made about the FFM and personality as a construct in general. Unlike "g" or other abilities, personality does appear to change regularly within individuals over time. McCrae et al. (1999) found that there was a systematic decline in neuroticism, extraversion, and openness between college age and middle adulthood (30–49), whereas agreeableness and conscientiousness increased during the same time period. More recent research (Srivastava, John, Gosling, & Potter, 2003) confirms that individuals do change as they age, and that these changes may also be related to gender. We use the term "caution" because, before we can reach any strong conclusions about personality change, we need to know more about the predictive value of the FFM factors within an individual over time. This is particularly true when young applicants are being considered for a career track that will stretch over a decade or more.

VOCATIONAL INTERESTS

Measures of **vocational interest** have been around for almost 80 years, but they have received only passing attention from I-O psychologists. Two reasons explain this lack of attention. The first is the belief that vocational interests do not predict job performance. The second is that they were often thought to be in the domain of vocational counseling and only useful for advising students about careers and occupations. As we will see below, there are reasons to reconsider these measures of individual differences.

SELF-EFFICACY
The belief in one's capability to perform a specific task or reach a specific goal.

VOCATIONAL INTEREST
Preference or liking for a particular activity or setting (as in a job or occupational setting).

In their simplest form, vocational interests are expressions of *liking* about environments (including social environments) and activities. When someone expresses a liking for "mechanical things" or "science" or "being around people," they are expressing an interest. An interest is less a behavior than a vision of oneself in a desired environment. You may like working with mechanical things but have a job as an accountant, or like working with people and have a job as a snow cat operator. If your interest (i.e., working with mechanical things or working with people) is strong and you find yourself in an environment that is not aligned with that interest, we would expect there to be some consequences. One of those consequences should be dissatisfaction with your occupation and probably with your job. We can see the prevalence of this dissatisfaction in the workforce when we consider how many people retire at the first opportunity in order to devote themselves to a second career that they "always wanted to do." A second, related consequence ought to be tenure in the occupation or job. People whose job and occupation are compatible with their interests should, all other things being equal, stay in that occupation (and possibly that job) for long periods of time. Finally, there ought to be some consequences for performance. If you are in a job or occupation that does not match a strong interest, it is more likely that your performance will be poorer than if you were in a job that matched your interests. This assumes, of course, that interests also reflect abilities to some extent.

It is important to remember that there are a myriad of other factors that can affect your job tenure, satisfaction, and performance. But that does not deny the possibility that interests add information that is not covered by ability or personality (Hogan & Blake, 1996; Lubinski, 2000). It is also intriguing that vocational interests appear in early adolescence (as early as age 13) and remain relatively stable over long periods of time (Lubinski, 2000).

Parents are often an important influence on their children's choice of occupation.

The reluctance of I-O psychologists to consider vocational interests in their decomposition of work behavior is unfortunate. To be sure, the associations with performance are not as high as one finds when considering mental ability or personality, but they exist and are reliable (Barge & Hough, 1988). Because one of the primary activities for I-O psychologists, whether employed within organizations or as consultants, is maintaining or enhancing profitability, it is not hard to understand why they would be more interested in predictors (such as "g" or conscientiousness) that are more strongly associated with performance. Similarly, it is not surprising that they would be less interested in predictors of satisfaction or occupational tenure than of performance. But that is a narrow view of work and workers. As we will see in Part III of this book, the investigation of satisfaction and organizational, occupational, and job tenure is beneficial not only to workers but also to employers.

As was the case with intelligence and personality, the area of vocational interests is dominated by one model. Developed

TABLE 3.3 Holland's Adjectival Descriptions of Six Personality Types

Realistic	Asocial, conforming, frank, genuine, hard-headed, materialistic, natural, normal, persistent, practical, self-effacing, inflexible, thrifty, uninsightful, uninvolved
Investigative	Analytical, cautious, critical, complex, curious, independent, intellectual, introspective, pessimistic, precise, rational, reserved, retiring, unassuming, unpopular
Artistic	Complicated, disorderly, emotional, expressive, idealistic, imaginative, impractical, impulsive, independent, introspective, intuitive, nonconforming, original, sensitive, open
Social	Ascendant, cooperative, patient, friendly, generous, helpful, idealistic, empathic, kind, persuasive, responsible, sociable, tactful, understanding, warm
Enterprising	Acquisitive, adventurous, agreeable, ambitious, domineering, energetic, exhibitionistic, excitement-seeking, extroverted, flirtatious, optimistic, self-confident, sociable, talkative
Conventional	Careful, conforming, conscientious, defensive, efficient, inflexible, inhibited, methodical, obedient, orderly, persistent, practical, prudish, thrifty, unimaginative

SOURCE: Hogan & Blake (1996).

RIASEC

Acronym for Holland's model of vocational interests, which proposes six interest types of people: realistic, investigative, artistic, social, enterprising, and conventional.

and presented by Holland (1985), it is known by the acronym **RIASEC** (see Table 3.3). The model proposes six interest types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional; the label RIASEC comes from the first initial of each interest type. An additional feature of the model is that the types are arranged hexagonally, with each type occupying a particular position (see Figure 3.4). In that figure some types are adjacent to each other and some types are directly opposite each other. Thus enterprising is opposite from investigative. Practically speaking, this means that if you express interests that would be enterprising, it would be unlikely that you would express interests that are investigative. And when you look at the definitions in Table 3.3, that makes sense. Investigating interests include reserved, cautious, and retiring behavior. Enterprising interests, on the other hand, represent excitement-seeking, talkative, and adventurous behaviors. Types that are close to each other in the hexagon are more compatible; it would not be surprising to see someone who expresses social interests expressing enterprising interests as well. The social type includes friendly and sociable interests, and the enterprising type includes agreeable, extraverted, and social interests.

Recently, there has been interest in the relationship between Holland's occupational types and the FFM. Barrick, Mount, and Gupta (2003) found strong relationships (correlations of the magnitude of +.40) between the RIASEC types of enterprising and artistic, and the FFM dimensions of extraversion and openness to experience. In a follow-up analysis, Mount, Barrick, Scullen, and Rounds (2005) proposed that vocational interests and personality actually combine to produce broad motives. They suggested that social interests (a vocational type) and extraversion (a FFM component) produce motives to interact with others, whereas artistic interests (a vocational type) and openness to experience (a FFM component) produce motives to enhance personal growth, seek new experiences and challenge the status quo. This work is very preliminary and we will revisit it in Chapter 8 when we consider the motivation of entrepreneurs.

ADDITIONAL ATTRIBUTES

The collection of cognitive abilities, physical and motor abilities, personality, and interests covers the major categories of proposed individual differences. The patterns formed by their combinations describe much of the variation among individuals. Nevertheless, some scientists propose additional aspects of individual differences. Below we will briefly cover some of them.

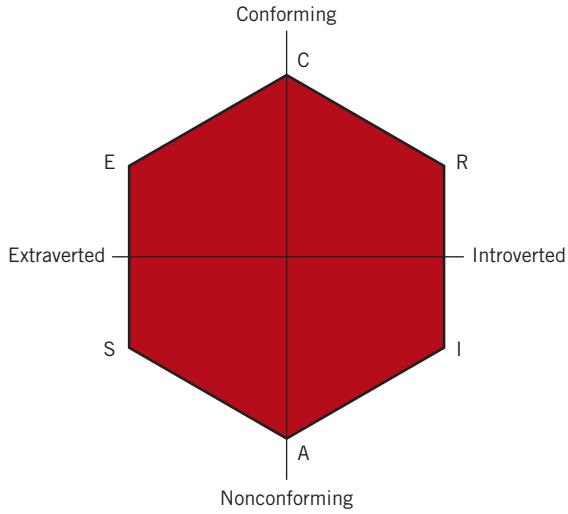


FIGURE 3.4 Interest Types Underlying the Hexagonal Representation of Holland's Vocational Typology
SOURCE: Hogan & Blake (1996).

SKILLS

Skills are practiced acts. Shooting a basketball, using a computer keyboard, and persuading someone to buy something are all examples of skills. They come with hours, days, and weeks of practice. It is unlikely that skills can be developed without certain abilities (eye–hand coordination, or memory, or reasoning), personality characteristics (persistence or agreeableness), and knowledge (understanding the controls that activate a piece of equipment). Although the skills depend on these ability, personality, and knowledge factors, the reason we call them skills is that they develop through practice. Technical and job-related skills are as varied as jobs and job tasks. There are other nontechnical skills that are more widespread than any technical skill. Examples include negotiating skills, communication skills, and conflict-resolution skills. These three are often lumped together by nonpsychologists and called **people skills**. Since they come into play most commonly in situations involving leader–follower and team member interactions, we will discuss these skills in the chapters that deal with teams and leadership.

SKILLS
Practiced acts, such as shooting a basketball, using a computer keyboard, or persuading someone to buy something.

PEOPLE SKILLS
A nontechnical term that includes negotiating skills, communication skills, and conflict resolution skills.

KNOWLEDGE

Knowledge can be defined as "a collection of discrete but related facts and information about a particular domain. It is acquired through formal education or training, or accumulated through specific experiences" (Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1999, p. 71). Knowledge is closely connected to skill when we are considering job-related skills (as opposed to psychomotor skills like shooting a basketball). Knowledge supports skill development and it comes in many varieties. It can be very basic (knowledge of mathematical operations or of vocabulary), or it can be sophisticated (knowledge of the circuitry of a notebook computer). Representative categories of knowledge as identified in the comprehensive occupational information network that has come to be known as **O*NET** (Peterson et al., 1999) are too detailed to present here but we have included the description on the text website for the interested reader. The O*NET architecture presents the name of the knowledge domain, the definition of the knowledge, and examples of what someone with a great deal or very little of the knowledge might be capable of doing. Perhaps the most immediate example of individual differences in knowledge is the distribution of test grades in your class. Although many variables may play a role in this grade distribution, one of those variables is certainly knowledge of the course material as presented in the text and lectures.

O*NET
Collection of electronic databases, based on well-developed taxonomies, that has updated and replaced the *Dictionary of Occupational Titles*.

Another kind of knowledge that has been proposed is called **tacit knowledge**, studied by Sternberg and his colleagues (Sternberg & Wagner, 1986; Sternberg, Wagner, & Okagaki, 1993). They distinguish between "academic" and "tacit" knowledge, the latter described as "action oriented knowledge, acquired without direct help from others, that

TACIT KNOWLEDGE
Action-oriented, goal-directed knowledge, acquired without direct help from others; colloquially called "street smarts."

allows individuals to achieve goals they personally value" (Sternberg, Wagner, Williams, & Horvath, 1995). They describe tacit knowledge as "knowing how" rather than "knowing that." A more formal way of distinguishing these two types of knowledge is **procedural knowledge** (knowing how) in contrast with **declarative knowledge** (knowing that).

PROCEDURAL KNOWLEDGE

Familiarity with a procedure or process; knowing "how."

DECLARATIVE KNOWLEDGE (DK)

Understanding what is required to perform a task; knowing information about a job or job task.

These researchers give an example of how tacit knowledge about getting along with your boss might affect your behavior. If you need to deliver bad news, and it is Monday morning, and you know the boss's golf game was rained out the day before, and the whole staff is nervous and walking on eggs, tacit knowledge would tell you that it would be best to deliver the bad news later. A common nonscientific term for tacit knowledge might be "street smarts." One of the important distinctions researchers make between formal or academic knowledge on the one hand and tacit knowledge on the other is that tacit knowledge is always goal-directed and useful, while academic knowledge may not be. People develop tacit knowledge about environments and processes that are personally valuable to them. Research seems to indicate that tacit knowledge is something above and beyond intelligence (Sternberg et al., 1995). Learning little tricks to perform better might be considered the light side of the tacit knowledge coin, and learning how to manipulate people might be the dark side.

EXPERIENCE

EXPERIENCE

Direct participation in, or observation of, events and activities that serves as a basis for knowledge.

The concept of tacit knowledge leads directly to a consideration of **experience** as an aspect of individual difference. Although experience does not always lead to tacit knowledge, tacit knowledge depends on experience. Just as most people would agree that individuals often differ in knowledge, they would also agree that individuals often differ in experience. This experience can be with a task, a job, an organization, or an occupation. Experience is often confused with seniority, but doing the same thing 100 times (seniority) is not the same as doing 100 things one time (experience). Jacobs, Hofmann, and Kriska (1990) suggested that experience on a given job is valuable up to a point, but then its value declines as the same work tasks and challenges begin to appear with greater frequency over time, making them less valuable "learning" experiences.

MEASUREMENT MODES

Unit of measurement used to assess experience.

LEVEL OF SPECIFICITY

Method used to gauge experience according to task, job, and organizational characteristics.

Two refined models of experience have been presented in the last few years. Quinones, Ford and Teachout (1995) proposed that experience can be considered along two dimensions: **measurement modes** and **level of specificity**. Measurement modes

Sometimes, the skill in reading a blueprint can be very important. NON SEQUITUR © 2004 Wiley Miller. Dist. By UNIVERSAL PRESS SYNDICATE. Reprinted with permission. All rights reserved.

BOX 3.2 AN EXAMPLE OF TACIT KNOWLEDGE

A postal worker gets on an elevator in a 25 story building and pushes the button for the 18th floor. Just before exiting the elevator at that floor, she pushes the button for the 25th floor, puzzling those left on the elevator who are going no higher than the 21st floor. The postal worker drops off mail and picks up mail from a central location on the 18th floor in less than 60 seconds, returns to the elevator, pushes the down button, and re-enters the elevator she just left making its way down from the 25th floor. She has learned that if she does not follow this routine, the elevator may not go to the 25th floor and she may have to wait several minutes for another elevator to travel up the 18 floors to retrieve her. This is tacit knowledge at its finest.

refer to the unit of measurement we use to assess experience. They propose that there are three modes: "amount" or the number of times a person has actually performed the task; "time," which would represent the length of time an individual has been performing a task or job; and "type," which captures some qualitative aspects of the experience related to task difficulty or job complexity. The second dimension of experience in their model addresses the issue of how specific the experience was. There are three levels of specificity: task, job, and organizational.

Tesluk and Jacobs (1998) expanded on the Quinones et al. (1995) model and suggested ways of combining the alternative measures suggested by the latter (amount, time, and type) to get a more complete index of experience. They also suggested that experience has a direct impact on increased work knowledge and skills, motivation, values, and attitudes, as well as indirect effect on job performance. Much of the emphasis in the Tesluk and Jacobs work experience model is on shaping experiences to make them of maximal value. We will return to the issue of shaping work experience in Chapter 7.

COMPETENCIES

In the past decade it has been common for I-O psychologists to talk about combinations of knowledge, skills, abilities, and other personality characteristics (KSAOs) in terms of **competencies**. Many different definitions and sets of competencies have been suggested. Kurz and Bartram (2002) have defined competencies as "sets of behaviors that are instrumental in the delivery of desired results or outcomes." Following from that definition, it is reasonable to assume that people can differ in the extent to which they possess competencies. But competencies are different from knowledge, or a skill, ability, or personality characteristic, in that a competency is really a collection of all of these specific individual difference characteristics. The essence of a competency is the *combination* of these characteristics and is not dominated by any one (Harris, 1998a). We will review a model of competencies called "The Great Eight," as proposed by Bartram (in press), in Chapter 4 when we discuss performance models.

COMPETENCIES

Sets of behaviors, usually learned by experience, that are instrumental in the accomplishment of desired organizational results or outcomes.

Competencies are unique in another way as well. Abilities can be defined and measured in the abstract, as can personality characteristics. But competencies only have meaning in the context of organizational goals. For example, you could distinguish between two individuals based on their measured conscientiousness, their reasoning ability, or their skill with a word processing program. But a competency of organizing and executing a business plan would require a combination of these three individual elements, in addition to various aspects of technical and procedural knowledge (Kurz & Bartram, 2002), and would have relevance only to that series of actions. Thus, competencies are really collections and patterns of the individual difference attributes we have already covered, rather than separate characteristics. We will return to competencies and how they are identified

(competency modeling) in Chapter 4, as a new way of thinking about analyzing jobs—a process called **job analysis**).

JOB ANALYSIS

Process that determines the important tasks of a job and the human attributes necessary to successfully perform those tasks.

EMOTIONAL INTELLIGENCE

In the 1980s Howard Gardner (1983, 1993) proposed a novel theory of intelligence. Rather than a unitary approach to intelligence such as "g," he posited seven different types of intelligence, including logical-mathematical, bodily-kinesthetic, linguistic, musical, spatial, interpersonal, and intrapersonal. He described the latter two intelligences as follows:

> Interpersonal intelligence is the ability to understand other people: what motivates them, how they work, how to work cooperatively with them. Successful sales people, politicians, teachers, clinicians, and religious leaders are all likely to be individuals with high degrees of interpersonal intelligence. Intrapersonal intelligence, a seventh kind of intelligence, is a correlative ability turned inward. It is a capacity to form an accurate veridical model of oneself and to be able to use that model to operate effectively in life. (1983, p. 9)

Gardner's notion of inter- and intrapersonal intelligence was popularized by Goleman (1995) using the label **emotional intelligence** (**EI**). EI is a relatively new concept with little in the way of an empirical data base at this point, but two questions about it have emerged. The first and perhaps simpler question is whether this actually represents a kind of intelligence, a skill developed and honed with practice, or a personality characteristic (Barrett, 2001). In many respects, this becomes more a semantic battle than a theoretical one. Nevertheless, the studies that have been done on the **construct** have been disappointing, failing to identify EI as something different from attributes with which we are already familiar (Davies et al., 1998; Roberts, Zeidner, & Mathews, 2001). In the past few years, there has been a flurry of debate regarding the value of EI as a "new" human attribute. Recently, the *Journal of Organizational Behavior* devoted an entire section to the debate (Ashkenazy & Daus, 2005; Conte, 2005; Daus & Ashkenazy, 2005; Landy, 2005a; Locke, 2005; Spector, 2005). In addition, Murphy (in press) has recently published what is destined to be the "authoritative" volume on EI. There are vigorous advocates on both sides. It would appear that as a unique attribute, EI may qualify for a minor role in the arsenal of human characteristics. Nevertheless, even if EI can be conclusively demonstrated to exist, it appears to occupy a minor role in predicting work success above and beyond the roles of "g" and personality. We will return to a discussion of EI measurement later in this chapter.

EMOTIONAL INTELLIGENCE (EI)

A proposed kind of intelligence focused on people's awareness of their own and others' emotions.

CONSTRUCT

Psychological concept or characteristic that a predictor is intended to measure; examples are intelligence, personality and leadership.

MODULE 3.2 SUMMARY

- Fleishman and his associates developed a taxonomy of 52 abilities, divided into the broad categories of cognitive, physical, and perceptual-motor abilities.
- "Intelligence" (or "g") is a very general mental capability that describes a person's ability to learn from experience.
- Meta-analyses of the relationship between "g" and job performance demonstrated that the more complex the job, the stronger the predictive value of general intelligence tests.
- Carroll proposed that intelligence had three layers, or strata. The highest layer is "g"; the next layer down consists of seven more specific abilities: fluid intelligence, crystallized intelligence, memory, visual perception, auditory perception, information retrieval, and cognitive speed.
- Physically demanding jobs require strength, flexibility, and stamina or aerobic endurance. Hogan proposed a seven-measure taxonomy of physical abilities, and combined these seven measures to form three higher-order physical

abilities: muscular strength, cardiovascular endurance, and movement quality.

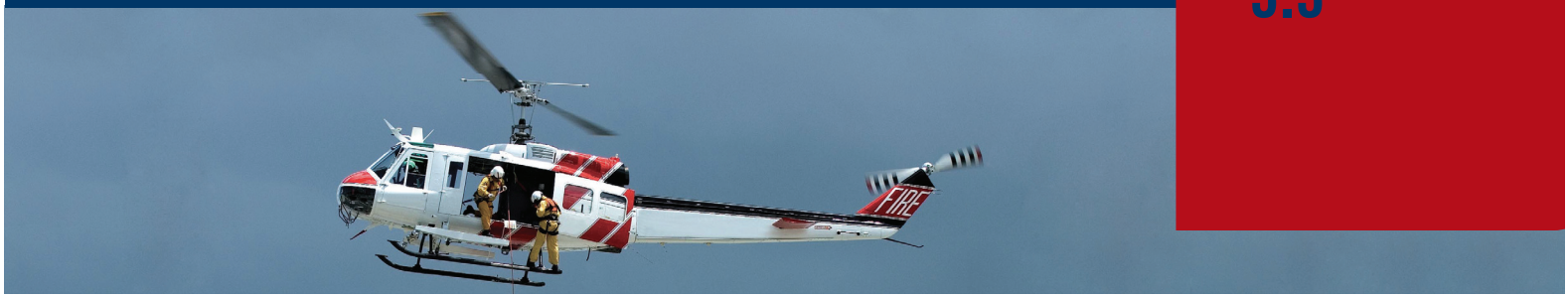
- It is important to determine whether employers' physical ability tests are fair to female applicants and older applicants, since both of these groups tend to have less strength than young men do. One way of enhancing the performance of females and older applicants on these tests is to encourage applicants to train ahead of time. It is also important that these tests relate to job performance prediction rather than injury prediction.
- There are clear connections between aspects of personality and various work behaviors, both productive (e.g., job performance) and counterproductive (e.g., dishonesty, absenteeism). I-O psychologists studying personality use a taxonomy labeled the Big 5 or the Five Factor Model (FFM).
- Of these five factors, the one that has attracted the most attention from I-O psychologists is conscientiousness. Barrick and Mount concluded, on the basis of a meta-analysis, that conscientiousness was positively related to success in all aspects of work for all occupations.
- Hough proposed nine basic personality factors rather than five; she suggested that the FFM factor of conscientiousness should be broken down into two discrete factors called achievement and dependability.
- The "dependable" (or "rule-bound") aspect of conscientiousness might actually be counterproductive in professions where the employee is expected to "think outside the box." The same can be true of "g" when success in the job depends on action rather than thought. Problems can arise when an individual has plenty of "g" but lacks other attributes.
- Barrick and Mount found through FFM research that in jobs where the employee had a great deal of control or autonomy, personality was much more predictive of performance than in jobs where the employee had little or no control.
- Tellegen has proposed an intriguing seven-factor model of personality that includes the five dimensions of the FFM plus two other dimensions: positive and negative valence.
- The area of vocational interests is dominated by Holland's model, known by the acronym RIASEC. The model proposes six interest types of people: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional.
- Skills are practiced acts. Although skills depend on ability, personality, and knowledge factors, what makes us call them skills is that they develop through practice.
- Knowledge can be defined as "a collection of discrete but related facts and information about a particular domain. It is acquired through formal education or training, or accumulated through specific experiences." Another proposed kind of knowledge is tacit knowledge, described as "knowing how" rather than "knowing that." A more formal way of distinguishing these two types of knowledge is procedural knowledge (knowing how) compared with declarative knowledge (knowing that).
- Although experience does not always lead to tacit knowledge, tacit knowledge depends on experience. Experience is often confused with seniority, but doing the same thing 100 times (seniority) is not the same as doing 100 things one time (experience).
- Competencies are "sets of behaviors that are instrumental in the delivery of desired results or outcomes." Competencies are different from knowledge, or a skill, ability, or personality characteristic, in that they are really a collection of all of these specific individual difference characteristics.
- Those who invoke the concept of emotional intelligence suggest that there is a unique kind of intelligence that is focused on our awareness of our own and others' emotions.

KEY TERMS

taxonomy
perceptual-motor abilities
affect
IQ
Intelligence Quotient
meta-analysis
Flynn effect
mean
standard deviation
stamina
muscular tension
muscular power
muscular endurance
sensory abilities
Americans with Disabilities Act
psychomotor abilities
sensorimotor abilities
motor abilities
Big 5
Five Factor Model (FFM)
conscientiousness
functional personality at work
agreeableness
emotionally stability
integrity
achievement
dependability
positive valence
negative valence
self-efficacy
vocational interest
RIASEC
skills
people skills
O*NET
tacit knowledge
procedural knowledge
declarative knowledge
experience
measurement modes
level of specificity
competencies
job analysis
emotional intelligence (EI)
construct

Foundations of Assessment

MODULE 3.3



THE PAST AND THE PRESENT OF TESTING

Yvonne felt as if she had been preparing for this day forever. There had been similar days, sure: the SAT exam to get into college, and the civil service test she took to get her summer job in the State Personnel Department. But this was show time. A high GRE score would be the ticket she needed for getting into a good graduate program. And that was exactly the problem. Yvonne choked up on standardized tests—always had and probably always would. Even though her SAT score had been low, she would finish with a 3.26 overall GPA and a 3.5 in her major. But getting into graduate school was not going to be as easy as it had been to qualify for her undergraduate program. The thing that really annoyed her was that these tests measured such a narrow band of who she was and what her capabilities were that it was a joke. How would they know that Yvonne was funny, loyal, and friendly, and had learned to read music in a weekend? Did they even care that she took hard courses rather than "cruisers"? She understood that there had to be *some* standard way of selecting among applicants, she just wished that it was not a standardized test.

Society seems to have a love–hate relationship with psychological testing, a practice almost as old as psychology itself. The term "**mental test**" was introduced by Cattell in 1890. As we described in Chapter 1, in the First World War over a million soldiers were tested for intelligence in order to determine which were best suited to be officers and which infantry. Up to that point, intelligence testing had been done on an individual basis, and this first trial of group testing was considered a massive success for the testing enterprise.

MENTAL TEST

Instrument designed to measure a subject's ability to reason, plan, and solve problems; an intelligence test.

But with this success came an embarrassment; soon after the war, psychological testing began to be used as the justification for limiting immigration. The army testing program discovered that immigrants and their offspring, who did not speak English as a first language, scored lower on these intelligence tests. Fearing that unchecked immigration would reduce the national intelligence level, Congress enacted immigration quotas. Although social critics were quick to point out the potential unfairness of intelligence testing, advocates saw it as a way to avoid the class system that had characterized industry and education in the 19th century. In their view, a test was "objective" and thus freed decisions (about jobs or education) from the grasp of favoritism and nepotism.

Private industry, like the government, was impressed by the success of the army testing programs and moved to implement testing as a way of selecting the most promising candidates from a pool of job applicants. Soon, however, the Great Depression of the 1930s arrived, drastically reducing the need to select from an applicant pool. There were no jobs to be had. When America entered the Second World War, the country returned to a full employment mode and virtually every able-bodied and motivated worker, male or female, either had a job or was serving in a branch of the armed services. Ships and airplanes were being built in record numbers, requiring one of the first "24/7" industrial environments.

Now there was no need for selection for the opposite reason: There were many more jobs than people.

On the military front, commanders quickly realized that war was now much more technologically advanced than it had been a generation earlier. Personnel needed to operate many different types of aircraft and ships with complex maintenance and repair demands. The task of the armed forces was no longer simply distinguishing between officers and infantry. The war effort needed pilots, bombardiers, artillery personnel, radar and sonar operators, and an enormous training and administrative staff. Psychological testing was once again pushed to the forefront as a tool in the war effort, this time with more sophisticated tests for the placement of recruits.

By the end of the Second World War, developers of tests had virtually glutted the market, offering ability, personality, interest, and knowledge tests. Neither the government nor the psychological profession exercised much control over the quality of the tests or the meaning of the test scores. A thriving and competitive testing industry operated without constraint until the early 1960s, when two societal forces converged to rein in testing. The first was a new wave of criticism about the value of testing from social observers (Gross, 1962; Whyte, 1956). These critics pointed out that employers were expecting job applicants to submit to a range of tests that had little apparent relationship to the job for which they were applying. Many of the tests, particularly the interest and personality tests, asked questions of a personal nature—topics like religion, sex, and politics. The second force was the passage of the Civil Rights Act of 1964, which prohibited discrimination in employment, including testing. If a test had the effect of reducing the employment opportunities of protected subgroups (e.g., African Americans, women), then the employer would need to provide evidence of the validity of that test. Since many of the tests available at that time had little validity evidence, employers saw this as a difficult hurdle to overcome.

As a result of the questions about the invasion of privacy and the possible discriminatory effects of tests, there was a marked reduction in test use for selection purposes, particularly intelligence and personality tests. The reticence lasted well into the 1970s, by which time more evidence of validity for tests had become available and the courts had clarified what was acceptable evidence for validity. At this time, research began to emerge showing that tests of cognitive ability were just as valid for minority test takers as for majority test takers. By the mid-1980s, testing was back in full swing and both intelligence and personality testing began to appear with greater frequency.

As we will see in the modules that follow, the content and process of employment testing is varied and encouraging. I-O psychologists have identified many different attributes that appear to contribute to work performance. Furthermore, I-O psychologists have identified many different methods for assessing these attributes.

But concerns about the "fairness" of testing continue to arise in many different settings. To mention just a few, some universities have decided to abandon standardized testing for applicants and introduce nonstandardized techniques that will permit motivation, interests, and values to play a greater role in student admissions. In both teacher and student testing in K-12 environments, there is a vigorous debate about the value of standardized tests for teacher certification and the awarding of high school diplomas. For example, many school districts require the successful completion of a series of content-specific tests (e.g., in mathematics or biology) as well as more general tests (e.g., knowledge of liberal arts) before granting teachers a permanent teaching certificate. These requirements occasionally result in lawsuits by unsuccessful teacher candidates (e.g., Gulino, 2002). In the wake of scandals such as the Enron and WorldCom accounting fraud cases, MBA programs are now considering the use of new "tests" of ethics, morality, and integrity to determine whom to admit to their MBA programs (Jackson, 2002).

Underlying all of these debates, the issue of fairness remains: Are standardized tests both effective and fair instruments for selecting among individuals? For every standardized test, there will be critics suggesting that the standardization prevents an illumination of the "essence" of the person. For every nonstandardized suggestion, there will be critics who will argue that the lack of standardization permits favoritism. Psychological testing will always have a values component to it in addition to the issues related to content and process.

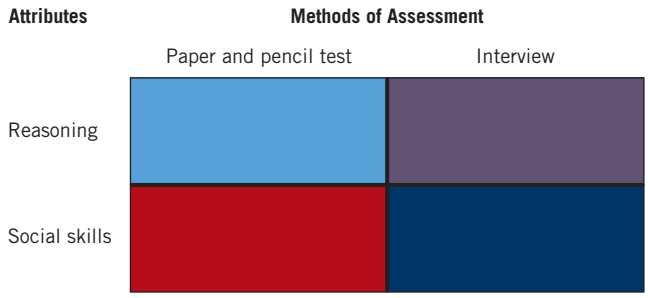


FIGURE 3.5 Two Attributes Measured Using Two Different Procedures

WHAT IS A TEST?

Robert Guion (1998) defined a **test** as "an objective and standardized procedure for measuring a psychological construct using a sample of behavior" (p. 485). Seventy years earlier, Clark Hull (1928) had proposed a virtually identical definition. Few definitions in psychology have remained so constant for such a long time. One of the appealing characteristics of this definition is that it is broad enough to cover a wide variety of tests and testing procedures. It encompasses paper and pencil tests, interviews, actual attempts to perform a piece of work (a work sample test), and even an application blank. The definition is also broad enough to cover many different types of content, including cognitive ability, personality, values, communication skills, interpersonal skills, and technical knowledge. In the modules that follow, we will review various content categories, as well as various techniques for assessing that content. As an example, if we were interested in the technical knowledge of an applicant for a word processing position, we could give the applicant a paper and pencil test and an interview, check with previous employers, have the applicant complete an actual word processing task at a workstation, or examine the applicant's formal education credits. Each of these techniques could be used to assess the same attribute: technical knowledge. Similarly, we might be interested in a number of different attributes of the applicant beyond technical knowledge, including communication skills, personality characteristics, interests, integrity, and career plans. We might use one or more interviews to assess each of these additional attributes. As you can see from Figure 3.5, in most practical testing situations, we are looking at the combination of attributes to be assessed (content) and ways to assess those attributes (process). Most employers look at several attributes using several techniques. Earlier in this chapter, we introduced the acronym KSAO (knowledge, skill, ability, other characteristics) to summarize the attributes of a worker. In one way or another, every test is an assessment of one or more of these content areas.

TEST
An objective and standardized procedure for measuring a psychological construct using a sample of behavior.

What Is the Meaning of a Test Score?

As Guion (1998) suggested, the term "objective" in his definition of a test implies quantification. When someone takes a test, he or she expects to receive a score on that test. It may be a simple pass–fail score (e.g., you may pass or fail a driver's license examination) or a score on some graded continuum (such as an 88 percent on a midterm examination or a B+ on a term paper). But the simple process of assigning a score is quite different from interpreting the meaning of that score. For example, if your instructor curves exam scores, and the exam was a tough one, an 88 might very well be in the A range and signal excellent performance. If, on the other hand, the test was an easy one and

virtually everyone got a 94 or above (except you), your 88 might be in the B range or lower.

NORMING
Comparing a test score to other relevant test scores.

Meaning is usually assigned to test scores through a process known as **norming**. Norming simply means comparing a score to other relevant test scores. In many employment settings, we compare individuals to one another, so the rules we use for making these comparisons should be unambiguous and fair. Test scores are often interpreted relative to some set of norms. In the classroom example above, your score of 88 percent is given meaning, or interpreted, by comparing it to the grades of your fellow students (the **norm group**). Instead of being compared to others in your class who took the same test you did, the instructor could have compared your score (and the scores of your classmates) to those of earlier classes who took midterms in the same content area. Or the instructor may not have curved the test at all but held to some previously determined comparison scale (90 to 100% = A; 80 to 89% = B, etc.). The development of test norms is very technical; excellent discussions of the process are presented in Guion (1998) and Cohen and Swerdlik (2002). For our purposes, it is simply important to be aware that while a test produces a "score," there is a need to interpret or give meaning to that score. As you will recall from our earlier discussion of validity in Chapter 2, validity is about inference: What can we infer from a test score about future performance? The meaning of a test score is a question of validity (Messick, 1995).

NORM GROUP
Group whose test scores are used to compare and understand an individual's test score.

Test Users and Test Interpretation

The issue of validity and meaning of a test score brings us to the more practical issue of who will interpret the test. Suppose you had been in an auto accident and were concerned about possible long-term effects from a blow to the head that you experienced in that accident. You might go to a neurologist for sophisticated testing to look for any impairment. Suppose further that the results of that testing arrived in the mail filled with numbers, diagnostic categories, and technical descriptions of the results of that testing. After poring over the results for an hour, you still might not know if there was permanent damage or not. Anyone without formal training in neurology would have difficulty understanding the meaning of the numbers and narrative from a standard neurological test battery.

Similarly, an individual not formally trained in the area of psychological assessment will have a difficult time interpreting the results of many psychological tests. Furthermore, individuals who lack suitable training are prone to making erroneous interpretations and, consequently, inappropriate decisions and actions. Fortunately, several documents are available that spell out proper and ethical procedures for test score interpretation and use (e.g., American Educational Research Association et al., 1999; Moreland et al., 1995; Society for Industrial and Organizational Psychology, 2003; Turner, DeMers, Fox, & Reed, 2001). Table 3.4 presents a list of the competencies that would be expected of those responsible for administering and interpreting psychological tests. As you can see, psychological testing, if it is to be done ethically and effectively, is no simple process.

What Is a Test Battery?

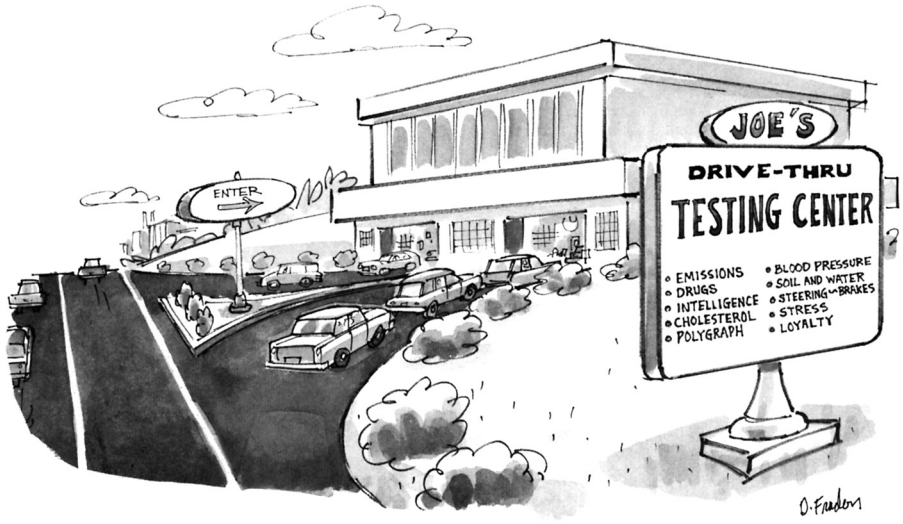
TEST BATTERY
Collection of tests that usually assess a variety of different attributes.

A **test battery** is a collection of tests rather than a single test. The tests in a battery are usually of different attributes. These attributes may be within a single area, such as a cognitive battery including subtests of reasoning, memory, and comprehension; or the attributes may be from conceptually different areas, such as a battery that includes a measure of cognitive ability, a personality test, a physical ability test, and a test of vocational interests. The term "battery" usually implies that all of the tests will be taken either in a single testing period or over a very short period of time. But whether the information being considered

TABLE 3.4 Twelve Minimum Competencies for Proper Use of Tests

ITEM NO.	COMPETENCY
1.	Avoiding errors in scoring and recording.
2.	Refraining from labeling people with personally derogatory terms like *dishonest* on the basis of a test score that lacks perfect validity.
3.	Keeping scoring keys and test materials secure.
4.	Seeing that every examinee follows directions so that test scores are accurate.
5.	Using settings for testing that allow for optimum performance by test takers (e.g., adequate room).
6.	Refraining from coaching or training individuals or groups on test items, which results in misrepresentation of the person's abilities and competencies.
7.	Willingness to give interpretation and guidance to test takers in counseling situations.
8.	Not making photocopies of copyrighted materials.
9.	Refraining from using homemade answer sheets that do not align properly with scoring keys.
10.	Establishing rapport with examinees to obtain accurate scores.
11.	Refraining from answering questions from test takers in greater detail than the test manual permits.
12.	Not assuming that a norm for one job applies to a different job (and not assuming that norms for one group automatically apply to other groups).

SOURCE: Moreland et al. (1995).



SOURCE: © The New Yorker Collection 1987 Dana Fradon from cartoonbank.com. All Rights Reserved.

is from several different assessment devices administered at one time or over a lengthy period of time, the critical issue is how to combine that information. Will it be combined to yield a single score with weights assigned to individual tests using a statistical equation of some type, or will the evaluator combine the individual test scores using a logical or nonstatistical process to yield a final recommendation? We considered the issue of statistical combination in Chapter 2 in the section on regression, but we will consider the broader issue of how test information may be combined at greater length in Chapter 6 when we deal with staffing decisions.

Where to Find Tests

At various points in the text, we mention some specific tests by name. There are literally thousands of psychological tests available on a broad range of topics. If you wanted to find a test, how would you do it? Textbooks on testing provide lists and examples of tests. For example, Anastasi and Urbina (1997) presented an extensive list of tests covering a range of topics, as well as a listing of test publishers. A more complete listing of tests, as well as reviews of those tests, can be found in two established sources. The first is known as the ***Mental Measurements Yearbook*** (**MMY**). This was first published in 1938 and has been updated an additional 14 times. The 15th edition (Plake, Impara, & Spies, 2003) was published in 2003. Buros Institute (named after the founder of the MMY, Oscar K. Buros) also publishes a companion volume without reviews called *Tests in Print*. The most recent edition of this companion text will be released in 2006 (*Tests in Print VII*, Murphy, Spies, & Plake, in press). Table 3.5 presents a typical entry in the MMY.

MENTAL MEASUREMENTS YEARBOOK

Widely used source that includes an extensive listing of tests as well as reviews of those tests.

Thomas (2004) has published an excellent handbook providing information about tests that are available for use in business settings. It covers both the content and format of all categories of modern industrial assessment and provides in-depth reviews of both

TABLE 3.5 Short Employment Tests, Second Edition

Purpose: "Measures verbal, numerical, and clerical skills."

Population: Adults.

Publication Date: 1951–1993.

Acronym: SET.

Scores: 4: Verbal, Numerical, Clerical, Total.

Administration: Group or individual.

Forms: 2 forms of each subtest (Verbal, Numerical, Clerical Aptitude).

Price: 2006: $249 per starter kit, including 25 each of the Verbal, Numerical, and Clerical Aptitude test booklets, scoring key, and manual; $95 per package of 25 test bookets (specify version and Form 1 or 2); $320 per package of 100 test booklets (specify version and Form 1 or 2); $30 per combination scoring key; $45 per manual.

Time: 5(10) minutes.

Comments: Distribution of Form 1 restricted to banks that are members of the American Banking Association.

Authors: George K. Bennett and Marjorie Gelink.

Publisher: Harcourt Assessment, Inc.

SOURCE: Murphy, Spies, & Plake, 2006

devices and assessment practices generally. It is an excellent resource for those with an interest in industrial assessment and a bit more focused than the other sources listed above because they deal with all tests (e.g., counseling, educational) rather than tests used exclusively in a business setting.

ADMINISTRATIVE TEST CATEGORIES

In descriptions of tests and testing, you may encounter several terms that require a brief explanation.

Speed versus Power Tests

Some tests have rigid and demanding time limits such that most test takers will be unable to finish the test in the allotted time. These are called **speed tests**. Your score on a speed test would be calculated by considering the number of items you were able to complete correctly in the time available. As Murphy and Davidshofer (2005) pointed out, if someone scores poorly on a speed test, it is not clear whether the person actually knew the answers but could not respond quickly enough, or would have been unable to answer correctly no matter how much time was allotted. **Power tests** have no rigid time limits. While some test takers may still not finish, enough time is given for a majority of the test takers to complete all of the test items. The items on power tests tend to be answered correctly by a smaller percentage of test takers than those on speed tests.

SPEED TEST

Has rigid and demanding time limits so most test takers will be unable to finish the test in the allotted time.

POWER TEST

Has no rigid time limits; enough time is given for a majority of the test takers to complete all of the test items.

Assessment professionals find that speed tests provide greater variability among candidates, allowing for more effective prediction, but they carry some vulnerabilities. The most obvious of these is whether the job actually requires such speed for successful performance. Few jobs have such demands. The second potential pitfall is the possibility of introducing unfairness to the testing process by emphasizing speed. One of the documented effects of the aging process is a decline in information processing speed. As we age, we take longer to complete cognitive operations. In many instances, this slowing process is irrelevant to the actual demands of a job; it won't matter that a worker took 10 or 20 seconds rather than three seconds to accomplish a task. Nevertheless, there are some professions (e.g., airline pilot, police officer, firefighter, bus driver) where speed of information processing or reaction might be critical. Disabled individuals, particularly those with learning disabilities, may also find themselves at a disadvantage on a speeded test. One of the most common requests for a testing accommodation made by individuals under the Americans with Disabilities Act (1990) is for additional time to complete a test. Thus, speed tests may increase the risk of legal challenge from many groups unless it can be shown that the type of speed required by the test is also required by the job.

The television show *Jeopardy* is an example of a speed test.

Group versus Individual Tests

Most standardized written tests, even if they are administered to single individuals, could be administered in group format. A cognitive ability test taken by applicants for a position in the police academy could be taken in a hall or a convention center with 20,000 other test takers, or the same test may be taken individually in a room on an army base where a candidate is stationed during a call-up of the

GROUP TEST

Can be administered to large groups of individuals; often valuable in reducing the costs (both in time and money) of testing many applicants.

INDIVIDUAL TEST

Test given only on an individual basis.

reserves. **Group tests** are efficient because they allow for the testing of many candidates simultaneously, resulting in rapid screening compared to individually administered tests. Group testing is also often valuable in reducing the costs (in both time and money) of testing many applicants.

Certain tests, however, can be given only on an individual basis. Examples include an interview, a test of hand–eye coordination, or an elaborate assessment of candidates for a high-level executive position based on interviews, work samples, and individually administered personality tests. **Individual tests** are also often more appropriate when the employer wishes to assess a candidate's *style* of problem solving rather than the simple *products* of the problem-solving process. Individual testing formats are also appropriate when the examiner needs to establish an interpersonal rapport with the test taker, as is commonly the case in certain clinical tests such as the Rorschach Inkblot test. Even though tests such as the Rorschach are often used in individual assessment, there is little evidence that they represent added value (i.e., validity) beyond information gathered using more structured devices (Murphy & Davidshofer, 2005).

Paper and Pencil versus Performance Tests

PAPER AND PENCIL TEST

One of the most common forms of industrial testing that requires no manipulation of any objects other than the instrument used to respond.

PERFORMANCE TEST

Requires the individual to make a response by manipulating a particular physical object or piece of equipment.

Paper and pencil tests are one of the most common forms of industrial testing. The paper and pencil test requires no manipulation of any objects other than the instrument used to respond. By extension, the modern version of the paper and pencil test might be the computer keyboard test where the keys and mouse are used only to choose the correct response or produce a narrative response to a question. Given the increasing popularity of computer- and Internet-administered tests, it might be better to adopt a term other than "paper and pencil testing"; a distinction such as nonmanipulative versus manipulative might be more apt. We will discuss computer and Internet testing in a later section of this chapter.

Performance tests require the individual to make a response by manipulating a particular physical object or piece of equipment. The score that the individual receives on the test is directly related to the quality or quantity of that manipulation. An example might be a test administered to a candidate for a dental hygienist position. The candidate might be asked to prepare a tray for cleaning or scaling teeth, to prepare a syringe of novocaine for administration by the dentist, or to prepare a mold for taking an impression of a row of teeth. In this case, the candidate's *skill* in performing these tasks may be as important as his or her knowledge of *how* to carry out the actions.

TESTING AND CULTURE

BIAS

Technical and statistical term that deals exclusively with a situation where a given test results in errors of prediction for a subgroup.

FAIRNESS

Value judgment about actions or decisions based on test scores.

CULTURE

A system in which individuals share meanings and common ways of viewing events and objects.

In the 1950s and 1960s, testing was largely lacking in controls, either legal or professional. As social critics pointed out, the quality of tests was therefore variable, and the potential for cultural influence and bias was substantial. An example would be a test that used a very high level of vocabulary to assess a relatively simple and straightforward skill. Instead of asking "How much is two plus two?" the item might have read, "If one were asked to calculate the arithmetic sum of the two integers that have been produced below, what would the resultant number be?" The second item would surely be more difficult for someone with a limited vocabulary or low reading comprehension to answer, even though both items are ostensibly assessing the same basic skill—arithmetic proficiency. Modern tests have eliminated most if not all of these reading level problems. What they may not have done, however, is to eliminate cultural influences.

Murphy and Davidshofer (2005) distinguished among three terms in discussing tests and testing: **bias**, **fairness**, and **culture**. They rightly pointed out that bias is a technical and statistical term that deals exclusively with the situation in which a given test results in

errors of prediction for a subgroup. Thus, if a test underpredicts the job performance of women (i.e., predicts that they will score lower on some criterion than they actually do) and overpredicts the job performance of men (i.e., predicts that they will score higher on some criterion than they actually do), then the test would be said to be biased. You will remember that earlier in this chapter, we described a case involving a strength test for female applicants in a meat packing plant. In essence, the judge in that case ruled that the strength test was *biased* since it predicted that a substantial percentage of women would perform poorly and almost all men would perform well at meat packing tasks. In fact, the test might have predicted injuries but was not effective in predicting actual performance on the job.

In contrast, fairness is a value judgment about actions or decisions based on test scores. Many employers base their decision to hire on a paper and pencil test of general mental ability. Many applicants believe that in addition to (or instead of) the cognitive ability test, dependability and motivation should play a role in a hiring decision. This was the view of Yvonne in the example at the beginning of this chapter. In the view of many applicants, the test and the method of hiring is unfair even though there may be no statistical bias in predictions of success.

Murphy and Davidshofer (2005) considered fairness to be a philosophical or political term and not a scientific one. They gave an example to make their point. A test of physical strength might predict job success equally for male and female firefighter applicants, yet eliminate most of the female applicants because they have less upper body strength than males. Many individuals would consider such a test unfair even though it was unbiased, because it prevents women from becoming firefighters. In contrast, a biased test might be used to increase the number of minorities in a particular job or company, but still be considered fair because it corrects for a past underrepresentation of those minority group members.

Culture is a third concept, separate in many respects from either fairness or bias. Culture addresses the extent to which the test taker has had an opportunity to become familiar with the subject matter or processes required by a test item (Murphy & Davidshofer, 2005). In many tests for teacher certification, there is a component that addresses the general cultural literacy of the candidate—for example, how well he or she knows works of art and music, variations of modern dance, and the deeper meaning of literary passages (National Evaluation Systems, 2002). Minority candidates often do poorly on these tests of cultural literacy and argue that they are at a disadvantage compared to candidates (particularly majority candidates) who have been exposed to this knowledge through home and school environments. In 1972 Williams proposed a test called the BITCH (Black Intelligence Test of Cultural Homogeneity). The test was composed of items that utilized the black ghetto slang of the time. Williams was not actually suggesting that the BITCH be used to replace other tests of intelligence. Instead, he published it in an attempt to highlight the influence of culture and subculture on language, and therefore on test scores. One item asked for the meaning of the phrase "running a game." Which of the following was the correct interpretation?

a. Writing a bad check.
b. Looking at something.
c. Directing a contest.
d. Getting what one wants.

The correct answer in 1972 (and still today in many black neighborhoods) was "d." But how many white test takers would have known that? Most whites got that answer wrong. Herlihy (1977) made the same point with a test called the CRUST (Cultural/Regional Uppercrust Savvy Test) (Cohen & Swerdlik, 2002). Consider the following CRUST item.

The Blue Book is:

a. The income tax guidelines.
b. A guide to pricing used cars.
c. A booklet used for writing essay exams.
d. A social register listing 400 prominent families.

Again, the correct answer is "d" from the perspective of the person who constructed the test, but only those in the highest levels of society would be likely to know this.

Greenfield (1997) presented examples of difficulties in "transporting" North American cognitive ability tests to other cultures. In one example from Cole, Gay, Glick, and Sharpe (1971), Liberians were asked to engage in a cognitive sorting task. They were asked to sort 20 objects into categories. According to the test developers, the objects divided evenly into four categories: food, food containers, clothing, and implements. Rather than sorting the 20 objects into these four categories, however, the Liberian participants made functional pairings of the objects. For example, they paired a potato with a knife, reasoning that a knife is used to cut a potato. When they were asked why they did that, they would reply that this is how a "wise man" would complete the task. After repeated attempts to get them to use the four neat categories of items, and repeatedly getting the "wise man" response, the researchers asked the participants to sort the items as a "fool" would do it. The subjects promptly sorted the items into the four categories that the researchers preferred. As Greenfield noted, "the researchers' criterion for *intelligent* behavior was the participants' criterion for *foolish*; the participants' criterion for *wise* behavior was the researchers' criterion for *stupid*" (p. 116). Greenfield concluded that to use a test developed in one culture for another culture, there must be agreement on the value of particular responses to particular questions, as well as agreement that the items mean the same thing in the different culture. Note that neither of these requirements has anything to do with the quality of the test's linguistic translation; instead, they relate to meaning in a deeper sense. Sternberg (2004) has argued vigorously that intelligence cannot be understood without taking into account the culture in which it is measured. He cites the example of the Taoist culture, in which intelligence includes the importance of humility, freedom from conventional standards of judgment, and full knowledge of oneself; in contrast, the Confucian perspective emphasizes the importance and persistence of life-long learning with enthusiasm.

A recent study of East Indian and American workers underscored Greenfield's caution (Ghorpade, Hattrup, & Lackritz, 1999). Although the researchers found few differences between Indian and American men or women with respect to the measurement of the personality variable Locus of Control, there were substantial differences between Indian and American women in the meaning of a self-esteem measure. Indian women were much more likely to feel guilt over individual activities, such as seeking opportunities to succeed and achieve, that might be seen by Americans as evidence of self-esteem. This was likely the result of differences between Indian and American women on Hofstede's collectivism–individualism dimension. In addition, it is likely that American women were more likely to identify with the masculine end of Hofstede's masculinity–femininity dimension, favoring accomplishment rather than interpersonal relations. Thus, if Indian and American women were compared on self-esteem, a researcher might see the Indian women as having "less" esteem when, indeed, what Americans view (and admire) as self-esteem had a far less positive connotation to the Indian women.

As Americans from different ethnic groups increasingly mingle in public schools, universities, other public institutions, and work settings, they are becoming more familiar with each other's subculture today than was the case 30 years ago. As a result, the concept of the cultural content in current tests is becoming less of an issue in explaining differences among ethnic groups. At the same time, cultural content is becoming an increasingly important issue in the workplace because of the growing multicultural nature of work and the increasing cultural diversity of applicant populations.

International Assessment Practices

Earlier in the chapter, we reported research that found that tests of mental ability were used more commonly in Europe than in the U.S. This is just one example of the differences that can be found worldwide in assessment practices. Variations in global assessment practice will become increasingly important in the next decade for both multinational employers and applicants to multinational organizations. Several reviews of assessment in other countries help to illustrate the differences between assessment in the U.S. and assessment elsewhere (Bartram & Coyne, 1998; Muniz et al., 2001; Oakland, 2004; Roe & van den Berg, 2003). Highlights from these reviews include the following:

- European psychologists would like to have a more structured role for professional institutions in developing and monitoring good testing practices. In response to that expressed need, the International Test Commission developed the *International Guidelines for Test Use* (International Test Commission, 2000).
- In industrial settings, the countries in which tests were most frequently administered by psychologists were Croatia, Bulgaria, Finland, Slovakia, Denmark, Japan, and Slovenia. The countries in which tests were most frequently administered by non-psychologists included Canada, Sweden, Cyprus, Norway, Switzerland, the U.K., and Germany.
- The greatest amount of information about test quality could be found in the U.S., the Netherlands, Japan, the U.K., Canada, Spain, Finland, Belgium, and Slovakia; the least amount of information was available in China, Denmark, Ukraine, and South Africa.
- In India and China, testing is largely unregulated; many countries are moving toward the certification and training of non-psychologists who use tests (Bartram, 2005).

In general, it would appear that the various guidelines available for evaluating tests, test use, and test users in the U.S. (APA Standards, 1999; SIOP Principles, 2004; Turner, DeMers, Fox, & Reed, 2001; Uniform Guidelines, 1978) are ideals to which many other countries aspire.

MODULE 3.3 SUMMARY

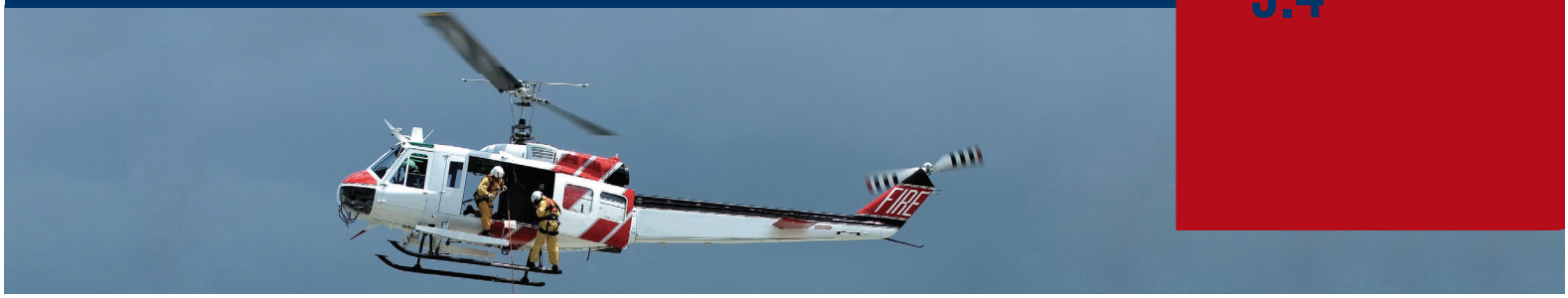
- Employment testing was first widely used after the First World War and has been heavily influenced by the Civil Rights Act of 1964. I-O psychologists are interested in determining how effective various tests are in predicting work performance. They have identified many different attributes that appear to contribute to work performance and many different methods for assessing these attributes.
- The definition of a test encompasses paper and pencil tests, interviews, actual attempts to perform a piece of work (a work sample test), and even an application blank. The definition is also broad enough to cover many different types of content, including cognitive ability, personality, values, communication skills, interpersonal skills, and technical knowledge.
- In Module 3.2 we introduced the acronym KSAO (knowledge, skill, ability, other characteristics) to summarize the attributes of a worker. In one way or another, every test is an assessment of one or more of these content areas.
- Tests can be described or differentiated according to categories that include speed versus power tests, individual versus group tests, and paper and pencil versus performance tests.
- In discussing tests and testing, it is important to consider three factors: bias, or errors of prediction; fairness, a value judgment about decisions based on test scores; and culture, the extent to which a test taker has the opportunity to become familiar with the subject matter.

KEY TERMS

mental test
test
norming
norm group
test battery
Mental Measurements Yearbook
speed test
power test
group test
individual test
paper and pencil test
performance test
bias
fairness
culture

Assessment Procedures

MODULE 3.4



ASSESSMENT CONTENT VERSUS PROCESS

It is common for employers and applicants to confuse the content of testing with the process of testing. As we suggested earlier in this chapter, it is possible to distinguish between *what* attribute is being assessed and *how* it is being assessed. For example, after applying for a job with a local company, an applicant might describe the process as including a personality test, a cognitive test, an interview, and a background check. The terms "personality" and "cognitive" describe the content of the assessment and the terms "interview" and "background check" describe the process of the assessment. The reason why this content–process distinction is important is that you will often see claims for the "validity" of the interview or work sample. But the validity depends not so much on the process by which the information was gathered as on the content of that information. In the sections that follow, we will consider information gathered in various formats, ranging from a paper and pencil test to an interview. But as we discussed earlier, many of these methods can be used to gather many different kinds of information. For example, an interview could assess communication skills, knowledge, ability, or personality—or, as is most often the case, a combination of those "content" categories—depending on what questions are asked and how it is scored. First, we will consider the content of assessment, and then the process by which this content may be gathered.

ASSESSMENT PROCEDURES: CONTENT

Cognitive Ability Tests

Guion (1998) defined **cognitive ability tests** as those which:

> allow a person to show what he or she knows, perceives, remembers, understands, or can work with mentally. They include problem identification, problem-solving tasks, perceptual (not sensory) skills, the development or evaluation of ideas, and remembering what one has learned through general experience or specific training. (p. 486)

Even though Guion identified what seem to be a variety of cognitive abilities (e.g., remembering, problem identification), as we saw earlier in this chapter, there is still a vigorous debate regarding whether there is only one overarching cognitive ability—"g" or "general mental ability"—or several distinct facets or abilities (Ackerman, 1992; Ackerman et al., 2002, 2005; Carroll, 1993; Olea & Ree, 1994; Ree, Earles, & Teachout, 1994).

COGNITIVE ABILITY TEST

Allows individuals to demonstrate what they know, perceive, remember, understand, or can work with mentally; includes problem identification, problem-solving tasks, perceptual skills, the development or evaluation of ideas, and remembering what one has learned through general experience or specific training.

Which would be the better shears for cutting metal?

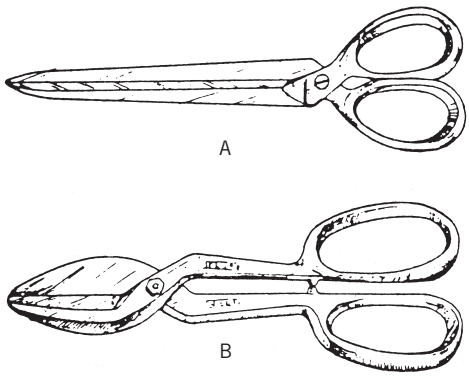


FIGURE 3.6 Sample Item from Bennett Mechanical Comprehension Test
SOURCE: Bennett Mechanical Comprehension Test, Form BB, Item Y.

In more than a century of cognitive ability testing, there have been tests that produce a single number intended to represent cognitive ability, tests of specific abilities, and test batteries that purport to measure several different facets of cognitive ability.

Tests that Produce a Single Score An example of a test intended to produce a single score representing general mental ability is the Wonderlic Personnel Test. It includes 50 items that assess verbal, numerical, and spatial abilities. Because its administration time is 12 minutes and most applicants cannot finish the test in the allotted time, the Wonderlic is considered a speed test. There are elaborate norms for the Wonderlic, making its interpretation relatively simple. Its ease of administration and scoring make it a popular device for many organizations. Murphy and Davidshofer (2005) endorsed the use of the Wonderlic, pointing to its high reliability and strong correlations with other, more elaborate, tests of intelligence.

Tests of Specific Abilities As implied by Guion's definition of a cognitive ability test, many tests concentrate on only one aspect of cognitive ability. An example of such a test is the Bennett Test of Mechanical Comprehension. Figure 3.6 presents an example from one form of the Bennett Test (1980). The item asks the test taker to examine the two different cutting instruments and to deduce, from either experience or logic, that the shears labeled B would be more effective at cutting metal than A. One can imagine that such a test item might be well suited for choosing applicants for the trade position of sheet metal worker or plumber.

Another example of a specific mental ability is spatial relations. Consider the item in Figure 3.7. It requires the test taker to do some actual mental manipulation of the factory shown from the front by "turning" the factory in his or her mind and then choosing the response below that would most closely resemble how the factory would look from the back. This ability to manipulate objects in one's mind is particularly useful for many hardware repair or "trouble shooting" professions, such as an auto mechanic or computer repair technician, where it is necessary to visualize a component buried deep under the hood of a car or in a hard drive. There are many other examples of specific cognitive abilities, such as clerical and perceptual accuracy, memory, and reasoning. Most testing texts (Anastasi & Urbina, 1997; Cohen & Swerdlik, 2002; Guion, 1998; Murphy & Davidshofer, 2005; Thomas, 2004) provide detailed descriptions of these tests. Mumford, Baughman, Supinski, and Anderson (1998) presented a sophisticated treatment of how to measure complex cognitive abilities such as reasoning and creative problem solving.

Cognitive Test Batteries Multiple aptitude test batteries have a long history in psychological testing in industry. In part, this is a historical accident because some of the early batteries were developed before the move toward the construct of general mental ability. Thurstone (1938) introduced a test of Primary Mental Abilities (PMA) that assessed numerical ability, verbal ability, reasoning, spatial relations, perceptual speed, and memory. More recent examples of multiple aptitude test batteries include the Armed Services Vocational Aptitude Battery or ASVAB (Katz, 1987; Murphy & Davidshofer, 2005), the Differential Aptitude Test Battery or DAT (produced by the Psychological Corporation, 1973, 1974), and the General Aptitude Test Battery or GATB (Hartigan & Wigdor, 1989). The ASVAB, as implied by its name, is used exclusively by the armed services for selection and placement. The GATB is used exclusively by the federal government to assist in the selection and placement of civilian workers. The DAT is commercially available for employer use. Students are more likely to be more familiar with the Scholastic Aptitude Test (SAT)

or Graduate Record Examination (GRE), both examples of cognitive test batteries. Although each of these batteries is slightly different, in one way or another they all measure verbal, numerical, spatial, and reasoning abilities. Although **cognitive test batteries** take longer to administer than a "single score" test like the Wonderlic, or any test of an individual facet of cognitive ability, they do have the advantage of providing more detailed information about particular manifestations of cognitive ability that may be more important in one job than another.

COGNITIVE TEST BATTERY

Collection of tests that assess a variety of cognitive aptitudes or abilities; often called Multiple Aptitude Test Batteries.

Knowledge Tests

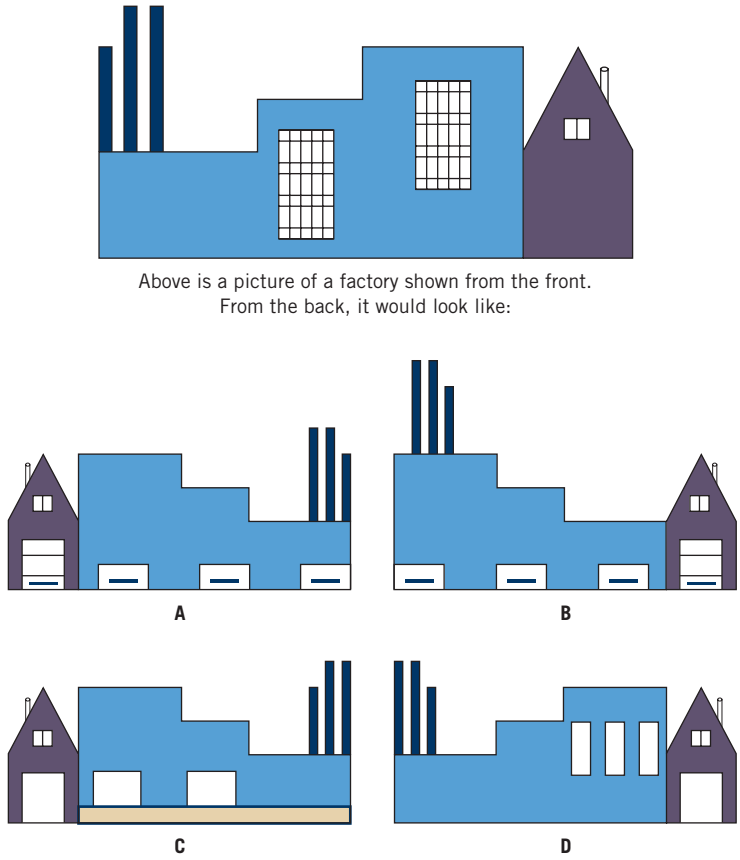
Tests you will take in this and other courses are **knowledge tests**. They assess the extent to which you know course material. These types of tests are typically tailored to course or training material. Knowledge tests are also administered for licensing and certification purposes, including teacher certification, nuclear power plant operator licensing, and licenses to practice law or medicine, or to sell investments. Knowledge tests are like any other type of test and require the same care in development, norming, and administration. We will discuss non-paper and pencil forms of knowledge tests later in this chapter.

KNOWLEDGE TEST

Assesses the extent to which individuals understand course or training materials; also administered for licensing and certification purposes.

TESTS OF PHYSICAL ABILITIES

As we saw earlier in the chapter, there are seven basic physical ability attributes (Hogan, 1991a). These include static strength, explosive strength, coordination, and stamina or aerobic endurance. While it is possible to measure each of these physical abilities in isolation, most physically demanding jobs actually require combinations of these abilities. As a result, many physical ability testing procedures tend to use simulated pieces of work to assess the combined abilities. For example, consider a test frequently used to assess the physical abilities of firefighter candidates (see Table 3.6), composed of several events each of which requires multiple abilities. An excellent review of physical abilities and their measurement appears in a study of age and physical abilities conducted for the Equal Employment Opportunity Commission (EEOC) and the Department of Labor (Landy, Bland, Buskirk, et al., 1992). There is substantial evidence that measures of physical abilities can improve the prediction of job success for many physically demanding jobs (Arnold, Rauschenberger, Soubel, & Guion, 1982; Campion, 1983; Hoffmann, 1999; Reilly, Zedeck, & Tenopyr, 1979). Arvey, Landon, Nutting, and Maxwell (1992) provide a good description of the development and validation of an entry-level physical ability examination for police officers. The



Above is a picture of a factory shown from the front. From the back, it would look like:

FIGURE 3.7 Spatial Relations Item from a Test for Firefighters

TABLE 3.6 Physical Ability Tests for Firefighters

Stairway climb: Candidate wears fire protective clothing and air tank and carries seven pieces of equipment up three flights of stairs, one piece at a time. Each piece of equipment weighs between 25 and 55 pounds.

Hose pull: Candidate wears air tank, stands in one spot, and pulls 50 feet of fire hose filled with water using a hand-over-hand technique.

Ladder pull: Candidate wears air tank and pulls a16-foot ladder from the ladder bed of a fire truck, places it on the ground, picks it back up, and replaces it in the ladder bed.

Dummy drag: Candidate drags a 125-pound sandbag around a serpentine course of 40 feet. The candidate must keep one knee in contact with the ground and may not lift or carry the sandbag but must drag it.

Blind crawl: Candidate wears fire protective clothing and an air tank. After putting on a blackened face mask, the candidate must crawl through a plywood maze that has several turns in it. In addition, there are sandbags located strategically throughout the maze. The maze is approximately 40 feet in length.

Pike pole: Candidate wears an air tank and alternately pulls and pushes a 75-pound weight attached to a pole hanging from a frame. The candidate must complete as many repetitions as possible in a four-minute period. A repetition is defined as one push and two pulls.

Fan hang: Candidate wears fire protective clothing and an air tank and lifts a 50-pound fan from ground level, hanging it on a standard door frame.

caution, as we have seen earlier, is that the physical ability tests are most defensible when used to predict *performance* rather than *risk of injury*.

Psychomotor Abilities

PSYCHOMOTOR ABILITIES

Physical functions of movement, associated with coordination, dexterity, and reaction time; also called motor or sensorimotor abilities.

Tests of **psychomotor abilities** involve the coordinated movement of the limbs in response to situational factors. It may be a complex task in which the individual is required to move arms and legs in coordination, as in flying an airplane, driving a vehicle, or playing an organ; or it may be a simple or discrete action such as firing a weapon, pulling a lever, or administering an injection to a patient. For some jobs, psychomotor abilities represent characteristics of the individual that have some potential for contributing to successful job performance above and beyond cognitive abilities, physical abilities, or personality characteristics. Psychomotor abilities are usually assessed using a task that requires dexterity, such as placing pins in slots with tweezers, such as is depicted in Figure 3.8. Ackerman and his colleagues have developed some sophisticated computer-based psychomotor tests for the selection of applicants for jobs such as air traffic controllers (Ackerman & Cianciolo, 1999, 2002).

PERSONALITY

As we have seen earlier in the chapter, personality attributes are now widely recognized as contributors to job success. There are many commercially available instruments for measuring personality characteristics, many based on the Big 5 model described earlier. Table 3.7 lists some of the more commonly used personality instruments. The history of personality testing can be described in two general phases. The early foundation of personality testing was focused on the identification of the abnormal personality and evidence

of possible psychopathology (i.e., mental illness). Using personality testing for that purpose might be thought of as an attempt to screen *out* potentially problematic employees. With the advent of instruments intended to provide quantitative descriptions of the normal (rather than abnormal) personality, personality testing in employment shifted to a screen *in* process whereby employers sought to identify applicants with positive personality characteristics (e.g., conscientiousness, emotional stability, or agreeableness) that would contribute to effective performance.

FIGURE 3.8 The Crawford Small Parts Dexterity Test

As you can see, Table 3.7 is separated into two sections. The upper section includes tests that have been frequently used for purposes of identifying signs of psychopathology—**screen out tests**. The tests listed in the lower section have been more frequently used to identify variations of normal personality—**screen in tests**. There is an important distinction between these two different categories of tests. Tests developed or intended to identify psychopathology, or used commonly for that purpose, are considered "Medical Tests" under the Americans with Disabilities Act (1990), particularly if the test is administered by a clinical or counseling psychologist or a psychiatrist. As such, they may not be administered until *after* an offer of employment has been made, as is the case with physical examinations, because emotional disorders are considered covered disabilities under the ADA. Applicants might be placed at a disadvantage in the selection process if their condition were revealed through pre-employment testing. On the other hand, tests developed or intended to assess normal personality may be administered as pre-employment tests and used for purposes of choosing among applicants prior to an offer of employment. If an employer administers a test such as the MMPI-II in order to choose among applicants prior to an offer of employment, that practice can be challenged in court and the applicant will likely win that challenge.

SCREEN OUT TEST

Used to eliminate candidates who are clearly unsuitable for employment; tests of psychopathology are examples of screen out tests in the employment setting.

SCREEN IN TEST

Used to add information about the positive attributes of a candidate that might predict outstanding performance; tests of normal personality are examples of screen in tests in the employment setting.

There are many positions of public trust (e.g., public safety officers, nuclear power plant operators, air traffic controllers, commercial airline pilots) that warrant testing for possible psychopathology to guard against catastrophic pathological actions by the incumbent. But most job titles in industry do not directly involve the health and welfare of the public, and testing for personality abnormalities would be questionable in such jobs. Figure 3.9 presents some sample items from the Saville Consulting Wave test, which is frequently used to assess normal personality in job applicants.

Practical Issues Associated with Personality Measures

Up to this point, we have been dealing with the "science" of personality. But there are also practical questions that arise about the measurement of personality for making employment decisions. Hogan, Hogan, and Roberts (1996) addressed those larger practical questions, as summarized in Box 3.3.

Faking The final question in Box 3.3 brings up a controversial point about personality tests. Some tests, particularly some commercially available integrity tests, are very transparent. It is obvious how one should answer the test questions in order to appear to have high integrity. This is a bit different from a cognitive ability test where a candidate might fake "dumb" answers (which would probably mean that the person *is* dumb) but

TABLE 3.7 Some Commonly Used Personality Instruments

Minnesota Multiphasic Personality Inventory II (MMPI-II)
California Psychological Inventory (CPI)
Personality Research Form (PRF)
Edwards Personal Preference Schedule
Jackson Personality Inventory-Revised (JPI-R)
16 PF Select
NEO-PI
Hogan Personality Inventory
Saville Consulting Wave

You will see a range of responses varying from very strongly agree to very strongly disagree. Choose the response alternative that best describes how you feel. Some statements are about being good at something and others are about what you prefer, need, or are interested in. Read each statement carefully because there may be differences between what you are good at and what you may need. Try to answer the questions from a work perspective as much as possible.

	Very Strongly Disagree	Strongly Disagree	Disagree	Slightly Disagree	Unsure	Slightly Agree	Agree	Strongly Agree	Very Strongly Agree
a) I need to have a clear set of priorities			✓						
b) I am great at encouraging others								✓	

FIGURE 3.9 What the Saville Consulting Wave Looks Like
SOURCE: Saville Consulting.

cannot pretend to be "smarter" than he or she actually is. A candidate might bear the following "script" in mind when answering integrity test questions:

> I have never stolen anything since I was a young child, and even then, I don't think I ever stole anything. I do not have any friends who steal, or would even *think* of stealing anything. If they did, they could not be my friends anymore and I would tell the appropriate authorities that they had stolen something. I think that showing up for work late, not doing a complete job, leaving work early, and taking sick days when you are not sick is also stealing and I would not do any of those things or be friends with anyone who would. I would inform management if I ever found out that a co-worker was engaging in any of these behaviors.

This "script" is only partly facetious. It is amusing in its extremity, but it makes the point that it is possible to answer questions on a personality-like device in a way that gets the best result—that is, an offer of employment. But what about tests that are not so transparent? From a practical standpoint, there are actually three questions to answer: (1) How easy is it to fake personality tests? (2) How many people do it? and (3) How much does it matter whether people do or do not fake? Let's take these one at a time.

SELF-PRESENTATION
A person's public face or "game face."

How easy is it to fake personality tests? Not difficult. As Hogan et al. (1996) pointed out, some are easier to fake than others. But you can answer any personality test in a way that makes you look "good." The real question is whether doing that truly qualifies as "faking." From some perspectives, personality is all about **self-presentation**; it is your public

BOX 3.3 PERSONALITY TESTING FAQ

Q: There are many personality tests and scales available. How do you choose among them?
A: Use valid and reliable tests that cover at least the Five Factor Model dimensions.
Q: Why should you use a test that measures more than one aspect of personality when you are interested in only one?
A: Because behavior usually is a function of many different influences, not just one.
Q: What do personality tests measure?
A: A person's typical "style."
Q: Why use personality tests to make employment decisions?
A: Because most workers and managers use terms like "being a team player," "remaining calm under pressure," "being persistent," and "taking initiative" as critical for success in almost any job.
Q: Do personality tests predict job performance?
A: Yes.
Q: Do personality tests predict performance in all jobs?
A: Probably, but they are less predictive for jobs with little autonomy.
Q: Weren't personality tests developed to measure psychopathology and for use in clinical settings?
A: Many years ago, that was true. The tests available today are designed to assess normal personality.
Q: People's behavior changes constantly. Doesn't this invalidate personality tests?
A: By definition, personality is relatively stable over time and from one set of circumstances to another and continues to affect our lives in important ways. Even though behavior changes occasionally, stable aspects of personality are still effective predictors.
Q: Do personality measures discriminate against ethnic minorities, women, older individuals, and the disabled?
A: There is no evidence of discrimination against these groups in well-developed personality tests. People over 40 tend to receive more positive scores than those under 40. There are some differences between males and females (men have higher scores on emotional stability and women have higher scores on conscientiousness) but these are not significant enough to result in different hiring decisions.
Q: Do personality tests invade privacy?
A: Some appear to. Choose tests with the highest validity and reliability, and the fewest number of offensive-appearing questions.
Q: What is the best way to use personality measures for pre-employment screening?
A: In combination with measures of technical skills, experience, and the ability to learn.

SOURCE: Based on Hogan, Hogan, & Roberts (1996).

face, your "game face." So to the extent that the personality test is a paper-and-pencil form of self-presentation, it is not faking, nor is it distortion (Hogan et al., 1996; Mount & Barrick, 1995).

Some have suggested that the real issue is whether the test taker has the correct frame of reference (FOR) for taking the test. As an example, consider being asked to take a personality test and told to use one of three perspectives: at school, at work, or in general (Schmit, Ryan, Stierwalt, & Powell, 1995). Chances are that your personality at work differs from your personality in nonwork social settings. As a sales representative, for example, you could be outgoing but in nonwork settings might be more reserved because there are fewer demands for extraversion. In a study with customer service reps for an airline, Hunthausen, Truxillo, Bauer, and Hammer (2003) found that specifically instructing employees to adopt an "at work" FOR increased validity of personality test scores. If these findings are replicated in other studies, it may have an effect on both research and practice. In the research context, it may mean that many of the reported validities of personality tests for predicting performance may substantially underestimate those values because a FOR other than "at work" may have been adopted by the test takers. In practice, this finding suggests that when personality tests are administered in a selection context, the respondent should explicitly be told to adopt an "at work" FOR.

How many people fake personality measures? It is not clear what the prevalence of distortion is (Mount & Barrick, 1995) because the prevalence depends, as we have seen in the preceding paragraph, on how you define faking. The main line of evidence to suggest that faking may be occurring is that applicant groups often have significantly more positive scores on given personality measures than employed groups (Bass, 1957; Kirchner, Dunnette, & Mousely, 1960; Weekly, Ployhart, & Harold, 2004), and, not surprisingly, the tendency seems to be greater among American than non-American applicants (Sandal & Endresen, 2002). In addition, sophisticated statistical analyses of responses to personality questionnaires (Michaelis & Eysenck, 1971; Schmit & Ryan, 1993) show that there are different patterns of responses from applicants than from employees or students. Not surprisingly, some studies say the rate of faking is substantial, whereas others say it is minimal.

This brings us to a third question: How much does it matter? The answer is that it does not appear to matter much. In studies where participants were instructed to distort their responses to make themselves look good, the predictive validity of the personality measures remained the same (Hough et al., 1990). And if we return to the self-presentation view of personality, "distortion" could either increase or decrease the validity of the personality measures. If the job in question were a sales position, some have suggested that a desire to look "good" in the eyes of another might actually be a job-related attribute (Hogan et al., 1996). A meta-analysis (Viswesvaran, Ones, & Hough, 2001) seems to effectively rebut that hypothesis, at least for managers. There was essentially a zero correlation between a test taker's desire to look "good" and his or her supervisory ratings on interpersonal skills. On the other hand, if an individual is having a performance counseling discussion with a supervisor, a more realistic presentation of strengths and weaknesses by the individual would be more effective than trying to look good. The issue of faking is not "settled" yet (Mueller-Hanson, Heggestad, & Thornton, 2003), but there does seem to be some agreement that it is not a fatal flaw in personality testing (Hough, 1998; Hough & Ones, 2001; Ones & Viswesvaran, 1998; Salgado, Viswesvaran, & Ones, 2000; Viswesvaran & Ones, 1999; Weekly et al., 2004).

There is one additional cautionary note of some practical significance for test takers inclined to intentionally distort their responses. Most personality tests have a "lie" scale, which indicates if a person is trying to make himself or herself look "ideal" in some way. The test report for an individual will usually include a cautionary note indicating a lack of confidence in the resulting scores if the applicant scored too high on the lie scale. In addition, there is some research (Dwight & Donovan, 2003) that indicates that if an individual test taker is warned that (1) faking can be identified, and (2) faking will have negative consequences in terms of being selected for a position, the test taker will be less likely to fake.

Integrity Testing

Until recently, integrity testing meant honesty testing. Employers have always been concerned with dishonest employees. We will consider counterproductive employee behavior in depth in Chapters 4 and 10, but for now, note that employee theft can make the difference between profitability and failure for an organization. Employers are often vigorous in investigating incidents of employee dishonesty after the fact. Money or product is disappearing—who is taking it? But honesty and integrity tests were developed to *predict* who might act dishonestly in the future rather than who is actually responsible for a counterproductive act.

Although honesty and integrity tests have been around for more than 50 years (Ash, 1976), there has been more enthusiasm for them in the past 15 to 20 years for several reasons. The first reason is economic: More and more employers are concerned about the high cost of dishonest employees, and integrity tests are relatively inexpensive. In addition,

TABLE 3.8 Examples of Overt and Covert Integrity Test Items

Overt Items

There is nothing wrong with telling a lie if no one suffers any harm (True or False?)

How often have you arrived at work under the influence of alcohol?

Do your friends ever steal from their employers?

Covert or Personality-based Items

Do you like taking risks?

Would your friends describe you as impulsive?

Would you consider challenging an authority figure?

SOURCE: Spector (2000).

from the I-O perspective various meta-analyses have demonstrated the predictive power of such tests. Finally, the polygraph legislation passed in 1988 radically reduced the use of the polygraph for pre-employment honesty screening, making paper and pencil tests more attractive, particularly those shown to be valid for predicting important work behaviors such as theft and absence. In jobs where polygraphs are permitted, integrity tests are considerably cheaper than extensive background checks or polygraph tests.

There are two different types of integrity tests: overt and personality based. The **overt integrity test** asks questions directly about past honesty behavior (stealing, etc.) as well as attitudes toward various behaviors such as employee theft. The **personality based integrity test** measures honesty and integrity with less direct questions dealing with broader constructs such as conscientiousness, reliability, and social responsibility and awareness. Examples of both types of items are presented in Table 3.8.

OVERT INTEGRITY TEST

Asks questions directly about past honesty behavior (stealing, etc.) as well as attitudes toward various behaviors such as employee theft.

PERSONALITY BASED INTEGRITY TEST

Test that infers honesty and integrity from questions dealing with broad constructs such as conscientiousness, reliability, and social responsibility and awareness.

There have been many extensive and high-quality reviews of integrity test research, and these reviews have concluded that those who score poorly will be poorer employees for any number of different reasons. They may be more likely to lie or steal, be absent, or engage in other counterproductive behaviors (Murphy & Davidshofer, 2005; Ones, Viswesvaran, & Schmidt, 1993; Sackett & Wanek, 1996). In the abstract, this sounds promising, but in the concrete, there are some problems with integrity tests. Murphy and Davidshofer (2005) summarized these concerns as follows:

1. It is difficult to know exactly what any given test of integrity measures. For example, taking a long lunch hour may be considered an indication of "theft" (of time) on one test and not even mentioned in another test. Data from a study by Wanek, Sackett, and Ones (2003) suggest that there are four basic components to "integrity tests" in general, but all four do not appear in any one test. These components are antisocial behavior (e.g., driving violations, theft admissions), socialization (e.g., emotional stability, extraversion), positive outlook (e.g., safe behavior, acceptance of honesty norms), and orderliness/diligence.
2. Unlike ability or even personality tests, applicants are seldom informed of their scores or the results of an integrity test. This is particularly disturbing to a candidate who has been rejected for a position and can't find out why. Nor are applicants typically warned of the risks and consequences of even taking the test in the first place, raising an ethical issue of informed consent. However, any applicant who refused to take the

test would naturally be considered to have withdrawn his or her application for employment, so it is not clear what the practical value of informing applicants might be.

3. Often, integrity test scores are reported in a pass–fail or, more commonly, a recommended–not recommended format. As we will see in Chapter 6, the setting of pass–fail scores is very technical, and it is not clear that the test publishers take these technical issues into account. That raises the possibility of false negatives—the possibility that an individual would be erroneously rejected as a "risk."

Cascio (1998b) made an additional point about integrity as a concept. Many employers and test publishers treat honesty as a trait, much like intelligence. But it is much easier for a person to "go straight," by behaving more honestly and morally, than it is for a person with lower general mental ability to "go smart." Yet organizations treat an honesty or integrity score like a cognitive ability score: If a person gives honest answers to overt questions about past indiscretions, he or she may be rejected even though he or she may have reformed. Ironically, the only way for the reformed individual to pass the test might be to lie!

You will recall that we discussed the concept of integrity in the section on personality earlier in the chapter. The discussion was in the context of the "bandwidth" debate concerning the FFM of personality. Some argue for a "narrow" bandwidth (e.g., separate scores for separate dimensions such as conscientiousness or emotional stability), and others argue for a wider bandwidth, which would involve developing a complex test to assess a complex trait. Integrity is a perfect example of this debate. One might approach the measurement of integrity by using a "broad bandwidth instrument" such as an integrity test, or inferring integrity from the combination of scores on conscientiousness, agreeableness, and emotional stability. Although this debate is largely theoretical, it also has practical implications. If an employer wants to assess the integrity of an applicant, what is the best way to do so? On the one hand, there is the ease of administration of an instrument to get right at integrity—the dedicated integrity test—rather than combining scores from three different dimensions of a broader personality test, such as the NEO-PI. On the other hand, much more is known about the meaning of any of the FFM dimensions than the typical score on an integrity test. In addition, the information gathered using a traditional FFM instrument can be used for predicting many behaviors beyond honesty.

What, then, is the employer to do? A meta-analysis by Ones and Viswesvaran (2001) is very informative on this issue. They compared personality tests with integrity tests for predicting various work outcomes and behaviors. The results were compelling. Integrity tests did much better ($r = +.41$) than FFM personality tests at predicting overall job performance ($r = +.23$) but FFM-based tests did much better ($r = +.51$) than integrity tests ($r = +.32$) at predicting counterproductive work behaviors (e.g., theft, violence). Thus, it seems that if behaviors like theft or violence are central to the selection decision, the employer should use FFM-based tests. On the other hand, if overall work performance is more critical, then an integrity test might be a better choice. Optimally, the employer might cover all bases and use *both* types of tests.

Emotional Intelligence

EMOTIONAL INTELLIGENCE (EI)

A proposed kind of intelligence focused on people's awareness of their own and others' emotions.

As we saw earlier in the chapter, the concept of **emotional intelligence** (**EI**) has achieved some notoriety with the public, if not widespread acceptance by psychologists. As there is no general agreement on the definition of EI, there can be no agreement on how to measure it. Recall also from that earlier discussion that Davies et al. (1998) found little evidence for the reliability or validity of existing EI tests. That is not say that there are no instruments available that claim to measure EI. A score on a test of EI is often called an

TABLE 3.9 A Sample Item from the MSCEIT™

SECTION H

1. Ken and Andy have been good friends for over 10 years. Recently, however, Andy was promoted and became Ken's manager. Ken felt that the new promotion had changed Andy in that Andy had become very bossy to him. How effective would Ken be in maintaining a good relationship, if he chose to respond in each of the following ways?

Response 1: Ken tried to understand Andy's new role and tried to adjust to the changes in their interactions.

a. Very ineffective b. Somewhat ineffective c. Neutral
d. Somewhat effective e. Very effective

Response 2: Ken approached Andy and confronted him regarding the change in his behavior.

a. Very ineffective b. Somewhat ineffective c. Neutral
d. Somewhat effective e. Very effective

emotional intelligence quotient, or **EQ**, to parallel the notion of IQ. As an example, Multi-Health Systems Inc. (MHS) is marketing an array of products related to EI and EQ, including the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT™), a scale for measuring organizational emotional intelligence, a 360-degree measure of emotional intelligence, an emotional intelligence interview protocol, a youth version of the emotional intelligence test to be used with children between the ages of 7 and 18, and a series of books and videotapes intended to help people more fully develop their emotional intelligence. Table 3.9 presents a sample item from a popular test of EI. Conte (2005) presents an excellent review of the available measures as well as their strengths and weaknesses. Unfortunately, he finds more weaknesses than strengths. The best that might be said at this stage is that the advocates of EI have yet to make a convincing case for either the construct or its measurement (Murphy, 2006). Additionally, EI researchers seem to be ignoring a substantial body of historical research on social intelligence that is also discouraging (Landy, 2006).

EMOTIONAL INTELLIGENCE QUOTIENT (EQ)
Parallels the notion of intelligence quotient (IQ); a score on a test of emotional intelligence.

In 1966 Marvin Dunnette wrote "Fads, Fashions, and Folderol," a sobering piece about research, theory, and practice in I-O. Fads were defined as "practices and concepts characterized by capriciousness and intense but short-lived interest" (p. 343). As data accumulate, emotional intelligence may very well prove to be a useful addition to the testing toolbox, but to avoid the graveyard of the "fads," more concerted efforts of assessing emotional intelligence will be required.

Interests and Values

It has been traditional to assess vocational interests as a way of guiding vocational choice. The three most popular instruments used for this type of assessment have been the Strong Vocational Interest Blank (SVIB), which began its development in 1921; the Occupational Interest Survey, developed and revised by G. F. Kuder (Kuder & Diamond, 1979); and the Self-Directed Search test based on Holland's (1985, 1994) theory of vocational choice, which we presented earlier in the chapter. As we saw, recent research has begun to connect personality and occupational interest (Mount, Barrick, Scullen, & Rounds, 2005). Thus, current vocational interest tests may see less use because a personality test might provide

more detailed and specific information for both vocational counseling and selection purposes.

The logic of vocational interest testing is simple. A candidate's score is compared to various occupational norm groups. The norm group with the closest match in terms of expressed interests is assumed to be the occupation for which the applicant is best suited. Interest tests are seldom used for selecting applicants for particular jobs. They are used more frequently to assist individuals in choosing or changing occupations. It is assumed that the norm group is composed of individuals with interests that support satisfaction and basic levels of success in that occupation. Thus, interest tests might be thought of as a form of motivational assessment.

In earlier years, values were sometimes also assessed. One of the most popular instruments was the Allport–Vernon–Lindzey Study of Values (Allport, Vernon, & Lindzey, 1960). Traditionally, values inventories were very similar to interest inventories. Values were thought to be broader and more stable than interests (Murphy & Davidshofer, 2005). More recently, "values" have taken on a new meaning, closer to the concept of ethical or moral behavior. In the wake of the various accounting scandals (e.g., Enron, Arthur Andersen, WorldCom), there has been a renewed call for the assessment of the "values" of business leaders and MBA students who may become those leaders (Browning, 2002). In addition, as we saw in Chapter 1, at least one of the primary theories of national culture (i.e., the theory of Schwartz) is based on the Allport–Vernon–Lindzey Study of Values survey, so individual values may very well see renewed interest for theoretical as well as practical reasons.

INDIVIDUAL ASSESSMENT

By their design, most paper and pencil tests are intended to be administered to large groups of individuals. For some situations, however, only one candidate (or a very few) will be assessed on many different attributes. This type of assessment process is often described as **individual assessment**. To select a CEO for a Fortune 500 company, for example, an executive recruiting firm may be retained to create a short list of three to five candidates who will then undergo intensive assessment. This assessment often includes paper and pencil tests, but they are administered and scored individually and may be used for creating a profile of a candidate rather than comparing one candidate with another. Because the target populations are usually upper-level executives in an organization, individual assessment is sometimes referred to as executive or senior leader assessment (Howard, 2001). Although frequently used for selection, individual assessment can also be used to identify training needs, provide career counseling, or provide performance feedback to key organizational members. Because it is time intensive and requires skilled assessors, it is expensive and unlikely to be used for any other than key positions in the company.

INDIVIDUAL ASSESSMENT
Situation in which only one candidate (or a very few) is assessed on many different attributes.

Individual assessment is complex, involving a wide variety of content areas as well as a wide variety of assessment processes. The tools most frequently used include various interactive assessment tools rather than paper and pencil tests. A primary reason for this is that the nature of the position is usually so complex that no paper and pencil test would, by itself, provide sufficient information. Although more than one candidate may be undergoing assessment, each candidate is usually assessed in isolation from the others. This allows the organization to keep the identity of candidates a closely held secret for the protection of the reputation of both the company (should a chosen candidate reject an offer) and the candidate (should the organization ultimately reject a candidate).

The "typical" individual assessment is likely to include ability tests, personality tests, a personal history statement, and interviews. It may also include simulation exercises or work samples, and less frequently, a clinically based personality test such as the Rorschach Inkblot Test or the Thematic Apperception test (TAT). As we indicated earlier, there is not

much scientific support for the use of these clinically oriented tests, but they are still occasionally used.

Although we will not cover individual assessment beyond this description, Silzer and Jeanneret (1998) have provided rich detail on the typical process and content of individual assessment for the interested reader, and Highhouse (2002) has presented a history of individual assessment that incorporates a more critical evaluation of the role of individual assessment in I-O psychology.

INTERVIEWS

In one form or another, an interview plays a role in virtually every selection or promotion decision. This has been true for many decades; one of the first texts dealing with employment interviewing was written by Bingham and Moore in 1931. Over the years, there have been many fine texts (e.g., Webster, 1982) and reviews of the research on the interview (e.g., Guion, 1998; Landy, 1989; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Posthuma, Morgeson, & Campion, 2002; Salgado, Viswesvaran, & Ones, 2000).

Interview Content

Interview content is often dictated by the amount of structure in the interview. A structured interview consists of very specific questions asked of each candidate, often anchored in asking the interviewee to describe in specific and behavioral detail how he or she would respond to a hypothetical situation. This has been labeled the **situational interview**, a subcategory of the structured interview. In addition, **structured interviews** typically have tightly crafted scoring schemes with detailed outlines for the interviewer with respect to assigning ratings or scores based on interview performance. The situational interview can be contrasted with another form of structured interview known as the behavior description interview. The basic difference between them is a time orientation. The situational interview asks the applicant what he or she *would* do, whereas the behavior description interview asks the applicant what he or she *did* do in the past. Recent research seems to favor the behavior description format (Taylor & Small, 2002), particularly when the interviews are being used to fill very high level executive positions (Huffcutt, Weekly, Wiesner, & DeGroot, 2001). Huffcutt, Conway, Roth, and Klehe (2004) also found that as the complexity of a job increases, the value of situational interviews decreases, although this is simply a finding and not an explanation of why this might be the case. Huffcutt, Weekly, Wiesner, and DeGroot speculate that the prevalence of the behavior description format at higher levels may be because the behavior description interview allows for a greater influence from verbal/presentation skills than the situational interview. Day and Carroll (2003) suggest another possible explanation: The behavior description interview assesses experience to a greater degree than abilities or personal characteristics. It is also likely that as one moves up the organizational (and complexity) ladder, experience trumps ability or personality.

SITUATIONAL INTERVIEW

Asks the interviewee to describe in specific and behavioral detail how he or she would respond to a hypothetical situation.

STRUCTURED INTERVIEW

Consists of very specific questions asked of each candidate; includes tightly crafted scoring schemes with detailed outlines for the interviewer with respect to assigning ratings or scores based on interview performance.

An **unstructured interview** has much broader questions that may vary by candidate and allow the candidate to answer in any form he or she may prefer. In addition, unstructured interviews are more likely to have less detailed scoring formats, allowing greater discretion by the interviewer for scoring. An example of structured interview questions is presented in Table 3.10. The questions were developed to elicit behavioral skills from candidates for 911 emergency dispatcher positions.

UNSTRUCTURED INTERVIEW

Includes questions that may vary by candidate and that allow the candidate to answer in any form he or she may prefer.

For the most part, interviews cover one or more of the following content areas: job knowledge, abilities, skills, personality, and person–organization fit (Huffcutt, Conway, Roth, & Stone, 2001). Huffcutt et al. found that the most frequently assessed constructs in interviews were personality and applied social skills, followed by cognitive ability, job knowledge, and skills. Salgado and Moscoso (2001) provided more detail on content. In a

TABLE 3.10 **Examples of Structured Interview Questions and the Real-Life Incidents that Are the Foundation for These Questions**

These questions were used to interview applicants for emergency telephone operator positions.

INTERVIEW QUESTION	CRITICAL INCIDENT
1. Imagine that you tried to help a stranger, for example, with traffic directions or to get up after a fall and that person blamed you for their misfortune or yelled at you. How would you respond?	1. Telephone operator tries to verify address information for an ambulance call. The caller yells at them for being stupid and slow. The operator quietly assures the caller an ambulance is on the way and that she is merely reaffirming the address.
2. Suppose a friend calls you and is extremely upset. Apparently, her child has been injured. She begins to tell you, in a hysterical manner, all about her difficulty in getting baby-sitters, what the child is wearing, what words the child can speak, and so on. What would you do?	2. A caller is hysterical because her infant is dead. She yells incoherently about the incident. The operator talks in a clear calm voice and manages to secure the woman's address, dispatches the call, and then tries to secure more information about the child's status.
3. How would you react if you were a salesclerk, waitress, or gas station attendant, and one of your customers talked back to you, indicated you should have known something you did not, or told you that you were not waiting on them fast enough?	3. A clearly angry caller calls for the third time in an hour complaining about the 911 service because no one has arrived to investigate a busted water pipe. The operator tells the caller to go to _____ and hangs up.

SOURCE: Schneider & Schmitt (1986).

meta-analysis of the employment interview, they found interesting content differences between conventional interviews and tightly structured behavioral interviews. They discovered that the less structured or conventional interview seems to be more closely associated with personality and social/communication skills. On the other hand, the tightly structured behavioral interview is more closely associated with job knowledge and technical attributes, and, to a much lesser extent, personality characteristics. Similar results have been reported by Huffcutt et al. (2001).

These results take on more meaning when considered in the context of reviews of the validity of the interview. It has been generally found (McDaniel et al., 1994) that the highest validity coefficients are associated with structured and behavioral interviews (often in the range of +.60) compared to the more personality-based interviews, which have validity coefficients more often in the range of +.30. These results would seem to be a strong recommendation for tightly structured interviews based on task-based job demands over interviews intended to assess personality characteristics or personal style. But a note of caution should be sounded here. Many of the studies on which these meta-analyses were based were conducted in an earlier time, before the emergence of team environments and client-centered work. As a result, many of the criteria used in the validation studies were

task based. It is not surprising, then, that lower validity coefficients would be observed for interviews centered on personality characteristics. These "personality based" interviews were also done in a time when few sound personality tests were available. Schmidt and Zimmerman (2004) present some intriguing findings that seem to demonstrate that when three or four independent unstructured interviews are combined, the validity for that unstructured combination is as high as the validity for a structured interview conducted by a single individual. This is good news and bad news. The good news is that certain administrative steps can be taken to increase the validity of unstructured interviews. The bad news is that it might be necessary to conduct three or four independent interviews to accomplish that increase, thus increasing the time and money the interview process requires.

In the context of the current state of the field, it might be reasonable to use psychometric devices (e.g., the NEO-PI, the Hogan Personality Inventory, or the Saville Consulting Wave) to assess personality attributes and the structured behavioral interview to assess knowledge and skills. Guion (1998) concluded that the structured interview is a valuable tool in the assessment toolbag. We agree.

Paradoxically, however, it appears as if managers may not agree. They tend to prefer unstructured to structured interviews (van der Zee, Bakker, & Bakker, 2002). For example, the interviews President Bush conducted in 2005 for his Supreme Court Justice nominees reportedly took place in the residence quarters, with dogs underfoot and with little apparent structure, concentrating more on the personal backgrounds of the candidates (Bumiller, 2005). Lievens and De Paepe (2004) have shed some light on this paradox. It appears that managers avoid structure because they feel that it makes the process too impersonal; they want more control over the interview questions and process. Lievens and De Paepe also found that those managers with formal training in interviewing (e.g., through workshops) were much more likely to impose more structure on the format.

Interview Process

Independent of the actual content of the interview, there are many relevant process issues. How should interviews be conducted? How should interviewers be trained? What are some potential sources of bias in interviews? Table 3.11 presents information on many of these practical issues. Studies (e.g., Huffcutt & Roth, 1998; Latham & Skarlicki, 1996; Sacco, Scheu, Ryan, & Schmitt, 2003) appear to confirm, at least on a preliminary basis, that little adverse impact is associated with the structured interview, particularly when compared with more traditional paper and pencil tests of cognitive ability. Nevertheless, these studies have examined traditional domestic demographic characteristics such as race and gender. As applicant populations become increasing multicultural, the issues of bias in the interview may re-emerge due to the more dramatic cultural differences that may appear in applicant responses. For example, many Asian cultures value modesty in self-presentation. Thus, Asian applicants may be less comfortable than American applicants in extolling their virtues when asked by an interviewer to describe strengths and weaknesses (a common question in unstructured interviews).

Interviewees may have a completely different set of anxieties and priorities when it comes to the interview. McCarthy and Goffin (2004) have identified five aspects of the interview that may be associated with the "weak knees and sweaty palms" of the applicant. These anxieties revolve around communication, appearance, social skills, interview performance, and behavioral control (i.e., observable tension in the applicant). Anyone who has been an applicant has experienced one or more of these "anxieties." But after years of experience, the first author of this text can assure you that exactly the same anxieties afflict the untrained or inexperienced *interviewer*.

TABLE 3.11 Potential Influences on Employment Interviews

Nature of the Information: negative versus positive
Placement of Information: early or late in the interview
Presence of Interviewer Stereotypes (e.g. Ideal Candidate)
Interviewer Knowledge of the Job in Question
Method used by Interviewer to Combine Information
Nonverbal Behavior of Candidate: posture, gestures
Attitudinal or Racial Similarity of Candidate and Interviewer
Gender Similarity of Candidate and Interviewer
Quality of Competing Candidates
Interviewer Experience
Applicant Physical Appearance
Attention to Factual Detail by Interviewer
Extent to Which Interview Is Structured
Note Taking by Interviewer
Use of Same Interviewer(s) for All Candidates

SOURCES: Based on Landy (1989); Huffcutt & Woehr (1999).

ASSESSMENT CENTERS

ASSESSMENT CENTER

Collection of procedures for evaluation that is administered to groups of individuals; assessments are typically done by multiple assessors.

Even though the word "center" evokes an image of a physical place, **assessment centers** are collections of procedures for evaluation, no matter where these procedures are carried out. Assessment centers are very much like the individual assessment procedure we described earlier, except they are administered to groups of individuals rather than single individuals, and the assessments are typically done by multiple assessors rather than a single assessor. Assessment centers have a long and successful history, and there are many good books and articles describing variations on the technique (Bray, Campbell, & Grant, 1974; Finkle, 1976; Guion, 1998; Spychalski, Quinones, Gaugler, & Pohley, 1997; Thornton & Byham, 1982). In earlier years, there were as many variations of assessment centers as there were users. For this reason, a task force published *Guidelines and Ethical Considerations for Assessment Center Operations* (Task Force on Assessment Center Guidelines, 1989). These guidelines have done much to standardize the assessment center process and protect the rights of those being assessed.

Most assessment centers share the following characteristics (Finkle, 1976):

1. Assessment is done in groups. A typical group size is 12, although smaller subgroups may be formed for specific exercises. The group format provides opportunity for peer evaluation.
2. Assessment is done by groups. Unlike the usual evaluators in individual assessment, assessment center evaluators are usually managers chosen from the organization but unfamiliar with the candidates.
3. Multiple methods of assessment are employed. Like individual assessment, these might include paper and pencil tests, group exercises, interviews, and clinical testing. A typical group exercise might be a leaderless group discussion that is observed and rated by the assessors. An individual exercise might be an in-basket exercise in which a can-

didate is presented with the contents of a typical in-basket and asked to deal with each element in the basket by making a phone call, sending an e-mail, writing a memo, or starting a file for information.

4. Assessment centers invariably have a "feel" of relevance to them, both for assessors and for those being assessed. They are seen as much more "real" than interviews, paper and pencil tests, or even isolated work simulations.

As in the individual assessment procedure, the results of the assessment center may include a report, recommendation, and feedback to the participants. An excerpt from a typical report appears in Table 3.12. On the basis of assessment center results, the organization may make one or more of the following decisions (Finkle, 1976):

1. An assessee may or may not qualify for a given job or job level.
2. Assessees may be ranked on a series of attributes and placed into different categories representing anticipated speed of promotion (e.g., fast track versus normal progression groups).
3. Predictions of long-range potential may be made for one or more of the assessees.
4. Development and learning experiences for aiding the assessee in personal or professional growth might be recommended.

There is general agreement that assessment centers can be valuable procedures for selection, promotion, and training needs analysis (Arthur, Day, Mcnelly, & Edens, 2003; Bartram, 2002; Hunter & Hunter, 1984; Schmitt, Gooding, Noe, & Kirch, 1984). There is less agreement with respect to why they work (Klimoski & Brickner, 1987; Lance, Newbolt, Gatewood, Foster, French, & Smith, 2000; Landy, 1989; Sackett & Tuzinski, 2001). Although the "why" question may be an interesting one for scientific and research purposes, it is less important from a practical perspective. Assessment centers include many different types of exercises and assess many different attributes. The information is eventually combined to yield a decision or recommendation that will be as good or as poor as the information that went into it.

Decomposing the assessment center into its constituent elements and asking which part makes the greatest contribution is like decomposing a bouillabaisse and asking which ingredient made it taste so good. Nevertheless, I-O researchers cannot resist the tempta-

TABLE 3.12 Portion of a Report Based on Assessment Center Evaluation

There were several indications from his behavior that his strong desire to make a favorable impression promoted above average tenseness in the assessment situation. On several occasions, his behavior was characterized by nervousness and controlled quietness, as though he were reluctant to enter into a situation until he felt absolutely sure of himself.

The picture he created was that of a young man eager to cooperate, comply, and do his best in order to fulfill the expectations others had for him.

In most respects, the trainee's general abilities compare favorably with the total sample of men in the Management Progress study.

Most members of the staff anticipated a very successful career in the Bell System for the trainee. . . . There was a mild amount of disagreement concerning the speed with which he is likely to reach the district level of management. Everyone agreed that he presently displays the abilities and potential to perform effectively at the district level.

SOURCE: Bray, Campbell & Grant (1974).

BOX 3.4 THE APPRENTICE

Perhaps the most vivid popular version of the assessment center is the reality television show, *The Apprentice*. This show pits a number of candidates against each other to see who will become an "apprentice" for a year with Donald Trump. The candidates are assigned to teams, whose composition changes as one candidate is "fired" each week. Within the candidate pool, the situation is largely leaderless at the outset, although, as in more traditional assessment centers, a leader eventually emerges from the pack. In *The Apprentice*, the implicit leader of the assessor pack is easy to pick out by simply noting hair style (although efforts are made to identify multiple assessors). The candidates are asked to work in teams with (and often expected to work *against*) each other on practical tasks including design, manufacturing, marketing, distribution, and sales. Then there is the "board room" confrontation between and among assessors and assessees. This crucible often produces memorable moments. For example, in season two, one candidate who offered to give up his "immunity" from being fired is fired by "The Donald" for being "stupid, impulsive, and life threatening" (Stanley, 2004). The feel of relevance of the situation, at least to candidates and the viewing public, is apparent (hence, perhaps, the term "reality" show). As the number of candidates dwindles, it is clear that some assessees do not "qualify" (they can be identified as the ones getting into the taxi with luggage at the end of each episode), that the eventual "apprentice" has been identified as having long-term potential, and that many of the unsuccessful candidates view the experience as developmentally valuable.

tion to decompose. And the temptation seems to be yielding informative results. The rationale of the assessment center is to provide opportunities for candidates to display effective performance in some tightly constructed simulated environments. But it is appearing more likely that it is not the performance that strikes the assessors, but underlying abilities and personality characteristics illuminated by those simulated environments. Overall assessment ratings appear to be closely associated with assessee cognitive ability (Collins, Schmidt, et al., 2003; Hoeft & Schuler, 2001) and, to a substantial but lesser extent, to assessee personality characteristics, particularly extraversion and emotional stability (Collins et al., 2003; Hoeft & Schuler, 2001; Lievens, DeFruyt, & van Dam, 2001). These results would seem to indicate that the combination of a good cognitive ability test and personality test might do as well as, if not better than (and at considerably lesser expense), a full blown assessment center.

Still from *The Apprentice*, showing Donald Trump presiding over a group of would-be apprentices.

One intriguing finding from a German study (Schuler, Moser, & Funke, 1994) was that assessment center results are much more predictive when the assessors have known the candidates for more than two years than when they have known them for less than two years. This suggests that assessors may be considering much more than the results of the assessment exercises in making evaluations, most likely past observations of the candidate's performance. At the very least, this study brings into question the practice of choosing assessors who are unfamiliar with the candidates.

Assessment centers can be expensive and time consuming. They are likely to be of greatest value to large organizations that favor internal movement and promotions and invest heavily in the learning and develop-

ment of their members. In addition, candidates who are evaluated through assessment centers are often very enthusiastic about the process, and this enthusiasm likely translates into acceptance of feedback. This can be particularly important when the goal is employee development rather than employee selection. Nevertheless, many organizations can accomplish assessment more effectively with more traditional assessment procedures.

WORK SAMPLES AND SITUATIONAL TESTS

Work Sample Tests

WORK SAMPLE TEST
Assessment procedure that measures job skills by taking samples of behavior under realistic joblike conditions.

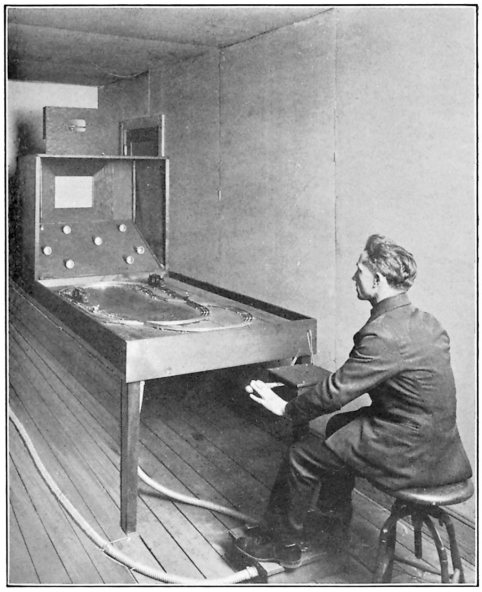
As the name implies, **work sample tests** measure job skills by taking samples of behavior under realistic joblike conditions. One of the earliest applications of this technique was in the selection of trolley car operators in Boston in 1910. Trolley cars frequently came into contact with people, horses, bicycles, and the newly introduced automobile. In 1913, Munsterberg set up a "work station" to simulate the controls of the trolley car and projected events onto a screen to see how potential operators would respond. Since this study was carried out a decade before the correlation coefficient became a common index of validity, Munsterberg's assertion that his workstation predicted operator success is anecdotal.

In today's work sample tests, the performance may or may not be assessed at an actual workstation, but the task assigned and the equipment used to complete the task are designed to be realistic simulations of the actual job. Consider the example of an individual applying for a position as an accounts payable clerk. The applicant might be given a checkbook in which to make entries, a work report from which to generate an invoice, a petty cash ledger to balance, and a payroll task. The results would then be compared against some standard and a score assigned representing the level of test performance. Table 3.13 illustrates some work sample elements.

Like assessment centers, work samples have a "real" feeling to them and usually elicit good reactions from candidates. Further, various studies have affirmed that work samples can be valid assessment devices (Hunter & Hunter, 1984). This is not surprising because work samples usually come directly from the tasks of the job in question, and it is easier to document their job relatedness. But like other formats, work samples are not intrinsically valid. Their job relatedness depends heavily on the attributes being assessed by the format. Using the example of the bookkeeping applicant, good performance may be the

TABLE 3.13 Some Examples of Work Sample Tests

MOTOR WORK SAMPLES	VERBAL WORK SAMPLES
Carving dexterity test for dental students	A test of common facts of law for law students
Blueprint reading test	Group discussion test for supervisor
Shorthand and stenography test	Judgment and decision-making test for administrators
Rudder control test for pilots	Speech interview for foreign student
Programming test for computer programmers	Test of basic information in chemistry
Map reading test for traffic control officers	Test of ability to follow oral directions



A 1920s example of work sample testing: an apparatus to test the skills of prospective trolley drivers.

result of specific knowledge (the candidate is familiar with the software), general knowledge (the candidate is familiar with check registers, invoicing, and so forth), or cognitive ability (the candidate is able to solve the problem presented by the task through trial and error). When work sample tests make unique contributions to test performance (e.g., above and beyond what might be predicted by a simple test of cognitive ability), it is likely due to general or specific knowledge. Callinan and Robertson (2000) suggested that work samples are best suited for predicting success in blue-collar jobs that involve skilled motor performance rather than jobs that deal with people. As Guion (1998) pointed out, the value of a work sample can be evaluated just as one would evaluate any assessment device: job relatedness, perceived fairness, and cost effectiveness. In Chapter 5, we will describe various techniques used to elicit knowledge from nuclear power plant operators, such as the "walk-through" method. This might also be considered an example of a work sample (Hedge, Teachout, & Laue, 1990).

Situational Judgment Tests

Recently, the notion of the work sample test has been expanded to cover white collar positions by creating what Motowidlo and colleagues (Motowidlo, Dunnette, & Carter, 1990; Motowidlo & Tippins, 1993) have referred to as low-fidelity simulations and others have referred to as **situational judgment tests** (**SJT**) (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). A situational judgment test is commonly a paper and pencil test that presents the candidate with a written scenario and then asks the candidate to choose the best response from a series of alternatives (see Figure 3.10).

SITUATIONAL JUDGMENT TEST

Commonly a paper and pencil test that presents the candidate with a written scenario and asks the candidate to choose the best response from a series of alternatives.

McDaniel et al. (2001) have reviewed the research on situational judgment tests and noted that in one form or another, such tests have been part of the assessment practice of I-O psychologists since the 1920s. In a meta-analysis of 102 validity coefficients, they concluded that there is substantial evidence of validity or job relatedness in these types of tests. They found that the single strongest component of these tests was general mental ability. Nevertheless, there appears to be more to SJTs than just "g."

Clevenger, Pereira, Weichmann, Schmitt, and Harvey (2001) evaluated the use of SJTs in hiring decisions for a government agency and a private sector transportation company. In addition to SJTs, they collected data on personality, cognitive ability, technical job knowledge, and job experience of the candidates. They found that SJTs were able to improve the prediction of performance even after the contributions of all of these other variables had been controlled, and even though the SJT scores were substantially correlated with the measure of cognitive ability. They suggested that SJTs are best used to measure procedural knowledge (what we referred to as tacit knowledge earlier in this chapter). In a more recent review, Chan and Schmitt (in press) have added adaptability (which we will cover in detail in Chapter 4) to their hypothesis about what is measured by SJTs. As you can see from Figure 3.11, it appears that various KSAOs produce competencies related to tacit knowledge and adaptability (which could also be labeled practical intelligence) and that these, in turn, produce positive and prevent negative job behavior. The relationship between KSAOs and practical intelligence helps explain why there are positive correlations between SJT scores, "g," and personality test scores. It also helps explain why SJTs predict performance beyond any one or combination of those attributes—namely because the attributes support the development of tacit knowledge and adaptability but are different from any of those supporting KSAOs. This model is supported by the research of

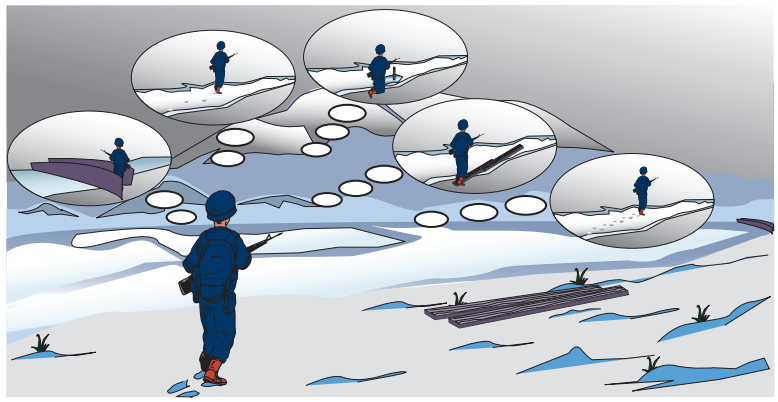


FIGURE 3.10 An Example of a Situational Judgment Exercise

A man on a very urgent mission during a battle finds he must cross a stream about 40 feet wide. A blizzard has been blowing and the stream has frozen over. However, because of the snow, he does not know how thick the ice is. He sees two planks about 10 feet long near the point where he wishes to cross. He also knows where there is a bridge about 2 miles downstream. Under the circumstances he should:

A. Walk to the bridge and cross it.
B. Run rapidly across on the ice.
C. Break a hole in the ice near the edge of the stream to see how deep the stream is.
D. Cross with the aid of the planks, pushing one ahead of the other and walking on them.
E. Creep slowly across the ice.

SOURCE: Northrup (1989).

McDaniel and Nguyen (2001), which shows an increase in SJT scores with increasing years of experience. It is plausible that tacit knowledge and adaptability increase with experience.

Another advantage of SJTs discovered in the study by Clevenger and colleagues was that the differences in scores between whites and both African Americans and Hispanics were considerably less than typically found in standard tests of cognitive ability. This may be a case of having your cake and eating it too. Not only did the SJT give a good assessment of general mental ability with lower adverse impact, it also measured

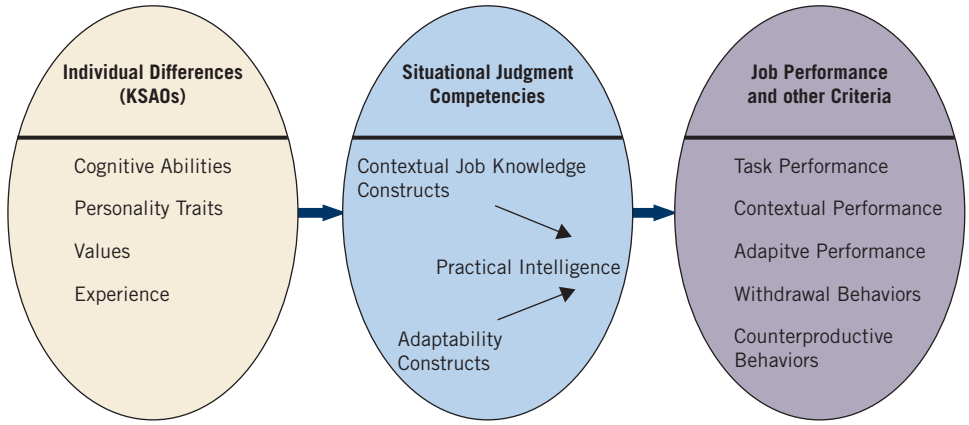


FIGURE 3.11 Framework for Relating the Multidimensional Nature of SJT to KSAOs and Job Performance

SOURCE: Chan & Schmitt (in press).

something in addition to "g" that was job related. This "something" was most likely practical intelligence as described above. In a follow-up study, Chan and Schmitt (2002) found once again that SJT scores contributed to the prediction of job performance for 160 civil service employees, beyond what could be predicted from cognitive ability, personality, and job experience. This study is particularly interesting because it was done in Singapore, suggesting that at least the format of the SJT can travel internationally.

SJTs have also been adapted for video presentation by using video vignettes, rather than a written description, to present the scenario. The results are encouraging. In two similar studies, Weekly and Jones (1997) and Chan and Schmitt (1997) found that black–white differences in SJT scores were smaller with a video than with a paper and pencil presentation, and that SJT produced more favorable attitudes toward the assessment process, particularly among African-American test takers.

The results of research on the SJT are very positive. They seem to possess three important characteristics for modern and practical assessment: They are job-related, they are well accepted by test takers, and they have reduced adverse impact compared to other traditional assessment devices. Further, the recent research on video presentations suggests that further advances are likely to occur in this area, particularly in terms of increasing the fidelity of the simulation from low to high and in further increasing the acceptance of the format by test takers.

MODULE 3.4 SUMMARY

- A vigorous debate continues over whether there is only one overarching cognitive ability—"g" or "general mental ability"—or several distinct facets or abilities. Psychologists have developed tests that produce a single number intended to represent cognitive ability, tests of specific abilities, and test batteries designed to measure several different facets of cognitive ability.
- Because most physically demanding jobs require combinations of physical abilities, many physical ability assessment procedures use simulated pieces of work (e.g., carrying a load up a ladder) rather than individual physical tests (e.g., sit-ups or bench presses). There is substantial evidence that measures of physical abilities can improve the prediction of job success for many physically demanding jobs.
- Personality testing in employment has shifted from a screen *out* process to a screen *in* process whereby employers seek to identify applicants with positive personality characteristics (e.g., conscientiousness, emotional stability, or agreeableness). There are many commercially available instruments for measuring personality characteristics, many based on the Big 5 model.
- Hogan, Hogan, and Roberts addressed practical questions about using the measurement of personality for making employment decisions.
- Practical considerations in personality testing include the use of integrity tests, on which "faking" is sometimes an issue, emotional intelligence tests, and tests of interests and values.
- It is important for employers and applicants to distinguish between the content of testing (*what* attribute is being assessed) and the process of testing (*how* it is being assessed). For example, the terms "personality" and "cognitive" describe the content of the assessment, and the terms "interview" and "background check" describe the process of the assessment.
- Individual assessment is complex, involving a wide variety of content areas and assessment processes. The tools used most frequently include various interactive assessment tools rather than paper and pencil tests, as the nature of the position is usually so complex that no paper and pencil test would, by itself, provide sufficient information.
- An interview plays a role in virtually every selection or promotion decision. Interviews vary in their structure and content. They can range on a continuum from very unstructured to very structured, and can cover one or more of the following content areas: job knowledge, abilities, skills, personality, and person–organization fit.
- Assessment centers have a long and successful history. They are administered to groups of

individuals rather than single individuals, and the assessments are typically performed by multiple assessors. There is general agreement that an assessment center can be a valuable procedure for selection, promotion, and training needs analysis.

- Other common assessment devices include work samples and situational judgment tests.

KEY TERMS

cognitive ability test
cognitive test battery
knowledge test
psychomotor abilities
screen out test
screen in test
self-presentation
overt integrity test
personality-based integrity test
emotional intelligence (EI)
emotional intelligence quotient (EQ)
individual assessment
situational interview
structured interview
unstructured interview
assessment center
work sample test
situational judgment test

MODULE 3.5 Special Topics in Assessment

INCREMENTAL VALIDITY

In the preceding modules, we have described quite a few tools that might go into the assessment toolbag. Until recently, assessment research often took on the flavor of a competition—which tool was better, a paper and pencil test of "g" or an interview? One study reported that "the validity" of a test of general mental ability was +.35, whereas another study reported that "the validity" of an interview was +.46, suggesting somehow that an interview is a more valid assessment device. Similarly, one might explore the differences in validity between a personality test and an interest test, or a work sample and a paper and pencil test. These are misleading questions for a number of reasons. First, we cannot answer these questions without answering another question: Better for what? Predicting satisfaction, or performance, or tenure, or management potential? Another reason why the questions are misleading is their implication that one is forced to choose a single instrument rather than developing a battery of assessment devices. Finally, the questions were misleading because they mixed test content with test process (e.g., test of "g" versus interview).

In the past few years, dozens of studies have purported to demonstrate the value of one or another device or test. Many of these studies compared the device of interest to another device. In addition, studies examined the predictive validity of particular combinations to demonstrate the added, or incremental, value of combining two devices. Thus, a study might show that the validity of a paper and pencil test of general mental ability was found to be +.35 but when it was combined with an interview, the validity of the two measures combined was +.51. Thus, one might conclude that the value of the interview is incremental; that is, it added to the validity of the paper and pencil test. Examples of **incremental validity** studies include:

INCREMENTAL VALIDITY

The value in terms of increased validity of adding a particular predictor to an existing selection system.

- Personality measures and biographical data (McManus & Kelly, 1999).
- Biodata and general mental ability (Mount et al., 2000).
- Personality measures and assessment centers (Goffin, Rothstein, & Johnston, 1996).
- Cognitive ability, interviews, and biodata (Bobko, Roth, & Potosky, 1999).
- Personality measures and mental ability (Bing, Davison, Whanger, & Van Hook, 2004; Kanfer & Kantrowitz, 2002).
- Situational judgment and cognitive ability/personality/job experience (Chan & Schmitt, 2002; Weekly & Ployhart, 2005).
- Situational judgment and Cognitive measures (Lievens, Buyse, & Sackett, 2005).

These studies point to an important principle: In assessment the issue is not *which* tool to use, but what *combination* of tools to use for the greatest predictive ability at the lowest cost.

As we saw earlier in the chapter when we discussed individual differences, and as we will see in greater detail in the chapters covering performance theory and prediction, industrial behavior is very complicated. It involves technical tasks as well as social ones. Successful performance in virtually any job depends on many different KSAOs. As a result, it makes little sense to limit the toolbag to one and only one tool. As Maslow said many years ago (1971), when the only tool in your bag is a hammer, you tend to treat everything as if it were a nail. As we continue to gather information about the incremental validity of various combinations of assessment tools, we will be better able to make practical recommendations about the most fair and effective assessment programs, as well as what tests and procedures might act as substitutes for other tests and procedures.

BIOGRAPHICAL DATA

It is common for organizations to gather personal information from applicants for positions. The best example is the type of information collected on an application blank: information about previous jobs, education, and specialized training. This type of information can also be used to predict job performance; the collection of it can be thought of as a "test" if the method used to collect it is standardized, the scoring is objective, and the sample of behavior examined is reasonable. This type of information has been variously labeled personal history, life history, biographical information, or—the simplest label—**biodata**.

BIODATA

Information collected on an application blank or in a standardized test that includes questions about previous jobs, education, specialized training, and personal history; also known as biographical data.

In the 1950s and 1960s, biodata predictors were based less on theory than statistics. If a particular piece of information (e.g., educational accomplishment) could be shown to predict success, it was included in an application blank. If no relationship could be found, it was not included. William Owens pioneered what has been called the rational approach to the use of life history data for the prediction of success (Mumford & Owens, 1982; Mumford, Snell, & Reiter-Palmon, 1994; Mumford & Stokes, 1991). Instead of simply looking at statistical relationships between individual history information items and success, Owens identified broader life history factors as a way of arranging all of the hundreds of pieces of information that could be gathered about someone. The underlying model for this type of biodata instrument is the **ecology model** (Mumford, Uhlman, & Kilcullen, 1992). In its simplest form, this model proposes that the events that make up a person's history are neither accidental nor random. They represent choices made by the individual to interact with his or her environment. As a result, these choices can signal abilities, interests, and personality characteristics. Thus, personal history data can be used as a surrogate or as an addition to other assessment information. As Figure 3.12 shows, there are many potential influences on the situations and actions that individuals choose from all available situations and actions. These precursors of situational choice and actions are the focus of biodata instruments. One implication of the ecology model is that as individuals mature (i.e., enter different life stages such as college, first job, parenthood, career), the predictability of biodata items might change. Dean and Russell (2005) have demonstrated such an effect.

ECOLOGY MODEL

Underlying model for life history biodata instruments. Proposes that the events that make up a person's history represent choices made by the individual to interact with his or her environment. These choices can signal abilities, interests, and personality characteristics.

Mael (1991) has suggested some characteristics of biodata items that distinguish them from other types of assessment, such as personality tests. Guion (1998) has summarized these characteristics as follows:

1. *Historical*. The item refers to events that have already occurred or are occurring rather than future, hypothetical events. For example, "Do you intend to develop computer skills?" is not a biodata item, but "How many times did you access the Internet in the past week?" is.
2. *External*. The events are observable and may involve others. They are not events that occur solely in one's own head. This would exclude items of the "how did you

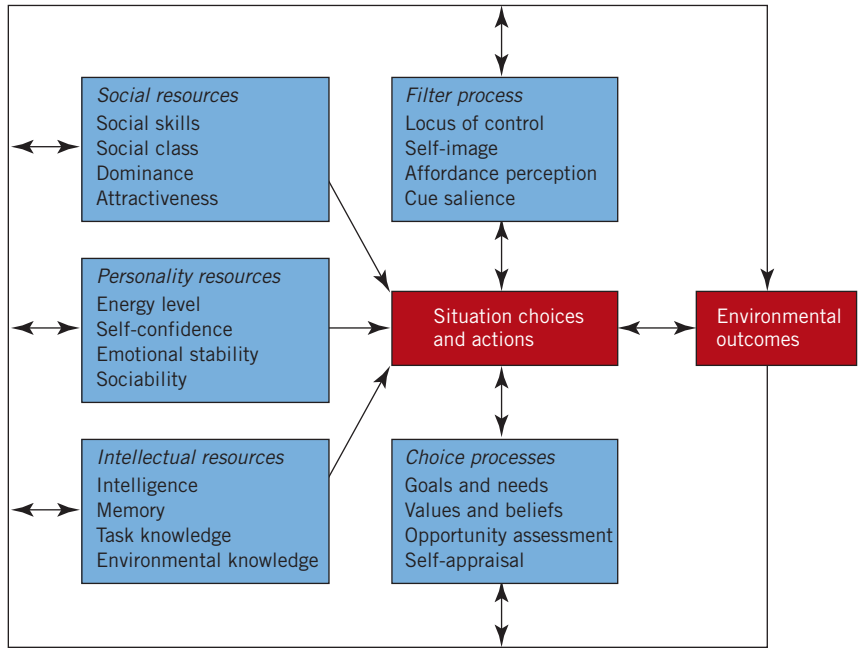


FIGURE 3.12

Construct Categories Drawn from an Ecology Model

SOURCE: Mumford & Stokes (1991).

feel . . ." variety. An example of an external item might be, "Have you ever been suspended from your job without pay for disciplinary reasons?"

3. *Objective*. The events are factual and do not involve interpretation. For example, "I think my last boss disliked me" would not qualify as a biodata item because it attributes an attitude to the supervisor that he or she may not have had. An objective item might be, "How many training courses have you taken outside of your company in the past five years?"
4. *Discrete*. The event described is concrete, with a beginning and/or ending, rather than open ended. An example might be, "When did you receive your permanent teaching certificate?"
5. *Control*. The event describes an action over which the applicant had control. An example might be, "How many times have you applied for a promotional opportunity in your company in the past five years?"
6. *Relevant and noninvasive*. The event should have at least the appearance of job relatedness and avoid overly personal aspects of a person's past. An example would be historical familial issues such as relations or income. "How many public presentations have you made in the past year outside of your company?"

Using these guidelines, Mael (1991) presented an illustration of acceptable and unacceptable biodata items (see Table 3.14).

In a later study, Mael, Connerly, and Morath (1996) identified four possible factors that determine whether a biodata item might be seen as invasive. Invasive items are those that might stigmatize someone, might remind the individual of traumatic events, were unduly intimate, or concerned religion.

Like other assessment instruments we have discussed, biodata instruments derive their job relatedness and value from the constructs they try to assess (e.g., cognitive ability, personality, experience, knowledge) and not from any magical properties of the format itself. Several studies have demonstrated that biodata items can improve prediction of success when added to other discrete assessment techniques such as the interview (Dalessio & Silverhart, 1994), personality tests (McManus & Kelly, 1999), and even general mental ability (Mount, Witt, & Barrick, 2000). But if one were to develop a comprehensive battery of

TABLE 3.14 Examples of Good and Poor Biodata Items

Historical	*Future or hypothetical*
How old were you when you got your first paying job?	What position do you think you will be holding in 10 years?
	What would you do if another person screamed at you in public?
External	*Internal*
Did you ever get fired from a job?	What is your attitude toward friends who smoke marijuana?
Objective	*Subjective*
How many hours did you study for your real estate license test?	Would you describe yourself as shy?
	How adventurous are you compared to your co-workers?
Firsthand	*Secondhand*
How punctual are you about coming to work?	How would your teachers describe your punctuality?
Discrete	*Summative*
At what age did you get your driver's license?	How many hours do you study during an average week?
Verifiable	*Nonverifiable*
What was your grade point average in college?	How many servings of fresh vegetables do you eat every day?
Were you ever suspended from your Little League team?	
Controllable	*Noncontrollable*
How many tries did it take you to pass the CPA exam?	How many brothers and sisters do you have?
Equal access	*Nonequal access*
Were you ever class president?	Were you captain of the football team?
Job relevant	*Not job relevant*
How many units of cereal did you sell during the last calendar year?	Are you proficient at crossword puzzles?
Noninvasive	*Invasive*
Were you on the tennis team in college?	How many young children do you have at home?

SOURCE: Mael (1991).

devices including cognitive ability, personality, interest inventories, and job knowledge, it is not yet clear how biodata information would add to what is assessed by those other techniques. There are also some lingering concerns about the fairness of certain types of biodata items. For example, Whitney and Schmitt (1997) found that compared with Whites, Blacks were more likely to describe past experiences that emphasized "the maintenance or restoration of long-standing traditions" and activities that were group rather than individually oriented. Even though the research was performed with college students, it certainly suggests that more research needs to be done in the area of ethnic and cultural

SOCIAL DESIRABILITY
Desire to be appealing to others.

influences on biodata instruments. There has been also some concern about whether biodata items are susceptible to faking, or at least **social desirability** influences in responding (Schmitt & Kunce, 2002). Schmitt, Oswald, Kim, Gillespie, Ramsay, and Yoo (2003) demonstrated that when the candidate was required to elaborate on various biodata responses, biodata scores did go down. But even though the scores changed, they seemed to change for everyone, since there was no reduction in the criterion-related validity of the biodata scores. Either people did not want to take the energy to elaborate, or everyone appears to claim experiences that they may not have had. The good news is that current research on biodata is considerably more theory based than it has been in the past (Mitchell, 1994). Once biodata instruments become more standardized and there is some agreement regarding what they measure most effectively, biodata may represent an attractive alternative to other available assessment devices.

GRADES AND LETTERS OF RECOMMENDATION

Employment applications, especially those for entry-level positions, often solicit high school and college grade point averages. In spite of their popularity, there is little careful research to support the use of grade point averages (GPAs) as a predictor, independent of measures of general mental ability, personality, or interests. In addition, substantial adverse impact against minority applicants appears to be associated with the GPA (Roth, BeVier, Switzer, & Schippmann, 1996; Roth & Bobko, 2000). There is some evidence that GPA and positive letters of recommendation can predict who will be *offered* a job, but not who will be successful in that job (Marshall, 1985).

Similarly, even though employers almost always ask job applicants for references or letters of recommendation, there has been little serious research on the validity or fairness of these devices (Loher, Hazer, Tsai, Tilton, & James, 1997). Since the threat of litigation from disgruntled rejected applicants looms large in the minds of many recommenders and reference sources, they are unlikely to provide negative information. In addition, without a great deal of imposed structure, and a clear understanding of the information base of the reference source, the information provided is often irrelevant, uninterpretable, or both.

Minimum Qualifications

Above, we indicated that the GPA is of largely unknown value as a predictor of job success and may have some vulnerabilities with respect to adverse impact. But that does not mean that education as a construct is of no value in a selection program. In public sector employment, there is a widely applied concept known as minimum qualifications (MQs). MQs usually involve a combination of education and experience and are used to make sure that individuals who wish to be considered as legitimate applicants have the prerequisite formal training and/or experience to assume a position without further training or experience. That is why they are called "minimum." That does not mean that the candidates will not be asked to complete additional assessment exercises such as tests or interviews. What it does mean is that applicants who lack the stated MQs will not be permitted to participate in subsequent assessment for the position, i.e., they will be rejected.

Typical MQ statements appear in Table 3.15. As you can see, MQs vary widely depending on the position in question. As you can also see, some MQs can be satisfied in several ways, which makes them less restrictive. Since MQs are selection devices in every sense of the word, they must be developed and validated just like any other selection device. The research on MQs is sparse (Levine May, Ulm, & Gordin, 1997), but a recent

TABLE 3.15 Exemplary MQ Statements

Communications Technician II

High school diploma/GED equivalency and 4 years of experience in the repair of two-way radios.

Graduation from an accredited technical/trade school in Electronics or a closely related field and 2 years of experience in the repair of two-way radios.

Civil Engineer–Construction Area

High school diploma/GED and 5 years of civil engineering experience in the construction area performing one or more of the following duties: project inspection, designing, drafting, or reviewing plans, contracts, and specifications, material calculations, and recordkeeping (checking contractor payrolls) and related duties.

High school diploma/GED and 8 years of civil engineering experience, 4 years of which must be in the construction area performing one or more of the following duties: project inspection, designing, drafting, or reviewing plans, contracts, and specifications, material calculations, and recordkeeping (checking contractor payrolls) and related duties.

High school diploma/GED and 8 years of experience as an engineering assistant in any work area.

Bachelor's degree in Civil Engineering or Civil Engineering Technology.

Senior Real Property Valuation Analyst

High school diploma/GED, completion of Appraisal Institute courses 110, 120, 310, and 510, and 7 years right of way specialist/real property valuation analyst experience with at least 3 of those years in real estate appraising.

Four year degree, completion of Appraisal Institute courses 110, 120, 310, and 510, and 3 years in real estate appraising.

demonstration of how MQs might be developed and validated provides the historical context, the methods that can be used, and the legal defensibility of MQs (Buster, Roth, & Bobko, 2005). Buster and colleagues suggest some principles for the development, use, and defense of MQs. These include:

1. Base them on a job analysis.
2. Direct them with a newly appointed job incumbent in mind.
3. Think about alternatives to formal course requirements that permit multiple avenues for meeting the MQ requirement.

Since MQs are used so frequently by public sector employers, particularly for promotional positions, they will continue to be challenged legally. As a result, particular care must be taken in their development.

Useless Assessment Practices: Graphology and the Polygraph

This section will be quite short. There are two practices, used largely by employers in countries other than the U.S. to assist in selection decisions, that are useless. We could devote several pages to demonstrating why they are useless, but we would rather devote those pages to practices that have some value. For this reason, we have placed the material showing that they are useless on the text website. The two practices that make our "useless" list

GRAPHOLOGY

Technique that presumes that traits can be assessed from various characteristics of a person's handwriting; also known as handwriting analysis.

POLYGRAPH

Machine that measures a person's physiological reactions. Approach assumes that when people are being dishonest, their physiological reactions will signal that they are being deceptive; often known as a "lie detector" test.

are polygraphs (electronic lie detection) and graphology (handwriting analysis). Although Gatewood and Feild (2001) estimate that 2,500 U.S. firms use graphology as a screening device, we are skeptical of that estimate. The polygraph is permitted only in certain job sectors, but is useless even in the job sectors for which it *is* permitted (e.g., screening for national security). As we describe on the text website, the research showing the uselessness of these practices is substantial and compelling. So let's move on.

DRUG AND ALCOHOL TESTING

There are several issues to address with respect to drug and alcohol testing in the workplace. The first is how acceptable the practice is to employees and prospective employees. As we will see in Chapter 11 when we consider fairness issues related to assessment, this type of screening is considered more acceptable by the public at large, as well as by employees, when the job in question involves the possible risk to the public (Murphy, Thornton, & Prue, 1991; Paranto, Truxillo, Bauer, & Leo, 2002).

The second issue relates to the legality of this type of screening. The courts have not yet finished deciding which practices impinge too greatly on an individual's right to privacy. Courts have upheld the right of railroads to test for the presence of drugs following an accident. In addition, it has been judged acceptable to test for drugs when screening applicants for drug enforcement posts with the federal government (Cascio, 1998a,b). In many areas of the private sector, drug screening is common for new hires. Harris (2000) reported that as many as two-thirds of large and medium companies screen new hires and as many as one-third of these companies screen current employees for drugs. With respect to alcohol use, current laws permit the random testing of individuals who work for commercial trucking companies, the aviation and rail industries, and mass transit; as well as nonrandom testing after an accident. Cascio (1998b) suggested several steps that an employer might take to enhance the defensibility and acceptability of a drug-testing program (see Table 3.16). In addition to the courts, employees and the public in general are very concerned about maintaining procedural justice when implementing a drug-testing program. In 1988 the federal government passed the Drug Free Workplace Act (DFWA) as a more affirmative approach to the problem of drugs in the workplace. As a result of DFWA,

TABLE 3.16 Ways to Enhance the Defensibility of a Drug Testing Program

To avoid legal challenge, companies should consider instituting the following commonsense procedures:

1. Inform all employees and job applicants, in writing, of the company's policy regarding drug use.
2. Include the policy, and the possibility of testing, in all employment contracts.
3. Present the program in a medical and safety context—that is, state that drug screening will help improve the health of employees and will also help ensure a safer workplace.
4. Check the testing laboratory's experience, its analytical methods, and the way it protects the security and identity of each sample.
5. If drug testing will be used with employees as well as job applicants, tell employees in advance that it will be a routine part of their employment.
6. If drug testing is done, it should be uniform—that is, it should apply to managers as well as nonmanagers.

SOURCE: Adapted from Cascio (1998b).

all federal contractors with contracts worth more than $25,000 are required to establish a drug-free policy. For nongovernment contractors, the DWFA allows for reduced worker compensation insurance premiums for employers, as well as access to an information network regarding drug testing (Gutman, 2000).

Extensive literature exists on the effects of alcohol and drugs on various aspects of physical and mental performance. As examples, it is well known that alcohol will slow reaction time, impair reasoning ability, induce drowsiness and clumsiness, and have a generally dulling effect on various senses. The same is true with various drugs, both illegally obtained and prescribed. Thus, from the performance perspective it seems clear that such substances will lead to lowered performance in a wide range of physical and mental tasks. But that is a "here and now" issue. Few would disagree that if an individual can be classified as intoxicated due to alcohol or drugs, he or she should not be permitted to engage in any work activity that might bring harm to the person, a co-worker, or the public. But the more intriguing question is whether a past history of use predicts future behavior. There is some evidence that drug use affects absenteeism and involuntary turnover. Normand, Salyards, and Mahoney (1990) reported that 5,500 applicants for postal positions were given drug tests. After 15 months, the new employees who had tested positive for drugs at the time of hire had an absenteeism rate almost 60 percent higher than those who had tested negative. In addition, almost 50 percent more of employees who had tested positive were fired during the 15 months than those who had tested negative (many for excessive absenteeism). Harris and Heft (1993) reviewed the drug-testing programs of large American companies and concluded that the outcomes most commonly predicted were absenteeism and involuntary terminations. But some have raised the issue of cause and effect. Galaif, Newcomb, and Carmona (2001) present evidence suggesting not only that drug problems lead to job instability, but also that job instability predicts later drug use. As we will see in subsequent chapters, both job instability *and* drug (and possibly alcohol) use may be tied together by a third variable—personality.

The problem of drugs or alcohol at the workplace must be kept in perspective. Most estimates suggest that fewer than 4 percent of applicants, and only 2 percent of incumbents, will test positive for drugs. Nevertheless, if one of those employees happens to be an airline pilot, the drug use poses a serious public safety problem. At 10:30 a.m. on July 1, 2002, both the pilot and the copilot of an America West flight were stopped before they were able to take off in a jet bound for Phoenix from Miami with 124 passengers on board. Each had a blood alcohol level that indicated impairment.

Given these concerns, why not simply test all employees all the time? Isn't that the safest policy? Perhaps it would be, if we could have perfect confidence in the accuracy of those tests. But large numbers of false positive indications (a person who fails a drug screen urinalysis who then tests negative using a different and more sophisticated protocol) have undermined confidence in the results of most mass administered drug-screening programs. As Harris (2000) suggested, these false positives come from sloppy procedures, inaccurate tests, or both. There are alternative types of drug tests, particularly those called immunoassay tests, that are considerably more reliable (Harris, 2000), but they are often prohibitively expensive for the employer. Unfortunately, the tests and testing procedures are outside the expertise of the I-O psychologist. What we can do is identify the performance areas most likely to be affected by the use of drugs or alcohol, and suggest occupations or jobs where such testing makes most sense.

COMPUTER-BASED AND INTERNET ASSESSMENT

Virtually every commercial test available in paper form is also available on the computer. Many are also available on the Internet, allowing for direct transmission and scoring of

the tests. This is a win–win situation for the employer and the applicant. It reduces time and effort for the applicant, and it permits the employer to process large amounts of data in very sophisticated ways in order to make selection decisions. It also cuts down on the time it takes to inform applicants of their status in the hiring sequence. As examples, via the Internet, an employer can administer a test world-wide in a matter of minutes; a candidate can receive a test score and an interpretive report within *seconds* of completing a test; a test can be revised in minutes at little cost; and scoring errors can be almost completely eliminated from the testing process (Naglieri, Drasgow, et al., 2004).

The variety of tests that can be presented on a computer platform is almost without limits. One might test for cognitive ability, personality, interests, and even psychomotor abilities. In addition, it is possible to present work samples, situational judgment tests, and very sophisticated and complex interactive cognitive tasks. The use of web cameras also permits some limited interpersonal testing, although there are still some elements of one-on-one interaction that are impossible to simulate by means of the computer. Aguinis, Henle, and Beaty (2001) provide fascinating examples of virtual-reality technologies for assessment. We have presented a sample of these in Table 3.17.

The topic of computer and web-based interaction brings up an often asked question regarding computer-based test presentation: Are we measuring the same thing as we are with a paper and pencil test or interview? The answer is, yes and no. For personality tests, Salgado and Moscoso (2003) present data showing not only that the psychometric properties of the tests in the two modalities are equivalent, but that test takers actually prefer the internet medium to the paper and pencil format. Ployhart, Weekly, Holtz, and Kemp (2003) present evidence suggesting that for some attributes (e.g., situational judgment, biodata, personality), web-based assessment produces *superior* psychometric results (e.g., reliability). For most measures of general mental ability and specific cognitive abilities, the answer is also yes, unless the tests are speed tests rather than power tests (Potosky & Bobko, 2004). Because it requires the test taker to be dexterous with the mouse or keyboard, speed adds a different dimension to the assessment (Mazzeo & Harvey, 1988; Mead & Drasgow, 1993). In addition, because web-based assessment has some built in system time, applicants often worry that they are being penalized on speeded tests for this "load time." Even though the system adds time to the testing period for this load time, and even though applicants are told this, they often forget that this is the case and fret over time "lost" (Potosky & Bobko, 2004). Finally, web-based testing often puts limits on test-taking "style." Many test takers prefer to go back and forth in a written test booklet completing items in different sequences, and the display screen seldom provides the amount of information that a test booklet page might. The screen might have room for one or a few items, whereas a test booklet might display a dozen or more items simultaneously (Potosky & Bobko, 2004).

In addition, the computer can be used to assess attributes that could never have been assessed by paper and pencil tests, such as reaction time and spatial and perceptual abilities (Murphy & Davidshofer, 2005). The following are some examples:

- Schmitt, Gilliland, Landis, and Devine (1993) described a computer-based system for assessing applicants for secretarial positions.
- We saw earlier that Ackerman and Cianciolo (1999, 2002) developed computer-based exercises for air traffic control positions.
- Olson-Buchanan, Drasgow, Moberg, Mead, Keenan, and Donovan (1998) developed a video-interactive test for assessing conflict resolution skills.
- Baron and Chaudry (1997) developed a computer-based interactive device for assessing customer relations skills.

There are many excellent reviews of the promise of computer-based testing, as well as its potential problems (Drasgow & Olson-Buchanan, 1999; McBride, 1998; Naglieri et al.,

TABLE 3.17 Illustrations of Potential Use of Virtual Reality to Assess Specific KSAOs *vis-à-vis* Traditional Techniques

KSAO/POSITION	TRADITIONAL SELECTION TECHNIQUE	VIRTUAL REALITY SELECTION TECHNIQUE
Job Knowledge/Civil Building Inspector	Paper-and-pencil knowledge test: License or Certification listed on résumé/ application	HMD: Move through virtual building and identify code violations
Skills		
Communication and Interpersonal/Manager	Assessment center in-basket	BOOM system: Interact with virtual subordinates
Handling of Hazardous Materials/Hazardous Materials Clean-up Team Member	Situational interview	HMD and data glove: Clean up hazardous materials at virtual accident scene
Abilities		
Cognitive: Visual-spatial/Mechanical Engineer	Paper-and-pencil ability test	Desktop system: Manipulate 3D job-relevant designs or models using 3D mouse
Psychomotor: Finger Dexterity/Manufacturing Position	Purdue Pegboard	Desktop system and data glove: Manipulate small virtual objects
Physical: Strength and Stamina/Firefighter	Simulation/obstacle course	CAVE system: Carry equipment in virtual fire scene
Perceptual: Vision and Depth Perception/Bus Driver	Driving record background check	Seated BOOM or vehicle-based system: Drive bus in a variety of traffic conditions
Other Characteristics		
Conscientiousness and Integrity/Office Administration	Paper-and-pencil conscientiousness or overt ability test	HMD and joystick: Perform tasks in a virtual office or play virtual reality game

SOURCE: Aguinis et al. (2001).

2004; Olson-Buchanan, 2001). At this point in development, the elegance and excitement of this medium is tempered by its costs. Well-developed and engaging computer (and particularly video) assessment exercises are extremely expensive, which means that they are out of the reach of the small to middle-sized organization. As an example, virtual reality testing environments can easily exceed $300,000 in cost (Aguinis, Henle, & Beaty, 2001). For large organizations that screen many applicants (e.g., federal and state agencies, large municipal and state police and fire departments, large manufacturing organizations), such a web-based or computerized testing format can be extremely powerful and cost-effective because, among other advantages, it does not require applicants to actually come to a central location for assessment. This is an exciting area for psychological assessment and substantially more data should be available in the next few years.

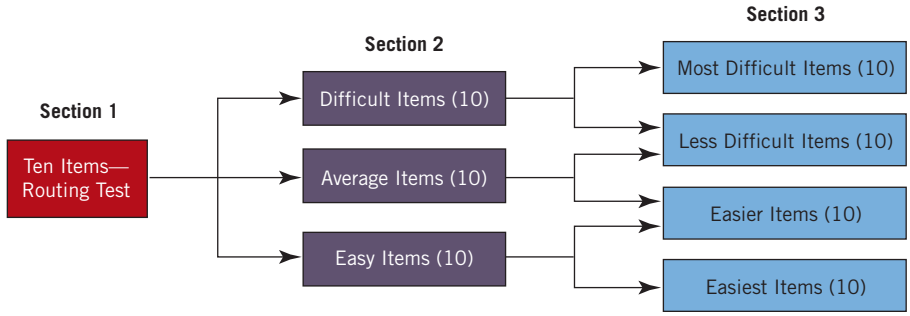


FIGURE 3.13 The Routing Test in Computer Adaptive Testing
In computer adaptive testing, test items are arranged in terms of their difficulty. Every candidate takes the same routing test, then one test from Section 2 and one from Section 3, based on his or her performance on the earlier section test.
SOURCE: Murphy & Davidshofer (2005).

Who Is a Candidate?

Until recently, there was some confusion and concern among employers regarding who was and was not an official "candidate" when Internet recruiting and job posting systems were used. The problem was that, depending on the definition, the employer might be charged with unfair discrimination based on the proportion of majority and minority applicants who were hired when recruited using Internet-based search engines to advertise openings. The Office of Federal Contract Compliance (OFCCP) clarified that issue in 2005. In order for an individual to be considered an applicant, four conditions must be met:

1. The individual submits an expression of interest in employment through Internet or related electronic data technologies.
2. The employer considers the individual for employment in a particular position.
3. The individual's expression of interest indicates that the individual possesses the basic objective qualifications for the position.
4. The individual at no point in the employer's selection process (prior to receiving an offer of employment from the employer) removes himself or herself from further consideration or otherwise indicates that he or she is no longer interested in the position (Business and Legal Reports, 2005b).

SIOP (Reynolds, 2004) considered these issues in some detail and generally approved the OFCCP's proposed definition. At this point, it is not clear that the OFCCP definition will be universally accepted by various federal agencies. EEOC has a competing, and less precise, definition in draft stage and it is not clear at this point whether the agencies will agree on a common definition. If they do not, employers are likely to be confused.

Computer Adaptive Testing

COMPUTER ADAPTIVE TESTING (CAT)
Presents a test taker with a few items that cover the range of difficulty of the test, identifies a test taker's approximate level of ability, and then asks only questions to further refine the test taker's position within that ability level.

ROUTING TEST
Preliminary test used in computer adaptive testing that identifies a test taker's approximate level of ability before providing additional questions to refine the test taker's position within that ability level.

An interesting innovation in computer testing is a "tailored" procedure known as **computer adaptive testing** (**CAT**) (Anastasi, 1982; Murphy & Davidshofer, 2005). In this procedure, a candidate does not need to answer every item on a test for adequate assessment. By presenting a candidate with a few items (e.g., ten) that cover the range of difficulty of the test, it is possible to identify a candidate's approximate level of ability and then ask only questions that will further refine the applicant's position within that category. The preliminary test, which every candidate takes, is called a **routing test**. The subsequent tests are the actual measurement tests (see Figure 3.13).

The potential advantages of CAT are obvious. Naglieri, Drasgow, and colleagues (2004) describe CAT systems for the assessment of musical aptitude, dermatology expertise, and conflict resolution skill. In addition, Drasgow (2004) provides illustrations of CAT for CPAs and architects. In addition to the power of stimulus presentation and manipulation, CAT can be done more quickly because each candidate takes fewer items than would appear on a paper and pencil test. There are some additional, less obvious, advantages (Anastasi, 1982). This type of test produces scores with equal or higher validity and reliability than conventional tests. In addition, CAT provides much finer discrimination among applicants at the high and low ends of the ability scale. The American Council on Education has published an informative set of guidelines related to the use of CAT in educational settings (Green et al., 1995), but these guidelines are also useful for industrial application. Because of the technical and empirical challenges of CAT, it is still out of the reach of most employers, but the armed services are currently using and refining such systems for recruit screening on the ASVAB (Murphy & Davidshofer, 2005). A CAT-based examination you may take if you consider pursuing a graduate career in I-O psychology is the Graduate Record Examination (GRE).

TESTING AND DEMOGRAPHIC DIFFERENCES

As indicated earlier in the chapter, the debate about the fairness of all types of assessment has raged for decades and is not likely to disappear in your lifetime. We will deal with the issues of justice and fairness from the psychological perspective in great detail in Chapter 11, and from the legal perspective in Chapter 6. Here, we will take a more descriptive approach and simply report statistics available regarding the extent to which majority and minority subgroups score differently, on the average, on various assessment devices and procedures. These data are excerpted from an excellent statistical and narrative review of these issues presented by Hough, Oswald, and Ployhart (2001).

General Mental Ability ("g") In comparisons of subgroups on "g," the following hierarchy seems to hold. Highest scores are obtained by East Asians, followed by whites, Hispanics, and finally African Americans. Older working adults score more poorly than their younger counterparts. There are no differences between men and women. Differences in reasoning ability follow the same pattern but are less pronounced as differences in "g."

Personality While whites and blacks score differently on measures of personality, those differences are minimal compared to "g," and it is not clear if these differences have any practical implications. Similarly, there are modest differences between men and women, and old and young, with the exception of a personality characteristic called "rugged individualism" in which women score dramatically lower than men. As this dimension reflects "masculinity" as we defined it when describing Hofstede's theory in Chapter 1, this finding should come as no surprise.

Physical Abilities As would be expected, men score substantially higher than women on muscular strength and stamina, moderately higher on coordination and balance, and moderately lower than women on tests of flexibility.

Mode of Presentation (Paper and Pencil versus Video) Whites obtain higher scores on paper and pencil tests than either African Americans or Hispanics; the difference is less pronounced on video tests. Women obtain higher scores than men on both paper and pencil and video-based tests.

Interviews and Biodata DeCorte and Lievens (2003) summarize earlier meta-analyses and estimate that compared to cognitive ability tests, majority–minority differences in structured interview scores and biodata scores are considerably less pronounced, although minority test takers do more poorly than their majority counterparts.

As you can see from this summary, minority and older test takers seem to be at a constant disadvantage in standardized assessment situations, more so in some forms of assessment than others. The exact magnitude of these differences is the subject of some debate (Bobko, Roth, & Buster, 2005; Roth, BeVier, Bobko, Switzer, & Tyler, 2001; Roth, Van Iddekinge, Eidson, Huffcutt, & Bobko, 2002), but the conclusion that the differences have practical importance and largely favor majority test takers is not.

By far the greatest differences between majority and minority test takers appear on paper and pencil tests of cognitive ability. What do I-O psychologists think about that? Murphy, Cronin, and Tam (2003) presented the results of responses of 700 I-O psychologists to questions related to this issue. This is what they found:

1. I-O psychologists generally accept cognitive ability tests as valid and fair.
2. They accept cognitive ability tests as good but incomplete estimates of intelligence.
3. They believe that a test of cognitive ability is not sufficient by itself to make accurate predictions of future job success.
4. They believe that different abilities will be more or less important for predicting success on various aspects of work behavior.
5. They believe that diversity in a workforce is good and that adverse impact should be avoided.
6. There is considerable disagreement among I-O psychologists regarding whether "g" is the *most* important human attribute in the work setting.

The data presented above highlight a dilemma: Our "g-ocentric" traditions appear to limit opportunities for minority applicants, but we value diversity and dislike adverse impact. The way off the horns of that dilemma seems to be the continued exploration of assessment practices that combine tests of cognitive ability (in both paper and pencil and more non-traditional forms) with tests of non-cognitive attributes. In addition, there are methods of combining information and making actual hiring decisions that can also be used to ameliorate the effect of cognitive ability tests on minority test takers. We will cover those methods in Chapter 6 on staffing decisions.

MODULE 3.5 SUMMARY

- An important issue in assessment is not *which* tool to use, but what *combination* of tools to use for the greatest predictive ability at the lowest cost. I-O psychologists have examined the predictive validity of particular combinations to demonstrate the added value of combining two or more assessment devices.
- Virtually every commercial test available in paper form is also available on the computer. The variety of tests that can be presented on a computer platform is almost without limits. At this point, however the elegance and excitement of the computer-based medium is tempered by its costs: Well-developed and engaging computer (and particularly video) assessment exercises are extremely expensive.

KEY TERMS

incremental validity
biodata
ecology model
social desirability
graphology
polygraph
computer adaptive testing (CAT)
routing test