CHAPTER 7
Measurement of Intellectual Capabilities: A Review of Theory

by John L. Horn

The publication of the Woodcock-Johnson in 1977 marked the advent of a cognitive battery significantly advanced in conceptualization and measurement of intelligence beyond the traditional batteries, such as the Wechslers and the Stanford-Binet. The revision of the Woodcock-Johnson in 1989 extended the match between new advances in intelligence theory and the measurement of intelligence.

The theory of intelligence upon which the WJ-R is based is generally known as *Gf-Gc* theory. A thorough description of this theory and its proper place in the history and broad array of conceptions of intelligence was, frankly, beyond the expertise and knowledge of the authors of this manual. It seemed, however, that such a discussion would be quite helpful for those who wish to broaden their understanding about the phenomenon they aim to measure when they use intelligence tests.

To provide this, an important contributor to the evolution and present status of *Gf-Gc* theory, John Horn, was asked to contribute a chapter to this manual. The purposes of the contributed chapter are to present an overview of intelligence theories and to place the *Gf-Gc* theory into the context of other theories and their historical underpinnings. John Horn is well-qualified to make this contribution because of his knowledge and because of his research with data from many intelligence batteries including the 1977 WJ and the 1989 WJ-R.

KSM
JKW
RWW

Purpose

This chapter provides a bird's-eye view of the knowledge base for the Woodcock-Johnson Psycho-Educational Battery — Revised (1989), or WJ-R. This test is an example of technology, or applied science. Useful technology is based on scientific understanding; the better the science, the more effective and efficient the technology can be. So it is with psychological tests. They should be based on the most dependable and current evidence of science.

In using psychological tests, it is important to understand the basic science on which the tests are based. With some technology, there is little need to understand the science. One can use a TV, for example, with little awareness of how it works. Effective use of measures of human cognitive capabilities, however, requires that one understand how the capabilities were identified and what the measures indicate about performance, adaptation, and adjustment. The aim of this chapter is to provide evidence upon which cognitive tests can and should be based and to promote better use of intelligence tests. Let us now turn to major issues that call for understanding.

Theories of Human Intelligence

The WJ-R is a model for scientific theory about cognitive capabilities; it is a set of operations (tests) that represent the concepts and measure the constructs of this theory. It is an attempt to express the concepts of the theory with measurements that are useful for making important decisions. The WJ-R also provides a means for evaluating the adequacy of the theory upon which it is based. On the one hand, the theory and its hypotheses provide a basis for the measurement operations of the tests. On the other hand, the tests provide a means for checking the hypotheses of the theory. This means that the WJ-R is an instrument for theoretical research as well as for applied work.

The question that has driven most scientific research on cognitive capabilities is "What is human intelligence?" That intelligence is a marvelous thing is universally agreed upon, but no one is able to agree on an adequate definition of what it is. The experts don't agree (Sternberg & Detterman, 1986).

Many kinds of behavior indicate intelligence, but identifying their essence defies our capacities for synthesis. Research, however, has indicated that the diversity seen in expressions of intelligence can be understood with a relatively small number of concepts. There is considerable agreement about these concepts among those who have studied intelligence, particularly when one allows for the fact that people often use different language to describe the same thing.

Dealing With the Impossible

Unresolvable issues should be recognized as unresolvable, and defining human intelligence is definitely one such issue. It can't be done because human intelligence is vast and dynamic and constantly changing. The outcomes of these changes cannot be predicted. Human capabilities that indicate intelligence expand with every new invention of the human mind, such as the printing press, the silicon chip, computer games, and nuclear energy. Old capabilities that once were important are no longer developed, perhaps because the culture no longer needs them. Efforts to define intellectual capabilities "once and for all" are doomed to failure because not only is the universe of these capabilities so vast that its boundaries are beyond comprehension, but also because it is constantly evolving into a new vastness.

Defining human intelligence is further complicated by our inability to divide human behavior into discrete segments. Ability, cognitive capability, and cognitive processes are not segments of behavior but abstractions that we have applied to an indivisible flow of behavior. One cannot distinguish between reasoning, retrieval, perception, and detection. The behavior one sees indicates all of these, as well as motivation, emotion, drive, apprehension, feeling, and more. Specifying different features of cognition is like slicing smoke — dividing a continuous, homogeneous, irregular mass of gray into . . . what? Abstractions. Terms like reasoning and retrieval are used to indicate that a task typically involves more of one than the other, but the behavior itself is continuous. And there are many ways to slice the smoke to indicate cognition (e.g., Cattell, 1957, 1971; Commons, 1985; Cronbach, 1970; Detterman, 1986; Gardner, 1983; Humphreys, 1979; Spearman, 1904; Sternberg & Berg, 1986; Sternberg & Detterman, 1986; Thomson, 1919; and Thurstone, 1931, 1935, 1938).

Given such unsolvable problems, we can never expect to know the precise nature of intelligence. We can know some and learn more. The problem of understanding intelligence is one of building construct validity (Cronbach 1977, 1987), which requires the building of a scientific theory.

Three Classes of Theories About Intelligence

There is wide diversity among scientific theories of intelligence (Sternberg & Detterman, 1986). This fact is likely to cause confusion. Some of this confusion can be reduced by recognizing that, at a broad level, the theories fall into three distinct classes: essence, compound, and mixture theories. These classes are analogous to the classes of substances in the field of chemistry.

Essence Theories

A theory of essence identifies a single characteristic that is responsible for the phenomena of interest. Essence theories of intelligence stipulate that all distinct intellectual abilities stem from one basic process or element — the essence of intelligence. For example, Eysenck (1982) has suggested that the essence of intelligence is the capacity to hold elements of stimulation in the span of immediate awareness. Depending on the theory, that essence may come from any of several sources. If the theory posits that intelligence is inherited, it is the essence that is inherited. If the theory argues that intelligence is the outcome of environmental influences, these influences determine the essence. Any alteration of the essence results in changes in its outward manifestations.

Compound Theories

In chemistry, a compound is a union of different elements that combine in a specified manner to form a unit. Water, for example, is a union of the elements of hydrogen and oxygen combined proportionally. Until a chemical compound is broken apart, which requires extraordinary action, it functions as a unit. Water behaves as an entity not easily divided into hydrogen and oxygen. It heats up and cools down as a whole, not as hydrogen and oxygen separately.

Similarly, compound theories of intelligence specify a union of different cognitive capacities, each always present and in a particular relation and proportion to other capacities. For example, intelligence might be considered a compound of apprehending, encoding, working memory, comprehending conjunctions and disjunctions, drawing inferences, and formulating conclusions (Horn, 1965). Because the theoretical components function together as a unit, whatever might affect the whole, such as inheritance or environment, affects the components equally. The components rise and fall in a lawful manner with stimulation and treatment.

A test of such a theory requires that a model be specified in which the exact role of each component is precisely indicated and that it be shown that this model provides a good fit to data (Horn & McArdle, 1980). The relationships, whatever they may be, among components and with other variables must also be precisely specified. The intercorrelations among the components can be anything, including zero. Large relationships are not required; but if the required systematic relationships are demonstrated, then the components can be said to represent a functional unity in the sense of a compound theory.

Mixture Theories
Mixture theories specify only a collection of different components. The components needn't stand in any particular proportion or relation to each other. Mixtures are analogous to dinners. Dinners may be made from the same ingredients and yet be quite different, and dinners made from different ingredients are even more variable. Most tests that measure general intelligence or IQ are mixture measures. Examples are the Wechsler and Stanford-Binet tests.

Mixture measures have no intrinsic features to indicate that one is more nearly correct than another. We know that some dinners are better than others; but the proof of the pudding is in the eating, not in the components of the mixture. Similarly, the features of a test that represents a mixture theory of intelligence must be judged in terms of the purposes for which it is intended. An observable feature in the intelligence of a human being must be judged in terms of what that person is expected to do or wants to do. The mixture of intellectual abilities that is best for creating sculpture is not necessarily the mixture that is best for creating theoretical physics.

Distinguishing Personal Theory and Scientific Theory

As one develops a theory about what intelligence is, it is tempting to believe that theory. Almost everyone has a theory about human intelligence, but there are few scientific theories. In order to know if a theory is scientific, the statements must be verifiable and verified; the concepts (constructs) must be operationally defined; the major features must be tested; and the theory must explain what is known about the phenomena that indicate intelligence. Without these qualities, a theory is merely thought or speculation, perhaps interesting or even largely correct, but not scientific. Without these qualities we cannot know how correct the theory is.

Many theories about intelligence are not based on the evidence of cognitive science. They are appealing because they contain interesting anecdotes and plausibility that fit with everyday experience. But plausibility is not enough. The history of science is eloquent testimony to the fact that what is plausible, is not often true. Plausibility does not necessarily take existing evidence fully and carefully into account; it lacks operational definitions and tests.

Historical Perspective on the Emergence of Theory About Human Intelligence

Current thinking about human intelligence derives from a complex set of influences that have operated over a long period of time. While this chapter is not intended to be a complete description of that development, the following is important to understanding the basic idea of general intelligence. This idea became prominent in the early part of this century and continues to be widely accepted. The idea of general intelligence is being replaced, however, in response to empirical evidence; the replacement is the idea of several intelligences, which is the foundation for the WJ-R.

Root Concepts

The origins of ideas about human intelligence are fuzzy. Although the modern concept of general intelligence came into our language only in the latter part of the last century, descriptive terms, such as wise, clever, and capable, can be traced to the Greeks during the Age of Pericles (Spearman, 1937). Through the ages, there have been attempts to describe features of thinking that distinguish humans from other animals and distinguish particularly wise and capable individual humans from others not so wise and capable. Features frequently identified include the ability to abstract, to adapt, to comprehend what covaries with what and thus form categories, concepts, and ideas, to label concepts with words and use language, to draw valid conclusions, to anticipate, to combine ideas in novel ways, to conceive universal ideas, to. . . . The list could go on because many, many features have been discussed. Spearman concluded that the essence of intelligence is a capacity for comprehending relations and correlates.

Concepts Based on Operational Definitions

Using repeatable operations to define objective tests, theories about human intelligence were translated into measurements. This became a reality only in this century. There were demonstrations in the 1890s of how responses to questions (test items) could be grouped together to provide indications of memory, knowledge, verbal comprehension, and reasoning. The first measure of intelligence was based on this research. At the request of the French Ministry of Education, Binet and Simon (1905) put together a set of items to form a single test designed to distinguish between slow and normal learners in the early years of schooling. The measure was based on the sum of the number of questions a child answered correctly relative to the average score of children the same age. This measure of a mixture of abilities came to be known as an intelligence quotient, or IQ, and then merely as intelligence.

The Binet-Simon test became the model for constructing almost all of what were called intelligence tests. It proved to be useful to the French Education Ministry for making decisions. It provided information about the abilities of children that otherwise had to come from judgments by teachers. Such judgmental measures were subjective and expensive while the measures obtained with the test were more objective and relatively inexpensive.

Word of the test and its success spread to the United States where, at Stanford University, Terman (1916) translated the Binet-Simon into English. The translation came to be known as the Stanford-Binet and as a measure of IQ. Immense effort was devoted to defining in words the IQ, or intelligence, that this test was said to measure. These definitions and applications of the test generated controversy. New tests were constructed to be improvements on the Stanford-Binet. The new tests generated more efforts to define intelligence and more controversy. The resulting multiplicity of tests and definitions led some to opine that intelligence is nothing more than whatever a particular intelligence test measures (Boring, 1923).

Theory, Tests, and Measurements Deriving From Spearman

In addition to the flurry to redo and redescribe tests of the Binet-Simon variety, many experiments were done to refine understanding of the capabilities that were measured by various IQ tests. The results of these experiments were used to develop further theories of intelligence. The theory put forth by Spearman (1904, 1927) was particularly influential.

Spearman's theory of *g* (to get away from unwanted connotations of the term intelligence) is based on a mathematical-statistical model that provides a test for hypotheses derived from the theory. If measures of cognitive capabilities, such as memory, reasoning, and knowledge, all indicate the capacity for eduction of relations and correlates but indicate no other basic capacity in common, then the intercorrelations among these measures can be reproduced mathematically with measures of only the one common factor, *g*. If such a model fits data to within chance variation, then there is statistical support for the substantive theory of *g*. The model indicates how tests can be constructed and grouped to measure the cognitive capabilities of *g*.

Spearman's theory is basically an essence theory in which the essence is comprehension of relations and correlates. This ability to abstract is at the base of concept formation, language formation, and comprehension. For example, by comprehending the various relations that define dogs, one can abstract the concept *dog*, identify entirely new instances of *dog*, affix a language label to represent this concept, and reason about various features of dogs and other creatures.

Spearman's essence theory has also been influential in the development of several compound theories of intelligence. Rimoldi (1948), for example, theorized that several capabilities could form a union that indicates a single factor. The essence of comprehending relations in Spearman's theory is specified by Rimoldi in terms of subprocesses of apprehension (encoding), holding in awareness (short-term memory), operations (of reasoning and classification), and generation of products. These processes may combine in particular proportions and perhaps in a particular order as postulated by, for example, Sternberg (1985). Such a theory is also similar to one for a particular form of intelligence, *Gf* (Horn, 1965), that is part of the theoretical framework for the WJ-R.

Many tests, and ideas about tests, were generated to explore the processes that contributed to the capacity for comprehending relations and correlates, and were later included in mixtures designed to measure intelligence. The best known are the matrices tests developed by Spearman (1927) but often referred to as Raven's (1938) Progressive Matrices, the Culture Fair tests (Cattell, 1940; Cattell & Cattell, 1977), and Koh's (1923) blocks — the basis for the block design test of the Wechsler scales (Wechsler, 1981). Other widely used tests of memory span, verbal and figural analogies, verbal and figural classification, and letter series also grew out of research based on Spearman's theory.

The fact that Spearman's theory of *g* was based on a mathematical-statistical model also spawned development of the methods of factor analysis. Much of modern structural theory of cognitive capabilities is based on these methods.

Threats to Theories of General Intelligence

From the moment Spearman (1904) proposed his theory there was opposition and a steady increase in the number of factors proposed to describe human intelligence. Burt (1909, 1911) showed that numerical (*N*) and verbal (*V*) factors, in addition to a general factor, were needed to account for the intercorrelations among the achievement tests that had been analyzed by Spearman. Burt's criticisms and results were replicated by several investigators between 1910 and 1920. In 1924, Burt reported substantial support for memory span (*Ms*), manual (*km*), and scholastic (*Sc*) factors, as well as for *V*, *N*, and a general factor. To this list were added, before

Thurstone's influence, spatial (*S*), perceptual speed (*P*), mechanical reasoning (*Mk*), visualization (*Vz*), and several less clearly indicated group factors (Alexander, 1935; Brigham, 1932; Brown, 1933; Corter, 1952; Cox, 1928; El Koussy, 1935; Kelley, 1928; Patterson & Elliot, 1930). The results of these and other studies (Cattell, 1940; Eysenck, 1982; Rimoldi, 1948; Willoughby, 1927) convincingly demonstrated that the model for *g* does not fit the data of human abilities. More than one common factor is needed to account for the intercorrelations among different measures of cognitive capability.

Efforts to modify *g* theory into a new essence or compound theory of abilities that would conform with the Spearman model (Alexander 1935; Cattell, 1940; El Koussy, 1935; Eysenck, 1982; Rimoldi, 1948; Willoughby, 1927) were unsuccessful. No single theory could be found to account for all the abilities that were assumed to indicate intelligence. There was also disagreement about what, in fact, constituted the essence of intelligence. It was suggested, for example, that a capacity for holding separate ideas and relations in the span of immediate awareness is the essence. There is little doubt that such memory is an important cognitive capability, but the evidence of many studies now makes it clear that such a capacity is not the essence of all other cognitive capabilities that comprise human intelligence.

No one-factor model could be found that would account for the interrelations among indicators of cognitive capability. To do so required several additional common factors. In defense of *g*-factor theory, Spearman argued that the additional common factors were only "swollen specifics." This argument was, however, not convincing. The additional factors indicated important intellectual capabilities that have different relationships to other variables such as education, neurological functioning, and development. Multiple factors are needed to explain the relationships among human abilities.

As scientific evidence on the relationships among abilities accumulated in the 20th century, it became clear that the idea of a single general intelligence was inadequate. The evidence indicated reliable independence among human abilities and theory needed to accommodate this evidence. Indeed, several theories were put forth in response to this evidence. The most prominent were mixture theories, concept-narrowing theories, hierarchical theories, and theories of multiple intelligences.

Response 1: Mixture Theories

Most of the tests that measure a cognitive ability have a positive correlation with all other such tests in almost any sample of people older than five years of age. The correlations may not be large, but they are almost always positive. This phenomenon is called the principle of positive manifold. The exceptions to this principle are highly speeded tests that are, except for speed, of trivial difficulty.

The principle of positive manifold has been used to support the idea that cognitive abilities are indicated by a single intelligence. Some have even speculated that intelligence is measured by all cognitive tests and that almost any mixture of such tests will provide a measure of such a single intelligence. Jensen (1982, 1984) is one of the leading proponents of this position. The sum of scores on any broad collection of cognitive tests, he argues, will provide a good "working definition" of Spearman's *g*. This mixture theory does not indicate what abilities should be represented in the combination.

One major problem with this speculation is that sums of scores on different sets of tests measure different collections of cognitive abilities. One sum does not measure the same collection as the other sums. Even if it were true that all sums measure a single intelligence, a supposition that the principle of positive manifold suggests but is not sufficent to support, each sum measures a different collection of cognitive abilities. So, the rank order of people's scores on mixture measures differ as a function of the collection of abilities measured. Of course, decisions based on the results of such mixture measures must also differ.

To hope to measure one intelligence with different mixture measures is rather like attempting to measure water intake with orange juice and vodka on the one hand, and milk and honey on the other. To be sure, water is common to each mixture measure, but because the components of the mixtures vary, the concoctions measure more than water intake and are unreliable for this purpose.

Similarly, the Stanford-Binet and the Wechsler tests might be said to measure *g*. Because of their components, however, they measure more than *g*; therefore, they indicate different things about cognitive functioning. This kind of measurement does not support Spearman's theory.

Some theorists have tried to specify the components that should comprise mixture measures of intelligence. Thomson (1919, 1948) attempted this in response to criticisms of Spearman's theory. Humphreys (1962, 1974, 1979) also attempted to deal with the proliferation of mixed-measure ability tests. Thomson and Humphreys tried to circumscribe the domain of capacities that indicate intelligence. Humphreys (1971), for example, suggested that intelligence is "... the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual that are available at any one period of time" (p. 31). He proposed that a good mixture measure of intelligence should be a representative sample of this repertoire (Humphreys, 1985, 1989).

Most tests of intelligence commonly used today are mixture measures, but they are not representative samples of the entire repertoire to which Humphreys referred. The abilities sampled in these tests represent the test authors' beliefs about what should be included. This can be seen clearly in Woodcock's (1990) comparison of commonly used tests. The varied mixtures of these tests could support the conclusion that intelligence is whatever one wants it to be.

It is probably impossible to draw a representative sample from the entire repertoire to which Humphreys referred. We can neither circumscribe the population for such sampling, nor specify criteria with which to evaluate what is representative. Thus, mixture measures of general intelligence are bound to be arbitrary.

The number of abilities needed to form a mixture that is representative in the sense Humphreys has specified is too large for sampling. Thousands of human abilities have been identified. Commons (1985) has estimated there might be 800,000 or more such abilities. New abilities are being discovered, and new tests to measure those abilities are being created every day (Humphreys, 1979). There is no way to representatively sample all these abilities.

In sum, mixture measures of intelligence measure different intelligences. Indeed, mixtures within a test used at different ages measure different things. The Stanford-Binet, for example, measures abilities in preschoolers that are different from those it measures in 13-year-old children. No mixture-measure test represents the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies that are considered to be intellectual in nature.

Response 2: Narrow the Concept

Rather than deal with the evidence that different abilities and different mixtures do not indicate the same intelligence, investigators have made up new tests to represent their particular ideas about intelligence. The Porteus (1946) maze, Raven's (1938) matrices, and the Kohs (1923) block tests are examples of this response to the evidence. All three of these tests relate to visualization ability, *Gv*. The Peabody Picture Vocabulary Test — Revised (Dunn & Dunn, 1981), widely used as a measure of intelligence, measures mainly verbal comprehension (*V*), one of the primary abilities that comprise mixture measures of intelligence.

Some investigators have narrowed the concept by specifying particular processes or abilities as the sine qua non of intelligence. For example, Kaufman & Kaufman (1983a, 1983b) specified that intelligence involves sequential and simultaneous processes. They designed tests intended to measure those two processes. The factor analytic evidence does not support this theory, but the tests of the Kaufman Assessment Battery for Children (K-ABC) provide good measures of *Gv* and short-term memory, *Gsm*.

Test samples can be restricted to approximate a one-common-factor solution. This requires careful selection of each test to ensure that it measures a single factor that is also measured by every other test but does not measure any other factors common to other tests. If the common factor represents a convincing theory of intelligence, a good fit for a one-factor model provides substantial support for the theory.

Building on the work of Rimoldi (1948, 1951), Horn (1965) closely approximated a one-factor solution for a battery of tests selected to sample only backward memory span and facets of reasoning. The reasoning sample was restricted to letter series, matrices, conjunctive concept formation, disjunctive concept attainment, common-word analogies, and figural analysis. Each of these tests has been described as an indicator of intelligence. The reasoning tasks require comprehension of relationships among elements and drawing conclusions from understanding those relationships. These features of comprehension of relations and correlates are the sine qua non in Spearman's theory of *g*. The memory task measures the ability to hold the elements of a problem in the span of immediate awareness, which is also prominent in theories of intelligence. The evidence that a one-factor theory nearly fits these data provides support for a limited theory of intelligence.

The theory truly is limited because abilities that are important descriptors of intelligence, such as verbal comprehension, visualization, and quantitative reasoning, do not fit this one-factor model. When the tests that do fit the one-factor model are included in batteries that contain other indicators of intelligence, they indicate a common-factor called *Gf*, or fluid reasoning (fluid intelligence). This factor accounts for only a small proportion of the common variance among cognitive tasks. Most of the tests of the common factor are significantly correlated with at least one other common factor. Thus, while the factor that meets the requirements of the Spearman model indicates an important feature of human cognitive capability, it represents just one feature. Other features are also important (Gustafsson, 1984, 1985; Undheim, 1987).

Response 3: Hierarchical Theories

The discovery that intercorrelations among abilities are almost always positive (made and repeatedly confirmed early in this century) was often regarded as evidence of general intelligence. Although this evidence of the principle of positive manifold is consistent with Spearman's theory, it does not prove it. Positive intercorrelations are required in Spearman's model, but they do not distinguish it from the models of most other theories. Measures can be positively correlated for many reasons represented by different common factors. Positive intercorrelations, as such, do not indicate one common factor. This was demonstrated by Spearman (1927) himself.

Often a general factor was extracted from batteries of cognitive tests and interpreted as intelligence without regard for whether or not it was representative of the variability in the battery. Second and third factors were, however, always needed to describe the residual covariability remaining after the first factor had been partialled out. The model for such calculations is hierarchical. If the factors thus calculated are regarded as indications of the organization among abilities, the implicit theory is also hierarchical.

Such factoring and interpretation was very common in much of the research of the first half of this century. It resulted in, as Burt (1949a, 1949b) recalled, a growing conviction that the mind was organized into a hierarchy of levels with a single general factor at the top and factors of varying degrees of generality further down. The more general factor includes the more specialized factors, as countries include counties. Many psychological and educational researchers accepted this kind of theory as a guide for research and use of ability tests.

Such a hierarchical system, however, has two principal limitations:

1. Results from different studies are not stable. The relationships of variables to the first factor vary from one study to another as the proportions of different kinds of tests change. Factors other than the first are bipolar, and the contrasts between positively and negatively related variables change dramatically in replication. For example, one study may contrast verbal and quantitative abilities; but in a second study the contrast is between verbal and spatial abilities; yet in a third study spatial and quantitative abilities form the contrast.

2. The factors, particularly the bipolar ones, rarely correspond to concepts suggested by other evidence, by logical groupings (conjunctions), or by theoretical considerations. For example, it is difficult to suppose that an absence of spatial ability contributes to verbal ability, but a bipolar factor in a hierarchical analysis may call for such an interpretation.

To deal with these problems, either positive or zero correlations of variables with a factor were forced. This was accomplished by redistributing the variance of the general factor and forcing arbitrary solutions.

More recently, hierarchical solutions based on lower level simple structure theory have gained favor (Gustafsson, 1984; Undheim, 1987). Solutions from these theories are in general agreement with the evidence suggesting the existence of several intelligences.

Response 4: Theories of Multiple Capabilities

Since Binet, the evidence of independence among abilities and the inadequacies of general intelligence theories nudged research toward the development of multiple ability theory. As developed by Thurstone (1931, 1935, 1938), this theory provided a dramatic alternative to the hierarchical theory described by Burt. The guiding principle of Thurstone's theory is a simple structure model of multiple common factors. This model represents an entirely different conception of cognitive processes and the organization of human capacities than does Spearman's theory or the hierarchical theories that derived from it.

The simple structure model stipulates that each common factor (cognitive capacity) relates primarily to only a few abilities, and each ability depends on only a few factors. Each of the several common factors in the Thurstone model is thus quite different from the one common factor in the Spearman model, which is expected to be related to all abilities. In contrast to the Spearman model, which stipulates that one factor describes all the intercorrelations among abilities, the Thurstone model stipulates that multiple common factors describe all the intercorrelations.

Building on the ideas of simple structure and multiple factors, research since the 1930s has been directed toward constructing a system of concepts, which are operationally indicated by factors, that describes human cognitive capabilities. This system accomplishes its purpose with two basic assumptions.

First, it is assumed that all the abilities that describe human intelligence are required in the tasks that are sampled in factor analytic studies. This does not require a test to measure every cognitive capacity, but it does require that a test sample these capacities in one way or another. A test, then, may measure several basic capacities of intelligence. A second assumption is that a simple-structure, reduced-space, common-factors system is a paramorphic model of the organization of human capacities.

Theories of Multiple Intelligences

Research based on these assumptions led to Thurstone's system of primary abilities and later to an expanded version of this system called Well Replicated Common Factor (WERCOF) primary abilities. Guilford (1967) also attempted to identify a comprehensive taxonomy. Inadequacies in these conceptualizations provoked efforts to develop the *Gf-Gc* theory.

Thurstone's Theory of Primary Abilities

Thurstone (1935) introduced his theory and tested his model with a reanalysis of Brigham's (1932) data. He located common factors of verbal comprehension (*V*), numerical calculation (*N*), and visual imagery with *g* being conspicuous by its absence. In his famous "Primary Mental Abilities" study, Thurstone (1938) added word fluency (*Fw*), inductive reasoning (*I*), general reasoning (*R*), and syllogistic reasoning (*Rs*) to the list of discovered primary abilities including *V*, *N*, *S*, *M*, and *P*. In 1940, Thurstone replicated the findings of the 1938 investigation.

WERCOF Primary Abilities

Comprehensive reviews of the work supporting the WERCOF system have been provided by Ekstrom, French, and Harman (1979), Horn (1982a, 1982b), and Hakstian and Cattell (1974). The tests used in these studies are similar in breadth to subtests used in mixture measures of IQ. Common-factor studies of many such subtests indicated distinct primary abilities that can be replicated. Each such ability was indicated by three or more somewhat different, but covarying, subtest measures. Brief descriptions of these abilities are provided in Table 7-1.

Table 7-1.
Tests Classified in Terms of the Primary Ability Best Measured

Test	Primary Ability	
Verbal: Vocabulary	V	Verbal Comprehension
Following Written Instructions	Ig	Integration
Following Spoken Instructions	CMS	Cognition of Semantic Systems
Reasoning: Verbal Analogies	CMR	Cognition of Semantic Relations
Problem Definition	NMS	Convergence in Semantic Systems
Assessing Everyday Arguments	Sp	Sensitivity to Problems
Deduction	Rs	Syllogistic Reasoning
Story Problem Representation	Es	Estimation
Numerical: Computations	R	General Reasoning
Math Bugs	ESS	Evaluation of Symbolic Systems
Computational Skill: Fluency	N	Number Facility
Inductive Reasoning	I	Induction
Set Recognition	ESC	Evaluation of Symbolic Classes
Set Relationships	ESR	Evaluation of Symbolic Relations
Visual Conceptualization	S	Spatial Orientation
Visual Manipulation	Vz	Visualization
Visual Constancy	Sc	Spatial Scanning
Analytic Perception	Cf	Flexibility of Closure
Perception: Moving Windows	Cs	Speed of Closure
Digit Span	Ms	Span (nonsense) Memory
Tone Span	Md	Auditory Span
Location Span	Mv	Visual Memory
Memory: Word Span	Mm	Meaningful Memory
Memory Acquisition	Ma	Associative Memory
Design Memory	Mo	Memory for Order
Spatial Organization	Mv	Visual Memory
Memory Retention & Learning	MMR	Memory for Semantic Relations
Verbal Productive Fluency	Fw	Word Fluency
Intermodal Transfer Fluency	Sr	Semantic Redefinition
Effective Problem Solving Strategy	EMI	Evaluate Semantic Implications
Word Parsing	NMC	Converge Semantic Classes
Phonetic Decoding	NMI	Converge Semantic Implications
Word Meaning Association	CMC	Cognition of Semantic Classes
Attention	P	Perceptual Speed

The most highly correlated subtests have been used to measure WERCOF abilities; therefore, the intercorrelations among the component subtests tend to be large. This narrows the definition of the abilities relative to the IQ measures discussed previously.

Memory span (*Ms*) is an example of a WERCOF ability. It is indicated by the following tests:

Digit Span. Immediately after numbers are presented one at a time on a TV screen, reproduce (using a number pad or keyboard) the numbers in the order in which they were presented.

Tone Span. Given that a high note is represented by one key and a low note by another key, press the keys in sequence to indicate a sequence of high and low tones heard just moments before.

Location Span. Immediately after a symbol (e.g., a fly) appears sequentially in the squares of a 3-by-3 matrix of squares, reproduce the order (the flight of the fly) by pressing keys corresponding to the 9 squares of the matrix.

The intercorrelations among measures of the *Ms* are typically in the .65 to .75 range when the reliabilities of the measures are of the order of .75 to .85.

Table 7-2 lists established WERCOF abilities. They vary in breadth. Not all of them are as narrow as *Ms.* In general, however, they are of roughly the same breadth, and they are narrow in comparison to most measures of IQ. Since we cannot confidently specify the best level of breadth on which to base studies of these abilities, Table 7-2 provides — at the primary level of breadth — a comprehensive indication of the cognitive capabilities humans possess.

Guilford's Structure-of-Intellect Organization

Guilford's attempt to organize human abilities can be described partly in terms of breadth (broad-to-narrow), partly in terms of conditions of administration (choose from among choices, open-minded, etc.), and partly in terms of other, less popular, features. Guilford (1967) attempted to identify basic features and develop a comprehensive taxonomy based on them. The following labels indicate the major categories of Guilford's system:

Contents	Products	Operations
F Figural	*U* Units	*E* Evaluation
S Symbolic	*C* Classes	*N* Convergent Production
M Semantic	*R* Relations	*D* Divergent Production
B Behavioral	*S* Systems	*M* Memory
	T Transformations	*C* Cognition
	I Implications	

According to Guilford, an ability test includes a combination of one operation, one content, and one product. For example, a multiple-choice vocabulary test is described as Cognition of Semantic Units (CMU). Guilford defines cognition as awareness, immediate discovery or rediscovery, or recognition of information in various forms. "Units are relatively segregated or circumscribed items of information having 'thing' character. ... Semantic information is the form of meanings to which words commonly become attached ... although we must recognize [also] that much semantic information is nonverbalized" (pp. 71 & 227).

The four contents, five operations, and six products can produce 120 3-way combinations. So, 120 tests could be constructed on the basis of Guilford's system. Is this the number of distinct abilities humans possess?

Table 7-2.
First-Order (Primary) Mental Abilities After Ekstrom, French, & Harman (1979)

		French-Thurstone Symbol	Guilford Symbol
Short-term Apprehension and Retrieval Abilities			
Associative Memory	When presented with one element of previously associated but otherwise unrelated elements, recall the associated element.	Ma	MSR
Span Memory	Immediately recall a set of elements after one presentation.	Ms	MSU
Meaningful Memory	Immediately recall a set of items that are meaningfully related.	Mm	MSR
Chunking Memory	Immediately recall elements by categories into which elements can be classified.		MMC
Memory for Order	Immediately recall the position of an element within a set of elements.		MSS
Long-term Storage and Retrieval Abilities			
Associational Fluency	Produce words similar in meaning to a given word.	Fa	DMR
Expressional Fluency	Produce different ways of saying much the same thing.	Fe	DSS
Ideational Fluency	Produce ideas about a stated condition or object — e.g., a lady holding a baby.	Fi	DMU
Word Fluency	Produce words meeting particular structural requirements — e.g., ending with a particular suffix.	Fw	DMR
Originality	Produce "clever" expressions or interpretations — e.g., titles for a story plot.	O	DMT
Spontaneous Flexibility	Produce diverse functions and classifications — e.g., uses for a pencil.	Xs	DMC
Delayed Retrieval	Recall material learned hours before.		
Visualization and Spatial Orientation Abilities			
Visualization	Mentally manipulate forms to "see" how they would look under altered conditions.	Vz	CFT
Spatial Orientation	Visually imagine parts out of place and put them in place — solve jigsaw puzzles.	S	CFS
Speed of Closure	Identify Gestalt when parts of whole are missing.	Cs	CFU
Flexibility of Closure	Find a particular figure embedded within distracting figures.	Cs	NFT
Spatial Planning	Survey a spatial field and find a path through the field — e.g., pencil mazes.	Ss	CFI
Figural Adaptive Flexibility	Try out in possible arrangements of elements of visual patterns to find one arrangement that satisfies several conditions.	Xa	DFT
Length Estimation	Estimate lengths or distances between points.	Le	
Figural Fluency	Produce different figures using the lines of a stimulus figure.		DFI
Seeing Illusions	Report illusions of such tests as Muller-Lyer, Sanders, & Poggenforff.		DFS
Abilities of Listening and Hearing			
Listening Verbal Comprehension	Show understanding of oral communications.		
Temporal Tracking	Demonstrate understanding of sequence of auditory information — e.g., reorder a set of tones.		
Auditory Relations	Show understanding of relations among tones — e.g., identify separate notes of a chord.		
Discriminate Patterns of Sounds	Show awareness of differences in different arrangements of tones.		
Judging Rhythms	Identify and continue a beat.		
Auditory Span Memory	Immediately recall a set of notes played once.		
Perception of Distorted Speech	Demonstrate comprehension of language that has been distorted in several ways.		

Table 7-2. *(Cont.)*
First-Order (Primary) Mental Abilities After Ekstrom, French, & Harman (1979)

		French-Thurstone Symbol	Guilford Symbol
Acculturational Knowledge Abilities			
Verbal Comprehension	Demonstrate understanding of words, sentences and paragraphs.	V	CMU
Sensitivity to Problems	Suggest ways to deal with problems — e.g., improvements for a toaster.	Se	EMI
Syllogistic Reasoning	Given stated premises, draw logically permissible conclusions even when these are nonsensical.	Rs	EMR
Number Facility	Do basic operations of arithmetic quickly and accurately.	N	NSI
Verbal Closure	Show comprehension of words and sentences when parts are omitted.		
Estimation	Use incomplete information to estimate what is required for problem solution.		CMI
Behavioral Relations	Judge interaction between people to estimate how one feels about a situation.		CBI
Semantic Relations: Esoteric Concepts	Demonstrate awareness of analogic relationships among abstruse bits of information.		CMR IMR
Mechanical Knowledge	Information about industrial arts — mechanics, electricity, etc.	Mk	
General Information	Science, Humanities, Social Sciences, Business	Vi	
Abilities of Reasoning Under Novel Conditions			
Induction	Indicate a principle of relationships among elements.	I	NSR
General Reasoning	Find solutions for problems having an algebraic quality.	R	CMS
Figural Relations	Demonstrate awareness of relationships among figures.		CFR
Semantic Relations: Common Concepts	Demonstrate awareness of relationships among common pieces of information.		CMR IMR
Symbolic Classifications	Show which symbol does not belong in a class of several symbols.		CSC
Concept Formation	Given several examples of a concept, identify new instances.		CFC
Speed of Thinking Abilities			
Perceptual Speed	Under highly speeded conditions, distinguish similar visual patterns and find instances of a particular pattern.	P	ESU
Correct Decision Speed	Speed of finding correct answers to intellectual problems of intermediate difficulty.		
Writing and Printing Speed	As quickly as possible, copy printed or cursive letters or words.		

There are three answers to this question. First, 120 is probably an underestimate of the number of separable human intellectual abilities. Second, it is doubtful that the 120 abilities of the Guilford system are reliably and validly distinct. Third, the Guilford system is probably not a good system for describing human cognitive capabilities. Let us examine the basis for these statements.

The Guilford system derives partly from logic and partly from empirical research. The main support for the system is in the Ekstrom et al. (1979) integration of the results of first-order factor analytic studies — the evidence of the WERCOF abilities. That research indicates 24 "well-replicated" ability factors and some 13 factors for which there is some, but not well-replicated, evidence.

It can be seen in Table 7-2 that some 25 of Guilford's 3-way combinations are associated with abilities for which there is empirical evidence from factor analytic studies. This is partly because Guilford based his system, *post hoc*, on these factor analytic findings. The theory is not supported by these findings but is an outgrowth of them.

Studies by Guilford and his co-workers designed to add evidence for the theory have been hampered by a lack of objectivity (Carroll, 1972; Horn, 1967; Horn & Knapp, 1973, 1974; Humphreys, 1962; Undheim & Horn, 1977). Their results do not provide unequivocable support for either retaining or rejecting the major hypotheses of Guilford's system.

Some of Guilford's 3-way combinations represent empirically based distinctions among primary mental abilities, but many are only logical indications of ways to construct tests, not indications of distinct human abilities. As Humphreys (1962) pointed out, the facets of the Guilford system "... are not psychological as defined. They should be useful to the test constructor, [but] they do not need to make a behavioral difference." (p. 480) The difficulties with Guilford's Structure-of-Intellect system are similar to those of other efforts to find an empirical basis for differentiating human abilities.

Preliminary Development of Gf-Gc Theory

The WERCOF system of primary abilities proved to be intractable and inadequate because very little evidence was developed to indicate that the abilities were developmentally and functionally distinct. While they represent distinct patterns of covariation, there is little firm evidence that those patterns correspond to distinct genetic influences, developmental histories, or organizations within the brain.

Similarly, Guilford's system proved to be inadequate because it lacked empirical support. Although there were many Structure-of-Intellect studies in the 1960s and 1970s, they do not provide a sound basis for rejecting or retaining the theory's hypotheses (Carroll & Horn, 1981; Undheim & Horn, 1977). The evidence suggests that, while the system provides a provocative basis for construction of cognitive tests, it is not a good representation of how abilities are organized in function or development.

Inadequacies in these primary ability systems provoked efforts to develop a better system. The metatheory and methods of Thurstone were applied to studies of the relationships among exemplars of the WERCOF abilities. A second-order system was defined that took into account the primary ability interrelationships. Linked with evidence from studies of age changes and age differences in abilities and studies of relationships between ability changes and brain damage, this system became the basis for a developmental theory of nine broad intelligences now known as *Gf-Gc* theory.

Cattell (1941) first stated the theory that guided this research in a symposium on age differences in abilities. The theory was based on evidence that two major classes of influences affected the normal development of cognitive abilities. First, there are influences associated with educational-cultural opportunities; and second, influences stemming from genetic factors and physiological-neurological functioning. (These two classes of influences are not entirely uncorrelated.) This evidence led Cattell to the view that observed individual differences in cognitive ability tests depend on the following:

G: variations in innate gene endowment,

dG: variations in environmentally-produced development of general ability,

C: variations in the closeness of the individual's cultural training and experience to the cultural medium in which tests are expressed,

t: variations in familiarity with tests and test situations,

fr: fluctuations in the effective expression or application of ability through varying strength and direction of volition,

s: variations in an ability specific to the test, and

c: chance errors of measurement.

In describing the *G* of this expression (not the *g* of Spearman's theory) Cattell (1941) stated a culture-fair concept of intelligence. The essence of the concept is that each individual has a particular capacity to perceive complex relations. This capacity exists independently of the field, skill, or knowledge in which it is most fully exercised. It is an abstraction from the field in which it is measured, as energy is distinct from the particular physical, chemical, or electrical system in which it happens to exist.

In describing the *dG* and *C* terms, Cattell introduced the notion of "crystallized" intelligence. First, he argued that *dG* and *C* probably would not be a measurement distinction. He reasoned that if the (*G*) detachment of the power from its manifestations is possible, then it is correct to ask how far that power, as such, can be impaired or augmented by environmental influences. This emphasizes that the environmentally produced change in intelligence, *dG*, is within the subject, and the potency of the influences represented by *dG* and *C* would be a function both of the time of exercise and the amount of culture encountered. If the information and skills were acquired very quickly, the effort on intelligence might not be so great as if they were exercised for a long time. Again, if they were acquired in early years when they seemed difficult, they would offer more exercise than to a mature intelligence. Finally, if they are acquired after the age at which mental capacity reaches biological maturity, their effect should be very small.

Cattell (1943) later argued that general ability was of two kinds: fluid ability, which manifests itself in relation perception, in speeded performances, and in new situations; and crystallized ability, which manifests itself only in relation perception in known material and speeded performance. Continued research (Cattell, 1950) demonstrated that these two abilities also show different developmental patterns of change; for example, fluid ability declines after about age 23, which leaves crystallized ability like a dead coral formation and maintains, except where brain injury occurs, the levels of the original fluid ability. The age-difference curves for intelligence refer to basic powers; the cumulative result of learning by these powers reaches a plateau. Consequently, where good performance is a matter of wide information, wisdom, and foresight founded in experience, of shrewd tactics, and, especially, of truths of living acquired from trial-and-error learning rather than didactic teaching, the older person has the advantage.

Gf-Gc theory derives, in part, from studies of how abilities are affected by brain damage and development, particularly in adulthood. The theory is based largely on studies of the covariability among abilities in factor analytic studies. It is a second-order system among the primary factors — a system of factors among factors. The intercorrelations that point to the system can be seen in many studies (reviewed in, for example, Horn, 1968, 1970, 1972, 1976, 1978, 1980, 1982a, 1982b, 1987, 1988, 1989a, 1989b; Horn & Donaldson, 1980).

In Horn's (1965) dissertation, reported also in Horn and Cattell (1966b), concepts of short-term apprehension and retrieval (*SAR*, later *Gsm*), fluency of retrieval from long-term storage (*TSR*, later *Glr*), processing speed (*Gs*), and visual processing (*Gv*) were introduced. Auditory processing (*Ga*) was added in the Horn and Stankov (1982) studies and correct decision speed (*CDS*) was identified in the Horn, Donaldson, and Engstrom (1981) studies. Quantitative ability (*Gq*) was indicated by the results of a number of studies (Horn, 1988, 1989b).

Major results indicating the structural features of the theory are also in Cattell and Horn (1978), Hakstian and Cattell (1978), Horn and Bramble (1967), Hundal and Horn (1977), Rossman and Horn (1972), Shucard and Horn (1972), and the recent studies of Carroll (1989), Gustaffson (1984), Undheim (1987), and Woodcock (1990). These recent studies indicate the generality of the system. Carroll's results stem from 461 separate studies done by almost as many investigators. Woodcock's findings are based on a standardization sample of 6,359 subjects spanning an age range from children to the elderly. Gustaffson's sample is Swedish, and Undheim's, Norwegian. The results of all this research indicate that the WERCOF system can be organized in terms of nine dimensions. Described in capsule form, these abilities are:

Fluid Reasoning (*Gf*)
Measured by tests that require inductive, deductive, conjunctive, and disjunctive reasoning to understand relations among stimuli, comprehend implications, and draw inferences.

Acculturation Knowledge (*Gc*)
Also called comprehension-knowledge, it is measured by tests that indicate the breadth and depth of the knowledge of the dominant culture.

Quantitative Reasoning (*Gq*)
Measured by tests that require understanding and application of the concepts and skills of mathematics.

Short-term Apprehension-retention (*Gsm*)
Also called short-term memory, it is measured with a variety of tests that require maintaining awareness of and recalling elements of immediate stimulation — i.e., events of the last minute or so.

Fluency of Retrieval from Long-term Storage (*Glr*)
Also called long-term memory, it is measured by tests that indicate consolidation for storage and require retrieval, through association, of information stored minutes, hours, weeks, and years before.

Visual Processing (*Gv*)
Measured by tests that involve visual closure and constancy and fluency in "image-ing" the way objects appear in space as they are rotated and flip-flopped in various ways.

Auditory Processing (*Ga*)
Measured by tests that involve perception of sound patterns under distraction or distortion, maintaining awareness of order and rhythm among sounds, and comprehending groups of sounds, such as chords, and the relationships among such groups.

Processing Speed (*Gs*)
Part of almost all intellectual tasks (Hertzog, 1989), it is measured most purely by tests that require rapid scanning and responding to intellectually simple tasks that almost all people would get right if the task were not highly speeded.

Correct Decision Speed (*CDS*)
Measured by tests that require quick answers based on thinking.

Almost all of the abilities measured by IQ tests and neuropsychological batteries are accounted for by these nine dimensions. Such tests may not necessarily claim to involve these abilities, but what they measure reliably can be predicted by the nine factors of *Gf-Gc* theory.

The *Gf-Gc* system differs from the primary mental abilities system in that each *Gf-Gc* factor is broader than the similar factor in Thurstone's system. That is, a *Gf-Gc* factor is comprised of and represents more elementary abilities. In a hierarchical system, the *Gf-Gc* factors are higher order organizations of the lower order primary mental abilities organizations (Hakstian & Cattell, 1978). For example, *Gc* includes the primary *V*, *Vi*, *Rs*, and *N* abilities. It also includes knowledge as measured by achievement tests (Woodcock, 1990). Other *Gf-Gc* factors similarly involve a combination of several of the primary mental ability factors.

Component abilities of a *Gf-Gc* factor are different from each other. This indicates breadth. But they are also similar relative to the abilities of other *Gf-Gc* factors. The similarity is responsible for the factor. Finding the factors and replicating that finding in different studies indicates the distinctiveness of a cognitive function.

The factors are construct independent. A best-weighted linear combination of any set of eight factors does not fully predict the reliable part of the ninth factor. This evidence shows that each factor measures a cognitive function not measured in the other factors. Additional evidence indicates the construct validity of the factors. They predict important different criteria, stem from different sets of determinants (including different sets of genes), and are affected differently by injuries, child rearing, education, drug use, and other environmental or lifestyle influences.

Each *Gf-Gc* factor is broad enough to represent a concept of intelligence and involves abilities that are important in defining intelligence. Each factor, however, is distinct from the others when viewed psychometrically and developmentally and in terms of neurology, predictability, and genetic analyses. Thus, each *Gf-Gc* factor represents a distinct concept of intelligence: *Gf-Gc* theory is a theory of several intelligences, rather than a theory of intelligence.

The Gf-Gc System

The different intelligences of *Gf-Gc* theory will be described first, followed by evidence that indicates their separate construct validities.

Knowledge or Crystallized Intelligence, Gc

This factor indicates the extent to which an individual has appropriated, for personal use, the intelligence of a culture. The following kinds of measures are indicative of this broad form of intelligence.

General Information
Measures of *Gc* involve knowledge about areas of scholarship, such as the humanities, business, history, the social sciences, the physical sciences, and mathematics, as well as knowledge about avocational aspects of culture like books, movies, music, and sports. The WJ-R tests of Science, Social Studies, and Humanities are examples of such measures.

Verbal Knowledge
Usually measured with vocabulary tests, verbal knowledge can be assessed with virtually any test that measures understanding of word meanings. The WJ-R Oral Vocabulary test is a measure of this ability.

Problem Definition/Representation
When given a verbal problem that includes possibly relevant information, the individual indicates what information is required, what is not required, what represents constraints, and what represents approaches. This indicates that *Gc* is an important aspect of metacognitive, planning, and executive abilities.

Assessing Everyday Arguments and Evidence
When given an argument, the individual decides if the conclusion is warranted or the reasoning is cogent. This ability is also an indication of *Gf* ability.

Analogies Reasoning as measured with Verbal Analogies
When given two related words, the individual identifies a word that has the same relationship to a given word. For example, *ax* is to *cut* as *shovel* is to.... This ability is also an indication of *Gf* ability.

Syllogistic Reasoning
Often regarded as a measure of deductive reasoning, such a test requires the individual to determine whether or not conclusions logically follow premises and arguments. Again, measures of this ability will usually involve *Gf* as well.

Story Problem Representation
When given a verbal problem in which numerical calculations, such as addition, subtraction, multiplication, and division, are needed, the individual chooses the series of calculations that leads to a correct solution. This type of measure is also an indication of *Gq* ability.

Those examples of tasks that measure *Gf* or *Gq* as well as *Gc* are indicative of alternative mechanisms or the idea that ability tasks can be dealt with by exercising one or more separate abilities. For example, to assess the validity of an everyday argument that might appear in a daily newspaper, an older adult is likely to use *Gc* abilities, whereas a younger person with less of the *Gc* knowledge would rely on *Gf* reasoning. Similarly, a person with well-developed *Gq* ability would use knowledge of mathematics to solve an arithmetic story problem while a person with more general *Gc* knowledge would do reasonably well solving the same problem using approximations derived from the "common sense" of general information.

Broad Reasoning or Fluid Intelligence, Gf

Gf is a mixture of distinct processes. While no compound theory of *Gf* has yet been accepted, evidence suggests that such a theory may be forthcoming (as discussed in Gustaffson, 1985; Undheim, 1987). *Gf* reasoning involves many mental operations such as identifying relations, drawing inferences, concept formation, concept recognition, identifying conjunctions, and recognizing disjunctions. Tests of *Gf* should not depend on knowledge that is available to some and not others because such a test would measure *Gc* (or *Gq*). In other respects, however, measures of *Gf* need not be highly similar. The following are examples of tests that indicate *Gf.*

Inductive Reasoning using Series Tests
In a Letter Series test, the task is to indicate the next letter in a series; for example, G H J M Q V _?_. Number Series and Figure Series tests are similar in construction.

Matrix Reasoning with Visual Patterns
When given a set of figures that change in systematic ways across columns and down rows of a matrix, the individual indicates what figure should appear in the lower-right cell of the matrix.

Interpreting Verbal Reasoning Pertaining to Visual Patterns
When given a figure in which circles, squares, and triangles overlap in complex ways, the individual makes a dot so that it is in, for example, the circles and triangles, but not in the squares.

Classification
The individual identifies an element (figure, word, letter set) that does not belong with other elements. For example: Which word in the following set of words does not belong with the others? rose, rock, carrion, perfume.

Common Features (known as Remote Associations)
The individual indicates a common feature in otherwise diverse elements. For example: What is common to the following? bathtub, boxing, wedding, rosy. The *Glr* alternative mechanism helps performance on such a test.

Conjunctive Reasoning with Set Recognition
The individual decides which items do, and do not, belong together. The WJ-R Concept Formation test is an example of this kind of measure.

Analogies Reasoning
This was described in discussing *Gc.* It exemplifies the concept of alternative mechanism. Verbal Analogies can be a reasonably good measure of *Gf,* rather than *Gc,* only if the words of the analogies are equally familiar or equally esoteric for all examinees. Thus, the relationships among the words, not knowledge, introduce variance in individual differences in correctly solving the problems.

The alternative mechanisms of *Gf,* rather than *Gc,* can be measured with several kinds of tests. The key is that the task require reasoning rather than knowledge. Syllogistic Reasoning and Assessing Everyday Arguments and Evidence are examples of such tests. Also, if word problems do not emphasize mathematical knowledge, then a test made up of such problems will measure *Gf* rather than *Gc* or *Gq.*

Broad Visual Intelligence, Gv

Tasks that call for fluent visual scanning, Gestalt closure, minds-eye rotations of figures, and ability to see reversals measure a kind of visual intelligence that is quite separate from *Gf* and *Gc.* The following abilities are indicative of *Gv:*

Visual Manipulation based on Paper Folding
The task is to perform mental operations that simulate the folding of a piece of paper, punching a hole through the folded paper, unfolding the paper, and identifying how the holes would appear.

Analytic Perception in Gottschaldt Figures or Hidden Figures
The task is to identify whether or not the outline of a particular figure can be traced when presented in conjunction with superfluous lines. In addition to being used as a measure of visualization, the test has been used to indicate Field Independence (Witkin & Goddenough, 1981). Field Independence was measured initially with Rod-and-Frame and Tilting-Room tests. These measures indicate inclination to use cues from one's own body, rather than environmental cues surrounding an object, as a basis for locating the object in space. Witken and his colleagues presented evidence suggesting that measures obtained with the Rod-And-Frame and Tilting-Room tests covary with measures based on the Hidden Figures test. Evidence is also presented to suggest that performance on these measures is associated with being male rather than female. Work by Bock and Kolakowski (1973) suggests that some of the individual difference in *Gv* stems from a sex-linked major-gene influence.

Visual Constancy
The task is to visualize how a figure looks as it rotates in space. In most studies of *Gv* and its components, a count of correct responses under slightly speeded conditions provides the measure. In the research of Shepard and Metzler (1971), reaction time (*RT*) to correctly visualize rotational outcomes was the principal measure; however, it is not known whether *RT* measures are indicative of *Gv.* Studies of visual abilities of pilot and navigator trainees in the US Air Force (Guilford & Lacy, 1947) suggest that *Gv* is not so much visualization speed as it is the ease or fluency of visualization.

Gestalt Closure
The task is to fill gaps to complete a view that is obscured as if seen through a fog or as if parts of a figure have been erased. The WJ-R Visual Closure test is an example of this kind of measure.

Design Memory
Although usually used to measure short-term memory, this test may be a good indicator of *Gv* because the task requires the individual to visualize the steps needed to draw a figure.

It is difficult to distinguish *Gv* from *Gf* if visual tests can be performed by reasoning. If tasks do not require fluent visual thinking as much as they do reasoning, they indicate *Gf.* If, however, tasks require fluent "seeing" of how figures can change in appearance as they move in space or as the perspective from which they are viewed changes, they indicate *Gv.*

Broad Auditory Processing or Auditory Intelligence, Ga

Abilities in this area are involved in almost every cognitive task that requires solving problems that are presented auditorily. Some of the tasks that indicate this factor are unique to the auditory mode; others are similar to visual tasks used to measure other *Gf-Gc* abilities (Horn & Stankov, 1982; Stankov & Horn, 1980). The following are examples of tests that may be used to indicate this factor:

Memory for Pitch (Wing, 1955)
In this test, a tune is presented twice. In the second presentation, one note has been changed. The task is to indicate which of the notes in the tune has changed.

Sound Blending
The task is to identify words that are presented auditorily with pauses between phonemes. For example, an item might be to identify the sounds /p/-/oo/-/t/ as the word *put*. This is one of the tests in the WJ-R.

Reordering Nonsense Syllables
The subject is given three nonsense syllables followed by a presentation of the same three syllables in a different order. The task is to indicate the difference between the two presentations.

Word Attack
The task is to pronounce nonsense words in accordance with the rules that govern pronunciation in English. This is a test in the WJ-R.

Maintaining Rhythms (Seashore, Lewis, & Saetvelt, 1960)
The subject is given pairs of rhythmic patterns of varying lengths. The task is to indicate whether the patterns in the pair are the same or different.

Tonal and Chord Series
The task in tonal items is to select from three choices a tone that will best continue a series of tones that has been presented. In chord series, the task is the same except that chords are presented rather than tones.

Short-Term Acquisition and Retrieval, Gsm

Studies of information processing relate most directly to this broad ability. Tests such as Digit Span, Tone Span, and Location Span, as well as the following tasks, indicate *Gsm*.

Word Span
If this task measures differences in an individual's familiarity with words, it indicates *Gc* rather than *Gsm*. If, however, the words are equally familiar or obscure to all respondents, the task indicates *Gsm*. The WJ-R Memory for Words test is an example of this type of measure.

Recency Memory
After being presented with stimuli in a particular temporal order, most individuals recall the most recently presented material better than they do items that were first or in the middle of the string. Individual differences in this phenomenon are associated with *Gsm*.

Design Memory
Usually a measure of *Gv,* this task can also indicate *Gsm* since good performance requires immediate memory as well as visualization skills.

Gsm includes several distinct processes with separate construct validities. Walsh (1986) has distinguished the distinct processes with an analytic visual backward-masking procedure. *Gsm* ability is affected by attentional processes (Craick & Byrd, 1982; McDowd & Birren, 1989). When the ability to attend is reduced, there is inevitable reduction in *Gsm* ability.

Long-Term Storage and Retrieval, Glr

Measures of *Glr* have been prominent in research on creativity (Cave, 1970; Cropley, 1972; Getzels & Jackson, 1962; Guilford, 1967; Mednick, 1956; Rossman & Horn, 1972; Torrance, 1972; Vernon, 1972). It is now recognized that creativity, like intelligence, is a complex of several functions, not simply the abilities of *Glr.*

There is good evidence indicating that the way information is organized at the time of encoding for memory storage predicts how it will be retrieved at a later date (Bower, 1972; Norman, 1979). A possible interpretation of this research is that *Gsm* should not be distinguishable from *Glr* — that *Gsm* should be highly indicative of retrieval from storage in long-term memory. The establishment of separate factors for *Gsm* and *Glr,* however, indicates that individual differences in immediate apprehension (*Gsm*) are independent of individual differences in the ability to retrieve information from memory storage (*Glr*). Different processes must be involved.

It can be difficult to distinguish *Glr* from *Gsm* if the measures of *Glr* do not require retrieval of information stored at least several minutes before or, preferably, several hours or days before. The retrieval ability of *Glr* can be measured with a variety of tasks.

Retention of Learning after Several Minutes or Hours
Tests of short-term memory can be used to measure *Glr* if they are used after short-term memory has faded. A test can be a good measure of *Glr* when the subject does not expect that recall will be requested after a lapse of time. The Delayed Recall — Memory for Names and Delayed Recall — Visual-Auditory Learning tests from the WJ-R are used to measure this ability.

Memory Acquisition
The task requires the individual to retrieve material, such as a page of words, that has been memorized verbatim. If the emphasis is on breadth of knowledge rather than fluent recall of knowledge, this task mainly measures knowledge abilities of *Gc* rather than *Glr.* To be primarily a measure of *Glr,* the task should require recitation of passages learned several minutes or hours previously. As with other measures of *Glr,* acquisition should not be highly speeded. It is not known whether time to reach a criterion in verbatim learning is a better indicator of *Glr* or *Gc.*

Intermodal Transfer Fluency
The task requires individuals to indicate whether or not a given word provides a good description of a picture. This is a measure of *Glr* because judging whether or not a given word is a good descriptor requires searching storage memory to find other words that might be better.

Verbal Production Fluency
The typical task is to write or say as many words as possible that begin with a particular letter. As is true for all fluency measures, the task must not be highly speeded because the subject must have enough time to sample fully from memory storage. Writing or speaking speed should not produce substantial variance in the measure. Otherwise, the task is mainly a measure of a broad speediness ability or inclination, which is indicative of *Gs*.

Expressional Fluency
The task measures an individual's facility for retrieving from memory storage appropriate and different expressions for an idea in writing or speaking.

Ideational Fluency
The task measures an individual's facility for finding different ways to interpret and write or talk about a particular event, such as a woman boarding a bus.

Associational Fluency
The task measures an individual's ability to retrieve words that are connotatively similar to a given word.

Quantitative Knowledge, Gq

By the time children have reached junior high school, individual differences in a broad range of quantitative skills differ from individual variance in the broad pattern of other knowledge that characterizes *Gc*. In terms of child development and prediction in academic and vocational settings, *Gq* has construct validity that is notably different from the construct validities of the other broad abilities in the *Gf-Gc* system. Thus, *Gq* represents an important feature of cognitive functioning. If quantitative thinking tasks are given under highly speeded conditions, they are likely to measure *Gs* as well as *Gq*. *Gq* can be measured with a variety of tests.

Calculation
Tasks that require the subject to add, subtract, multiply, and divide whole numbers, fractions, and decimals provide a good measure of the range of individual differences in *Gq*, if speed of performance is not emphasized. Among very well-educated people, problems requiring understanding of geometry, trigonometry, and basic calculus will distinguish *Gq* from *Gc*. The WJ-R includes a calculation test as part of the Achievement Battery.

Applied Problems
When more information is provided than is needed for solution, the task is to select the appropriate information and use the appropriate mathematical procedures in calculations. The tasks are similar to what are called word problems. To distinguish *Gq* from *Gf*, the problems should not require difficult or abstract reasoning. To distinguish *Gq* from *Gsm*, the problems should be easily available to the subject so as not to require the abilities of holding information in immediate awareness. The WJ-R includes a test of applied problems as part of the Achievement Battery. The Arithmetic subtests of the Wechsler scales are similar to applied problems, but require the subject to keep the relevant information in mind. For this reason, the Wechsler tests measure *Gsm* as well as *Gq*.

Cognitive Processing Speed, Gs

Speediness in scanning, comparing, inspecting, and becoming aware of the salient features of problems is a pervasive source of individual differences in cognitive tasks. It contributes to the observed variability in almost all timed tests. This contribution can, however, be small relative to the variance produced by other factors if the tasks are not highly speeded and score is not primarily determined by how many items are attempted. An isolated measure of *Gs* can be achieved by increasing the speed requirements of almost any cognitive task. An easier method, however, is to provide problems in which almost everyone would get all items correct if the task were not highly speeded. These kinds of measures relate to speed of talking and to other indicators of behavioral quickness. It is generally thought that speediness very broadly (across many kinds of behaviors) characterizes individual differences (Birren, 1965, 1974; Salthouse, 1982; Salthouse & Somberg, 1982). The following tests provide good examples of such tasks:

Cross Out
The task is to mark all of the drawings in a row of drawings that are identical to the first drawing in the row. The subject must do this for a number of rows and as quickly as possible. The WJ-R includes this test.

Finding *q*'s
On a page filled with printed letters or words, the task is to find and circle as many letter *q*'s as can be found in a very limited amount of time. Any letter could be used, of course, but *q* is distinctive and not as common as other letters.

Comparing Numbers
When given two columns of seven-digit numbers, the task is to decide whether the number in each row of the left column is the same as the number in the right column, and mark *Yes* or *No* to indicate the decision. The subject must, as with other tests of *Gs*, perform the task as quickly as possible.

Coding, Digit Symbol (from the Wechsler scales)
Under highly speeded conditions, the task is to indicate which figures on a page match figures that appear in a row across the top of the page.

Visual Matching
The task is to identify and circle the two identical numbers in a row of six numbers. The subject must do this for a number of rows and as quickly as possible. The WJ-R includes this test.

Correct Decision Speed, CDS

Speediness in finding correct solutions to problems of moderate difficulty has a low correlation with *Gs* and several other abilities. It is seen, therefore, to be construct independent. The feature of *CDS* that distinguishes it from *Gs* is that the tasks to measure it are the same as those used to measure other abilities: *Gf, Gc, Gq, Gv*, and so forth. However, the measure is not the number of correct answers, as it is in the other abilities, but the amount of time taken to generate the answers. Further, the problems are all of moderate difficulty, so the task does not tap the highest level of difficulty with which a subject can cope, as with a measurement of *Gf*.

CDS correlates positively with a measure of the quickness with which wrong answers are produced. This measure might be called Wrong Decision Speed (*WDS*). The measures have not been found to form a common factor. The correlations between *CDS* and other variables are different from the correlations between those variables and *WDS*. This suggests that *CDS* and *WDS* represent different psychological processes.

Other Variance at the Broad Level in the Ability Domain

Several tasks measure skills that are confined to understanding English language. Although related to *Gc* and *Glr*, these tasks represent a separate ability with notably different relationships to other factors than those for *Gc*. This ability is indicated by the following tasks. These may have important applications in education.

Word Parsing
Given sentences in which words have been run together, the task requires the individual to separate the words in order to make sense out of a statement and judge whether it is true or false.

Phonetic Decoding
The task requires the individual to indicate awareness of homonyms, different letter combinations that have the same pronunciation, and homophones, different pronunciations that use the same letter combinations. The Word Attack test in the WJ-R is a measure of this ability.

Particularly in young children, response to novelty (*RTN*) also appears to be an important feature of cognitive functioning (Campos & Sternberg, 1981; Fagan, 1984). The evidence suggests that a measure of the time an infant attends to complex, rather than simple, patterns of stimuli (an operational definition of *RTN*) is predicative of later measures of intelligence. Typical correlations for such measures are about .40.

RTN may be indicative of emerging motivation to deal with cognitive tasks. This motivation may, in fact, drive the development of cognitive capabilities. For example, if children are programmed to respond to rewards, their responses to simple patterns of stimuli may decline if they become bored. If they then discover that the novelty of dealing with new complexities alleviates the boredom, they may move increasingly to more complexities. If *RTN* is not punished during development by, for example, being coupled with demands for correctness, children may move to resolving ever-higher degrees of complexity.

The ability to resolve complexities is an important feature of human intelligence. In fact, measures of cognitive capability are largely measures that require resolving complexities. Thus, at least in infants, *RTN* appears to be an important feature of human thinking.

This represents a largely speculative account for the development of *RTN*. It does, however, rest on evidence that an infant's *RTN* is indicative of subsequent cognitive ability. Further research is needed to account for the emergence of individual differences in cognitive capabilities during infancy and childhood.

Applications and Understandings From the Gf-Gc Theory

As was stated earlier, the purpose of this chapter is to provide an overview of the development of the *Gf-Gc* theory and to examine applications of the theory to research and practice. Since the WJ-R is based upon this theory, the following are areas of investigation and application for which the WJ-R might be used.

Perspective on Gf-Gc Theory

Over the course of human development, from conception to death, there are both genetic and environmental influences that bring about the abilities we see and measure. These influences are predictors of individual differences in cognitive capabilities. Cognitive capabilities are also predictors themselves. The abilities present at any point in development determine emerging features of personality, including emerging cognitive capabilities. Thus, the capabilities are both products and predictors.

The interaction among predictors and outcomes might be considered analogous to strands in an n-dimensional web. Each kind of predictor is a causal strand. Different strands come together at connecting nodules. These represent cognitive capabilities developed up to a given point in time. From each nodule, the strands extend outward, along with the causal strands of genetic and environmental influences, to form more nodules in the network. Construct validity studies provide evidence that cognitive capabilties have different causes, or strands that lead to them, and different predictions of outcomes, or strands leading out to more products. This evidence demonstrates that cognitive capabilities have different functions. Construct validity supporting this phenomenon is extensive but not fully integrated into a theory of human personality. The evidence is increasing, probably at a faster rate than our ability to incorporate it into scientific theory. Any effort to capsulize this evidence is bound to be limited. But in examining this research, one can find some understanding of the complexity of human cognitive capabilities.

The next section begins with a summary of evidence indicating the development of individual differences in cognitive capabilities. The emphasis is on adulthood. Too often development is considered to be limited to childhood and adolescence. Over the last few decades, however, it has become clear that to understand human capabilities we must understand how these factors develop through adulthood as well as in childhood. In this context, the term *development* includes decreases in ability as well as increases. The summary of development is followed by consideration of the genetic and environmental influences on differences in cognitive capabilities.

Gf-Gc Theory Applied to Human Development

The *Gf-Gc* system provides a basis for evaluation of changes in human capability through adulthood. A concept of general intelligence does not provide a sound basis for understanding human cognitive functioning because different intellectual abilities have different patterns of change in adulthood. The nomological networks of relationships for different abilities are quite different. This can be seen clearly in analyses of age differences and age changes in adults.

The evidence suggests that longitudinal (follow-up) and cross-sectional (age comparison) studies yield consistent results. The averages (over broad samples of adults) for *Gf, Gsm*, and *Gs*, classified as "vulnerable" abilities, persistently decline with age from approximately age 25 years onward. In the same samples of people, the averages for the abilities of *Gc*, *Glr* and *Gq*, classified as "maintained" abilities, either increase with age or do not decline. The evidence of age differences and age changes for *Gv*, *Ga* and *CDS* is sparse, but suggests that decline in these abilities occurs at a later age than the decline for the vulnerable abilities and is not as steep.

If the abilities of *Gf*, *Gc*, *Gsm*, *Glr*, *Gq*, *Gv*, *Ga*, *Gs* and *CDS* are all regarded as indicators of IQ, but no specific combination of these capabilities is specified, then different measures of general intelligence involving different proportions of these basic abilities will be formed in accordance with the theory that any broad mixture provides a good working definition of *g* or IQ. Such different measures have been formed. Their use in studies of aging has yielded conflicting and confusing results.

Studies with mixture measures that include primarily vulnerable abilities showed decline in IQ with aging. When a mixture measure was comprised primarily of maintained abilities, however, no age-related decline or aging increase in IQ was found. When the mixture measures included evenly distributed vulnerable and maintained abilities, the conclusion was that IQ neither declines nor increases with age in adulthood. Results such as these were considered to be contradictory.

It can now be seen that such results are not contradictory. They indicate that IQ, conceptualized and measured as a mixture, is not a unitary concept. The results are contradictory only because it was assumed that different abilities indicate the same construct, IQ. With the 20/20 vision of hindsight, that assumption can be seen to be invalid.

Early studies showing increase in IQ with age were longitudinal in design, whereas early studies showing decline were cross-sectional. In the longitudinal studies, the times between retests were small, so little decline could be expected. The sample sizes were also small, so there was little power to show significant change. Thus, the effect sizes in the longitudinal studies usually were not sufficient to indicate an effect comparable to that found in cross-sectional studies. This led to the conclusion that the contradictory findings of cross-sectional and longitudinal studies were due to the differences in design and to the age-cohort differences seen in cross-sectional studies but not in longitudinal studies.

People born at approximately the same point in history are said to belong to the same age cohort. People born earlier in this century attended school for fewer years than people born later in the century. Education enhances intellectual abilities. Thus, the assumption emerged that education differences between young and old were responsible for cross-sectional findings showing that older people scored lower than younger people on measures of IQ. When people are followed in longitudinal study, the reasoning continued, they are all of the same age cohort so those influences are eliminated. The longitudinal findings showing that IQ does not decline thus indicate that the cross-sectional results showing decline are due to age-cohort differences.

This reasoning is not consistent with the findings and does not explain the differences. The longitudinal results indicating no decline were based on mixture measures comprised primarily of *Gc* and *Glr* abilities that generally do not decline. Aging decline was not found, but it was not expected. For example, when breadth of vocabulary (a *Gc* measure) and fluency in describing possible uses for a common object such as brick (a measure of *Glr*) were the principal definers of intelligence, improvement in intelligence over adulthood, not decline, was found (Bayley & Oden, 1955; Horn et al., 1981). If, however, measures of *Gf*, such as letter series and matrices, were used to measure intelligence in the same subjects, then both longitudinal and cross-sectional studies showed aging declines in intelligence (Bayley, 1966; Owens, 1966). When the age range and sample size in longitudinal studies were sufficient to show an effect size comparable to that found in cross-sectional studies and vulnerable abilities were measured, then decline comparable to that seen in cross-sectional studies was found (Horn, 1970; Horn & Donaldson, 1980). Similarly, when good indicators of the maintained abilities were used in cross-sectional studies, increases comparable to those found in longitudinal studies of maintained abilities were found.

Thus, the results of the two kinds of studies are largely consistent. Age cohort differences no doubt operate, but they appear to be small and not largely responsible for the results. From young adulthood to old age there is, on average, a monotonic decrease in the vulnerable abilities. Over much of this period, in the same samples of individuals, there are increases in the maintained abilities. The contradictory results attributed to differences between longitudinal and cross-sectional designs mainly indicate, instead, that different intelligences have different patterns of change through the period of adulthood development. One caveat needs to be added to this conclusion. There is evidence suggesting that, in very old age, the maintained abilities also decline and the rate of decline of the vulnerable abilities increases (Schaie & Baltes, 1977). The former result very likely does not reflect age cohort differences.

Further research has been directed at the nuclear cognitive processes involved in the decline of vulnerable abilities using multiple partialling control methods. Several studies have indicated that the averages for both *Gf* and *CDS* decline with age and to about the same extent. It is reasonable to suppose that the decline in *CDS* is responsible for at least part of the decline of *Gf*; that decline of *Gf* reflects a loss of speed in arriving at correct decisions. If this were the case, and a statistical estimate of the variability in *Gf* that is associated with variability in *CDS* were subtracted (partialled) from *Gf*, the result would be a new measure of *Gf* in which the part due to *CDS* was removed. For example, let *F* symbolize this new measure. If the hypothesis that *Gf* decline is due, in part, to *CDS* decline is correct, this new measure will show less decline than *Gf* because it is purged of the decline caused by *CDS*. The curve for *F* would, therefore, be similar to that of *Gf-Gsm*. If the hypothesis is not correct, the partialling will have no effect and the curve of decline for *F* would be similar to that of *Gf*.

Three separate studies have yielded results that indicate that the hypothesis is not correct. *CDS*, although it does decline with age, does not appear to be a process of *Gf* that accounts for notable aging decline of *Gf*. *Gsm*, on the other hand, does account for about one third of the decline of *Gf* over the period from 25 to 65 years of age. The reverse is also true. *Gf* accounts for some of the aging loss of *Gsm*. It is difficult to know which is the chicken and which is the egg.

Further analysis of these results suggest that what is called "loss of memory" with advancing age can be understood to be loss of reasoning capacities (*Gf*) that support memory (Horn, 1988; Horn et al., 1981). The separate partialling of components of *Gf* reasoning and memory indicate that reasoning accounts for most of the aging decline of the memory components, particularly encoding of novel information (Hultsch, 1971; Mandler, 1967). *Gf* accounts for most of the aging decline of *Gsm*, particularly encoding done spontaneously, as in measures of incidental memory. It is also true for forced encoding directed by explicitly activated metamemory and depth of processing (Botwinick, 1977; Craik, 1977). The general finding is that *Gf* accounts for much of the aging decline of each of the components of memory.

Partialling analyses indicate that elementary capacities for dividing attention, maintaining close attention, and avoiding preoccupation with irrelevancies account for a substantial portion of the aging decline of *Gf*. Each of these also relates to memory and declines with age in adulthood. These attentional processes are also involved in *Gs* speediness that declines with age, and *Gs* decline is also related to *Gf* decline. At the base of these processes is a capacity for maintaining and focusing attention. This capacity is measured with very simple tasks that require subjects to behave as slowly as possible. The measure is labeled Concentration on Slowness (*COS*). It declines with age in adulthood, and its decline accounts for the largest part of the aging decline of *Gf.* Moreover, this measure of ability to behave slowly accounts for much of the aging decline of the *Gs* ability to behave quickly. In particular, *COS* accounts for most of that part of *Gs* that is associated with aging decline of *Gf*. *COS* also accounts for much of the aging decline in measures of ability to divide attention (*ATD*). When both *COS* and *ATD* are subtracted from *Gf,* the decline of *Gs* that is associated with the decline of *Gf* is eliminated. Although the results are rather involved, they suggest that it is not speediness per se that declines with age, but the attentional capacities on which speed of performance is based. The same attentional capacities are also important parts of the aging decline of *Gf* reasoning.

Older adults are more cautious about giving answers when they are not sure, a variable labeled Carefulness (*CAR*). Further, they tend to work longer before giving up on a difficult problem, a variable labeled Persistence (*PRS*). *CAR* and *PRS* result in slowed performance on timed tests. It had been thought that aging decline in *Gf* might reflect these stylistic differences. The results from partialling analyses do not, however, support such hypotheses. The findings indicate that *CAR* and *PRS* do operate; however, they enable older adults to score higher, not lower, on measures of *Gf* abilities! When *CAR* and *PRS* are controlled in the decline of *Gf,* the decline is not reduced; it is actually increased. The negative slope for corrected curves is steeper than before control. Such findings indicate that carefulness and persistence are qualities that enable older adults to perform better on *Gf* tasks than they would perform if these qualities were not allowed to operate. When advantages associated with carefulness and persistence are removed by statistical control, there is significant increase in the aging decline of *Gf.*

One important conclusion from analyzing variables representing attention, carefulness, persistence, encoding, and speediness is that there is considerable overlap in the processes that are assessed. Many ostensibly different variables measure the same basic processes. For example, although measures of inspection speediness are operationally independent from measures of slowness and short-term memory, these variables involve a common process. They involve attention focusing, which is implicated in *Gf* decline. They do not carry entirely independent variance in accounting for the aging loss of *Gf.* The same can be said for several other combinations of the variables thought to represent distinct processes.

Genetic and Environmental Influences in Gf-Gc Theory

Variance in all the abilities we have considered in this chapter may be produced primarily by genetic or environmental influences. Most of the discussion of such matters, however, has centered on *Gf* and *Gc*. The hypotheses of Cattell (1941, 1957) are most frequently discussed.

According to Cattell's first hypothesis, *Gf* primarily reflects genetic influences. His second hypothesis is that *Gc* reflects mainly environmental influences. *Gc* stems from *Gf* and becomes independent from *Gf* as individual differences in environmental influences accumulate through childhood (Horn & Cattell, 1966b, 1982). It follows, therefore, that during the earliest period of life virtually no distinction between *Gf* and *Gc* can be measured. This is because there are few individual differences in environmental influences and little time for such influences to operate. As development proceeds beyond the early years, however, the distinction between *Gf* and *Gc* is clearer and more measurable. It follows, too, that other things being equal (e.g., the reliabilities), the evidence for the genetic basis of *Gf* should be stronger than for the heritability of *Gc*. Although the research results are not entirely clear, they generally do not support these hypotheses.

Two major conclusions are suggested by the results of research on the origins of *Gf* and *Gc*. First, the heritability of *Gf* is not larger than the heritability of *Gc*; and second, *Gf* and *Gc* both stem from sets of gene determiners, but the sets are different. Thus, different intelligences can be distinguished even in early childhood. This is partly because they stem from separate genetic determiners and because they are influenced by separate environmental determiners. The distinctions among different intelligences probably mirror distinctions in neurological functions.

It has been argued that evidence from behavioral genetics forces acceptance of a theory of general intelligence. These arguments have been put very forcefully. Jensen (1973), for example, has likened the heritability of IQ, based on different mixture measures, to the heritability of a polygenetic trait. He has referred to the quasi-normal distribution of IQ measures and the regression of IQ scores for related people as evidence in support of this theory. This information is not evidence to support such a theory, however.

There may be an attribute called intelligence that conforms to heritability theory, but a normal distribution for measures of IQ neither supports nor threatens such a theory. Gene determiners produce a normal distribution in the theory, but item responses produce the normal distribution of IQ scores. If all the items of a test are easy, for example, many people will score high and the distribution of scores will be skewed. If all items are difficult, an opposite skew will be found in the distribution. If items are evenly distributed between easy and difficult and are highly correlated, the distribution of scores will tend to be rectangular — about the same number of people will be found at each score. If the items are correlated around .20 to .35, as is common in measures of IQ, the distribution will be binomial, or approximately normal. The shapes of the distributions are a function of the item sample, not the genes.

There is no known isomorphism between alleles of genes and responses to items; there are no compelling reasons to suppose that their origins might be the same. The influences that determine which items appear in an IQ test are not at all similar to the influences that determine gene selection in reproduction.

Environmental influences, too, can combine independently to produce a quasi-normal distribution. In this case, finding a normal distribution for a composite measure neither supports nor refutes a claim that the trait is determined by environmental factors. There is no reason to expect that the items in an IQ test are isomorphic to environmental influences. The shape of the distribution of scores on an IQ test provides no evidence that the characteristic being measured is or is not determined by either genetic or environmental influences.

Similarly, regression to the mean of measures of IQ for children relative to similar measures for their parents neither supports nor threatens claims that the measures are genetically determined. Such regression is simply a restatement of the fact that the two measures are less than perfectly correlated. This independence could reflect the fact that genes are sorted independently in the two parents that transmit genetic potential to the child, but it could also reflect any of several other influences, not the least being that environmental influences affecting a child are somewhat independent of the comparable influences affecting parents. Regression, as such, is no more an indicator of genetic transmission of IQ than it is an indicator of enviornmental influences. It does not contribute to evidence for a polygenetic theory of IQ.

Claims that abilities are at least partially determined genetically are often supported by correlations between IQ scores for related people. The correlations follow roughly the order of the extent to which the people are genetically related. For identical twins, who basically have the same genetic structure, the correlations are about .8. For fraternal twins, who share genes to the same extent as ordinary siblings or as child and parent, the correlations are about .6-.7. For ordinary siblings and between parent and child, correlations are approximately .5. For half-siblings, the correlations are roughly .3-.4. The correlations for first cousins, uncle-nephew, aunt-neice, and child-grandparent are a bit lower than this. For unrelated people raised in the same home, the correlations range between 0 and .3; for unrelated people randomly assigned to pairs, the correlations are expected to be near 0 (Erlenmeyer-Kimling, 1972).

Such results suggest that some of the individual variability in IQ measures stems from genetic factors. A problem with this conclusion is that the order of similarity in genetic relationships is also the order of similarity for the environments in which people develop. The environments for identical twins are most similar, even when they are reared apart; for fraternal twins, environments are next most similar; for ordinary siblings, next after this, and so on. The similarities expected for environment decreases monotonically with decreases in similarity in genetic structure. Genetic and environmental influences are confounded. No research design or analysis methodology exists to remove these influences from the results.

Estimates of heritability based on such confounded, fallible data always reflect to some unknown degree both environmental and genetic determination of individual differences. In any case, they do not indicate whether the measure found to have a particular heritability is or is not unitary, or a polygenetic single attribute. The heritability can be the same for a mixture of attributes that are inherited quite independently as for a polygenetic unitary trait, such as skin pigmentation. An analogy to facial beauty may help illustrate this point.

Facial beauty is a mixture of features, such as the shape of the nose, the space between the eyes, the turn of the lips, and so on, and one might add up these features as in a mixture measure of intelligence. There is reasonable agreement in our society about who does and does not have facial beauty. These agreements would probably correlate positively with a mixture measure of facial features of beauty. Similarly, reasonable agreement can be found in ratings of who is and is not intelligent and these ratings correlate positively with mixture measures of intelligence.

But what goes into mixture measures of beauty — from physiology, biology, genetics, sociology, anthropology, and psychology? Is it likely that these components indicate a single, unitary, polygenetic trait of beauty? Some beautiful faces have long, thin noses like Meryl Streep's; other beautiful faces have short, wide noses like Sally Field's. The eyes are different in different, equally beautiful faces. So are the cheeks, the skin tone, etc. From genetics, we know that distinct features like noses, eyes, cheeks, and skin tone are inherited independently. These distinct features yield distinctly different physiognomies that, however, can be identified with the single label *beautiful face.* From an anthropological or sociological perspective, what defines beauty in one society does not do so in another society. Evidence of heritability for such mixtures does not indicate that all different beautiful faces represent a single trait. Similarly, evidence of the heritability of mixture measures of intelligence do not support a claim that the mixture is unitary — genetically determined as a whole. In neither case is the evidence of heritability relevant to the argument that the measures represent unitary, polygenetic traits. Beauty is not what is inherited. Features of beauty are genetically determined and mixtures of these traits can have high heritability. But it is not the totality that is inherited, only the features. Similarly, the different abilities in a mixture measure of intelligence are separately inherited. The heritabilities for different mixtures can be high and numerically similar, but this is not evidence that there is any functional unity among the components of the mixtures.

Continuing this analogy, the complexity of faces and their features can be compared to the complexity of brains that lie behind faces. Brains, like faces, may at first look the same, but on closer examination it can be seen that each one is different. The brain is regarded as the basis for intelligence. It is often thought that unitary action of the brain underlies and accounts for a unitary attribute called intelligence. But the evidence suggests that individual differences in brains and neurological function are not unitary. A theory of mass action of the brain is probably not a good analogic theory for intellectual functioning: it is no longer a major focus of neurological research. The brain has several separate features (Cowan, 1979; Dunant & Isreal, 1985; Eccles, 1977; Hubel, 1979; Iverson, 1979; Kety, 1979), and most current research is directed at understanding these features.

There are distinguishable neurotransmitters that are not equally distributed throughout the brain but are located in particular centers and along separate pathways. For example, the norepinephrine system centers in the locus coeruleus, which branches largely into the hypothalamus and adjacent areas. It is closely associated with arousal of neurological functions that appear to be manifested in *Gf* (Iverson, 1979; Horn, 1982a, 1985a; Raz, Millman, & Sarpel, 1990). The dopamine system seems to be centered in the substantia nigra and corpus striatum and is linked to complex events associated with such outcomes as Parkinson's disease. The serotonin system, also, has a distinct place of function in the brain and particular associations with behavior.

Anatomical analyses indicate distinct functions associated with different sections of the brain. The left hemisphere, for example, is associated with different aspects of intellectual function than is the right hemisphere, and a growing body of evidence suggests that the top-to-bottom and front-to-back divisions of the brain are even more important indicators of distinct ability functions (Blackwood & Corsellis, 1976; Bourne, Ekstrand, & Dominowski, 1971; Prohovnik, 1980).

Brains involve distinct components with different genetic determinants. Each component has a different role to play in sensation, perception, learning, and all that comprises intellectual ability. Different configurations of these features produce different capacities, different perceptions, and different ways to process the same information, just as faces involve distinct features with different genetic determinants that, in different configurations, exemplify a beautiful face. The features of the brain may be configured so that the word *intelligence* is applicable, just as the word *beauty* may be applied to a certain configuration of facial features. Intelligence and beauty, in this example, simply unite diversity in a single word. Such words, however, do not indicate scientific laws. Studies of brain function and of different brains do not support a theory of general intelligence any more than studies of facial features and different faces support a theory of general beauty. To the contrary, the evidence suggests that there are several intelligences.

Summary

There are so many kinds of behavior that are indicative of intelligence that identifying their essence has been virtually impossible. However, scientific research on cognitive capability indicates that much of the diversity in intelligence can be understood in terms of a relatively small number of concepts. Interrelations among human abilities currently indicate nine basic capacities: reasoning capabilities (*Gf*), knowledge from acculturation (*Gc*), visualizing capabilities (*Gv*), auditory capabilities (*Ga*), quantitative capabilities (*Gq*), fluency of retrieval of knowledge (*Glm*), processes of maintaining immediate awareness (*Gsm*), processes of speed of apprehension (*Gs*), and processes for quickly arriving at decisions (*CDS*).

Each of these capacities is broad enough to represent what has been described as intelligence. Yet high or low measures in one of these capacities does not indicate correspondingly high or low measures in the others. Each of these capacities or intelligences has its own distinct distribution. Each also has distinctly different relationships with other variables. The capacities have different construct validities.

Reasoning capabilities(*Gf*) and knowledge from acculturation (*Gc*) are most often discussed as types of intelligence. *Gf* is a collection of reasoning capabilities that is relatively independent of education and acculturation. This fluid reasoning can influence a wide variety of intellectual activities. It is measured in tasks that require discovery of a general rule that covers new incidents and identification of changes and differences in order to predict what will come next.

Most of what is most often referred to as intelligence is *Gc*. It involves the ability to make good use of language and to solve many of the complex problems required in everyday living. It is sometimes called "common sense" or "social intelligence." It develops from the experiences of acculturation and increases with experience in solving the problems of a society as well as with education that provides new methods for dealing with life. *Gc* is exercised when one quickly reads a chapter such as this one and grasps the essential ideas.

The two basic capacities, *Gf* and *Gc*, as well as other broad abilities in the *Gf-Gc* system, are composed of "primary" mental abilities. The number of primary abilities thus far discovered is less than 40. With them, however, it is possible to explain much of the person-to-person variation in reasoning, problem-solving, inventing, and understanding. Analogous to the way chemical elements are organized according to the Periodic Law, the primary abilities are organized in the broad capacities of *Gf, Gc, Gv, Ga*, and so on.

The different cognitive capabilities relate in different ways to age over the human lifespan. This is seen in both longitudinal research based on repeated measures of the same people, and cross-sectional research based on comparisons of people born at different times. The averages for *Gf, Gsm, Gs*, and *CDS* decrease steadily from the early 20s onward. The averages for *Gv* and *Ga* increase into the 30s or early 40s and then decrease gradually. The averages for *Gc, Glr*, and *Gq* increase into the 60s before decline begins.

Abilities that decline in adulthood are said to be vulnerable. They are adversely and irreversibly affected by brain damage. Abilities that do not decline in adulthood, or decline late and little, are said to be maintained. These abilities are intially affected by brain damage, but spring back to prior levels with recovery.

Individual differences in *Gc* have been found to increase with age, which seems to indicate that some individuals continue to devote considerable effort to learning after formal schooling ceases, whereas others do not. Individual differences in *Gf* have not been found to increase with age, which suggests that for most people these abilities reach an asymptote of development in early adulthood.

The declines of *Gf*, *Gsm*, and *Gs* in aging are related to loss of ability to maintain close attention and divide attention. Loss of these abilities results in loss of ability to comprehend complex relationships. Aging decline of *Gf* is registered mainly in loss of ability to deal with the most complex of relationships. Loss of abilities to maintain and divide attention reduces the ability to encode, which results in loss of short-term memory, speed of apprehension, and reasoning. Increases in carefulness and persistence during aging may partially compensate for these losses.

Little is known precisely about what produces declines and enhancements in cognitive capabilities in adulthood. Declines in the vulnerable abilities appear to result from accumulations of small losses in brain function. There are suggestions that these losses relate to lifestyle factors that have deleterious effects on the nervous system. For example, abusive use of alcohol seems to have such effects.

Whether in children or adults, measurement of these varied components that comprise what we know as intelligence is of significant importance. Assurance that such measures are reliable and valid increases their positive application to research and practice. In the WJ-R such measures reflect current scientific knowledge. The WJ-R thus provides measures of human cognitive capabilities that are based on modern science. It is considerably beyond other tests in the field.