

Chapter 12: Testing

Testing

Assumptions for a Test

Norms

Reliability

Validity

Objectivity and Efficiency

Ethics

Tests of Interests

Tests for Achievement and Aptitude

What is Intelligence?

Classic Tests of Intelligence

The Intelligence Quotient

Modern Tests of Intelligence

Definition and Measurement of Mental Retardation

What Impacts Intellectual Development?

USING PSYCHOLOGY: Can you Improve your IQ?

Intelligence and Creativity

Personality Tests and Errors

Interviews

Questionnaires and Self-Reports

Projective Tests

REVIEW

ACTIVITIES

INTERESTED IN MORE?

Testing

WHAT'S THE ANSWER?

In career counseling, what's wrong with just asking people what they'd like to be? Why wouldn't they know best what their own interests are?

"My grandfather was 63 last year. He got laid off from his job when his company lost a big contract. The dumbest thing happened to him. One place where he went for a job interview asked him to take an intelligence test before they'd hire him. He didn't get that job. He found out later that his intelligence rating wasn't high enough on that test to handle the job he wanted. Can you imagine that? I've lived with him all my life. He's raised three kids. He's been a construction foreman. He's done everything. The person he talked to said older people's IQ drops with age. The older they get, the

dumber they get is what I guess he meant. I don't believe it. Gramps is too smart for that." *What is the truth here? Will your intelligence grow, remain constant, or drop as you get older?*

Many studies of individual differences among humans have been conducted. The typical form of human test results is a normal distribution. A good test is based on standardized norms and is reliable and valid. Finally, good tests should be as objective and efficient as possible. They should yield the most useful information with the least time demands on the examiner and the person(s) taking the test.

Ethical principles guide the giving and interpreting of psychological tests. Ethical issues include the demonstrated value of the test itself, the fairness of the test for the person being tested, and the rights of privacy involved.

There are three types of tests most often used in career and vocational counseling. Interest tests help people identify potential careers. Career recommendations are based on a close similarity between the interests of the person being counseled and those of successful people already in various fields. Achievement tests are the most widely used, and possibly also the most widely abused, tests now available. Achievement tests stress content validity. You may not be able to separate aptitude tests from achievement tests on the basis of content, since they are quite similar. The main difference is in the intended use of the scores. Aptitude tests are used to project a person's potential success in future activities. Thus, they stress predictive validity.

Intelligence is usually defined in terms of the processes used to measure it. Some think intelligence is composed of a single general factor; others view it as clusters of abilities; still others think it is composed of many individual skills. Classic tests of intelligence tend to be based on the assumptions that intelligence changes with age and that as children grow up they are capable of doing more complex tasks. Our intelligence quotient is calculated from our measured mental age and our chronological age. Modern tests of intelligence, based on more recent conceptualizations of intelligence, place more emphasis cognitive processes and evolutionary pressures. All of the tests are designed to assess humans' relative strengths in the various cognitive and performance skills needed to survive in our everyday world. Heredity and environmental factors combine to affect our intelligence. From infancy onward, certain kinds of mental and physical activities lead toward increasing our intelligence. Even birth order and family size can make a difference. Family influence is instrumental as

well in providing the cradle for creativity. Creativity is difficult to measure.

Personality is measured most often in one of two ways: by direct observation and interviews, or by questionnaire/self-report tests, such as the MMPI or various projective tests. The projective tests (Rorschach, TAT) present the person taking the test with an ambiguous stimulus. In theory the person must inject his own perceptions into the situations to provide a response. Such projective tests can be subjective and thus not too valid.

Assumptions for a Test

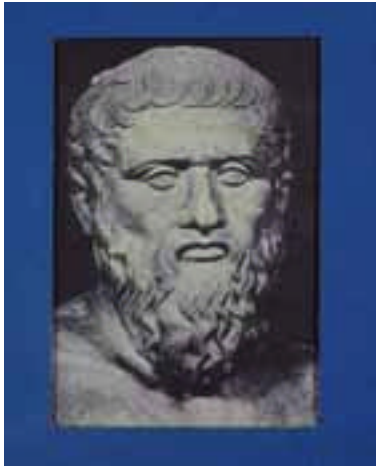
"No two persons are born exactly alike, but each differs from each in natural endowments, one being suited for one occupation and another for another."

This statement is one of the first times society recognized what is now widely accepted fact: there are individual differences among people -- differences most frequently assessed by various tests of interests, achievement and aptitude, intelligence, and personality. What is so interesting is that the statement was written some 2,400 years ago! It's contained in Plato's *Republic*. Plato identified three kinds of people, as seen in Table 12.1. He suggested that the way in which men performed certain actions that were requested of them indicated which ones would make good warriors. So we can credit Plato with developing what was really the first test to assess individual differences.

Table 12.1

FROM PLATO'S REPUBLIC: THREE KINDS OF INDIVIDUAL DIFFERENCES

PEOPLE	ROLE	CHARACTER
Workers	Produce material goods	Obedience, temperance
Warriors	Defend against attacks	Energy, courage
Philosopher-kings	Govern lower levels	Wisdom



Plato's brilliance extends beyond just the development of humankind's first test. In developing that test, he established a benchmark. Any member of the working class who desired to join the military could train so as to be able to achieve the required levels of performance. That is, experience -- or nurture -- could be used to modify human behavior. Moreover, members of

of the first two classes could ever become a member of the ruling class. You could gain membership in that class only by means of your heredity -- that is, your birth parents. Although he was not using hereditary influences as we do in modern times, still Plato identified hereditary limits on behavior. In short, with his Table, Plato anticipated the nature-nurture argument regarding whether experience (nurture) or heredity (nature) most influences our behavior.

The United States' Declaration of Independence says that "all men are created equal," but that refers to equal opportunity. It does not refer to inherited abilities. We all differ from one another. One branch of psychology deals solely with the study of individual differences among humans. However, we are simply going to assume that those differences exist. In fact, they exist in everything from our height and weight to our intelligence, our running speed, and the length of our arms. How do these individual differences show up? They show up almost every time we measure any human dimension or ability.

Individual differences may result from our hereditary background or from the environment in which we grow up and live. However, most of our differences result from the joint impact of both heredity and environment. Our age, our sex, our race, our mental skills and abilities, or even the prior training we've had may also cause such differences. This chapter focuses on how we measure these individual differences.



When human performance data are plotted on a graph, results often fall into what is called a normal distribution. Certain observations about that distribution are important here. In any range of scores, such as those from intelligence

tests, we find that most individual scores are grouped near the middle. Most of us are average or close to average in many abilities. The further we get from the midpoint of that distribution, the fewer individuals we find. In the matter of intelligence, for instance, much less than one percent of all humans have an IQ that is below 55 or above 145. So, knowing the existence of Individual differences clearly exist. How do we measure or test them?

A test is a systematic procedure for comparing the responses of two or more persons to the same series of stimuli. In other words, tests are used to present the same items to many people and to see how their performances compare. Such tests include a wide variety of stimuli and procedures. They may be true-false, multiple-choice, or open-ended questionnaires, or they may involve spoken or written responses to rather complex stimuli. What is considered a "normal" (typical) performance on such tests is almost always based on records of past performances accumulated over long periods of time.

A good test involves many, many features. All such tests share several features in common. The tests are based on norms, and such tests are valued if they exhibit high reliability and high validity. Good tests are usually objective and efficient. It is crucially important that tests be conducted (and the results be treated) in a manner consistent with nationally endorsed ethics for testing endorsed by the American Psychological Association and the Canadian Psychological Association.

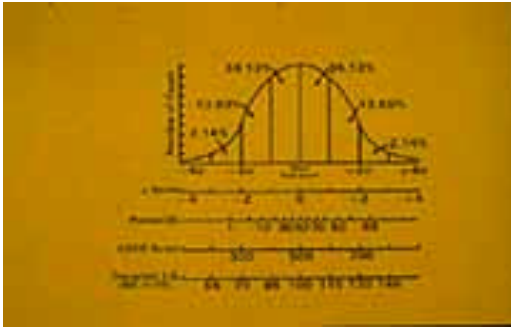


Norms

The key word in our definition of "test" is the word "compare." If I tell you, "Mikko has a test score of 200!", that's useless information by itself. It only

gains meaning as we learn how other people did on the same test. Such other scores will allow us to interpret Mikko's score as good, average, or poor.

By far the most widely found distribution of performance scores is the normal distribution, shown in graph form. This is found so frequently that statisticians and and psychologists have actually studied the distribution itself. As a result, we've learned many things about it. For instance, we know that the mean (or average) score in a normal distribution falls exactly at the mid-point. Half the scores are above this point; half are below it.



a student in your class, his test score of 200 would've been the best in class. More than 99 percent of the class would have scored lower than he did.

All of this is important any time we try to interpret one person's test score. To make the score meaningful, we must know how a large number of people perform on our test. These performances are used to establish "norms," which is the average performance of many people under specified conditions. So we establish norms by standardizing (specifying or operationalizing) the procedure under which a test is given. In your class tests, for instance, it wouldn't be fair for everyone else to have only an hour for the test if you are allowed two hours. The procedures would be different, and thus your score couldn't be interpreted in terms of the same graph of typical performance on the test. Standardizing such procedures increases a test's objectivity and efficiency.

In a nationally administered test, such as the Scholastic Assessment Test (SAT), the instructions are very precise. You will be allowed exactly the same amount of time to perform the test at your test-site as will students in Seattle, Houston, and New York, as well as Springfield, Ponca City, and Oliver Springs. The instructions are read to you, the timing is controlled, and everyone taking the test on that date is asked the same questions. Although not part of your instructions, the requirements for a proper test site by the Educational Testing Service even spell out proper light levels, temperature, and humidity! As a result, only individual differences in whatever abilities are being measured can influence the test score. This degree of specificity regarding the test situation increases the reliability and validity of the test results.

The test must be standardized by giving it to enough people that a distribution of responses is established. It's important that the people whose performance is used to standardize the test are similar to those who later take the test -- one of the basic elements of proper test ethics. Once the test scoring procedure has also been standardized, your test can be scored and recorded with everyone else's. Finally, your score can be

interpreted in terms of the test's distribution of scores as good, fair, or poor.

Reliability



Once we've established norms for our test, we're faced with another problem. We want to make sure our test is reliable. Reliability refers to the ability of a test to give consistent, stable, constant results each time it is given. It would be a useless test that

measured your intelligence as superior one time and at the lowest levels another. There are many ways to assure that a test is reliable. Most often the test is given, and then given again to the same -- or an identically matched second -- group. We expect that whoever scored at the top on the first test would also be at the top of the score distribution on the second test. The extent to which the scores of all who take the tests are identical is a measure of test-retest reliability. However, such an assessment technique has problems. If the test is boring the first time, it's likely to be even more boring the second time. If students remember the answers from the first time the test is administered, their score may be higher the second time, but the effects on the correlation of scores across tests is harder to predict. An individual's ability may change between tests -- perhaps because he or she has studied more. The motivation of the person taking the test may change from one administration to the next.

Other measures can also be used. Comparable-forms reliability involves presenting two different arrangements of the test items to two comparable groups of people. Problems with memory, boredom, and differing levels of motivation can be easily controlled in this way, but there is a problem demonstrating the comparability of the two groups. Still, we would expect that people who are well versed in the material being tested would do well on either form of the test. Again, the higher the correlation of the test scores, the more comparable are the forms. Instructors, for instance, who use different orders of the same set of test items on an in-class test, can assess comparable-forms reliability by randomly assigning students to take specific forms of the test, and then compare the class averages on each form of the test.

Finally, we can assess a test's split-half reliability. On any multi-item test, we should be able to split the items to compare a wide variety of different items against other items in a test. For instance, we could take a 50-item multiple-choice test and calculate each student's score on the odd-number items (1, 3, 5. . .) and compare those to their score on the even-numbered items (2, 4, 6. . .). Whoever scored best on the odd-numbered items should score best on the even-numbered items if we have a reliable test. That comparison also can be applied to the individual items in that we would expect that people who score high on the test as a whole are most likely to choose the correct answer for any given item. Occasionally a reverse-discriminator is created -- an item for which the best students consistently choose the wrong option. At worst, the lowest performers on the test may be most likely to select the answer that is coded as correct. Though rare, such items are typically discarded because they reduce the overall reliability of the test.

Reliability is one of the two most important features a test must exhibit. The other feature is validity. Additionally, a good test is one which yields accurate, replicable information obtained in an objective and efficient form, consistent with the nationally endorsed ethical principles concerning the testing process.

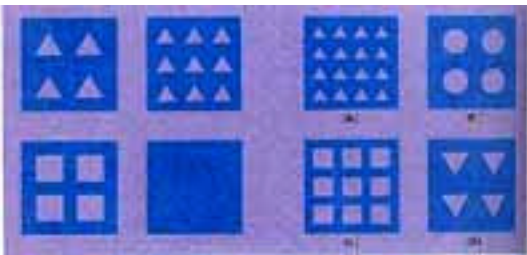
Validity

In addition to establishing the norms on which a test is based and assuring a test's reliability, another important factor in developing a good test is to assure that the test is valid. Validity refers to how accurately a test measures what it is supposed to measure. There are many different ways in which we can determine the validity of a test. All we do is to establish a criterion -- something we wish to predict. For instance, imagine you've developed a quickie-test of the willingness of members of the opposite sex to accept dates with you. You can measure the predictive validity of your test simply by giving it and then asking for a date. If the response to your request has been predicted by your test a high proportion of the time, then you have a valid test. It measures what it's supposed to measure, and it has predictive validity. Feature 12.1 discusses some of the problems we may have with validity. These problems have occurred in trying to develop culture-free tests that can be given without bias to anyone, regardless of race.

FEATURE 12.1**YO NO HABLO INGLÉS**

If you don't speak Spanish, then you had trouble with the title, which says, "I do not speak English." How would you like to take an intelligence test in Spanish to determine whether you were smart enough to stay in school? Unfair? Of course. But that's the problem faced by people who don't speak Standard English and who haven't been raised in our culture. It's not a simple problem. Most tests of intelligence and thinking skills depend on language. The instructions are almost always written in English, and a number of the tasks are word-games or involve working with words.

Now in some senses that's fair. Many of the jobs in the United States and Canada depend on someone with knowledge of our culture. Language is important to the performance of most jobs. Yet there are situations where not being able to speak English may cause a person to score low in tests measuring abilities where language use is not important. The Wechsler Scales of Intelligence—for adults, children, and preschool/primary age children is one of the most popular tests of intelligence. It is based partly on performance tasks involving little direct use of language.



It was an awareness of this problem caused by a lack of language or skills or cultural experience that led to the development of "culture-fair" tests. They are nonverbal, and the instructions can often be translated into any language to communicate with the test-taker. The Figure is a sample item from the Progressive Matrices test. It was developed in England and has been used in many Asian, African, and European countries. It is language-free and depends very little on cultural experiences. Presented with a series of cells, one of which is blank, the person taking the test must choose the correct answer to fill the blank cell. Thus the test can be used to assess general cognitive skills, without hindrances that might be caused by language difficulties. How well this test can predict future success is still uncertain.

Another measure of your test is to assess its concurrent validity. To do so, compare subjects' scores on your test to their performance on a test with previously established validity. Suppose there is a traditional date-test that has already had its validity assessed and established. If subjects on your test obtain scores that correlate well with the standard date-test, then your test has concurrent validity.

Another kind of validity is content validity. Sometimes that's easy to determine; other times it isn't. The content validity of the exams given by your psychology professor refers to how well the various parts of his or her tests cover the topics you've studied in the course. If the first question on the final exam is written in French, you can begin to suspect you're taking a test of limited content validity.

Finally, there's face validity -- the one with which the public is most familiar. Does a test look as if it measures what it's supposed to be measuring? If so, then it has face validity. Most of the "Sunday Supplement" tests of personality, or likeability, or suitability for marriage are high on face validity, but may be woefully low on assessed content- and predictive validity.

It is possible for a test to be highly reliable, yet totally invalid. If you speak Russian, what would happen if we tried to measure your knowledge of Russian using a test in Spanish? Obviously, we would confirm time and time again that you had no knowledge of Russian. It could be a very reliable test with no validity. And how about the other side of the coin? Can a test be highly valid without being reliable? No. A person can be female without being pregnant, but she can't be pregnant without being female. Having a test which is both reliable and valid, suggests the test is objective, but says nothing about its efficiency or the ethics governing its development or use.

Objectivity and Efficiency

An objective test is one in which each person scoring the test reaches the same conclusion about the person taking the test. Although they are far from perfect, multiple-choice and true-false tests are likely to be the most objective. People may argue about how to score a particular answer right or wrong, but once they've agreed, then everyone can/will score it the same way. Although objective tests tend to be both reliable, such tests may or may not be valid.

A subjective test is one that emphasizes the examiner's skills in interpreting test results. Subjective tests, like all

other tests, strive for accuracy, but they are less likely to achieve it. Essay exams are a good example of subjective tests. The examiner's opinions of the writer may bias the scoring, as may the writing skills of the test-taker. Writing skills may have nothing to do with what the test is trying to measure, but a skilled writer may be given a better score.

When you applied to college, the admissions officer making a decision about your application would make a better decision, the more information he or she had. Their decision would be better if they could first test you for a week, then interview you for a week, follow you around for a third week during your senior year in high school, and study your home life for a fourth week. More measures -- especially those based on proper norms -- would increase the accuracy of the total test results. However, it would also be a terrific waste of time, not to mention an invasion of your privacy. Admissions personnel can learn most of what they need to know to make an intelligent decision about you with only a day or two of testing.

As a result, the best tests are a compromise. We balance our needs for high validity and reliability against the costs in time, money, and materials to you and the examiner. The best tests are those with high efficiency that gain the maximum amount of useful information in the least amount of time, consistent with nationally recognized ethics governing proper testing procedures.

Ethics

There's been a lot of argument about tests and testing in past decades. There have been abuses. But we discuss five features that a good test should have. If those features are present, it is not likely that a test will lead to the wrong judgment.

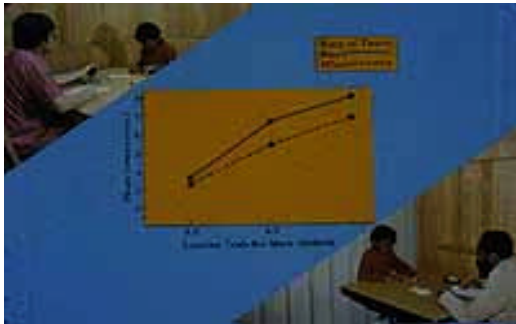
In highly industrialized societies such as those of the United States and Canada, psychologists must often aid in making important decisions about people -- decisions that impact their lives. Company executives should be chosen from among the best qualified for the job.



An airport traffic controller needs a special kind of stable personality. He or she should be resistant to stress because of the nature of the work. In years past, such decisions might have been based solely on seniority or who-

knows-who; nowadays the norms for such are quite different. These decisions can't be delayed or avoided. The more relevant, accurate, reliable, and valid the available information is, the better will be the decision. Testing yields one kind of descriptive data about an individual, but as some psychologists have pointed out, it is only one kind.

At least tests are somewhat impartial -- they pose the same question for all to answer, and they are usually objective and efficient. Yet, we also know that each of us has had different experiences. At what point does that experience begin to count? Should urban blacks be denied access to college simply because their skill with language is not measured by the usual college entrance tests? Should a medical school admit only on the basis of achievement, or can aptitude play some role in the admission decision? None of these questions are easy to answer, but they all border on the ethical issues surrounding testing.



One issue concerns the testing of minorities on tests that have been standardized on white, middle-class citizens. This violates our elsewhere-stated need for standardized norms in order to interpret properly any test score. The typical experiences of various

minority group members may differ from those of middle-class citizens, as indicated by the sample questions in Feature 12.2.

FEATURE 12.2

TESTING AND CULTURAL BIAS

Let's set up an imaginary situation to illustrate a very serious, real-life problem. You're applying for a job, but you must first take an intelligence test to qualify. Answer the following typical items from two different tests:

Test One

- 1) How many seconds are there In an hour? _____
- 2) Animal is to oxygen as plant Is to: _____
- 3) What state borders Ohio on the west? _____
- 4) How many red stripes are in 'the American flag? _____

Test Two

- 1) Who did "Stagger Lee" kill in the famous blues song? _____
 - a. His mother
 - b. Frankie
 - c. Johnny
 - d. His girlfriend
 - e. Billy

- 2) January 15 has been made a national holiday because that was the day when:
 - a. the slaves were freed in the U.S.A.
 - b. the slaves were freed in Texas
 - c. the slaves were freed in Jamaica
 - d. Martin Luther King was born
 - e. Booker T. Washington died

- 3) "Go Tell It on the Mountain" is the title of a spiritual;
it is also:
 - a. the name of James Brown's first hit
 - b. the first line of a speech by H. Rap Brown
 - c. an expression of disbelief
 - d. the name of a poem by Langston Hughes
 - e. the title of a novel by James Baldwin

- 4) Which of the following terms is out of place here?
 - a. Splib
 - b. Blood
 - c. Gray
 - d. Spook
 - e. Black

The answers are printed on the next page at the end of the Feature. Score, yourself separately on each test, giving yourself one point for each correct answer. Now, suppose we could convert your test performance into an estimate of your IQ as follows:

Points	Estimated IQ
0	Less than 70
1	85
2	100
3	115
4	More than 130

Judging from these typical questions, which test would you rather take if you knew beforehand that they expected to hire

someone with above average IQ? The first test is based on facts many middle-class persons would know. The second test, however, is based on knowledge you would be more likely to gain in a low socioeconomic neighborhood. Each test is culture-fair for the people on whom it is "standardized." But how would you feel if asked to take and be scored on the test standardized on a group to which you do not belong? Would it be fair?

Answers: Test One: 1) 3600, 2) carbon dioxide, 3) Indiana, 4) 7; Test Two: 1) Billy, 2) Martin Luther King was born, 3) novel, 4) gray

Another controversy concerns the invasion of privacy.



A secretary must know how to type -- there's little debate about that. And a test of typing skill would certainly not be questioned in hiring a secretary. But what about tests that assess the honesty of an applicant? At what point does the right of a company or organization to know certain

personal information exceed the right of the individual to decline to provide such information? This is the complex issue of privacy -- not easily resolved, but more and more debated in modern times.

You do have many rights with regard to testing. First, you have a right to the best test -- in terms of the criteria we listed elsewhere -- that can be constructed for any testing situation. Second, you have a right to the most educated use of those results that can be obtained. Third, you have the right to absolute privacy of the test results -- they are your scores, and they can be released only with your permission. Finally, you have a right to have those scores interpreted for you by the most professionally skilled interpreter who is available. These conditions are important in the effective use of a test. If met, it is unlikely that an error in judgment will be made, whether that involves a diagnosis of personal abnormalities, advice regarding career interests and decisions, intelligence, or the formulation of educational or vocational plans based on past achievements or current aptitudes.

Tests of Interests

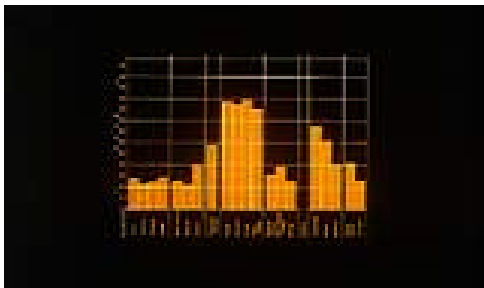
We have an interesting problem with interest tests. They are, of course, very important during the early years of a person's life as he or she is trying to make important decisions about a career and schooling. Isn't it enough just to ask people what they want to do? Sounds simple enough, but it doesn't work. Why? Because in reality the life of a doctor or a psychologist, a teacher or a researcher is seldom as pictured on television or in popular reading sources. Such media capitalize on the unusual, not the typical. As a result, you may be disappointed in your choice of career by the time you've gained enough experience to know what is actually involved. Then it may be too late to change.

It was realized even as early as 1920 that testing people's current interests would yield useful information on possible careers for them, for reasons discussed in the Think About It. Obviously, having the interests that would lead to a particular career is no guarantee that it would make a good career choice. You may or may not have the aptitude to achieve in the field you select. However, such interest inventories do indicate probable areas of success.

Think About It

The question: In career counseling, what's wrong with just asking people what they'd like to be? Why wouldn't they know best what their own interests are?

The answer: People do know their own interests. That information turns out to be very valuable in guiding people into a good career. What people taking such inventories don't know are the interests of successful people in a variety of fields, and they also don't know whether or not they share interests with those successful people. There are no "right" answers in an interest inventory test. The typical test is standardized in terms of interest patterns—hobbies, skills, sports, even reading and entertainment preferences. People taking the test can compare their interests with those of people who have succeeded in certain careers.



To what extent do interest tests steer our thinking about various occupations? More and more women are entering the work force in widely varying careers. As a result, although the interest inventories themselves have

not changed, the interpretation of test results on one inventory was expanded in 1981 to include 162 occupations, 99 of which had been developed in the preceding four years. Although some careers are still occupied almost entirely by men (airline pilot, for instance) or women (nurse), only four occupations listed give norms for only one sex. As you can imagine, it has been difficult to keep the inventories up-to-date when career patterns are changing rapidly.

Two tests are most often used to assess interests. One is the *Campbell Interest Inventory*. This is composed of eight parts, covering 317 items. Sometimes you must react directly (Like, Indifferent, or Dislike) to specific school subjects (from acting through zoology, including psychology), leisure activities (from poetry through skiing, including organizing a play), and so forth. Other scales require you to choose (do you like one or the other or are they equal) between two activities (such as going to a play vs. a dance), and to describe yourself -yes, no, or can't you decide whether you start activities in your group?

Another widely used interest test is the *Kuder Occupational Interest Survey*. In this inventory you must choose -- from a sample of 100 sets of three activities -- the one you like most and the one you like least. The person's forced choices in this inventory are collapsed onto ten interest scales such as Outdoor, Mechanical, Persuasive, Artistic, and so forth. Various patterns of interest scores are then related to a wide variety of occupations. As your interests vary, the occupation for which you may be best suited will show up as a unique pattern of interests.

It is possible to cheat on these tests, but why bother? There are scales built into the tests that detect inconsistent or sloppy answering patterns, but again the only loser is the test-taker. Such tests will not lead you out of the wilderness. If you are confused about what career you should select, but unprepared to make decisions about which of your interests are most important or dominant, then the test will simply confirm you are confused. With dedication to making decisions in which you believe, the inventories can be very helpful -- not only in identifying careers you should consider, but also in identifying fields of study you probably should not consider.

Tests for Achievement and Aptitude

Achievement tests are the most widely used, but also the most widely abused, of all tests. Such tests are most often used



to measure how much you've learned in class or in a training program. They may also be used to assess how well teachers are teaching.

Most of the published achievement tests are carefully standardized. Tests such as the Iowa Tests that you may have experienced in Grades 1-8 or the *Stanford Achievement Tests* for Grades 1-9 are widely tested themselves before being

given to students. The *California Achievement Tests* (Grades 1-12) are another example. Each of these tests, as well as the mid-terms and final you'll have in this course, are aimed at assessing how much you've learned (achieved) up to the time you take the test. Thus, achievement tests stress content validity. They are measuring how much you already know.

Aptitude tests are very similar to achievement tests. You'd be hard pressed to distinguish between achievement- and aptitude tests only on the basis of their contents. The difference between them is based mainly on the intended use of the scores. Like achievement tests, aptitude tests try to assess what you know now. The difference? Aptitude tests use your performance to project your future performance. But such projections can only be based on abilities you've already developed.

Some of us have a tremendous range of abilities; others of us have very few abilities. For such people aptitude tests will be of little use. The tests would only confirm the obvious -- the person would be good in everything, or nothing. However -- back to the normal distribution again -- most of us have an intermediate range of abilities. For us aptitude tests can be very useful in identifying our major strengths so that we can develop them.

Two major kinds of aptitude tests exist. One type measures our scholastic abilities. These are the tests with which you may be most familiar, such as the Scholastic Assessment Test (SAT) or the ACT (American College Testing) Assessment taken near the end of high school and required for admission into many colleges. The Graduate Record Examination (GRE) serves a similar purpose for college graduates who wish to enter graduate school.

In addition to these, there are a wide variety of more specialized tests of specific abilities. These tests may assess your mechanical aptitude, your skill in logic, or your aptitude

in art or music. Whereas the achievement tests emphasize content validity, it should be no surprise to you that aptitude tests stress predictive validity.

As you can see, tests to measure your interests, your achievements, and your aptitudes all relate to your career. In other parts of the chapter we look at two other major uses of tests -- assessing intelligence and personality.

What is Intelligence?

This is a very important question to ask about any test and the assumptions on which the test is based. Most definitions of intelligence focus only on the tools used to measure it, not on the concept itself.



One way to define intelligence is to say that intelligence is what an intelligence test measures. This is not as silly as it sounds, since it emphasizes the tests themselves, which are crucially important. And it really is a very good operational definition, as you saw in the Psychology: Its Nature and Nurture Chapter.

Suppose you take a test that involves only your mathematical skills, or general knowledge, or civic responsibilities. That test is not measuring all aspects of your intelligence. In that sense, then, intelligence is, or is limited to, what's measured by a test.

But there's another way to go about it. As most of us use the term "intelligence," we're really talking about behavior. We tend to think of it as a personal characteristic that can be gauged or measured by a test. One psychologist has suggested



that intelligence involves four conditions: First, awareness -- intelligent behavior is conscious, rational, and controlled. Second, goal-direction -- intelligent behavior is not random. Rather, it is targeted at accomplishing

some definable goal. Third, rational -- intelligent behavior is consistent. Finally, intelligent behavior is worthwhile or *constructive*, not destructive, aimless, or worthless. In sum,

intelligent behavior is conscious, goal-directed, rational, and constructive. If that is its nature, and if the tests used to assess it are critical, then what are some views of intelligence itself?

One view is that intelligence is a single, general trait, and that tests are simply various ways to measure it, as advocated by Charles Spearman. If this view is correct, then all tests of intelligence should be positively correlated. That is, if they're all measuring the same thing, the more of that trait that is present, the higher all test scores should be. However, research does not support this view very strongly; there is only minimal evidence of what Spearman called our g-factor, standing for general intelligence. If the g-factor accounted for much or most of our intelligence, then assessing it would be like drilling for oil on oil-rich land -- any test of any skill would yield as assessment of your general intellect.

A second view suggests that our intelligence may be composed of groups, or clusters, of skills. One suggestion, by Louis Thurstone, listed seven such Primary Mental Abilities, as seen in Table 12.2.

Table 12.2

HUMAN PRIMARY MENTAL ABILITIES

PRIMARY ABILITIES	SKILLS TESTED
Verbal comprehension-V	Vocabulary, reading
Word fluency-W	Anagrams, rhyming, naming categories
Number-N	Simple arithmetic ability
Space-S	abstract visualization, perception of spatial relations
Associative memory-M	Rote memory of paired words
Perceptual speed-P	Visual details: Similarities, differences
General reasoning-R	Inductive, deductive reasoning

This means that all tests assessing the same skill would be related. There is evidence to suggest the factors listed in the Table exist. Any test tapping into your number skills would correlate positively with any other test assessing another facet of your same number skills -- even though such test results

might correlate poorly, if at all, with measures of your memory skills.

A third view is exactly opposite the original general-factor proposal. This suggests that each test by its very content determines what is being measured. One theory states that our intellect is composed of 150 mental abilities, composed of all possible combinations of three types of factors. These include five kinds of contents, five kinds of mental operations, and six resulting products. If this assumption were true, there'd be little correlation between the results on any of these tests -- we might be very high on some and still be quite low on other subtests.

These three views define the classic views of intelligence and the tests based on these views. Such logic was also responsible for developing the intelligence quotient. More modern views have broadened our conceptualization of mental capacity or incapacity to include many skills -- such as sports and interpersonal skills -- not traditionally included in our concept of intellectual development. Which theory is correct? No simple answer will do. One psychologist has even suggested aspects of all three views are correct! Argument is fruitless. Intelligence tests should be chosen -- as is true of any test -- with an eye toward the ultimate use of the results. There are numerous factors also thought to be correlated with our intelligence, including creativity.

Classic Tests of Intelligence



Alfred Binet, in France, was the first person to develop a test of intelligence. The original work was based on two assumptions, both of which later proved to be true. First, it was assumed that to test intelligence it would be best to present problems that allowed the children to function normally. Second, it was also assumed that all else being equal, as a child got

older he or she should be able to solve a broader, more complex variety of problems -- that is, that intelligence increases in complexity with age. We also now know that our intelligence is frequently impacted by the culture in which we live; creating a

culture-free test has been a very difficult process. Some would argue that it is not possible to create a truly culture-free test.



The result of all this work was a 30-item test -- the predecessor of the still-available *Stanford-Binet Intelligence Scale*. To administer the Stanford-Binet test requires a highly trained examiner. For each age level up through early adolescence there are tests of varying

content, but roughly equal difficulty. More and more challenging tests are offered as age level increases. For decades the Stanford-Binet test was among the top five most frequently used tests; in the last decade it has dropped out of that elite group -- a sure indication that it no longer fits modern psychologists' definitions of a test of intelligence.

David Wechsler developed another popular series of tests, very different in conceptualization and format from the Stanford-Binet test.



The Wechsler scale for adults is constituted of 14 subtests -- each organized to test different abilities and arranged in increasing order of difficulty. Eight of the subtests listed in Table 12.3 are utilized -- with adjustments for decreasing

difficulty -- in tests later revised to assess adolescents and children, including infants.



Selected other subtests are used only in WAIS-III and WISC-III or WISC-III and WPSSI-R -- some as supplementary subtests, some as optional subtests depending on how or for whom IQ is being measured. The Wechsler tests measure both verbal skills -- the only thing

assessed by the Stanford-Binet test -- and performance skills, yielding two different IQ scores. The Wechsler scales can be used to aid in the diagnosis of psychiatric disorders. Finally,

the original WAIS was developed just for use with adults, to complement the Stanford-Binet that works only up to late adolescence.

Table 12.3

ASSESSING THE DEVIATION IQ			
WPPSI	WISC	WAIS	Subtest
			Verbal Scales
x	x	x	Information
x	x	x	Comprehension
x	x	x	Arithmetic
x	x	x	Similarities
x	x	x	Vocabulary
	s	x	Digit Span
		x	Letter-Number Sequencing
o			Sentences
			Performance Scales
	x	x	Digit Symbol-Coding
	x	x	Picture Completion
x	x	x	Block Design
x	x	x	Picture Arrangement
x	x	x	Object Assembly
		x	Matrix Reasoning
	s	x	Symbol Search
x	s		Mazes
o			Animal Pegs

x = Scale used on indicated Test; o = optional subtest;
s = supplementary test



The Wechsler series includes three tests now available as: (1) the third edition of the *Wechsler Adult Intelligence Scale (WAIS-III)*, (2) the third version *Wechsler Intelligence Scale for Children (WISC-III)*, and (3) the second edition of the *Wechsler Preschool and the Wechsler*

Preschool and Primary Scale of Intelligence (WPPSI-R). Both the WAIS and the WISC have separate verbal and non-verbal performance scales that yield verbal and performance IQs that can be combined for comparison with a Stanford-Binet-measured IQ. The WPPSI-R -- having fewer tests -- yields only a single estimate of an infant's or young child's IQ.

The Intelligence Quotient

The intelligence test probably most familiar to you is the "IQ" test. What is an IQ, or Intelligence Quotient? Though sometimes mistakenly credited to Alfred Binet, German psychologist William Stern developed the original concept of the intelligence quotient.

Three terms are important: mental age, chronological age, and intelligence quotient.

The first concept is mental age (MA), or the average age, expressed in months, of children who answer correctly certain items on a standardized test.

$$IQ = \frac{MA}{CA} \times 100$$

is then compared with the second concept -- her actual, or chronological age (CA), in months, in this case 96 (8 x 12).

Age	Results	Credit
10	Pass all	120
11	Pass 3/6	96
12	Pass 2/6	72
13	Pass 1/6	48
14	Pass 0	24

CA = 10 yrs. 4mo. = 124mo. MA = 132mo.

IQ = $\frac{MA}{CA} \times 100 = \frac{132}{124} \times 100 = 106$

For example, if a girl exactly eight years old passes those items passed by the average ten-year-old, her mental age is 120 months (10 years x 12 months). This score

is then compared with the second concept -- her actual, or chronological age (CA), in months, in this case 96 (8 x 12). To determine the IQ, you divide the mental age by the chronological age and multiply by 100 -- to yield a whole number: $IQ = (MA / CA) \times 100$. Our eight-year-old would thus have an IQ of 125, that is $(120 / 96) = 1.25 \times 100 = 125$. Thus, any IQ over 100 indicates above average ability; any IQ below 100 indicates below average ability in school subjects.

average ability; any IQ below 100 indicates below average ability in school subjects.

Why do we use the IQ concept? The original test of intelligence was developed by a committee, of which Binet was a member, to measure the school abilities of Paris school children and their ability to profit from instruction. IQ scores can help to identify students who require specialized attention, such as remedial help or more challenging programs. Scores can

often be useful in predicting future academic success. But some questions about their usefulness remain.

Modern Tests of Intelligence

At one time theories of intelligence were defended in fierce turf wars. With competing theories, psychometricians (measurement experts) such as Charles Spearman and Louis Thurstone led their armies of graduate students and other believers into lengthy, difficult battles. For psychologists, each skirmish was a major event.

Eventually other leaders and other campaigns surpassed these psychological leaders. The battlefield lay still, littered with indecisive relics of various statistical battles. In more recent years, other battles have been fought, but the attention of both psychologists and graduate students alike has turned to other fronts -- to concerns with cognitive psychology and neuropsychology -- frontiers where the battles are more intense, more interesting, and potentially more rewarding (adapted from Sternberg, 1977, pp. ix).

Intelligence declined as a topic of intellectual interest. Brief rallies occurred around topics such as heritability of intelligence and racial differences in intelligence. However, by 1970 one psychologist had remarked that ". . .to the theoretical psychologist. . .intelligence seems to have outgrown its usefulness" (Vernon, 1971).

Two events occurred to alter the intellectual battlefield. First, Robert Sternberg (1985, 1988, 1996) proposed a new triarchic componential theory of intelligence. Assuming theories of intelligence should be able to specify the mechanisms by which intelligent behavior is accomplished, he defined a component as an elementary information process that operates upon internal representations of objects or symbols (Sternberg, 1985, pp. 97).

The main components of Sternberg's theory of intelligence serve at least three kinds of functions: (1) Metacomponents are executive processes used in planning, monitoring, and making decisions as a task is being performed. (2) Performance Components are used to execute various strategies for task performance which may include encoding of stimuli, combining or comparing stimuli, and responding. (3) Knowledge Acquisition Components are concerned with acquiring declarative and procedural knowledge: Selective-encoding, -combination, and -comparison.

The executive processor is the unique feature of Sternberg's theory -- the component that allows you when working

on an algebra problem to step back at some point, realize that your work is headed nowhere and initiate a decision to start over or develop another strategy. The performance components are similar to performance skills assessed in other tests of intelligence such as Wechsler's, though there is more emphasis on the cognitive underpinnings of such actions in Sternberg's theory. Verbal skills and knowledge, as assessed by classic tests of intelligence are similar to Sternberg's Knowledge Acquisition Components.

Second, Howard Gardner (1983) has proposed a Theory of Multiple Intelligences. It is Gardner's view that our mental skills are based upon ". . . a number of different intellectual strengths, each of which may have its own developmental history" (Gardner, 1985, pp. 59). Such intellectual competencies involve problem solving skills which enable us to resolve genuine problems or difficulties that we may encounter

Under some conditions, these problem solving activities may lead to creating an effective product. Also involved is the potential for creating or finding problems -- which is the basis for acquiring new knowledge. To qualify as an intelligence, there are eight "signs" for which Gardner looks including each of the following: (1) Potential isolation by brain damage. (2) The existence of idiot savants, prodigies, and other exceptional individuals. (3) An identifiable core operation or set of operations. (4) A distinctive developmental history, along with a definable set of expert "end-state" performances. (5) An evolutionary history and evolutionary plausibility. (6) Support from experimental psychological tasks. (7) Support from psychometric findings. And finally, (8) susceptibility to encoding in a symbol system. There are eight intelligences which qualify according each of these eight signs, as detailed in Table 12.4.

Table 12.4

HOWARD GARDNER'S EIGHT INTELLIGENCES

Type of Intelligence	Skills Drawing Upon the Intelligence
Linguistic	Reading, writing a paper, listening to/understanding speech
Logical-mathematical	Solving math or balancing checkbook, logical reasoning, or doing a proof in mathematics

Spatial	Moving from one place to another, reading maps, packing a suitcase
Musical	Singing, composing music, playing a piccolo, understanding music structure
Bodily-kinesthetic	Playing a competitive sport (football, soccer, basketball), throwing a discus, running a foot race, dancing, swimming
Interpersonal	Understanding and relating to other people; running a business, understanding people's motives and/or emotions
Intrapersonal	Understanding yourself. Knowing who you are, what is important to you, what motivates you, knowing your potential for change
Naturalist	Understanding regularities or patterns in the physical world

(Adapted from Gardner, H., 1983, 1993.)

Whereas Sternberg emphasizes the interaction among the three components of his view of intelligence, Gardner stresses the uniqueness of his multiple intelligences. Gardner's work is the first to give intellectual credence to skills such as those demonstrated by NBA basketball players (Bodily-kinesthetic) and the skills of corporate executives and leaders (Interpersonal). With such modern theories, the traditional definition of IQ is in decline, though there is a continuing concern with problems of mental retardation.

Definition and Measurement of Mental Retardation

In the sections on ethics and Feature 12.2 discussing Testing and Cultural Bias you read about one of the major concerns voiced by critics of intelligence tests -- that such

tests are culturally biased. Still other studies have emphasized another factor -- that IQ, once thought to be fixed at birth by heredity, is also the result of interaction between parents and child. The kinds of experiences infants have can mean significant differences in various intellectual areas. People in great poverty, for example, would probably have fewer opportunities to provide infants with stimulation. In fact, studies of children raised in slums (and orphanages) have shown that infant education can dramatically increase a child's IQ score. Does a better environment, then, actually make children smarter? This may be so. However, Edward Zigler, of Yale University, believes rather that the environment of deprivation makes children less smart than their inborn potential would dictate under better circumstances.

Level of mental capability	IQ	%
Idiot - Cretinoid	Profound MR	0.12
	Severe MR	
Imbecile - Trainable	Moderate MR	2.14
	Mild MR	
Moron - Educable	Dull normal	13.89
	Average	
	Bright normal	68.26
	Superior	
Gifted	146-160	2.14
	161-175	
	Above 176	0.12

Notice here again an interaction between heredity and intelligence. As you can see from the above appraisal of IQ testing, labeling a person mentally retarded because of low IQ scores may be misleading. While persons with IQs below 70 are usually considered retarded, factors

in the course of development may be inhibiting their performances on tests. Careful observation and examination are necessary to determine whether a person is profoundly, severely, moderately, or mildly retarded, or indeed retarded at all. Sometimes shocking mistakes in labeling have been made. Only proper diagnosis can indicate the type of training best suited for the individual. The profoundly retarded must be cared for all their lives, whereas the severely retarded can (with complete supervision) function at a minimal level.



The moderately retarded are considered to be trainable -- to do semiskilled or assembly -- line work. Those who are moderately mentally retarded are often referred to as trainable mental retardates.

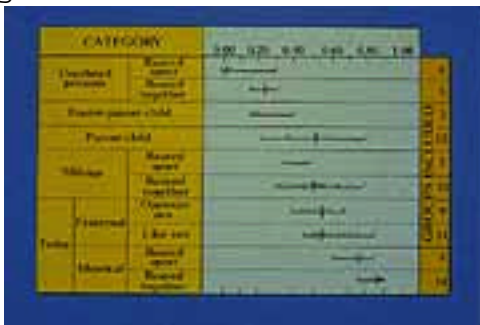
The educable mental retardates are only mildly retarded; with proper education they can function reasonably well both socially and economically.

What causes retardation? There are two types of causes: Functional retardation is caused by cultural deprivation with no

known organic cause. Organic retardation is caused by inherited conditions such as Down's syndrome, phenylketonuria (PKU), and cretinism (or hypothyroidism). Other types of retardation can be caused by poisons in the environment, including the prenatal environment. Brain injury and even malnutrition can cause retardation. Ongoing research is still uncovering influences on the development of the brain and human intelligence.

What Impacts intellectual Development?

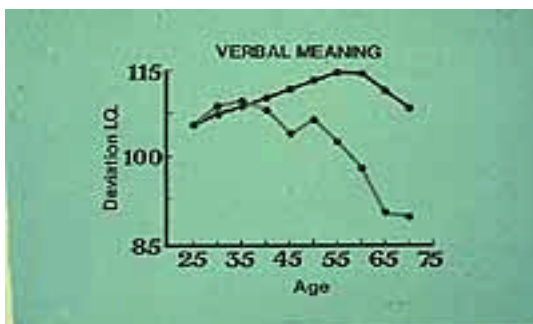
What's your IQ? What is the IQ of a person in your neighborhood who is the same age and sex as you? Considering any two people at random, there will be essentially no correlation between their IQs. What factors, then, do cause a correlation?



If you know your parents' IQs, that is useful information. There is a correlation between your

intelligence and that of your parents -- usually positive -- especially if you all live in the same house. If you have a fraternal twin who is the same sex as you, your IQs will be even more positively correlated. And, if you have an identical twin living with you, your IQs are even more likely to be almost the same. As the degree of genetic relationship increases, so does the similarity of IQs. Heredity counts. But since children of comparable genetic relationships reared together will have IQs more alike than those not reared together, we've also confirmed the influence of environment.

One oft-debated question about IQ concerns whether or not it changes with age.



In one cross-sectional study, two psychologists gave some 500 people, 20-70 year olds, an intelligence test. Seven years later they gave 302 of the original 500 (60 percent) many of the same tests. These data were combined in a longitudinal study. Both types of studies

are described in our discussion of cross-sectional and longitudinal studies.

Does IQ change with age? The answer depends on the type of study. Read the Think About It for some relevant results.

Think About It

The question: Will your intelligence grow, remain constant, or drop as you get older?

The answer: In longitudinal studies, a person's intelligence will show a slight, steady increase during the adult years through (roughly) age 60 of his or her lifetime. In cross-sectional studies older persons will seem to be slightly less intelligent than younger ones. This occurs because education today is better, and intelligence tests assess knowledge. Over the last 50 years, as the Stanford-Binet and the Wechsler tests have been restandardized, the average IQ of Americans has climbed almost 15 points. As the knowledge base grows, each new generation knows more at any given age.

What comes with age is experience in knowing when to apply knowledge, an ability not well measured by most tests of intelligence. Older people do not get dumber as they age. They are often working with more limited knowledge, but doing so with greater skill.



Another factor that influences intelligence is one's birth order in a family. An only child is born into an intellectually stimulating environment. The other members of the family are all adult and already educated. For a child is born into a large family,

with many brothers and sisters close to him or her in age, the average level of intellectual stimulation will be lower for the later-born children than was the environment in which the older children were born. The parents must also cater to the needs of the brothers and sisters. And talking and doing things with brothers and sisters may not educate you as well as would the same amount of instruction from your parents.

Birth order may affect not only intellectual level but also motivation -- for instance, for achievement. Of the first 23 United States astronauts, 21 were first-born or only children; one had a single older brother who died as an infant, and one was 13 years younger than his older brother. Sally Ride, the first American woman astronaut, was also a first-born child.

USING PSYCHOLOGY: Can you Improve your IQ?

Within limits, yes, it is possible to improve your IQ. Several factors are involved, since both heredity and environment determine your intelligence. Thus, it is possible to create an environment that will do the most for your intellectual development. Since intelligence is most often measured in a formal testing situation, experience with the testing format is clearly helpful. It pays to know what kind of testing procedure will be used.

A second factor concerns motivation. Prior classroom experiences may sometimes discourage you from trying to do your best in a formal testing situation. A relaxed, confident attitude is helpful. Keep in mind that no test at any time in your life is an absolutely critical event from which there is no recovery if things don't go well. An error some students make is to get too keyed up for the tests. They run their motivation level up so high that they are not at their peak performance range. Do something you enjoy the night before the examination! Third, some might hope to "cheat" on the tests by trying to find out beforehand what questions will be asked. Of course, that isn't really raising your IQ. The tests are given on the assumption that you are a "naive" participant taking the test. This means that you've not already seen the specific problems being posed for you in the test. If that's not true, then the test is not a valid one.

Fourth, most commonly used intelligence tests in North America are based on middle-class values and typical experiences. Anything that will increase your exposure to such an environment is likely to increase your measured IQ. Friends, clubs, social organizations, and participation in community affairs are some examples. So, it is possible to affect your IQ within a limited range by focusing on environmental factors affecting your experience and performance in the test.

Finally, recent reports have suggested that drinking a generous glass of orange juice before taking a test -- any test -- will improve your performance on the test. Most likely this is because of the additional energy available to the brain as you are taking the test.

While you may not be concerned with this matter now, there is also some helpful information available about how you may be able to maximize (when the time comes) your children's intelligence. This involves many factors in development that we discuss in the Early Development: From Creation to Adolescence Chapter. Also review the interesting facts about birth order, which are summarized in Table 12.5.

Table 12.5

FACTORS AFFECTING YOUR INTELLIGENCE

<u>FACTOR</u>	<u>IMPACT</u>
BIRTH ORDER	Earlier born tend to be smarter (because they are in a richer intellectual environment).
FAMILY SIZE	Larger total family size tends to mean lower IQ for comparable birth order (#2 child of 2 siblings smarter than #2 child of 7 siblings).
SPACING	More distance between children means less adverse effects of later birth order (because the average intellectual climate is richer).

*Note: The studies on which this table is based involved differences in tested IQ of only 1-2 points through an entire generation.

Intelligence and Creativity

Does having high intelligence make people more creative? Intelligence and creativity seem to be related, but creative ability has proven difficult to measure reliably and validly. Creativity involves originality and flexibility in problem solving. For instance, take a look at the inversion of Inversions. Intelligence tests usually measure convergent thinking, which a person uses to organize various facts to arrive at the single (correct) answer. By contrast, tests of creativity often assess divergent thinking, which a person uses to search for many possible answers given a limited number of stimuli. Some have asserted that standard tests of intelligence -- requiring a search for the answer -- penalize those offering creative answers.

A variety of attempts to measure creativity have been developed. One is a Remote Associates Test: What word relates to cake, blue, and cottage? What relates to *iron*, *hot*, and *bell*? The answers are cheese and bar. Another approach involves an open-ended assessment of creative thinking. For

example, one test lists objects a person might use to produce something and challenges the participant to suggest ways to improve the product. Scoring creativity tests poses a number of problems. How unusual must an answer be to be scored "creative"? Obviously, much more research is needed.

We do know that, while most of us have some sort of creativity, exceptional creativity seems to run in families, as you saw in analyzing the Bach family tree in the Early Development Chapter. To nurture creativity, a family must provide the freedom and encouragement that allows such talents to emerge and develop.

Personality Tests and Errors

There is a fifth, and very common, use of testing we want to discuss -- that of assessing personality. Basically, there are three choices in the testing method to use. We can observe behavior and try to make decisions about someone. We can administer various formal tests to try to convert people's behavior into numbers, about which we can then make decisions. Or, we can administer projective tests which are more subjective measures of human behavior. There are problems with all three measure techniques. Personality tests are most often used either in assessing job applicants or in diagnosing problems of personal adjustment.

All these observation techniques, including behavioral observation techniques such as interviews, tend to be subject to certain types of errors of assessment. One such error is stereotyping, which is attempting to fit the applicant's performance into categories so broad as to be ineffective. As we discuss in the You and Groups Chapter, not all redheads have fiery tempers, not all Chinese are earnest and loyal only to family ties.

A second problem is insensitivity. Sometimes an interviewer may not interpret an applicant's answer correctly. Thereby the interviewer misses an opportunity to pursue what might otherwise be a fruitful line of questioning or observation.

A third problem is the halo effect -- meaning that prior knowledge of the applicant (good or bad) influences interviewers' observations. They may record data only so as to confirm opinions and beliefs formed before the interview and contrary to answers or evidence being offered by the applicant.

Because of these potential error sources, there has been a long effort to try to render the assessment of personality more objective and precise. This can be accomplished using

standardized questionnaires and self-report assessment devices or by interviews structured in various ways.

Interviews



Suppose you are being interviewed for a job. You are asked several general information questions about your life history. Then, when the interviewer asks you what college or university you're going to (or have graduated from), he or she makes an insulting comment about your

school. Without your knowing it, the interviewer may be stressing you to see how you react under pressure. The stress interview is one type of testing situation that is available.

Another is the structured interview that covers a specific series of questions and topic areas -- the interviewer may even be working from a sheet of questions. It reduces biases -- both yours and the interviewer's -- but it may be stressful if it doesn't allow you to expand on your answers.

A third type is the unstructured interview, where you can ask questions of whoever is conducting the interview. Of interest here is how the interview is controlled. A skillful job applicant may manage to talk only about those abilities in which he or she excels, especially if interviewed by an inexperienced personnel manager. In the unstructured interview, the lack of structure permits various forms of error to creep into the process; this lowers the validity and reliability of decisions based on such interviews.

The fourth type of interview is the exhaustive interview, where you may be interviewed for many hours and/or by more than one person at a time. This can also be very stressful, but not unlike the type of day that would be faced by a very busy business or academic person. All of these interview styles represent variations of standard personality and performance assessment tests.

Questionnaires and Self-Reports

As we discuss this means of collecting data for assessing personality, we'll describe the procedure briefly. Then we'll select a test to show you the concepts that apply. Questionnaire/self-report tests are easy to give to large groups

of people. Most such personality tests involve a large number of questions, each of which can be answered yes/no or true/false. These test results are often used as an initial screening device -- a rough indicator of present or possible future problems. Tests such as these are usually supplemented by additional tests, possibly including interviews or projective tests. They are used to verify the initial findings, or follow-up to find more information about conditions revealed by the first tests.

Probably the most widely known of all personality tests is the Minnesota Multiphasic Personality Inventory, or MMPI. There are now more than 6,000 published references to and studies of the MMPI, released in 1989 in a revised form called MMPI-2. The MMPI-2 is composed of 567 statements about you -- simple statements about aspects of you, to which you can easily respond "true," "false," or "cannot say."

These statements can be scored and analyzed by computer in terms of 10 scales of tendency toward such disorders as hypochondriasis (Hs, or health anxiety), depression (D), and Schizophrenia (Sc, suggesting withdrawal into a private world). Buried within the multiple hundreds of statements are some items used to detect lying or faking socially desirable answers -- the Lie (L), Faking (F), and Correction (K) scales. An example? Responding "true" to "I like everyone I know" will elevate your score on the Lie scale.

Although the MMPI has rather poor reliability, it is still a critical tool used in assessing personality. Other tools include the interview, the intelligence test, and the projective test. The MMPI identifies possible diagnoses to be verified or rejected using other sources of data. The test is so firmly based in supportive research findings, that participants' responses are analyzed by computer and the report generates both an etiology (what caused the problem?) and a prognosis (What is the likelihood of recovery?).

Projective Tests

As we discuss this means of collecting data for assessing personality, we'll describe the procedures for using projective tests briefly. Then we'll select a test to show you the concepts that apply.



The projective tests offer the person taking the test a seemingly formless, unstructured task. Very few cues are offered as to what is expected or what is a "good" or "correct" answer. The theory behind these tests suggests that given a specific, but vague stimulus,

participants will have to draw on their own perception of the situation to respond. It is thought that when we are forced to do this, it will cause us to reveal our own personality or views of the important controlling events in the world.

The theory seems sound, but there are some problems. Projective tests are subjective. Their validity and reliability have been questioned by psychologists because of the limited number of norms provided for some of these tests. As a result, psychologists tend to use these types of tests only for guidance, not as a sole source of information about a client or patient

One classic projective test is the Thematic Apperception Test (or TAT), which has been available since the early 40s. The TAT consists of detailed pictures that are open to different interpretations. Persons taking the test are asked to make up a story about each one. They are to relate what led up to the situation, what is going on, what the characters are thinking or feeling, and what is likely to follow.

Scoring is not standardized. However, the examiner might be looking for themes that keep appearing in the stories. There may be a tremendous range of responses, but the results for the TAT are often determined by the examiner's experience with giving the test. For any individual, it is the unusual responses -- those that depart from the expected -- which are most important and the most revealing.

How are we to evaluate the projective tests? The underlying assumption is that unstructured test situations can be used to detect enduring personality traits. Especially with the TAT there is evidence that temporary, everyday things may affect it. Hunger, lack of sleep, or social factors such as frustration or failure in other test situations may affect the results. Perhaps the most effective use of projective tests is as an "ice-breaker" early in the therapeutic situation. Despite widely recognized poor reliability, the TAT remains a very popular test because it offers a nonthreatening means to introduce participants to psychotherapy.

REVIEW

ASSUMPTIONS FOR A TEST

1. What is the purpose of tests? Why have they been developed?
2. What is a "normal distribution"?
3. Why are standardized norms for tests important?
4. What is meant by saying a test is "reliable"?
5. What is "test validity?" Explain three ways in which a test may be shown to be valid.
6. Compare and contrast the scoring of objective and subjective tests.
7. If a test is efficient, what do we know about it? Why is efficiency important in Testing?
8. What precautions help assure that tests will be administered and interpreted in an ethical manner?

TESTS OF INTERESTS, ACHIEVEMENT, AND APTITUDE

1. What types of tests are used in career and vocational counseling? On what are such tests based?
2. Compare and contrast achievement and aptitude tests.

TESTS OF INTELLIGENCE

1. Give at least two definitions of intelligence.
2. Describe three different views concerning the nature of intelligence.
3. Name and describe briefly the most commonly used intelligence tests.
4. How is IQ calculated?
5. What is now known about retardation?
6. Why is creativity difficult to measure?
7. What are some influences on the development of intelligence?

TESTS OF PERSONALITY

1. Describe briefly how your behavior in an interview might be used to evaluate your personality. What errors of assessment might be made and why?
2. What are projective tests? Name those most commonly used.
3. Are projective tests well validated? Are they reliable?
4. What would you want to know about any test you took?

ACTIVITIES

1. Examine back issues of Reader's Digest, Family Circle, or any magazine that is likely to run "tests" of marital happiness or personal adjustment. Select a test and list for yourself what skills you think are being tested. Then take the test, score your performance, and see how well the article says you did. Now, reread the article closely and see if you can find out how the "normal" performance was defined. What group of people was used to standardize the test? How well did the article and test specify the norms, reliability, validity, objectivity, and efficiency of the test?

2. If your college/university has a counseling center, find out whether you can take either of the traditional tests of vocational interest (the -Campbell or Kuder tests). If you take one of the tests, make an appointment with a psychologist to have your scores interpreted. Did you find any interest revealed that you hadn't known you had? If this service is not available, colleges and universities often teach a course on "tests and measurement." Students in the course are learning how to administer and interpret vocational interest tests. Through your [professor (or directly) you might contact a member of the faculty in your psychology department or in your institution's counseling center to find out if they need volunteers to take these tests.

3. Does a school district or county near you use an achievement test such as those available from Iowa, California, or Stanford? If so, call or visit one of the school psychologists or central-administration personnel (call ahead for an appointment!). Ask him or her to review with you the national norms for the test as well as (time permitting) the performance data for the students from the whole district or county. Discuss the psychologist's interpretation of the data, including any explanations for the performance of the local students as compared with the national norms.

4. Sit down with someone who loves you -- a friend, parent, brother or sister, or significant other. Have this person help you analyze what things you do best. Are you super-quick in mathematics? Fast with a smile? Easy to get along with? Do you work especially well with people in a noisy or crowded situation? Find out your strengths. Then decide how you can

best convince an interviewer of those strengths. Now you're ready. Do you need a job for next summer? Start making contacts now. See how many times you can get yourself interviewed. If you don't get a particular job, try to find out why.

5. Make some symmetrical inkblots. Fold a piece of heavy white construction paper in half. Into the crease drop several drops of ink; now press the two halves together so as to squeeze the ink out into a variety of symmetrical forms. You may have to do quite a number before you get a satisfactory variety of forms. Now show each form, one at a time, to a friend. Ask your friends (who wish to participate) to write down two or three things that they see in the form -- either in its parts or as a whole. Then collect the papers and analyze the responses with the friend or (anonymously regarding who supplied the answers) with fellow students. How much agreement is there? Why might people respond very differently to the same form?

INTERESTED IN MORE?

ANASTASI, A. (1982). *Psychological Testing*, 5th ed. Macmillan. A classical text covering many aspects of testing. Describes limits on procedures, and gives examples of intelligence, achievement, aptitude, and personality tests.

COON, K. (1977). *The Dog Intelligence Test*. Avon. Don't take this one too seriously. Gives instructions for administering and scoring ten dog intelligence tests. Fun to read.

CRONBACH, L. J. (1970). *Essentials of Psychological Testing*, 3rd ed. Harper & Row. Thorough coverage of all the major issues involved in developing and administering a test of skill or potential. Includes a wide array of tests.

EYSENCK, H. J. (1968). *Know Your Own IQ*. New York, NY: Penguin Books. Eight tests, with answers and scoring instructions, that allow you to make an estimate of your IQ as it might be measured on the major standard tests.

KAMIN, L. J., ed. (1974). *The Science and Politics of IQ*. Halsted Press. Gives a variety of views on a series of controversial aspects of intelligence: its existence, causes, and assessment.

THORNDIKE, R. L. & HAGEN, E. P. (1977). *Measurement and Evaluation in Psychology and Education*, 4th ed. New York, NY: John Wiley and Sons. Heavy reading. Covers the basics of testing and some related social and political issues.

TYLER, L. F. (1971). *Tests and Measurements*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall Publishers. A brief, tough, solid review of the assumptions underlying tests. Includes a review of test development, statistics, and a discussion of issues of fairness and accuracy.

WRIMBEY, A. & WRIMBEY, L. S. (1975). *Intelligence Can Be Taught*. New York, NY: Bantam Books. Suggests that many abstract skills measured by tests can be improved with training. Thus, you can raise your IQ?!?