

# INDIVIDUAL DIFFERENCES IN COGNITIVE FUNCTIONS

*Phillip L. Ackerman*  
*Georgia Institute of Technology*

*David F. Lohman*  
*University of Iowa*

---

## PREAMBLE

---

The study of individual differences in cognitive functioning dates back almost to the beginning of the modern psychology (e.g., Cattell, 1890; Cattell & Farrand, 1896). Intelligence testing in the service of education has been a hallmark of differential psychology since the earliest development of the Binet-Simon (1905) scales, and the English translations and adaptations by Goddard (1908), Kuhlman (1912, 1922), Terman (1916), and others. In the decades that followed these early efforts, thousands of studies have been reported on many different aspects of intellectual abilities and achievements. For example, Wright (1968) documented 6,736 articles and books concerning human intelligence. A quick survey (using Psych-Info) indicates there has been no slowing of research in the field; from 1969 to 2004, there were 36,036 articles, chapters, and books that at least mentioned "human intelligence" and another 4,567 dissertations, though there is certainly some overlap between these sets. Given this enormous literature, it is impossible to provide a comprehensive review, even if the search is limited to investigations with direct implications for educational psychology.

Thankfully, there have been several recent reviews of the field. Most notable is Carroll's (1993) reanalysis and

integration of more than 460 factor-analytic studies of cognitive abilities. In addition, in the first edition of this series, Gustafsson and Undheim (1996) provided a remarkably thorough and thoughtful review of the central issues of individual differences in abilities, with special attention to educational implications and applications. More recently, Lubinski (2000) reviewed the incremental validity of ability, interest, and personality variables in the prediction of intellectual development, vocational adjustment, work performance, and other life outcomes. Jensen (1998) provided a summary of research on *g*. And McGrew and Evans (2004) summarized revisions and extensions to the Cattell-Horn-Carroll theory since the publication of Carroll's (1993) book.

Because the study of human abilities is a comparatively mature field, there is no need to reiterate the points made repeatedly in each of these reviews. Instead, our goal for this chapter is to provide an overview of a few salient areas that have evidenced either a marked increase in research activity or new substantive discussion over the past decade. For example, although cognitive psychology continues to affect in useful ways our understanding of ability constructs, most of the research in the past decade that bears this influence has focused on relationships between constructs (such as relationships among working memory, knowledge, and reasoning) rather than

continued, in-depth analyses of particular tasks thought to be good markers for particular abilities. Thus, we make no effort to update previous summaries of this literature (Lohman, 2000; Pellegrino, Chudowsky, & Glaser, 2001; Snow & Lohman, 1989). Instead, we focus on those research questions that have sparked widespread interest and that have the most direct implications for educators. Readers interested in broader surveys are encouraged to consult other summaries (e.g., Ackerman, Kyllonen, & Roberts, 1999; Mislevy, in press; Sternberg & Pretz, 2005). We trust that the topics reviewed in this chapter give the reader a taste of the current directions in the field, that provide a good representation of what Lakatos (1978) referred to as "progressive research programmes" and "degenerating research programmes."

---

### RECENT DEVELOPMENTS IN THEORIES OF COGNITIVE ABILITIES

---

#### A Brief Synopsis of the Cattell-Horn-Carroll (CHC) Theory

Of the many recurring misconceptions about human abilities, one of the most harmful for educational researchers is the belief that they can adequately represent ability in their research by some measure of *g*. This belief is abetted in part by the fact that *g* is the single best predictor of many important educational and social criteria, especially those that are also averages over diverse performances. However, more specific criteria (e.g., grades in a particular class rather than GPA) are better predicted when ability constructs other than *g* are added to the mix (Gustafsson & Balke, 1993; Wittmann & Süß, 1999). Indeed, as behavior becomes more specific and contextualized, measures of *g* become less useful and measures of more specific knowledge, skills, and abilities become more useful for prediction and explanation. Educational researchers who include ability constructs in their studies need to represent abilities of different levels of breadth or generality. But to do this well requires knowledge of the abilities at different levels in the ability hierarchy and of the tests that can be used to estimate each. Therefore, we briefly summarize the key features of the major contemporary model of human abilities.

The Cattell-Horn-Carroll (CHC) theory of cognitive abilities is the best validated model of human cognitive abilities. The theory integrates Carroll's (1993) three-stratum theory with the Gf-Gc theories of Cattell (1971/1987) and Horn (1991; Horn & Noll, 1997). CHC theory posits a three-strata ability hierarchy. Stratum I contains more than 70 primary or narrow cog-

nitive abilities (e.g., memory span, mechanical knowledge, closure speed, phonetic coding, associative memory, reading speed, speed of articulation). These are the most psychologically transparent abilities. Many are critical aptitudes for success in particular educational tasks (e.g., phonetic coding abilities for early reading). Stratum II consists of nine broad group factors: Fluid Reasoning (Gf), Comprehension-Knowledge (Gc), Short-term Memory (Gsm), Visual Processing (Gv), Auditory Processing (Ga), Long-term Retrieval (Glr), Processing Speed (Gs), Decision/Reaction Time Speed (Gt), Reading and Writing (Grw), and Quantitative Knowledge (Gq). Finally, general ability or *g* sits alone at stratum III.

McGrew and Evans (2004) have recently provided a review of this theory, with particular attention to changes to the theory that have occurred since the publication of Carroll's (1993) seminal book. Table 7.1 is abstracted from this review. Readers are encouraged to consult these publications for more detailed exposition of each of the factors in the model.

#### Recent Developments in CHC Theory

Five of the general factors in the model have been studied extensively in recent years. We briefly note some of these efforts.

**Visual Processing (Gv).** There has been a resurgence of interest in the utility of *Gv* abilities for the prediction of educational and occupational choice. For example, Shea, Lubinski, and Benbow (2001) found that spatial abilities measured in early adolescence contributed importantly to the prediction of subsequent educational and vocational preferences of academically gifted students, even after verbal and mathematic abilities had been entered in the regression models. Others have continued to explore the extent to which different spatial primary factors require different cognitive processes, particularly factors that differ in the apparent complexity of processing required (Juhel, 1991) or the involvement of executive processes in working memory (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001).

**Auditory Processing (Ga).** Contrary to earlier assertions by some reading researchers that phonetic coding (or phonemic awareness) involved two distinct abilities, several investigations now show that phonetic coding is a unidimensional construct (e.g., Wagner et al., 1997). Other researchers have explored relationships between phonological awareness, music perception, and early reading (Anvari, Trainor, Woodside, & Levy, 2002).

TABLE 7.1. Broad (Stratum II) and Narrow (Stratum I)  
Cattell-Horn-Carroll (CHC) Ability Definitions  
(After McGrew & Evans, 1994)

---

Fluid Reasoning (*Gf*): The use of deliberate and controlled mental operations to solve novel problems

- General Sequential (deductive) Reasoning (RG)
- Induction (I)
- Quantitative Reasoning (RQ)
- Piagetian Reasoning (RP)
- Speed of Reasoning (RE)

Crystallized Knowledge (*Gc*): Verbal declarative and procedural knowledge acquired during formal schooling and general life

- Language Development (LD)
- Lexical Knowledge (VL)
- Listening Ability (LS)
- General (verbal) Information (KO)
- Information about Culture (K2)
- Communication Ability (CM)
- Oral Production and Fluency (OP)
- Grammatical Sensitivity (MY)
- Foreign Language Proficiency (KL)
- Foreign Language Aptitude (LA)

General (domain-specific) Knowledge (*Gkn*): Breadth and depth of acquired knowledge in specialized domains

- Knowledge of English as a Second Language (KE)
- Knowledge of Signing (KF)
- Skill in Lip-reading (LP)
- Geography Achievement (A5)
- General Science Information (K1)
- Mechanical Knowledge (MK)
- Knowledge of Behavioral Content (BC)

Visual-Spatial Abilities (*Gv*): The ability to generate, retain, retrieve, and transform well-structured visual images

- Visualization (Vz)
- Spatial Relations (SR)
- Closure Speed (CS)
- Flexibility of Closure (CF)
- Visual Memory (MV)
- Spatial Scanning (SS)
- Serial Perceptual Integration (PI)
- Length Estimation (LE)
- Perceptual Illusions (IL)
- Perceptual Alternations (PN)
- Imagery (IM)

Auditory Processing (*Ga*): Abilities involved in discriminating patterns in sounds and musical structure

- Phonetic Coding (PC)
- Speech Sound Discrimination (US)
- Resistance to Auditory Stimulus Distortion (UR)
- Memory for Sound Patterns (UM)
- General Sound Discrimination (U3)
- Temporal Tracking (UK)
- Musical Discrimination and Judgment (U1 U9)
- Maintaining and Judging Rhythm (U8)
- Sound-Intensity/Duration Discrimination (U6)
- Sound-Frequency Discrimination (U5)
- Hearing and Speech Threshold factors (UA UT UU)
- Absolute Pitch (UP)
- Sound Localization (UL)

---

TABLE 7.1. (Continued)

---

Short-term Memory (*Gsm*): Ability to apprehend and maintain awareness of information in the immediate situation

- Memory Span (MS)
- Working Memory (MW)

Long-term Storage and Retrieval (*Glr*): The ability to store and consolidate new information in long-term memory and later fluently retrieve that information

- Associative Memory (MA)
- Meaningful Memory (MM)
- Free Recall Memory (M6)
- Ideational Fluency (F1)
- Associational Fluency (FA)
- Expressional Fluency (FE)
- Naming Facility (NA)
- Word Fluency (FW)
- Figural Fluency (FF)
- Figural Flexibility (FX)
- Sensitivity to Problems (SP)
- Originality/Creativity (FO)
- Learning Abilities (L1)

Cognitive Processing Speed (*Gs*): The ability to perform relatively easy or over-learned cognitive tasks, especially when high mental efficiency is required.

- Perceptual Speed (P)
- Rate-of-Test-Taking (R9)
- Number Facility (N)
- Speed of Reasoning (RE)
- Reading Speed (fluency) (RS)
- Writing Speed (fluency) (WS)

Decision/Reaction Time or Speed (*Gt*): The ability to react and/or make decisions quickly in response to simple stimuli

- Simple Reaction Time (R1)
- Choice Reaction Time (R2)
- Semantic Processing Speed (R4)
- Mental Comparison Speed (R7)
- Inspection Time (IT)

Psychomotor Speed (*Gps*): The ability to rapidly and fluently perform body motor movements independent of cognitive control.

- Speed of Limb Movement (R3)
- Writing Speed (fluency) (WS)
- Speed of Articulation (PT)
- Movement Time (MT)

Quantitative Knowledge (*Gq*): Wealth of acquired store of declarative and procedural quantitative knowledge (not reasoning with this knowledge)

- Mathematical Knowledge (KM)
- Mathematical Achievement (A3)

Reading/Writing (*Grw*): Declarative and procedural reading and writing skills and knowledge

- Reading Decoding (RD)
- Reading Comprehension (RC)
- Verbal (printed) Language Comprehension (V)
- Cloze Ability (CZ)
- Spelling Ability (SG)
- Writing Ability (WA)
- English Usage Knowledge (EU)
- Reading Speed (fluency) (RS)
- Writing Speed (fluency) (WS)

---

TABLE 7.1. (Continued)

Psychomotor Abilities ( <i>Gp</i> ): The ability to perform body motor movements with precision, coordination, or strength
Static Strength (P3)
Multi-limb Coordination (P6)
Finger Dexterity (P2)
Manual Dexterity (P1)
Arm-hand Steadiness (P7)
Control Precision (P8)
Aiming (A1)
Gross Body Equilibrium (P4)
Olfactory Abilities ( <i>Go</i> ): Abilities that depend on sensory receptors of the main olfactory system
Olfactory Memory (OM)
Olfactory Sensitivity (OS)
Tactile Abilities ( <i>Gb</i> ): Abilities that depend on sensory receptors of the tactile system for input
Tactile Sensitivity (TS)
Kinesthetic Abilities ( <i>Gk</i> ): Abilities that depend on sensory receptors that detect bodily position, weight, or movement of the muscles, tendons, and joints.
Kinesthetic Sensitivity (KS)

**Short-term Memory (*Gsm*).** Much research in recent years has focused on the construct of working memory and its overlap with measures of memory span (MS) and fluid reasoning (*Gf*). (We review some of this research in a subsection section.) Carroll (1993) included both primary factors for both memory span and working memory in his theory. However, McGrew & Evans (2004) concluded that more recent evidence shows that measures of working memory are invariably factorially complex, do not coincide with the primary factor called Memory Span (MS), and do not define a primary factor of their own.

**Crystallized Intelligence/Knowledge (*Gc*).** Several studies (some of which we review below) show convincingly that *Gc* is not to be equated with general (domain) knowledge (*Gkn*) (Rolfhus & Ackerman, 1999). *Gc* is typically assumed to estimate general (verbal) cultural knowledge, whereas *Gkn* represents the breadth and depth of knowledge in particular domains. Many erroneously assume that the sorts of general achievement surveys (e.g., NAEP) or tests (e.g., ITED) administered to all students sample both kinds of educational outcomes. However, most of these tests sample little domain knowledge. (In fact, tests commonly used to define factors that are labeled *Gc* often would load substantially on *Grw*.) In part, this is because students in different schools often do not study the same content or study it in different grades. However, most measures of educational achievement sample only a small part of students' domain knowledge, even in domains that all students study.

**Mental Speed (*Gs*, *Gt*, and *Gps*).** Although researchers sometimes speak of mental speed as if it were a single

construct, research shows that measures of mental (and psychomotor) speed form a complex hierarchy of their own (Stankov, 2000). For example, studies by Ackerman, Beier, and Boyle (2002), O'Connor and Burns (2003), and others suggest that broad cognitive speed (*Gs*) may be divided into Perceptual Speed (*P*) and Rate of Test Taking (*R9*). The *P* factor can be further differentiated into at least four primary factors (such as Pattern Recognition; see Ackerman & Cianciolo, 2000). Similarly, the *R9* factor subsumes several primaries (such as Speed of Reasoning).

Studies by Roberts and Stankov (1999) and their colleagues show that the broad decision speed factor (*Gt*) subsumes intermediate level Reaction Time and Movement time factors, which in turn subsume several primary factors such as Simple Reaction Time (*R1*) and Choice Reaction Time (2).

Finally, Broad Psychomotor Speed (*Gps*) subsumes at least four primary factors (including Speed of Writing [*WS*]). Strength and accuracy of psychomotor skills are subsumed under a separate broad group factor (Broad Psychomotor Ability [*Gp*]).

**Extensions of CHC Theory (*Gb*, *Gk*, & *Go*).** Primarily through the work of Stankov, Roberts, and their collaborators, CHC theory has been extended to include tactile and kinesthetic factors. Although spatial visualization accounts for much of the variance on complex tactile and kinesthetic tasks, separate Tactile (*TS*) and Kinesthetic (*KS*) elements have been identified. On the basis of several such studies, Stankov (2000) suggests that broad kinesthetic (*Gk*), tactile (*Gt*), and olfactory (*Go*) factors should be added to CHC theory.

#### Summary and Reflections on CHC Theory

In summary, the hierarchical model that Carroll (1993) proposed has been elaborated and extended several important ways in the past decade. Probably the most important message for most educational researchers in this work is that understanding how abilities moderate success in learning requires multiple measures that are chosen to represent each of multiple abilities at multiple levels in hierarchy. Deciding which abilities to measure requires knowledge of prior research, but also of the demands and affordances of the learning tasks for the study's participants. On the one hand, one need not represent the whole of the hierarchy to do this well. On the other hand, a single measure of *g* (or of some more specific ability) will not allow one to understand much about the unique demands of the task, or of the ways in which those demands may change systematically across individuals or over time.

Researchers who accept a hierarchical model continue to debate the magnitude of influence of *g* on the range of human intellectual abilities and even whether it is a psychologically meaningful construct. For example, Vernon (1950) estimated that *g* accounts for anywhere between 20 and 40 percent of the total ability variance in most test batteries. Others have argued for greater or lesser role for *g*, depending on both their theoretical orientation and the nature of the empirical samples under investigation (e.g., in an unselected sample, *g* tends to account for a greater degree of variance than in a sample with restricted range-of-talent, such as a sample of college students). Some theorists (e.g., Horn & Noll, 1997) question whether *g* is psychologically meaningful. Instead, these researchers emphasize ability factors at the primary or broad-group level (Strata III and II, respectively, in Carroll's 1993 model). For other researchers who accept *g*, the critical issue is the relative emphasis that is placed on ability factors at different levels in a hierarchical model.

A hierarchical model gives precedence to the general factor. This is not an inherent property of the model so much as the application of the principle of parsimony to interpretations of the model. Parsimonious interpretation means that one does not attribute variation to lower order factors that can be "explained" by the general factor. What is left over is divided among broad group factors. Once these have had their say, then the primaries are admitted. Those interested in understanding what abilities might be have always been troubled by this. The primary factors at the base of the hierarchy are invariably psychologically more transparent than higher order factors. In fact, psychological clarity decreases as one moves up the hierarchy. There is far greater agreement on what phonemic awareness might be than what verbal ability might be; greater agreement on what verbal ability might be than on general crystallized ability (*Gc*); and there is least agreement on what *g* might be. A hierarchical model thus gives parsimony precedence over psychological clarity.

On the criterion side of the equation, a similar precedence for parsimony over educational clarity is shown when educational performance is assessed with GPA or other aggregate measures. When high levels of aggregation are adopted for both ability predictors and educational criteria, there is substantial Brunswik Symmetry (Wittmann & Süß, 1999), and a substantial degree of predictive validity for general ability measures (see, e.g., Gustafsson & Baulke, 1993).

#### Other, Non-CHC Theories of Abilities

In their 1996 chapter, Gustafsson and Undheim briefly reviewed two theoretical approaches that have contin-

ued to attract considerable attention from educational practitioners. These are the frameworks proposed by Gardner (1983) and Sternberg (1985), respectively. Next we provide an update on the status of these approaches to intelligence.

#### Gardner's Theory of Multiple Intelligences

Gardner introduced the theory of theory of multiple intelligences in a popular book in the early 1980s. The theory posited seven different intelligences: logical-mathematical, linguistic, spatial, musical, bodily-kinesthetic, interpersonal, and intrapersonal (Gardner, 1983). Since then, naturalist, spiritualist, and existential intelligences have also been proposed by Gardner (1993, 1999), although only the naturalist intelligence has survived (Gardner, 2003). In their chapter Gustafsson and Undheim (1996) pointed out that several of these intelligences are well accounted for by traditional ability constructs. For example, some abilities proposed and investigated by Cattell and Horn (e.g., Horn, 1965, 1989) map reasonably well to some intelligences proposed by Gardner; linguistic intelligence is essentially *Gc*; logical-mathematical intelligence maps to *Gf/Gq*; and spatial intelligence maps to *Gv*. There have also been measures developed for assessing musical abilities (e.g., Seashore, 1919; Vispoel, 1999), but the remaining abilities have eluded standardized assessment. For example, many attempts to assess social intelligence have resulted in relatively little in the way of measures that show discriminant validity with other more traditional intelligence measures (see Schneider, Ackerman, & Kanfer, 1996, for a review).

Publications in recent years have focused on many different criticisms of Gardner's theory (e.g., Brody, 1992; Klein, 1997; Lohman, 2001; Lubinski & Benbow, 1995; Messick, 1992; Willingham, 2004). These criticisms are far ranging. Some concern the implications of the theory for school curricula; others focus on the theory itself and the extent to which it has been tested empirically. Here, we limit our discussion of Gardner's theory to issues about individual differences in cognitive functioning. With respect to the traditional approaches to intellectual abilities, Gardner makes several key claims, as follows:

1. Intelligence is plural, not singular.
2. Individuals differ in their profiles of intelligences or abilities.
3. Intellectual assessment should not be limited to paper-and-pencil tests, but should include "more humane methods—ranging from self-assessment to the examination of portfolios of student work" (Gardner, 1999, p. 73).

4. "Intelligence should not be expanded to include personality, motivation, will, attention, character, creativity, and other important and significant human capabilities" (Gardner, 1999, p. 74).

There is little scientific objection to some of Gardner's objections to traditional conceptions of intelligence, but much disagreement about the alternatives he has proposed. In this context, we briefly review each of these points.

**Intelligence Is Plural.** This claim can be strident or prosaic. In the strident version, Gardner (1993) says that *g* is a methodological consequence of using short-answer, multiple-choice, paper-and-pencil tests of the sorts of linguistic and logical intelligence that are at best useful for predicting success in the narrow domain of conventionally structural schools (Gardner, 1993, p. 39). Needless to say, this view is not warmly received by those who study human abilities. The less controversial, even prosaic version of the claim is that intelligence is more than *g*. This allows something like *g* to exist, but emphasizes the importance of broad group factors. This claim is quite uncontroversial. Indeed, even the most ardent supporters of *g* do not claim that intelligence is wholly made up of a single general intellectual ability (see, e.g., Jensen, 1998). Carroll's (1993) comprehensive review of human cognitive abilities lists dozens of different abilities and has generated very little controversy in the field.

**Individuals Differ in Their Profiles of Abilities.** Again, there is little debate on this topic. Reporting of individual profiles of abilities has been a major component of ability testing in the schools and for vocational purposes for nearly as long as there has been large-scale testing (e.g., E. L. Thorndike's CAVD test, see Thorndike, Bregman, Cobb, & Woodyard, 1927; Thurstone & Thurstone's PMA test, see Thurstone, 1957). Ability profile reporting is so ubiquitous that nearly every schoolchild and parent in the United States is familiar with seeing a bar graph of multiple abilities in annual assessments. However, the use of ability profiles to describe an individual's strengths and weaknesses requires qualification, as noted by Cronbach (1960). For most multiscore tests, the several scores are often significantly correlated with one another. Because the reliability of difference scores is typically quite low when the scores are themselves substantially correlated, differences between two scores for an individual must be relatively large before it can be concluded that two scores differ dependably. In many educational contexts, however, the negative consequence of not detecting a difference (e.g., a relative weakness in reading comprehension) outweighs the negative consequences that might

attend reporting a difference where none existed. In this case, the most likely consequence would be that the student would be given additional assistance in developing reading skills. In such situations, confidence intervals that more nearly equate the probabilities of Type I and Type II errors are probably preferred over the more conservative intervals that control only for Type I errors (see Feldt, 1967).

From the perspective of a hierarchical model, however, the issue is not only whether individuals differ in their profile of abilities but whether an estimate of *g* (something like the overall height of the profile) also conveys useful information. Clearly, the *g* score is most informative when the individual scores from which it is estimated do not differ significantly and least informative when they differ markedly. Indeed, if there is considerable scatter among the scores, then a weighted average that estimates *g* may mislead. Importantly, even when correlations among tests are substantial and evidence for *g* is strong, profiles may still carry useful information for most students—and critical information for some. For example, the nine reasoning tests on the multilevel edition of the Cognitive Abilities Test (Lohman & Hagen, 2001) are grouped into Verbal, Quantitative, and Nonverbal (Figural) batteries, each with three tests. The general factor accounts for approximately 81% of the common variance in these three scores. However, only about one third of students who take the test show a profile in which the three reasoning scores do not differ significantly from each other. How can this be? Although the general factor captures most of the common variance, each battery has a substantial specific component of variance. For example, although 64% of the variance in the Verbal Battery is explained by *g*, 36% is not. Approximately 5% of the variance can be attributed to errors of measurement and so 31% of the variation in Verbal scores is independent of *g* and errors of measurement. The Quantitative and Nonverbal batteries also have substantial components of reliable but specific variance. When all three are considered simultaneously, there is substantial variability in score profiles in addition to the variability in the composite score (see also Lubinski, 2004).

**Intellectual Assessment Should Not Be Limited to Paper-and-Pencil Tests.** Again, there is little argument with this proposition. In fact, Binet's method of intelligence assessment (and the modern instantiations such as the Stanford-Binet and Wechsler tests) did not involve the use of paper and pencils by the examinee. Numerous examples of "performance" tests of intelligence can be found in the traditional literature (e.g., Goodenough's [1926] draw-a-person test). In addition, self-assessments of abilities tend to be relatively reasonably correlated with objective

measures of intelligence (e.g., see Ackerman, Beier, & Bowen, 2002; also see Baird, 1969). However, self-assessments cannot be used effectively in applied situations, when the goals of the individual being assessed are not closely aligned with the organization's goals, for the simple reason that the individual may be motivated to give spurious responses to the questions, in order to attain his/her goals. For example, as partly evidenced by SAT prep course enrollments, many college applicants are more interested in being admitted to the college/university of their choice than they are in matching the demands of the school curriculum to their abilities. Thus, one can reasonably expect that self-assessments of abilities in a college selection context will be highly inflated by all but the most honest or the most gullible of candidates.

Gardner (1999, 2003) suggests using portfolios and other nonstandardized procedures for intellectual assessment. Portfolios and other types of performance assessment are commonly used in the measurement of educational attainment. Portfolios are often used when it is impractical to expect a student to produce a product within a limited time (such as in art), when the goal is to assemble a collection of the student's best work, or to document growth in the performance of particular tasks (e.g., writing of essays) (Nitko, 2001). However, portfolios are difficult to score reliably (Koertz, Stecher, Klein, & McCaffrey, 1994) and, like other performance assessments, may advantage students whose parents and teachers have the time and expertise to help them develop and refine their work. Further, although virtually all measurement people see value in assessments that mimic the criterion behavior (Lindquist, 1951), few would agree that standardized assessments should be abandoned altogether! Further, the assertion that performance assessments are "more humane" assumes that traditional assessments are inhumane. The claim that little children are distressed by tests, although commonly made and supported by anecdote in the popular press, actually has little empirical support. For example, in one study, Frisbie and Andrews (1990) observed the test-taking behavior of more than 600 kindergarten children in 17 schools as they took Level 5 of the ITBS. Although they note that administering a test requires more planning and care with young children than with older children, most of the kindergarten children in their study encountered no problems in taking the test. In fact, young children commonly enjoy taking these sorts of tests. What is not warranted are high-stakes decisions about children on the basis of a low test score, especially when the test is administered in a group by an inexperienced examiner, and when the test does not provide caution indices for students whose patterns of item and subtest scores depart markedly from

the patterns expected under the scaling model (Lohman & Hagen, 2002, p. 62). On the other hand, not assessing children until third or fourth grade misses opportunities for intervention that could assist at least some students.

There is another, even more basic difference between traditional tests and portfolios. That difference is best articulated as the distinction between maximal and typical intellect (e.g., see Ackerman, 1994, for a discussion). Traditional tests of intelligence have focused on maximal intelligence—it was a major component of Binet's procedure to elicit the performance of the individual with maximal motivation to succeed. Binet's justification for these conditions was that he wanted to minimize, at least as far as possible, the influences of prior privilege (e.g., socioeconomic status, prior educational opportunities) from the test scores. Although portfolios often aim to collect a student's best performances, evaluating abilities by looking over a much longer time frame than the traditional intelligence test raises the influence of an individual's typical investment of intellectual effort in the particular domain. (If Gardner's goal is to eliminate "motivation" from the conceptualization of intelligence—see later discussion—the assessment of portfolios is clearly working in the opposite direction.)

*"Intelligence should not be expanded to include personality, motivation, will, attention, character, creativity, and other important and significant human capabilities."* Paradoxically, here Gardner parts company with many researchers, by narrowing the construct of intelligence in a major fashion. Previous researchers have taken different approaches to the integration of these constructs into their conceptualization of intelligence, albeit with different degrees of success. For example, Spearman and his students considered "Factor W" (e.g., Webb, 1915), a factor of conscientiousness or conative propensity as an integral determinant of performance on intellectual tasks, and W. Alexander (1935) identified factors X (interests) and Z (achievement) as similarly integral to an ability test performance. Guilford's (1959) framework put intelligence as only one of seven different aspects of personality. Most notably, Wechsler (1950) remarked that intelligence is much broader than simply knowing the correct answer on a reasoning test. Specifically, he stated "Actually, intelligence is an aspect of behavior; it has to do primarily with the appropriateness, effectiveness, and worthwhileness of what human beings do or want to do" (p. 135). Later, Wechsler refined his orientation, as follows: "What we measure with (intelligence) tests is not what tests measure—not information, not spatial perception, not reasoning ability. These are only a means to an end. What intelligence tests measure, what we hope they measure, is something much more

important: the capacity of an individual to understand the world about him and his resourcefulness to cope with its challenges" (Wechsler, 1975, p. 81).

The attempt to separate the cognitive from the affective and conative components of cognition runs counter from much recent work that seeks to integrate them. Corno et al. (2002) argue that it is particularly important for educational researchers to be aware of the many complex but important interactions among cognitive, affective, and conative constructs in school learning. Artificially separating cognition from, say, volition has impoverished efforts to understand both domains.

In general, the attempt by Gardner to separate intelligence from personality, motivation, attention, and the like seems to fail in an attempt to "carve nature at its joints" (Plato, in *Phaedrus*), as these constructs are integral to intellectual performance. Without motivation, for example, an individual is unlikely to accomplish any "intellectual" task. Attempting to arrive at a depiction of intelligence without motivation (or many of these other constructs) denies what is both obviously and empirically true.

**Concluding Comments on Gardner.** Far from standing in stark contrast with Gardner's major arguments, the preceding discussion illustrates that there is value in what he has suggested regarding the construct and measurement of intelligence. However, it would be erroneous to say that his arguments represent something new in the field of intelligence. All of these issues have been well represented in prior theoretical developments and empirical work and are present in many modern theories of intellectual abilities and several existing assessment instruments. Nonetheless, the science of intelligence research parts ways with Gardner in two ways.

First, we do not think it useful to overly restrict the construct of intelligence. Perhaps this is easier to see if one considers estimating individuals' readiness to learn or perform in a particular situation (i.e., their aptitude) rather than their intelligence. Readiness to learn algebra, for example, clearly has affective and conative components as well as cognitive components. Artificially separating the information processing from the affective and conative components of cognition (e.g., interest, motivation, anxiety) underrepresents the construct of readiness. Indeed, one of the main implications of recent research on individual differences is that understanding the joint action of constellations traits or aptitudes can significantly enhance the ability to predict and, at times, to alter outcomes for individuals (Ackerman, 2003; Corno et al., 2002; Snow, 1978).

Second, given that the foundation of science of psychology is empirical observation, until such time as there are reliable and valid measures of a construct, there must

be a high level of skepticism regarding the existence of the construct. The Interpersonal, Intrapersonal, Naturalist, Spiritual, and Existential intelligences proposed by Gardner simply do not meet this criterion, and as such cannot yet be considered scientific constructs, from either verificationist (Carnap, 1987) or falsificationist (Popper, 1963) perspectives. In the final analysis, Gardner may have legitimate doubts about the feasibility of paper-and-pencil testing procedures to provide adequate measures of these proposed intelligences, but these doubts do not invalidate E. L. Thorndike's dictum that "Whatever exists at all, exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality" (quoted in Joncich, 1968). Without means for measurement, the constructs are not well suited for scientific discussion.

Third, Gardner's model of parallel abilities (or "intelligences") falls short not simply because it fails to acknowledge or explain why abilities are correlated, but because it cannot explain why the observed correlational structure implies a hierarchy (Lohman, 2001). In other words, the theory does not explain why some intelligences are more intelligent than others. As Gustafsson (1988) has argued, *g* overlaps substantially with *Gf*, which in turn is virtually synonymous with the primary factor called Inductive Reasoning (*IR*). Although some *g*-theorists eschew efforts to explain psychological basis of *g* (Jensen, 1998), it is clear that a substantial portion of the variation in *g* can be attributed to individual differences in reasoning, particularly inductive reasoning. Further, as several researchers have now shown, a significant fraction of the variation in reasoning abilities is shared with measures of working memory (see later discussion). However, Gardner's extreme modular view of cognition dismisses the notion of a common working memory. Thus, one of the chief complaints cognitive psychologists have leveled against Gardner's theory (i.e., its rejection of working memory) turns out to be the basis of the differential psychologists' complaint as well (i.e., the rejection of *g*) (Lohman, 2001). An extreme modular view of either working memory or ability fails to explain the empirical observations in either domain.

### Sternberg's Triarchic Theory of Intelligence

Sternberg (1985) introduced his triarchic theory as an attempt to expand the conceptualization of intelligence beyond the traditional construct. He argues that the first part of his framework, *analytic (or academic) intelligence*, represents the essence of the traditional view of intelligence. However, the other two parts of the framework, *creative intelligence* and *practical intelligence*, purport to go beyond the traditional conceptualization. Creative



intelligence is said to be required when responding intelligently to a relatively novel task or situation. In Sternberg's (1993; 2004) yet unpublished test battery, the Sternberg Triarchic Abilities Test (STAT), creative intelligence is assessed through novel analogies, number problems with fictitious operations, and figural series items that require transformations. Practical intelligence is proposed to represent adaptation to, shaping of, and selection of real-world environments. In the STAT, practical intelligence is assessed by tests made up of "everyday reasoning," "everyday math," and "route planning" items. As operationalized on the STAT, both creative and practical intelligence appear to differ in degree rather than in kind from abilities measured by the analytic scale, and to overlap considerably with abilities measured in non-Gf branches of the hierarchical model.

The most recent version of STAT contains 45 multiple-choice items (nine scales of five items each). Each five item scale is defined by the application of a particular "intelligence" (analytic, creative, or practical) to a particular content (verbal, numerical, or figural). A recent, as yet unpublished study that uses the scale scores contains additional details (Sternberg & the Rainbow Project Collaborators, 2004). In the reported study, scores were averaged across the content facet for all three intelligences. (The content distinction was tenable only for the analytic scale.) Total score on STAT creative scale added to the prediction of undergraduate GPA after SAT scores and high school GPA were entered in the regression model. STAT analytic and practical did not. An extensive discussion of the validation of the first edition of the STAT is provided by Brody (2003a, 2003b, though see Sternberg, 2003, for a response to Brody's criticisms.)

Although we can report only modest progress in the validation of the Sternberg triarchic theory since the review provided by Gustafsson and Undheim (1996), perhaps it is as yet too early to be able to provide a critical assessment of the utility of the theory for expanding the conceptualization of intelligence to the domains of practical and creative intelligence. For example, the finding that STAT Creative (and other performance measures of creativity in Sternberg et al., 2004) predicted college GPA awaits interpretation, especially a disentangling of confounds with conventional measures of verbal reasoning, verbal fluency, and writing abilities. As noted by McNemar (1965), developing a valid measure of creativity has been a difficult task for psychometric researchers over the past century, and perhaps it will yet be another decade before significant progress is made on this front by Sternberg and his colleagues. It should be mentioned, especially in contrast to Gardner's theory of multiple intelligences, that Sternberg agrees that testing his theory requires the development of defensible ways to measure the constructs

he has proposed, and establishing their similarities and differences with measures of other constructs (though see Messick, 1992, for a contrasting view).

---

## DEVELOPMENTS THAT FOCUS ON KNOWLEDGE

---

### Intelligence, Learning, and Knowledge

Early theorists and modern researchers generally subscribe to the proposition that individual differences in intelligence correlate strongly with individual differences in learning (e.g., see Buckingham, 1921). Indeed, the major success of the Binet-Simon scales and subsequent omnibus tests of intelligence is based on the high predictive validities for academic success of children and adolescents. As noted in the earlier review of individual differences in cognitive functioning (Gustafson & Undheim, 1996), several decades of controversy have revolved around establishing the relationship between intellectual abilities and individual differences in learning, especially when specific learning tasks are considered (in contrast to overall academic achievement). Investigations of individual differences in learning generally find relatively small predictive validities for intellectual abilities measures, suggesting to some researchers that the relationship between intelligence and learning may be markedly overstated (e.g., see Ackerman, 2000b; Lohman, 1999, for reviews).

A critical issue in these studies is the distinction between the measurement of learning as gain or as accumulated competence (i.e., final status). Gain and final status scores show similar correlations with other variables only if the variance of individual differences in initial status is small (and thus overshadowed by the variability in final status) or is uncorrelated with final status. In most educational contexts, initial differences in performance both are substantial and are correlated with final status. In such circumstances, learning as measured by gain is unreliable and typically shows small correlations with other variables. However, if the variability in performance increases with practice, then gain scores can not only be much more reliable, but show substantial correlations with other variables as well (Lohman, 1999).

However, even when learning is measured by final status, correlations with measures of ability are often small. An approach that offers resolution to this issue is the principle of Brunswik Symmetry (see Wittmann & Süß, 1999). The basic idea is that predictor-criterion validities are maximized (a) when there is a match between predictor breadth and criterion breadth, and (b) when the mapping between predictors and criteria is direct. Generally speaking, when broad criteria are to be

predicted, Brunswik Symmetry dictates that the best set of predictors will also be broad; when narrow criteria are to be predicted, narrow predictors have the potential for maximum validity. And, the mapping must be direct—such that the wrong narrow predictors may yield very low or zero validity for narrow criteria. In the context of individual differences in learning, the principle of Brunswik Symmetry would imply that one will not likely find large correlations between broad ability measures (such as IQ) and narrow learning tasks, such as those examined in laboratory studies. Instead, the prediction of individual differences in learning for narrow tasks requires a much more nuanced task analysis to determine which abilities (if any) are requisite for learning and for performance after practice or instruction. In contrast, when broad learning indicators serve as the criteria (e.g., grade point average), broad predictors such as IQ will represent an optimal match, in terms of Brunswik Symmetry, and in turn will evidence high validity coefficients. The implications of these considerations are that investigators seeking to establish the general proposition that intelligence is the ability to learn must limit their consideration to broad indicators of learning and academic achievement.

#### Abilities and Domain Learning

Although school serves many different purposes (e.g., see Alexander & Murphy, 1999), two major goals of educational programs are to impart general skills for problem solving and to impart domain knowledge to students. Partly because of explicit overlap between test content and criterion task content, there is relatively little controversy regarding strong relationships between intellectual abilities (such as reasoning and problem solving) and educational criteria in the areas of critical thinking (e.g., see Herrnstein, Nickerson, de Sánchez, & Swets, 1986). What is less clear is the relationship between abilities and the acquisition of domain knowledge. The main reason for this shortcoming is that traditional intellectual ability theory has focused on either general abilities or broad content abilities that are less associated with domain knowledge.

There have, of course, been exceptions to this pattern of investigations (e.g., see Vernon, 1950). In recent years, however, some researchers have turned from a focus on process and critical thinking to a focus on knowledge and the attainment of expertise (e.g. Stanovich & West, 1989). Theories proposed in this domain cover a wide range of approaches, from those that are mainly concerned with how prior knowledge and interests affect domain learning (e.g., see Alexander, Kulikowich, & Schulze, 1994) to theories that propose that individual differences in abilities

are largely irrelevant to the acquisition of expertise (e.g., Chi, Glaser, & Rees, 1982; Ericsson, Krampe, & Tesch-Römer, 1993).

In contrast, other theories such as the one proposed by Ackerman (1996) suggest an important role of individual differences in intellectual abilities in determining the direction and level of cognitive investment in the acquisition of domain knowledge. Ackerman's theoretical framework, called PPIK, for intelligence-as-Process, Personality, Interests, and intelligence-as-Knowledge, builds on the investment hypothesis of Cattell (1971/1987), but represents both a refinement and a simplification of Cattell's unified perspective. The PPIK approach takes Cattell's notion that fluid intellectual abilities (Gf) (especially those of basic processes, such as memory and abstract reasoning) are generally casually antecedent to the development of crystallized intellectual abilities (Gc).

The PPIK approach indicates that the direction and level of cognitive effort devoted to acquisition of domain knowledge are influenced by a relatively small set of trait complexes (that involve personality traits, interests, and domain self-concept). The trait complexes are generally or specifically facilitating (e.g., science/math and intellectual/cultural trait complexes) of domain knowledge acquisition, or are generally or specifically impeding of domain knowledge acquisition (e.g., clerical/conventional and social trait complexes). In this framework, both Gf and Gc are important determinants of individual differences in domain knowledge. Individual differences in Gf directly influence domain knowledge in mathematics and physical sciences, and Gc directly influences knowledge in most domains, but Gf only indirectly influences knowledge in these other domains through its influence on Gc. Consistent with Cattell's (1971) investment hypothesis, the PPIK approach is developmental, in that knowledge is acquired throughout the life span. In contrast to Cattell's investment hypothesis, and in contrast to traditional approaches to intellectual abilities, the PPIK theory specifically addresses individual differences in knowledge that is not "common" to the wider culture. (Note that in CHC theory, this knowledge would help define the Gkn factor, not the Gc factor. See Table 7.1). That is, the PPIK approach attempts to bridge Binet's approach to intelligence assessment (which attempts to focus only on processes and knowledge that are common to the dominant culture) with a perspective that gives individuals credit for knowledge that may be relatively uncommon or specific to a particular area of study or interest (e.g., current events, arts, technology, social sciences, and so on; see Ackerman, 2000a; Ackerman & Rolfhus, 1999; Beier & Ackerman, 2001, 2003). This focus on an expanded consideration of domain knowledge is especially important when one cannot expect a common school curriculum or occupational and avocational experiences, such as occurs

in the study of late adolescence through adulthood. From this life-span perspective, investigations of Gf, Gc, and domain knowledge have shown that although middle-aged adults have lower Gf scores, on average, when compared to young adults, they have higher Gc scores (as expected), but also on average, have much higher levels of domain knowledge in nearly all domains assessed to date (except for knowledge in math and physical sciences, two broad domains that are most highly associated with Gf abilities; Ackerman, 2000a).

With the notable exception of advanced placement tests (see e.g., Ackerman & Rolhus, 1999) and the subject tests of the GRE, most standardized measures of educational achievement contain relatively few items that directly sample examinees' store of declarative knowledge—at the level of either general concepts or more specific facts. For example, what students may know about the periodic table or the function on the respiratory system has, at best, an indirect impact on most measures of science achievement. In part, this is an ongoing reaction of educators to methods of instruction that emphasize the acquisition of low-level factual knowledge at the expense of more general problem-solving and thinking skills. It also reflects the lack of a common curriculum in American education. Test developers are reluctant to ask questions about the concepts that many students have not encountered in their studies. However, one of the most salient features of expertise is the attainment of vast, well-organized systems of conceptual and factual knowledge in particular domains (Glaser, 1992). Failure to assess student's progress in attaining such organized knowledge bases not only results in underrepresentation of the achievement construct (Messick, 1989), but makes it much more difficult for those who study the effects of variations in instructional methods on student learning to observe significant changes in student learning. It is generally harder to obtain improvements in the sorts of general verbal problem solving required by Gc-loaded achievement tests than improvements in domain knowledge and skills.

---

### WORKING MEMORY

---

At the interface between experimental psychology and the study of individual differences, there has been considerable interest and research in the past decade on the construct of working memory (WM), especially in terms of the relationship between WM and general intelligence. A full review of this literature is beyond the scope of this chapter (see Ackerman, Beier, & Boyle, 2005). Here we provide some historical background, a brief review of the current literature, and a discussion of the implications for educational psychology.

Individual differences in immediate memory (later called short-term memory) were first investigated by Jacobs (1887), who observed that older children performed better than younger children when asked to repeat back sequences of numbers read aloud. Binet later investigated immediate memory for sentences and for unrelated words, and found similar results. Terman's (1916) introduction of the Stanford-Binet scales included both number span and sentence span tests. Terman also included a backward digit-span test, which requires examinees to repeat back the number sequence in the reverse order of presentation, which Terman claimed makes "a much heavier demand on attention" (p. 208) than the forward digit span test. Current omnibus intelligence tests, such as more recent versions of the Stanford-Binet and the Wechsler tests, have continued to incorporate span memory tests as an essential part of the overall evaluation of intellectual ability. In the most comprehensive review of the correlational literature conducted to date, Carroll (1993) identified five immediate memory factors, namely: Memory Span, Associative Memory, Free Recall Memory, Meaningful Memory [or Memory for ideas], and Visual Memory, along with one higher-order memory factor that was composed of these five lower-order group factors (see Table 7.1).

Theory and empirical research in experimental psychology have converged on a representation of immediate memory that has the characteristics of a central executive (e.g., Norman & Shallice, 1986), and two slave systems that represent a phonological loop (for "speech-based information") and a visuo-spatial sketch pad, which "is responsible for setting up and manipulating visual images" (Baddeley, 1998, p. 52). The initial attempt to bridge the construct of working memory to individual differences measures was presented by Daneman and Carpenter (1980), where they reported high correlations between a measure of WM and a measure of reading comprehension. Subsequent investigations (e.g., see Baddeley, 1986, and Daneman & Merikle, 1996) have supported a conclusion that there are indeed significant correlations between WM measures and reading comprehension, but not quite a large as initially suggested by Daneman and Carpenter.

An investigation by Kyllonen and Christal (1990) suggested a much more central role for WM ability in the larger context of intelligence. In a series of experiments, these authors suggested that reasoning ability is little more than working memory capacity, based on high correlations between latent variables of WM and a general reasoning ability. Subsequent investigators (e.g., Conway, Kane, & Engle, 1999) have taken an even more extreme view, namely that WM is "closely associated with general fluid intelligence" and "maybe isomorphic to, general intelligence and executive function" (Engle, 2002,

pp. 21-22). If it would be possible to fully account for general intelligence by a small set of WM tests, there would be great potential for revolutionizing assessment methods and practice all across the educational spectrum. However, these claims have proven to be both unwarranted, and, based on a more thorough analysis of the data (e.g., Ackerman, et al., 2005), they have been largely retracted (e.g., Kane, Hambrick, & Conway, 2005). Current estimates of the overlap between WM and general intelligence are in the neighborhood of 25 percent to 50 percent of shared variance—about the same as the shared variance between other memory abilities identified by Carroll (1993) and general intelligence.

Other investigations have indicated that there are substantial correlations between more traditional span memory abilities and the more complex measures of WM, suggesting that there may not be much “value added” by WM measures above and beyond the kinds of span measures that have been used for nearly a century of intelligence testing (e.g., see Beier & Ackerman, 2004).

Beyond the issue of shared variance between WM and intelligence are other important concerns that might have relevance for understanding the components of intellectual abilities. Several investigations have been conducted that attempt to determine whether components of working memory (e.g., the phonological loop and the visuo-spatial sketchpad) represent differentiable abilities (e.g., Shah & Miyake, 1996). Other investigations have attempted to map out the developmental course of WM capacity during childhood and adolescence (Fry & Hale, 1996) and in the course of adult aging (e.g., Phillips & Hamilton, 2001). Still other investigations have attempted to separate the executive or attentional aspects of individual differences in WM from the processing and storage components of WM (for a review, see Oberauer, in press).

To date, there are too few comprehensive data and too much controversy to allow us to draw substantive conclusions from this line of investigation. The most promising domain of inquiry appears to be the narrower, but still critically important area of individual differences in reading comprehension. It may very well be that the key role played by WM abilities, beyond what can be assessed by extant measures of span memory and general intellectual abilities, lies in the specific relations between the attentional, storage, and processing aspects of WM and the underlying determinants of individual differences in reading comprehension. An answer to the question of whether individual differences in WM are a cause, consequence, or concomitant correlate of reading comprehension is currently not known.

Regardless of the outcome of these investigations, individual differences in working or short-term memory capacity will surely continue to be a fruitful area of inves-

tigation for educational researchers. Students commonly fail to learn because task demands exceed the span of information that they can maintain in an active state in working memory. For example, it is difficult to infer relationships between ideas unless both can be held in attention simultaneously. One of the most effective ways to improve performance of students who cannot do this, then, is to reduce the information processing or storage requirements of tasks, by automatizing component skills, by redesigning the tasks, or by altering instruction to offload non critical, attention-demanding processes (Merriënboer, Kirschner, & Kester, 2003). For example, even adults find it difficult to learn complex skills if they must monitor their own behavior (Kanfer & Ackerman, 1989). Learning is more efficient if the monitoring function is offloaded (to another learner, a coach, or the computer) until the student has reduced the attention demands of task performance through practice. Indeed, probably the most well documented way to improve the performance of less able, younger, or novice learners is to reduce the information processing burdens of the task (Snow, 1978). Unfortunately, this is often done by offloading problem-solving or reasoning rather than ancillary task demands. Temporary improvements in performance thus come at the expense of the development of higher-level thinking skills, thereby not developing the more critical aptitudes for academic learning (Martinez, 2000; Snow, 1996).

---

## OTHER INFLUENCES IN COGNITIVE FUNCTIONING

---

### Stereotype Threat and Group Differences in Intellectual Abilities

In 1995 Steele and Aronson (1995) introduced the concept of stereotype threat (“being at risk of confirming, as self-characteristic, a negative stereotype about one’s group,” p. 797) into the broader discussion of intellectual test performance. Specifically, these authors claimed that when African-American students at Stanford University were told that a test they were about to take was “diagnostic” of their “reading and verbal reasoning abilities” (p. 799), they performed comparatively worse than a control condition, when compared to White students in the same conditions. Steele and Aronson (1995) suggested that activating a negative stereotype (in this case, that African-Americans perform worse on intelligence-like tests) caused an “inefficiency of processing” that would lead to depressed performance scores. Although the results of their initial studies were weak (of the two

reported studies, only one reached statistical significance), some distance from the original research has increased the magnitude of the reported effect. For example, the nonsignificant result (" $p < .19$ ") in Steele and Aronson (1995) was reported to be "strong evidence of stereotype threat; Black participants greatly underperformed White participants in the diagnostic condition but equaled them in the non diagnostic condition"; Steele, 1997, p. 620.) Steele also expanded the construct to other identifiable groups that have negative stereotypes (e.g., women and math performance). Other investigators have further expanded the construct to include a wide variety of other negative stereotypes that could presumably affect intellectual ability performance (e.g., Hispanics, low-income people [Good, Aronson, & Inzlicht, 2003], and even psychology students, in the context of a proofreading task [Seibt & Förster, 2004]). To date, there have been several empirical research studies that have purported to replicate the Steele and Aronson (1995) studies or expand the framework for stereotype threat. However, there have also been several studies that have purported to show that the results of Steele and Aronson (1995) cannot be replicated (e.g., see Cullen, Hardison, & Sackett, 2004), and there have been several articles that have outlined several methodological and statistical objections that potentially threaten the validity of the results that show negative effects of stereotype threat (e.g., Sackett, Hardison, & Cullen, 2004). It is not our aim to provide a complete review of this literature here, but there are two important issues regarding stereotype threat and intellectual abilities that merit discussion. The first issue concerns the conditions of intellectual ability testing, and the second issue is the potential of stereotype threat for understanding race, gender, and other group differences in intellectual abilities.

**Conditions of Testing.** In the hundred or so years of modern intellectual ability assessment, it seems to have been long forgotten (or rendered a quaint anachronism in the modern educational system where frequent testing takes place), that the mental test is essentially "an experiment" (e.g., see Terman, 1924). That is, as noted by E. G. Boring (quoted in Terman, 1924) "methodologically there is no essential difference between a mental test and a scientific psychological experiment." (p. 98). At a narrow level, this refers to the fact that each test item is a stimulus, and each answer is a response, in the classic behaviorist representation. But, at a broader level, this kind of consideration applies to the entire testing milieu as well. Several early investigators easily demonstrated that the conditions of testing (especially in terms of the rapport established between examiner and examinee) can be important determinants of performance on intelligence

tests (e.g., Rothney, 1937; Sacks, 1952; and Sears, 1943). Moreover, the hypothesis that rapport was partly responsible for race differences in intelligence test scores had been reasonably well researched in the 1920s and 1930s (e.g., see Canady, 1936; Klineberg, 1928). Far from being something new and interesting, the idea that the motivation and attitude of the examinee might adversely affect intelligence test scores is as old as the Binet-Simon scales (e.g., Binet & Simon 1905; Rothney, 1937; Sears, 1943; see also Ackerman, 1996, for a more recent discussion of Binet's instructions on the conditions of intelligence testing).

As noted by many researchers over the years, what the examinee (or research participant) brings with him or her to the test can be an important determinant of test performance. Experimental research has long pointed to stressors, such as the physical environment (e.g., noise, heat/cold, vibration) as having relatively generic effects on performance—namely marked impairment of cognitive processes for tasks that are demanding or novel, and relatively little impairment for well-practiced skills (e.g., see Hancock, 1986). There is also a substantial list of psychological stressors that have been identified as having potential for impairing cognitive performance. These include general anxiety, performance evaluation apprehension, interpersonal competition, and so on. Other factors also enter into possible mediators of cognitive performance, and some have been shown to have interactive effects, such as time-of-day of testing, personality traits such as introversion/extroversion and impulsivity, and even caffeine intake (e.g., see Revelle, Humphreys, Simon, & Gilliland, 1980). Many cognitive and intellectual abilities tests are especially susceptible to these influences, because they are designed to elicit the "maximal" performance of the examinees—that is, the level of performance attained by an individual who is completely focused and optimally attentive to the task, and because such tasks have been further designed to eliminate as far as possible, any transfer of skills from outside the testing situation (Ackerman, 1994; Cronbach, 1949).

However, there are other tests that are designed to measure knowledge and skills that have been developed outside of the testing situation (e.g., some crystallized intelligence measures, such as vocabulary, general information, and domain knowledge) that should be much less susceptible to these various stressor effects because they are less dependent on limited attentional resources in the testing situation. To date, there has been virtually no research that has established whether or not the effects of stereotype threat interventions are moderated by the type of tests, but in order to understand the underlying mechanisms of such phenomena, such research is needed.

Ultimately, though, it is most certainly true that at the individual and group levels, some conditions of testing are likely to be more or less facilitating or impeding of performance on some intellectual ability tests. On the one hand, the fact that there may be interactions between these conditions of testing and group differences suggests perhaps that greater care must be taken to optimize the conditions for each group in order to maximize overall scores. On the other hand, it must be kept in mind that the primary goal of testing in admissions and employment contexts is to measure individual differences in knowledge, skills, and other characteristics required for success in school or the workplace. Equalizing the scores of identifiable groups also raise (or at least do not reduce) the overall test validity. For example, using nonstandard testing procedures that raise the scores of an identifiable group of examinees but do not simultaneously raise the scores on the criterion variable (such as grade point average) will not yield a psychometrically useful result. On the other hand, identification of factors that systematically moderate test-criterion relationships suggests ways in which the criterion situations might be modified better to accommodate the needs of different individuals. For example, a high correlation between spatial ability and performance in mathematics in a particular institution may reflect a general fact of nature. More likely, it reflects something about the way in which mathematics is taught at that institution.

In the final analysis, what is perhaps most surprising about the recent studies on stereotype threat is how small the effects are, in the larger context of other conditions that can depress intelligence test performance. Although there have been no studies of omnibus intelligence testing under stereotype threat, the current evidence suggests that the relatively weak results demonstrated so far would result in only a few points in overall IQ test performance under stereotype threat, at least in the laboratory and field contexts in which it has been investigated.

### Gender Differences

The study of gender differences in cognitive/intellectual functioning has been a recurring theme in the past 90 or so years. The issue was first raised by Yerkes, Bridges, & Hardwick (1915), in their development of a "point scale" for measuring intelligence of children and adolescents. In their norming sample, they noticed that at some ages, girls tended to have higher average raw scores, while at other ages, boys had higher average scores. The differences in overall scores were small, but these authors concluded that separate intelligence norms should be used for girls and boys, such that each individual could be evaluated

with respect to his or her own gender reference group. Terman (1916) in his development of the Stanford-Binet intelligence test also noted small gender differences, on the order of 2 or 3 points. For his test, girls tended to perform slightly better, on average, than boys up to age 13, after which boys tended to perform slightly better, on average, than girls. Terman's approach to gender differences was, however, entirely different from that adopted by Yerkes et al. That is, Terman created a single scale, which was used for both boys and girls. In later revisions of the Stanford-Binet (e.g., Terman & Merrill, 1937), the authors describe how they explicitly removed items from the test that showed large mean gender differences, in order to yield a result of no overall gender differences in intelligence test scores.

Because the specific selection of individual items and scales for an omnibus intelligence (IQ) test, such as the Stanford-Binet or the Wechsler, is somewhat arbitrary (as long as the items meet the basic properties of criterion-related validity and construct validity), few objections have been raised over the ensuing decades that the IQ test is essentially unfair to boys or girls. Whether there is a basis for arriving at a different conclusion about equivalent scores for the genders lies beyond the scope of this chapter and most likely falls into the political or public policy domain. On the one hand, Terman's decision to "make" the genders equivalent on IQ renders impossible any substantive discussion of overall intelligence differences between genders. On the other hand, when abilities that are lower in the hierarchy than general intelligence are considered, there is ample evidence for gender differences. It is these differences that we turn to for further treatment. A comprehensive review of the literature on gender differences in intellectual abilities lies beyond the scope of this chapter. Interested readers should consult some of the more extensive treatments of this research domain (e.g., Halpern, 2000; Linn & Peterson, 1986; W. Willingham & Cole, 1997).

*Verbal, Spatial, and Math Content Abilities.* Although hundreds of studies of gender differences in abilities have been reported in the literature, the most thoroughly researched areas of inquiry have generally proceeded along the lines of different test content, in contrast to the underlying processes of cognitive functions. The consensus opinions of researchers in the field are that, on average, girls tend to perform better than boys on a wide range of verbal tasks including fluency, reading comprehension, and vocabulary (e.g., see Halpern, 2000; Hyde & Linn, 1988). At the individual test level, the gender differences are typically small to nonexistent for reading comprehension, vocabulary, and verbal reasoning tests; intermediate for tests of grammar, syntax, or style; and

largest for verbal fluency, clerical speed, and spelling tests (Cleary, 1992; Halpern, 2000). Thus, it is not so much the verbal content that generates the sex difference as the particular type of processing that is required (Lohman, 1994b). Tests that show the largest female advantage require rapid, sequential processing of arbitrary sequences of letters or phonemes. A critical requirement is keeping track of order information. For example, on a spelling test, knowing the correct letters is not enough; sequence is crucial. Anderson (1983; Anderson et al., 2004) posits that order information for such stimuli is represented in memory by a particular type of mental code (the linear order or string code). Tests that demand this sort of thinking are more likely to have their primary loading on one of the primaries that define the Ga, Gs, or Grw factors in the CHC model than the Gc factor (see Table 7.1). Indeed, one way to predict whether a task will elicit sex differences is to examine the extent to which such processing is required, especially when it must be done so fluidly and flexibly (as in tongue twisters and secret languages).

Spatial tasks often show large differences that favor males. Because of the diversity of spatial ability factors (e.g., see Lohman, 1988), and the fact that these abilities tend to be less frequently measured in standard academic testing situations, there remains some controversy about the overall magnitude of gender differences in spatial abilities, but estimates typically run in the neighborhood of about  $d = .4$  to  $.8$  (e.g., see Voyer, Voyer, & Bryden, 1995), which would be considered a moderate to large effect in Cohen's framework. For adolescents and adults, differences are largest on tests that require mental rotation of three-dimensional stimuli and smallest on those that can be solved by reasoning. Therefore, as with the sex difference on verbal tasks, the difference is probably best characterized not by the stimulus as by the type of mental representation that must be generated, and the nature of transformations that must be performed on that representation. High-spatial individuals evidence the ability to combine and recombine visual images at will (Lohman, 1994a).

For math abilities, boys tend to perform, on average, better than girls on several kinds of math tests, but especially in the domain of problem solving (on the order of  $d = .29$ ; see Hyde, Fennema, & Lamon, 1990), but there are negligible differences between the genders on computational math tests (e.g., those that involve typically highly speeded basic arithmetic functions). When differences in computation skills are observed, they tend to favor girls. Differences in mathematical problem solving emerge at high school ages and persist into adulthood.

Although there is some evidence that the gender differences in these three broad content domains (especially mathematical abilities) have diminished

in magnitude somewhat over the past few decades (e.g., see Feingold, 1988), there is little consensus that these differences have entirely disappeared or are likely to disappear in the near future. Further, because the interindividual differences in all of these abilities within gender groups are much larger in magnitude than the between-group differences, generalizations from group means to individual scores are not warranted. In addition, there is a well-established documentation of a divergence between gender differences in ability test scores and gender differences in academic grades, such that women tend to obtain higher grades, on average, than boys do, even when boys tend to do better on ability tests (e.g., see Willingham & Cole, 1997). Recent decisions to include measures of writing abilities on college entrance tests may alter this state of affairs, however. Girls typically outperform boys on measures of writing abilities, and so they may outperform boys both on grades and on the verbal and writing portions of the entrance tests. Nonetheless, the larger differences between gender groups on spatial abilities, some math abilities, and writing abilities have implications for selection into courses of study at the postsecondary level, especially in the physical sciences and engineering domains (Shea et al., 2001). Further, as noted by Stanley and Benbow (1982), small group differences at the mean, can result in substantial differences in the ratios of the respective group representation at the tails of the distribution of abilities.

Like all abilities, spatial abilities are amenable to training and practice. Certainly performance on particular spatial tasks can be improved substantially (e.g., Lohman, 1988; Lohman & Nichols, 1990). The extent to which improvements after small amounts of practice transfer to other spatial tasks is less clear. To the degree that training on spatial abilities may reduce the overall magnitude of individual differences, such interventions might have the effect of also reducing gender group differences. Thus, it might be possible that targeted instructional programs may ameliorate or at least diminish gender differences in spatial abilities that are critical to academic success in physical sciences and engineering. The alternative, of course, is to modify instruction in such disciplines to reduce at least some of the demands on spatial abilities.

Finally, simultaneous consideration of trait clusters or complexes often gives much more information than the examination of traits in isolation. In the case of gender differences, it is the profile of analog spatial versus sequential verbal abilities that shows the greatest relationship with both physiological variables—such as hormone levels (Bock, 1973; Nyborg, 1983)—and personality variables (Riding & Boardman, 1983). Explorations of relationships between abilities and learning styles seem

more fruitful when studies attend to the spatial-sequential profile rather than to each variable separately, especially when the "verbal" domain is represented by spelling or verbal fluency rather than a more general verbal skill.

### AGE DIFFERENCES

With about 25% of all students enrolled in postsecondary education in the United States over the age of 30 (*Chronicle of Higher Education Almanac*, 2004), and the proportion expected to rise in the foreseeable future, considerations of adult aging and intelligence have become more important concerns for educators than perhaps they were in the past. First, we provide a brief review of the general patterns of aging and abilities, followed by a review of the relations between abilities and learning in the context of aging. We conclude this section with a discussion of some key implications of aging, abilities, and education.

#### Aging and Intellectual Abilities

Cross-sectional studies generally show that intellectual ability peaks around the ages of 18–25. Cohort analyses typically reveal higher scores for each succeeding generation. Coupled with longitudinal data that reveal somewhat later peaks (in the late 20s), the general sense of adult intelligence is that declines in abilities can be expected as early as the 30s, though large differences in intelligence are not usually found until the 40s and later. Examination of general intelligence scores addresses only one salient aspect of the effects of aging on abilities. Different abilities have different patterns of growth, stability, and decline as adults enter middle age and beyond (e.g., see Schaie, 1970, 1996). From a content ability perspective, math and spatial abilities show peak levels at the youngest ages, followed by substantial declines in the 30s and 40s. Abstract reasoning and immediate memory (short-term memory and working memory) also show similar early peak performance, followed by substantial declines with increasing age. In contrast, verbal abilities tend to show a pattern of growth throughout early and middle adulthood, followed by stability well into middle age. Marked declines in verbal abilities are often not seen until age 60 or later.

The pattern of adult ability changes with age tends to be consistent with theory of fluid and crystallized intelligence outlined by Cattell (1943). That is, fluid intellectual abilities (e.g., abstract reasoning, immediate memory) tend to peak in early adulthood, followed by declines into middle and late adulthood. In contrast, crystallized abilities such as vocabulary and verbal comprehension,

which are most highly associated with educational and experiential influences, show growth and stability well into the middle-adult years. The declines in fluid abilities with increasing age are typically larger than the increases in crystallized abilities, so that the overall pattern (as noted earlier) is for an overall decline in composites of general intelligence that give equal weight to the two types of abilities.

There are two important points that need to be made about this pattern of ability changes with age: First, as far as educational applications are concerned, there is no *inherent* reason why there should be an equal weighting of fluid and crystallized abilities. For domains such as physical science and mathematics, fluid intellectual abilities are more predictive of educational outcomes than crystallized abilities, especially for young children or novices of any age. For a wide range of other educational domains, crystallized abilities are more predictive of educational outcomes. (Note that in all domains, prior knowledge and skill in the domain are generally the best predictors of future success, especially when the demands of future learning are similar to those of learning to date.) Interestingly, adults have a relatively good sense of their own abilities, and of the nature of changes in fluid and crystallized kinds of abilities with increasing age (e.g., see Ackerman et al., 2002). Colleges and universities that offer programs of study specifically aimed at older, nontraditional students tend to match the content of course offerings to these changes in abilities (e.g., see Ackerman, 2000a). That is, the postsecondary courses most frequently available to older adults (in terms of night and weekend courses) are in the domains of the humanities, social sciences, and business; the courses least frequently offered are in the physical sciences and mathematics, especially at advanced levels of study.

The second point regarding age changes in fluid and crystallized abilities with age is that there are substantial interindividual differences in intraindividual change with age. That is, the average trends of growth, stability, and decline are relatively stable, but many individuals may have earlier or later peaks of ability than others, and more or less substantial declines in the various abilities with age. Prediction of these different aging patterns is a topic of keen interest in the field of life-span developmental psychology, but at present there are few, if any, diagnostic indicators for such differences (see Li et al., 2004).

The age-related changes in intellectual abilities raise important concerns for both selection programs and the design of instructional methods, especially in postsecondary education. Traditional selection programs that depend on omnibus tests (such as the SAT and GRE) have good validities for prediction of academic performance, especially for the first semester or year of college/university or



graduate/professional school performance (e.g., see Lin & Humphreys, 1977). Given the age-related changes in performance on these tests, one can expect that as adults reach age 30 or 40, they will perform at less competitive levels on these selection tests (because both cohort differences and age-related changes indicate that the older adults will perform more poorly on omnibus ability tests than younger adults). Whether these lower ability levels match the lower expected performance on academic indicators such as grade point average remains an open question. A review by Kasworm (1990) of adult undergraduates reveals just how sparse the literature is on these issues. Kasworm identified only 11 studies in this domain, and no definitive results, though several studies suggested that older adults perform at a level comparable to younger adults. Most likely, the nature of the outcome measure moderates the type of relationship observed. Adults typically perform better on essay tests and other tasks that allow them to show how well they have integrated new concepts into existing conceptual networks.

When one considers both the direction of educational interests of middle-aged and older adults (e.g., towards domains that depend more on crystallized intellectual abilities and less on fluid intellectual abilities) and the pattern of intellectual ability changes with age, it may very well be that middle-aged adults can be expected to perform at a higher level than younger adults in the programs where both are likely to be found. Selection procedures that more specifically match the ability demands of the educational program to the tests used for selection might find that higher weights for crystallized abilities might be in order in some domains. Under this scenario, middle-aged adults may find that they are much more competitive for selection, and that the selection procedure may turn out to have higher criterion-related validity than a system that uses only an omnibus intelligence or aptitude measure for selection. There is at present, though, far too little empirical research in this area to be able to provide specific recommendations.

In considering how to integrate changes in intellectual ability with the nature of instruction as far as middle-aged and older adults are concerned, Lorge and Kushner (1950) suggested that the main issue has to do with the "tempo" of instruction. That is, given that middle-aged and older adults tend to show the largest deficits (in comparison to young adults) on speeded intellectual tasks, the key toward optimizing instruction is to reduce the speed demands of the educational situation. Schaie and Willis (1978), in their review of life-span development and implications for education, agree with the conclusion of Lorge and Kushner, in that the pace of instruction should be slower for older adults than for young adults. However, Schaie and Willis also suggest that educators

should consider two different approaches to education with older learners, namely "teaching to the weakness vs. teaching to the strength" (p. 131). In the former case, this could mean evaluating the patterns of cognitive decline with age, and designing educational interventions that attempt to remediate or ameliorate such declines (e.g., in terms of providing refresher courses on math and spatial tasks, or in terms of providing explicit training on memory improvement methods). In the latter case, building to strengths means that the instructional design could be tailored to that which the middle-aged or older adults bring to the learning situation that are not as well developed in young adults. Given the pattern of crystallized ability increases with age, one possibility is to structure instruction so that there is a greater dependence on transfer of knowledge, or to depend more on verbal strategies for problem solving than for nonverbal reasoning. In addition, Schaie and Willis (1978) suggest taking account of different learning strategies that appear to be better developed in older adults than younger adults (such as self-study as opposed to classroom learning), as a means toward further facilitating educational achievement.

---

#### CONCLUDING REMARKS

---

After 100 years of research in the field of human intelligence, and nearly 100 years of practical assessment of human intelligence in education, much is known about the structure and functions of abilities. Developments in the past decade fall into four different categories: (1) new theories that are largely reintroductions of previously investigated constructs; (2) relatively modest increments in knowledge about key abilities (what Kuhn, 1970, described as "normal science"); (3) revolutionary theoretical proposals with little or no empirical support; and (4) shifts of focus in the field, partly in reaction to the changing face of higher education. Although the study of aging and intelligence in adults is a domain that has had substantial research over the past 80 years, we believe that the last category of developments, in terms of application of intelligence theory to adult education, and in terms of aging and individual differences in domain knowledge, represents an important new area that will provide both opportunities and challenges for research and practice. Because much of what has been learned in the field depends on long-term or longitudinal research studies, this kind of research has been difficult and expensive to conduct (e.g., see Learned & Wood, 1938). Researchers have also been slow to recognize the importance of building (or using) tests that have score scales that better support interpretations of longitudinal changes than can be squeezed out of raw scores (see, e.g.,

Embretson & Hershberger, 1999; Kolen & Brennan, 2004). The paucity of longitudinal studies in the literature suggests that there are great opportunities for researchers in the future to make marked contributions to the field. We hope that, by the time the next review of the field is conducted, there will be much more to say about the interactions between intellectual abilities, other constructs, and the nature of adult intellectual development.

We also believe that intuitive theories about ability mislead not only those unfamiliar with research on abilities but, in a subtler way, the rest of us as well. Probably the most common naïve theories about ability are that (a) intelligence is unidimensional, not multidimensional, and (b) a good ability test would reveal the examinee's innate potential or capacity on this dimension. By definition, innate capacity is independent of education, culture, motivation, and other contaminants. These beliefs may explain the recurring appeal to the educators of figural reasoning tests that purport to be culture-fair measures of intelligence (e.g., Naglieri, 1997). But professionals are misled as well. If given a choice, most would choose a measure of Gf over a measure of Gc or Gkn as the better test of intelligence, even for adults. The fact that Gc and Gkn tests sample what people know and can do and are generally much better indicators of readiness for learning or performance in that domain than a measure of Gf is easily overlooked because the Gf tests (especially figural reasoning tests) appear to measure something untouched by education, culture, and the like. However, Gf-like measures are most relevant when one knows nothing about the domain (or when prediction is over much longer intervals; see Horn & Noll, 1997). Thereafter, what one knows and can do in the domain become measures of functional intelligence in the domain. These measures of acquired knowledge and skill reflect the product not only

of past investments of Gf and other abilities, but also of interest, motivation, and other affective and conative variables. In such cases, Gf is more a measure of historical intelligence than of functional intelligence. It is no more the real intelligence than one's childhood home is one's current abode (except, of course, for those who never went anywhere).

Those anxious to embrace a multidimensional view of educational outcomes might attend more to the measure of Gkn—that is, the development of organized systems of conceptual and factual knowledge in domains—than to continued refinements of general measures of Gc and Grw abilities. We also believe that Gkn measures would be more sensitive to instructional interventions than tests that depend on more general thinking and problem-solving skills that define Gf. Indeed, although most educators are quite sensitive about the extent to which intelligence and other ability tests measure the direct products of schooling, few seem to care that items on many achievement tests are really better measures of general reasoning abilities (or, in the case of performance assessments, the ability to follow directions) than of achievement in the domain itself.

Lastly, abilities are, as Snow and Yalow (1982) noted, education's most important product as well as its most important raw material. Abilities are thus best viewed as both inputs and outcomes of good education. As inputs, they not only predict but moderate the success different learners experience in different instructional environments (Corno et al., 2002). As outcomes, they increase students' readiness for new learning in yet-to-be-experienced environments. More than any other construct, ability is at the heart of education. We hope that it will also continue to occupy a central role in educational psychology.

## References

- Ackerman, P. L. (1994). Intelligence, attention, and learning: Maximal and typical performance. In D. K. Detterman (Ed.) *Current topics in human intelligence: Vol. 4: Theories of intelligence* (pp. 1-27). Norwood, NJ: Ablex.
- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, 22, 229-259.
- Ackerman, P. L. (2000a). Domain-specific knowledge as the "dark matter" of adult intelligence: gf/gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences*, 55B(2), P69-P84.
- Ackerman, P. L. (2000b). A reappraisal of the ability determinants of individual differences in skilled performance. *Psychologische Beiträge*, 42, 4-17.
- Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist*, 38, 85-93.
- Ackerman, P. L., Beier, M. B., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, 34, 587-605.
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131, 567-589.
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30-60.
- Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual

- differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 6, 259-290.
- Ackerman, P. L., Kyllonen, P. C., & Roberts, R. D. (Eds.). (1999). *Learning and individual differences: process, trait, and content determinants*. Washington, DC: American Psychological Association.
- Ackerman, P. L., & Rolffhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and non-ability traits. *Psychology and Aging*, 14, 314-330.
- Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). The influence of topic knowledge, domain knowledge, and interest on the comprehension of scientific exposition. *Learning and Individual Differences*, 6, 379-397.
- Alexander, P. A., & Murphy, P. K. (1999). Learner profiles: Valuing individual differences within classroom communities. (pp. 413-436). In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants*. Washington, DC: American Psychological Association.
- Alexander, W. P. (1935). Intelligence, concrete and abstract: A study in differential traits. *British Journal of Psychology Monograph*, 6(19).
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111, 1036-1060.
- Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing and early reading ability in preschool children. *Journal of Experimental Child Psychology*, 83, 111-130.
- Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.
- Baddeley, A. (1998). *Human memory: Theory and practice* (Rev. Ed.). Needham Heights, MA: Allyn & Bacon.
- Baird, L. L. (1969). Prediction of accomplishment in college: A study of achievement. *Journal of Counseling Psychology*, 10(3), 246-253.
- Beier, M. E., & Ackerman, P. L. (2001). Current events knowledge in adults: An investigation of age, intelligence and non-ability determinants. *Psychology and Aging*, 16, 615-628.
- Beier, M. E., & Ackerman, P. L. (2003). Determinants of health knowledge: An investigation of age, gender, abilities, personality, and interests. *Journal of Personality and Social Psychology*, 84(2), 439-448.
- Beier, M. E., & Ackerman, P. L. (2004). A reappraisal of the relationship between span memory and intelligence via "best evidence synthesis." *Intelligence: A Multidisciplinary Journal*, 32, 607-619.
- Binet, A., & Simon, T. (1905/1973). *The development of intelligence in children*. (El. Kite, Trans.). New York: Arno Press.
- Bock, R. D. (1973). Word and image: sources of the verbal and spatial factors in mental test scores. *Psychometrika*, 38, 437-457.
- Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.
- Brody, N. (2003a). Construct validation of the Sternberg Triarchic abilities test: Comment and reanalysis. *Intelligence*, 31, 319-329.
- Brody, N. (2003b). What Sternberg should have concluded. *Intelligence*, 31, 339-342.
- Buckingham, B. R. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, 12, 271-275.
- Canady, H. G. (1936). The effect of "rapport" on the IQ: A new approach to the problem of racial psychology. *Journal of Negro Education*, 5, 209-219.
- Carnap, R. (1987). The confirmation of laws and theories. In J. A. Kourany (Ed.), *Scientific knowledge: Basic issues in the philosophy of science* (pp. 122-138). Belmont, CA: Wadsworth.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, J. McK. (1890). Mental tests and measurements. *Mind*, 15, 373-380.
- Cattell, J. McK., & Farrand, L. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review*, 3, 618-648.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153-193.
- Cattell, R. B. (1971/1987). *Abilities: Their structure, growth, and action*. [Revised and reprinted as *Intelligence: Its structure, growth, and action*.] Amsterdam: North-Holland.
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-76). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chronicle of Higher Educational Almanac, 2004-2005* (2004). Washington, DC: Author.
- Cleary, A. T. (1992). Gender differences in aptitude and achievement test scores. In *Sex equity in educational opportunity, achievement, and testing: Proceedings of the 1991 ETS Invitational Conference* (pp. 51-90). Princeton, NJ: Educational Testing Service.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (1999). Is Spearman's G determined by speed or working memory capacity? *Psychology*, 10, #74. Deposited July 11, 2002, at <http://psycprints.ecs.soton.ac.uk/archive/0000709/>.
- Corno, L., Cronbach, L. J., Lohman, D. F., Kupermintz, H., Mandinach, E. B., Porteus, A. & Talbert, J. E. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mohwah, NJ: Lawrence Erlbaum Associates.
- Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper.
- Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper.
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, 89, 220-230.

- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19(4), 450-466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3, 422-433.
- Embretson, S. E. & Hershberger, S. L. (1999). *The new rules of measurement: What every psychologist and educator should know*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19-23.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, 43, 95-103.
- Feldt, L. S. (1967). A note on the use of confidence bands to evaluate the reliability of a difference between two scores. *American Educational Research Journal*, 4, 139-145.
- Frisbie, D. A. & Andrews, K. (1990). Kindergarten pupil and teacher behavior during standardized achievement testing. *Elementary School Journal*, 90, 435-448.
- Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, 7(4), 237-241.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Gardner, H. (1993). *Multiple intelligences: The theory into practice*. New York: Basic Books.
- Gardner, H. (1999). Who owns intelligence? *Atlantic Monthly*, 283, 67-76.
- Gardner, H. (2003, April). *Multiple intelligences after twenty years*. Paper presented at the meeting of the American Educational Research Association, Chicago. Retrieved Dec. 30, 2004 from [http://www.pz.harvard.edu/Pis/HG\\_ML\\_after\\_20\\_years.pdf](http://www.pz.harvard.edu/Pis/HG_ML_after_20_years.pdf)
- Glaser, R. (1992). Expert knowledge and processes of thinking. In D. F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 63-75). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goddard, H. H. (1908). The Binet and Simon tests of intellectual capacity. *Training School Bulletin*, 5, 3-9.
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24, 645-662.
- Goodenough, F. L. (1926). *Measurement of intelligence by drawings*. Yonkers-on-Hudson, NY: World Book.
- Guilford, J. P. (1959). *Personality*. New York: McGraw-Hill.
- Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, (Vol. 4, pp. 35-71). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Gustafsson, J.-E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407-434.
- Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. Chapter in D. C. Berliner, & R. C. Calfee (Eds.) *Handbook of educational psychology* (pp. 186-242.) New York: Simon & Schuster Macmillan.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hancock, P. A. (1986). Sustained attention under thermal stress. *Psychological Bulletin*, 99, 263-281.
- Herrnstein, R. J., Nickerson, R. S., de Sánchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist*, 41, 1279-1289.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. Ann Arbor, MI: University Microfilms International.
- Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences. Advances in theory and research* (pp. 61-116). New York: W. H. Freeman.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual*. Chicago: Riverside.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, 107(2), 139-155.
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: a meta-analysis. *Psychological Bulletin*, 104(1), 53-69.
- Jacobs, J. (1887). Experiments on "prehension." *Mind*, 12, 75-79.
- Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.
- Joncich, G. (1968). *The sane positivist: A biography of Edward L. Thorndike*. Middletown, CT: Wesleyan University Press.
- Juhel, J. (1991). Spatial abilities and individual differences in information processing. *Intelligence*, 15, 117-137.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 66-71.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology—Monograph*, 74, 657-690.
- Kasworm, C. E. (1990). Adult undergraduates in higher education: A review of past research perspectives. *Review of Educational Research*, 60, 345-372.
- Klein, P. D. (1997). Multiplying the problems of intelligence by eight: A critique of Gardner's theory. *Canadian Journal of Education*, 22, 377-394.
- Klineberg, O. (1928). An experimental study of speed and other factors in "racial" differences. *Archives of Psychology*, 15(93), 1-109.

- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.
- Koertz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont Portfolio assessment program: Findings and implications. *Educational Measurement Issues and Practice*, 13, 5-16.
- Kuhlman, F. (1912). A revision of the Binet-Simon system for measuring the intelligence of children. *Journal of Psycho-Asthenics Monograph Supplement* (Vol. 1, No 1).
- Kuhlman, F. (1922). *A handbook of mental tests*. Baltimore: Warwick & York.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389-433.
- Lakatos, I. (Ed.). (1978). *The methodology of scientific research programmes: Philosophical papers* (Vol. 1). Cambridge, UK: Cambridge University Press.
- Learned, W. S., & Wood, B. D. (1938). *The student and his knowledge*. New York: The Carnegie Foundation for the Advancement of Teaching.
- Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, 15, 155-163.
- Lin, P. C., & Humphreys, L. G. (1977). Predictions of academic performance in graduate and professional school. *Applied Psychological Measurement*, 1, 249-257.
- Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In E. F. Lindquist (Ed.), *Educational measurement*. Washington, DC: American Council on Education.
- Linn, M.C. & Peterson, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde & M. C. Linn, (Eds.) *The psychology of gender: Advances through meta-analysis*. Baltimore: Johns Hopkins University Press.
- Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181-248). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lohman, D. F. (1994a). Spatial ability. In R. J. Sternberg (Ed.), *Encyclopedia of intelligence* (pp. 1000-1007). New York: Macmillan.
- Lohman, D. F. (1994b). Spatially gifted, verbally inconvenienced. In N. Colangelo & S. Assouline (Eds.), *Talent development: Proceedings of the Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 251-264). Dayton, OH: Ohio Psychology Press.
- Lohman, D. F. (1999). Minding our p's and q's: On finding relationships between learning and intelligence. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds), *Learning and individual differences: Process, trait, and content determinants* (pp. 55-76). Washington, DC: American Psychological Association.
- Lohman, D. F. (2000). Complex information processing and intelligence. In R. J. Sternberg (Ed.) *Handbook of human intelligence* (2nd ed., pp. 285-340). Cambridge, UK: Cambridge University Press.
- Lohman, D. F. (2001). Fluid intelligence, inductive reasoning, and working memory: Where the theory of Multiple Intelligences falls short. In N. Colangelo & S. Assouline (Eds.), *Talent Development IV: Proceedings from the 1998 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 219-228). Scottsdale, AZ: Gifted Psychology Press.
- Lohman, D. F., & Hagen, E. (2001). *Cognitive abilities test (Form 6)*. Itasca, IL: Riverside Publishing Company.
- Lohman, D. F., & Hagen, E. (2002). *Cognitive abilities test (Form 6) Research Handbook*. Itasca, IL: Riverside Publishing Company.
- Lohman, D. F., & Nichols, P. D. (1990). Training spatial abilities: Effects of practice on rotation and synthesis tasks. *Learning and Individual Differences*, 2, 67-93.
- Lorge, I., & Kushner, R. (1950). Characteristics of adults basic to education. *Review of Educational Research*, 20, 171-184.
- Lubinski, D. (2000). Assessing individual differences in human behavior: "Sinking shafts at a few critical points." *Annual Review of Psychology*, 51, 405-444.
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities 100 years after Spearman's (1904) "'General Intelligence,' Objectively determined and measured." *Journal of Personality and Social Psychology*, 86, 96-111.
- Lubinski, D., & Benbow, C. P. (1995). An opportunity for empiricism [Review of H. Gardner, *Multiple intelligences: The theory in practice*]. *Contemporary Psychology*, 40, 935-938.
- Martinez, M. E. (2000). *Education and the cultivation of intelligence*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McGrew, K. S. & Evans, J. J. (2004). Internal and external factorial extensions to the Cattell-Horn-Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's seminal 1993 treatise. Retrieved Dec. 30, 2004 from <http://www.iapsych.com/CHCPP/4.CHCTheoryExtensions.html>: Institute for Applied Psychometrics.
- McNemar, Q. (1965). Lost: Our intelligence? Why? *American Psychologist*, 871-882.
- Merriënboer, J. J. G. van, Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, 38, 5-13.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). NY: Micmillan Publishing Co.
- Messick, S. (1992). Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Gardner's *Frames of Mind* and Sternberg's *Beyond IQ* in the context of theory and research on the structure of human abilities. *Psychological Inquiry*, 3, 365-384.
- Mislevy, R. J. (in press). Cognitive psychology and educational assessment. In R. Brennan (Ed.), *Educational Measurement* (4th Ed.). Phoenix, AZ: Greenwood.
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, P. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent variable

- analysis. *Journal of Experimental Psychology: General*, 150, 621-640.
- Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test: Multilevel technical manual*. San Antonio, TX: Harcourt Brace.
- Nitko, A. J. (2001). *Educational assessment of students*. Upper Saddle River, NJ: Merrill.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1-18). New York: Plenum.
- Nyborg, H. (1983). Spatial ability in men and women: Review and new theory. *Advances in Behavior Research & Therapy*, 5, 89-140.
- Oberauer, K. (in press). The measurement of working memory capacity. To appear in O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 393-407). Thousand Oaks, CA: Sage.
- O'Connor, T. A., & Burns, N. R. (2003). Inspection time and general speed of processing. *Personality and Individual Differences*, 35, 713-724.
- Pellegrino, J. W., Chudowsky, N., Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Phillips, L. H., & Hamilton, C. (2001). The working memory model in adult aging research (pp. 101-125). In J. Andrade (Ed.), *Working memory in perspective*. New York: Psychology Press.
- Popper, K. R. (1963). *Conjectures and refutation*. London: Routledge and Kegan Paul.
- Revelle, W., Humphreys, M. S., Simon, L., & Gilliland, K. (1980). The interactive effect of personality, time of day, and caffeine: A test of the arousal model. *Journal of Experimental Psychology: General*, 109, 1-31.
- Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learning and Individual Differences*, 11, 1-120.
- Riding, R. J., & Boardman, D. J. (1983). The relationship between sex and learning style and graphicacy in 14-year old children. *Education Review*, 35, 69-79.
- Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of Educational Psychology*, 91, 511-526.
- Rothney, J. W. M. (1937). The new Binet—A caution. *School and Society*, 45, 855-856.
- Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American—White differences on cognitive tests. *American Psychologist*, 59, 7-13.
- Sacks, E. L. (1952). Intelligence scores as a function of experimentally established social relationships between child and examiner. *Journal of Abnormal and Social Psychology*, 47, 354-358.
- Schaie, K. W. (1970). A reinterpretation of age related changes in cognitive structure and functioning. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology* (pp. 423-466). New York: Academic Press.
- Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study*. New York: Cambridge University Press.
- Schaie, K. W., & Willis, S. L. (1978). Life span development: Implications for education. *Review of Research in Education*, 6, 120-156.
- Schneider, R. J., Ackerman, P. L., & Kanfer, R. (1996). To "act wisely in human relations": Exploring the dimensions of social competence. *Personality and Individual Differences*, 21, 469-481.
- Sears, R. (1943). Motivational factors in aptitude testing. *American Journal of Orthopsychiatry*, 13, 468-493.
- Seashore, C. E. (1919). *The psychology of musical talent*. Boston: Silver, Burdett, & Co.
- Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, 87, 38-56.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125(1), 4-27.
- Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93, 604-614.
- Snow, R. E. (1978). Research on aptitudes: A progress report. In L. S. Schulman (Ed.), *Review of Research in Education* (Vol. 4, pp. 50-105). Itasca, IL: Peacock.
- Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.) *Educational measurement* (3rd ed.) (pp. 263-331). New York: MacMillan.
- Snow, R. E., & Yalow, E. (1982). Intelligence and education. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 493-585). New York: Cambridge University Press.
- Snow, R. E., & Yalow, E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law*, 2, 536-560.
- Spearman, C., & Jones, L. L. W. (1950). *Human ability: A continuation of "The abilities of man."* London: Macmillan & Co., Ltd.
- Stanley, J. C., & Benbow, C. P. (1982). Huge sex ratios at upper end. *American Psychologist*, 27, 972.
- Stankov, L. (2000). Structural extensions of a hierarchical view on human abilities. *Learning and Individual Differences*, 12, 35-51.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24, 403-433.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.

- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge, UK: Cambridge University Press.
- Sternberg, R. J. (1993). *The Sternberg Triarchic Abilities Test*. Unpublished test.
- Sternberg, R. J. (2003). Issues in the theory and measurement of successful intelligence: A reply to Brody. *Intelligence*, 31, 319-329.
- Sternberg, R. J. (2004). *The Sternberg Triarchic Abilities Test*. Unpublished test.
- Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J., & Grigorenko, E. L. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test in three international samples. *European Journal of Psychological Assessment*, 17, 1-16.
- Sternberg, R. J., & Pretz, J. E. (Eds.). (2005). *Cognition and intelligence: Identifying the mechanisms of the mind*. New York: Cambridge University Press.
- Sternberg, R. J., & the Rainbow Project Collaborators (in press). Arguing the SAT through assessment of analytical, practical, and creative skill. In W. Camara & E. Kimmel (Eds.), *New tools for admission to higher education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Terman, L. M. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.
- Terman, L. M. (1924). The mental test as a psychological method. *Psychological Review*, 31, 93-117.
- Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence*. Boston: Houghton Mifflin.
- Thorndike, E. L., Bregman, E. O., Cobb, M. V., & Woodyard, E. (1927). *The measurement of intelligence*. New York: Teachers College Columbia University, Bureau of Publications.
- Thurstone, T. G. (1957). The tests of primary mental abilities. *Personnel and Guidance Journal*, 35, 569-578.
- Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.
- Vispoel, W. P. (1999). Creating computerized adaptive tests of music aptitude: Problems, solutions, and future directions. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 151-176). Mahwah, NJ: Lawrence Erlbaum Associates.
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250-270.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, 33, 468-479.
- Webb, E. (1915). *Character and intelligence: An attempt at an exact study of character*. Unpublished doctoral dissertation, Cambridge University, Cambridge, UK.
- Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist*, 5, 78-83.
- Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist*, 30, 135-139.
- Willingham, D. (2004). Reframing the mind. *Education Next*, 3, 18-24. Retrieved December 30, 2004, from <http://www.educationnext.org/2004/3/18.html>
- Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 77-108). Washington, DC: American Psychological Association.
- Wright, L. (1968). *Bibliography on human intelligence*. Washington, DC: U.S. Department of Health, Education, and Welfare.
- Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. (1915). *A point scale for measuring mental ability*. Baltimore: Warwick & York.

---

---

◆

HANDBOOK  
OF EDUCATIONAL  
PSYCHOLOGY

SECOND EDITION

EDITED BY

PATRICIA A. ALEXANDER

PHILIP H. WINNE



LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS  
2006 Mahwah, New Jersey London