Reasoning Abilities[1]

David F. Lohman

University of Iowa

Draft of January 9, 2002

To appear in R.J. Sternberg, J. Davidson, & J. Pretz (Eds.), *Cognition and intelligence*

The topic of human intelligence exceeds the span of any one discipline or method of inquiry. Different aspects of intelligence are best understood from disciplines as diverse as evolutionary biology, neuropsychology, cognitive psychology, anthropology, and education. At its core, however, intelligence is defined by differences between individuals or species. To say that one individual is more intelligent than another is to make a value judgment. Theories of human intelligence must therefore be able to explain those behaviors or accomplishments that societies value as indicants of intelligence (Sternberg, 1985). Such explanations may, at one extreme, invoke the action of neural mechanisms (Garlick, 2002) or, at the other extreme, the importance of social processes (Vygotsky, 1978). Ultimately, however, the theory must explain individual differences in those complex human behaviors that are most commonly understood as indicants of intelligence. Thus, the central facts to be explained by a theory of intelligence must go beyond faster or more efficient processing of elementary tasks, for example, or the efficiency of biological processes and inherited structures, or the influence of schools, environments, or even cultures. Rather, a theory of intelligence must explain the writing of novels, the solving of complex mathematical problems, the designing of skyscrapers and microchips, and the myriad other forms of complex cognition valued by society. In short, an understanding of how individuals solve complex tasks, and an explanation of why they differ so markedly in their ability to do so are central facts for any theory of intelligence.

Cognitive Tests as Cognitive Tasks

But which tasks should we study? There are many thousands of complex tasks, each of which might be considered an indicant of intelligence. Correlational studies of human abilities offer a reasonable starting place, since they (a) identify dimensions of individual differences that cut across tasks, (b) show which of these individual difference constructs best predict performance in non-test situations, such as success school (Brody, 1992) or work (Hunter & Schmidt, 1996), and

(c) identify tasks that repeatedly emerge as good measures of particular constructs. Estes (1974) was one of the first to suggest that careful examination of the processes test-takers use when solving items on ability tests might give an initial purchase on a process model of intelligence. Although such tasks commonly lack authenticity, efficient measures of ability constructs tend to make plain critical cognitive processes that are typically less transparent in more authentic, everyday tasks. To be sure, something is lost, but something is also gained.

In addition to identifying tasks that define ability constructs that predict valued non-test performances, correlational studies also show how ability factors are related to one another. This is useful because it helps investigators know how the ability construct they are studying relates to other ability constructs. There is now broad consensus that these relations can be represented hierarchically (Carroll, 1993; Gustafsson & Undheim, 1996). Even more suggestive for the present discussion, however, was the demonstration that hierarchical factor models are conformable with a radex model. The radex is produced by treating test intercorrelations as distances, which are then scaled in two or three dimensions using nonmetric, multidimensional scaling. The resultant scalings show three important features, (see Snow, Kyllonen, & Marshalek, 1984). First, tests cluster by content, which typically appear as verbal, spatial, and symbolic/quantitative slices of a two-dimensional radex pie. Second, tests and test clusters that define broad factors tend to fall near the center of the radex plot. More specific primaries fall near the periphery. Indeed, in a well-balanced battery of tests, tests that define G fall near the center of the plot. Third, task complexity is roughly related to distance from the center (or G). This suggests that one key to a theory of G, then, may be an understanding of the complexity gradients that emanate like spokes from G to more peripheral or specific abilities.

In this chapter, I briefly survey research on test-like tasks modeled after item-types commonly used on intelligence tests. I focus especially on measures of reasoning, particularly inductive reasoning, in part because reasoning tests have been studied extensively and in part because inductive reasoning is the primary ability most commonly associated with G. Gustafsson (1988) claims, for example, that general mental ability (G) can be equated with general fluid ability (Gf), which in turn can be equated with inductive reasoning (I). Sternberg (1986) makes a similar point:

> An interesting finding that emerges from the literature attempting to relate cognitive task performance to psychometrically measured intelligence is that the correlations of task performance and IQ seems to be a direct function of the amount of reasoning involved in a given task, independent of the paradigm or label given to the paradigm.... Thus, reasoning ability appears to be central to intelligence. (pp. 309-310)

Even though there is more to intelligence than reasoning, reasoning is a crucial aspect of any understanding of human intelligence.

Measures of reasoning and their uses

Although many different tasks have been used to measure reasoning, a few are used much more commonly than others: analogies, matrix problems, series completions, and classification tasks. Some test batteries also measure verbal reasoning through sentence completion tests, sentence comprehension tests, and even vocabulary. Others include more specific spatial tasks, such as form boards or paper-folding tests. And others use quantitative tests that require examinees to make relational judgments (such as greater than or less than) between quantitative concepts, or to determine how numbers and mathematical operators can be combined to generate a product.

Reasoning tests have important uses in many applied fields, particularly education. When administered to children, the main uses of such tests are: (a) to provide an estimate of students' general cognitive development that usefully supplements measures of achievement and teacher observations, (b) to provide an alternative frame of reference for interpreting academic achievement, and (c) to guide efforts to adapt instruction. Each of these uses is discussed in considerable detail elsewhere (Lohman & Hagen, 2001a,b; 2002).

Although tests of reasoning abilities have important uses, they are widely misunderstood – both by their critics and their supporters. An all-too-common misunderstanding is that a good ability test of any sort measures (or ought to measure) something like the innate potential or capacity of the examinee. A less common but equally extreme view is that reasoning is nothing more than knowledge, and knowledge is nothing more than experience. As in other domains, such personal theories are often difficult to change. An analogy to physical skills can be helpful. Cognitive skills have much in common with physical skills. Indeed, some models for the acquisition of cognitive skills are taken directly from earlier models of physical skills. My analogy begins with the commonplace distinction between cognitive abilities that are clearly tied to education and experience versus those that are less obviously tied to specific experiences. In the domain of reasoning, the former are sometimes called general crystallized abilities and the latter general fluid abilities. Crystallized abilities are like knowledge and skill in playing different sports. These skills are developed through years of practice and training. Athletes show different levels of competence across sports just as students show different levels of competence in various school subjects. But athletes also differ in their levels of physical fitness. Physical fitness is aptitude for acquiring skill in any sport. Athletes who have higher levels of physical fitness or conditioning will generally have an easier time learning new skills and will perform those that they do learn at a higher level. But physical fitness is also an outcome of participation in physically demanding activities. Further some sports—such as swimming —are more physically demanding than other sports and result in higher increments in physical conditioning for those who participate in them. In a similar manner, reasoning abilities are both an input to as well as an outcome of good schooling (Snow, 1996; Martinez, 2000). Indeed, expecting a measure of reasoning abilities to be independent of education, experience, and culture is like expecting a measure physical fitness to be uninfluenced by the sports and physical activities in which a person has participated.

The role of knowledge in reasoning

Reasoning well in domains of non-trivial complexity depends importantly on knowledge. Expertise is rooted in knowledge, and experts reason differently about problems than do novices. Because of this, some have erroneously assumed that good reasoning is nothing more than good knowledge. This does not take into account the importance of good reasoning in the acquisition of a well-ordered knowledge base. Nonetheless, an increasingly sophisticated knowledge base supports increasingly sophisticated forms of reasoning. For example, experts form problem representations that are more abstract than those of novices. Markman and Genter (2001) argue that the formation of moderately abstract conceptual relations may be a precurser to the detection of coherent patterns. Further, moderately abstract, principle-based concepts are easier to retain and manipulate in working memory, thereby freeing attentional resources for higher-level processes. There is thus an important synergy between good knowledge and good reasoning.

Studies of tasks modeled after item-types on intelligence tests often ignore these contributions of knowledge – particularly domain-specific knowledge – to reasoning. The loss is probably most obvious in the domain of verbal reasoning. The verbal reasoning skills of lawyers or scientists go well beyond the sort of decontextualized reasoning abilities assessed on most mental tests. A rich understanding of a domain and of the conventions of argumentation in that domain are needed in order to identify relevant rather than irrelevant information when understanding the problem, to decide which alternatives are most plausible and need to be considered, and then to decide how best to marshal evidence in support of a position. Strong warrants for an argument are considered highly plausible by those evaluating it. Plausibility judgments reflect both the beliefs of listeners and their assessment of the logical consistency of the argument. Standards for evaluating arguments are thus necessarily somewhat subjective. Nevertheless, some types of arguments are widely recognized as logically unsound. Toulmin, Rieke, and Janik (1984) classify these as (1) missing grounds (e.g. begging the question), (2) irrelevant grounds (e.g., red herring), (3) defective grounds (e.g., hasty generalization), (4) unwarranted assumptions, and (5) ambiguities.

Careful studies of reasoning in knowledge-rich contexts also show processes that generalize across domains. Newell and Simon's (1972) distinction between strong and weak methods of reasoning is helpful in especially helpful here. *Strong methods* of reasoning rely heavily on knowledge whereas *weak methods* depend less on content and context. Weak (or domain-general) methods describe what people do when they do not know what to do. Strong (or domain-specific) methods describe what they do when they know what to do. Therefore, children and novices are more likely to use domain-general methods. Further, as Markman and Gentner (2001) observe, many instances of domain-specific thinking result from domain-general processes operating on domain-specific representations. They also note that an exclusive focus on domain-specific thinking can result in a psychology of "particularistic descriptions" (p. 225) rather than of general processes and underlying dimensions. For example, domain-general structural alignment and mapping processes describe how people reason analogically in particular domains. Everyday reasoning depends heavily on the efficacy of past reasoning processes (stored as knowledge) as well as the efficacy of present reasoning processes. Indeed, the ability to adopt a decontextualized reasoning style is considered by some to be the sine-qua-non of good reasoning (Stanovich, 1999). Such thinking is often quite deliberate and open to introspection. Contextualized reasoning processes, however, often operate outside the realm of conscious awareness.

<u>Tacit and Explicit Processes</u>

Human reasoning occurs at several different levels. Most cognitive scientists distinguish between tacit and intentional reasoning processes (Evans & Over, 1996; Stanovich, 1999). *Tacit processes* that facilitate reasoning occur without conscious intervention and outside of awareness. They typically do not require attention. Such thinking is sometimes described as associative because it depends on the network of ideas and associations in memory. Tacit processes are typically used when we make a decision in a quick or intuitive way because it feels right, rather than because we have a clearly articulated set of reasons. We are aware of the outcome of these tacit processes but not of the processes themselves. Tacit processes are particularly important in focusing attention and in building a mental model of a problem. Effective problem solvers typically attend to different features of the problem than do less-effective problem solvers. Effective problem solvers know what to look for and what to ignore. In part, this is due to greater experience, in part, to better use of past experiences. Others describe this automatic attention as the extent to which the person is attuned to certain aspects of a situation and not to others. By temperament or training, some people are more attuned to the distress of others, to the beauty in a painting, to the mathematical properties of objects, or to the alliteration in a poem.

Tacit processes are also importantly linked to feelings which seem essential for solving ill-structured problems that have no single answer. This runs counter to the belief that emotion interferes with reasoning. Yet without ready access to the affective associates of memories, problem solvers seem to drown in a sea of equally plausible but equally bland alternatives (Damasio, 1994).

Intentional reasoning processes, on the other hand, occur within the sphere of conscious awareness. Individuals are aware not only of the outcome of their thinking, as with tacit processes, but also of the processes themselves. It is this type of reasoning that is most distinctly human. Such thinking is often described as strategic, or rule based. It typically requires effort. It allows one to bypass the relatively slow accumulation of experiences that underlie tacit learning. We can thereby transfer principles (e.g., One should always capitalize the first letter of the first word in a sentence) rather than an accumulation of varied experiences (e.g., I have seen many sentences, and it feels like it is probably okay to capitalize the first word). Put differently, tacit processes are generally fast, but limited to the range of contexts repeatedly experienced. Intentional reasoning processes, on the other hand, are slow and effortful, but extremely flexible.

Thus, reasoning involves both conscious (explicit) and unconscious (tacit) processes. Although some psychologists refer to both explicit and tacit reasoning processes, others argue that situations elicit reasoning only to the extent that they require conscious application of particular mental processes (Elshout, 1985; Sternberg, 1986). In this chapter, I speak of unconscious processes that facilitate reasoning but reserve the term *reasoning* for certain types of conscious, attention-demanding, non-automatic thinking.

<u>Reasoning and working memory</u>

One of the more important controversies about reasoning abilities is the extent to which individual differences in reasoning abilities overlap with individual differences in working memory capacity. Kyllonen and Christal (1990) sparked the controversy with their finding that latent variables for working memory and reasoning factors correlated **r** = .80 to .88 in four large studies with U.S. Air Force recruits. Other researchers also found large path coefficients between measures of working memory and measures of fluid reasoning abilities (Süβ, Oberauer, Wittman, Wilhelm, & Schulze, 2002; Conway, Cowan, Bunting, Therrialt, & Minkoff, 2002). However, critics complained that some tasks used to estimate working memory in these studies were indistinguishable from tasks used to estimate reasoning. For example, the *ABC Numerical Assignment* test requires examinees to solve for C in problems such as the following: $A = C + 3$, $C = B/3$, and $B = 9$. The task is thought to measure working memory because only one equation is visible at a time and the computations are relatively simple. But it is certainly possible that at least some individuals must use reasoning abilities to solve such tasks. Other critics (e.g., Fry & Hale, 1996) have argued that processing speed accounts for most of the relationship between the reasoning and working memory constructs in these studies. Ackerman, Beier, and Boyle (2002) note that processing speed is itself a multidimensional construct. They conclude that, although there is little doubt that measures of working memory are significantly associated with measures of general intelligence, the two are not synonymous. Indeed, raw correlations between measures of the different construct are typically in the range of r =.2 to .4.

In part, this is a problem of words. *Working memory* connotes too small a construct; *reasoning* connotes too large a construct – especially given the way each is typically measured. Consider first the reasoning construct. In the best of these studies, reasoning is estimated by performance on a series of short, puzzle-like tasks. More commonly, it is estimated by a single test such as the Raven Progressive Matrices (Raven, Court, & Raven, 1977) that uses a single item format. As Ackerman et al. (2002) note, "if the Raven is not an exemplary measure of general intelligence (or even Gf), any corroborations between experimental measures (such as [working memory]) and Raven … are apt to miss important variance … and result in distortion of construct validity." p. 586. Indeed, figural reasoning tests such as the Raven are typically much poorer predictors of both real-world learning and academic achievement than measures of verbal and quantitative reasoning. Whether measured by one task or several short tasks, the reasoning construct is under-represented.

On the other hand, the construct measured by the series of "working memory" tests is much more complex than its label suggests. These tasks generally require participants to understand and follow a sometimes complex set of directions, assemble and then revise a strategy for performing a difficult, attention-demanding task, maintain a high level of effort across a substantial number of trials, and then repeat the process for a new task with a new set of directions. By design, many working memory tasks require individuals to process simultaneously one set of ideas while remembering another set. These processes, while generally thought to be easy, are certainly not trivial, especially when performed under memory load. Verbal working memory tasks commonly require reading comprehension, mathematical tasks require computation, and spatial tasks require transformations such as mental rotation. Tasks are also designed to elicit executive functions such as the monitoring of processes, controlling their rate and sequence of operation, inhibiting inappropriate response processes, coordinating information from different domains, and integrating ideas into a coherent mental model.

Therefore, another way to express the conclusion that individual differences in working memory and reasoning overlap would be:

> A substantial portion of the individual difference variation in the limited and somewhat artificial set of reasoning tasks included in our study can be accounted for by individual differences in the ability to assemble a strategy for the simultaneous storage and transformation of ideas, to monitor the success of this strategy and change it as needed, to coordinate information from different sources, to inhibit some mental operations and to activate others, to sequence these mental operations, and to integrate ideas in to a coherent mental structure or model.

<u>Reasoning tests as cognitive tasks</u>

There is now an extensive literature that examines how individuals solve items on reasoning tests, particularly analogy, seriation, and classification tasks. For reviews, see Sternberg (1985); Snow and Lohman (1989); and Lohman (2000). However, constructs such as reasoning ability are defined not by particular tasks, but by the common co-variation in several tasks. Put differently, understanding individual differences in solving matrix problems or letter-series problems or analogy problems is not the same as understanding individual differences in reasoning ability. Every well-constructed test measures something that it shares with other tests designed to measure the same construct and something unique to the particular test. There are no exceptions. The main sources of uniqueness are the idiosyncrasies of the particular sample of items contained in the test and the format in which they are administered. This is shown clearly in factor analyses of wide-ranging test batteries. The loading of a test on the factor it helps define is often only slightly greater than its loading on a test-specific factor. A well-grounded theory of reasoning ability, then, must look beyond the sources of individual differences in particular reasoning tasks to those that are shared by several reasoning tasks.

There are two aspects of constructs to be considered, which are nicely captured in Embretson's (1983) distinction between construct representation and nomothetic span. Construct representation refers to the identification of psychological constructs (e.g., component processes, strategies, structures) that are involved in responding to items on tests. Processes of most interest are those that are common across families of tests that collectively define individual difference constructs such as inductive reasoning ability.

Nomothetic span, on the other hand, concerns the correlates of individual differences on a test. Of the many processes that are involved in performance on a particular task, only some will be shared with other tasks, and of these common processes, an even smaller subset will be responsible for individual differences that are common across tasks. In other words, even processes and structures that are common to all tests in a family of reasoning tasks may contribute little or not at all to individual differences in reasoning ability.

Construct Representation of Reasoning Tests

With these caveats in mind, then, I briefly summarize investigations of the processes test-takers use when solving items on reasoning tests (for a more detailed summary, see Lohman, 2000). In the subsequent section, I summarize hypotheses about which processes are most responsible for generating observed individual differences in reasoning abilities..

Pellegrino (1985; see also Goldman & Pellegrino, 1984) argues that inductive reasoning tasks such as analogies, series completions, and classifications all require four types of processes: encoding or attribute discovery, inference or attribute comparison, relation or rule evaluation, and decision and response processes.

Encoding processes create mental representations of stimuli on which various inference or attribute-comparison processes operate. The nature of these processes differ across tasks. In an analogy [A is to B as C is to D], the inference process must determine how various terms are related to each other. In classification problems [Given the set *apple, pear, banana*, which word belongs: *orange or pea*?], the inference process must identify a rule or category that is shared by all the terms. In series problems [Given *3,4,6,9,13*. What comes next?], the inference process must identify the pattern in a sequence of letters or numbers. Inference processes are usually not sufficient for problem solution, however. One must also determine relationships among two or more first-order relationships in the problem. In an analogy, for example, the relationship between A and B must be identical to the relationship between C and D. In a matrix problem, the relationship among elements in one row must be the same in the other two rows. Pellegrino (1985) argues that one of the most important aspects of inductive reasoning is the ability to create complex relationship structures in memory and to determine their consistency. Errors occur when working-memory resources are exceeded.

Sternberg (1986) claims that there are three kinds of reasoning processes, any one of which define a task as a reasoning task. The three processes are (a) selective encoding (distinguishing relevant from irrelevant information), (b) selective comparison (deciding what mentally stored information is relevant for solving a problem), and (c) selective combination (combining selectively encoded or compared information in working memory). Furthermore, the three processes define a reasoning situation only to the extent that they are executed in a controlled rather than in an automatic fashion. This implies that the extent to which a task measures reasoning depends on the relative novelty of the task for the individual.

These processes are implemented by various sorts of inferential rules. Procedural rules include operations called performance components in earlier theories (Sternberg, 1977; Sternberg & Gardner, 1983). Declarative rules vary by problem content and specify the type of semantic relations allowed in a problem. (For verbal analogy problems, for example, the set of possible semantic relations includes equality, set-subset, set-superset, static properties, and functional properties). Not all rules are rules of reasoning; reasoning rules are those that serve the functions of selective encoding, selective comparison, and selective combination. Thus, mnemonic strategies and computation algorithms are not reasoning rules.

The theory also claims that the probability that particular influential rules will be used in the solution of a reasoning problem and will be influenced by mediating variables, such as the individual's subjective estimate of the likelihood of the occurrence of a rule, the individuals' prior knowledge, working memory capacity, and ability to represent certain types of information (e.g., spatial versus linguistic).

Sternberg claims that the major difference between inductive and deductive reasoning is that the difficulty of the former derives mainly from the selective encoding and comparison processes, whereas the difficulty of the latter derives mainly from the selective combination process. Thus, for verbal analogies, the primary difficulty is determining which of the many features of the A term are relevant to the B term as well. For example, in the analogy paper:tree::plastic:?, one must decide which of the many attributes of the word paper (that we write on it, that it sometimes comes in tablets, that printers use it, that it is a short form of the word "newspaper," that it is made from wood, etc.) also overlap with what one knows about the word tree. In contrast, figural analogies tend to emphasize selective encoding. A key difficulty of such problems is deciding which features of the stimuli to attend to in the first place.

Series completion problems not only require many of the same processes as analogies (Greeno, 1978; Pellegrino & Glaser, 1980; Sternberg & Gardner, 1983), but also emphasize selective comparison. In a typical series problem, there are many possible relations that could be obtained between successive pairs of numbers or letters. For example, in the series 1, 3, 6, 10,..., the relation between the first two digits could be plus 2, times 3, next odd number, etc. The relation between 3 and 6 could be plus 3, times 2, etc. Problem difficulty is highly related to the obscurity of the rule. However, when multiple rules account for a series, the "best" rule is typically the most specific rule. A similar set of arguments apply to the analysis of classification problems.

For deductive reasoning tasks such as categorical syllogisms, however, the main source of difficulty lies not in encoding the terms or even in selectively comparing relations among them, but rather in keeping track of the ways in which terms can be combined. Consider, for example, a categorical syllogism such as "Some A are B. All B are C." Is the conclusion "Some A are C" valid? Information processing models of syllogistic reasoning all share four stages of information processing, which Sternberg (1986) calls encoding, combination, comparison, and response. In the encoding stage, the individual must create a mental representation of each premise that is amenable to mental transformation. The large number of combinations between representations of premises taxes processing resources. For example, the problem "Some B are C. Some A are B." involves 16 combinations (four for each of the two premises). Further, the exact inferential rule used also appears to be a major source of difficulty, although there is controversy as to exactly what these rules are. More important, however, has been the recurring finding that many other factors (what Sternberg calls mediators) influence performance as categorical syllogisms. For example, subjects show flagrant biases in solving such problems as a function of the emotionality of the premise, subjects' agreement with the content of the premises, abstractness of the content, and even the form in which the problems are presented. Some strategies simply facilitate performance; others completely bypass the reasoning process (e.g., Yang & Johnson-Laird, 2001). This suggests that, although such problems may be interesting candidates for research, they are probably not good candidates for assessments of individual differences in reasoning abilities.

Another type of deductive reasoning task that has been extensively studied is the linear syllogism. These are problems of the sort "Bill is taller than Mary. Mary is taller than Sue. Who's tallest?" Problems of this sort have anywhere from two to four terms, with the most typical number being three. As in other deductive reasoning problems, the major source of difficulty is not in encoding the terms or in comparing them (for example, to know that "short" is the opposite of "tall"), but rather to combine the information in the premises into a single mental model. Unlike linear syllogisms, however, there are fewer content-induced biases to cloud performance. Indeed, the most likely bias occurs when the premise contradicts one's personal knowledge, such as when one knows that Mary is shorter than Sue, whereas the problem asks one to envision the opposite. Such contra-factual reasoning can be deliberately introduced into problems (e.g., "imagine that mice are larger than elephants," etc.). For an introduction to recent investigations of this type of deductive reasoning, see Johnson-Laird (1999).

Johnson-Laird (1999) argues that mental models are useful for predicting performance on these types of tasks. Although models often give rise to images, they are distinct from images because models can contain abstract elements, such as negation, that cannot be visualized. Yang and Johnson-Laird (2001) showed how the thory of mental models could explain some sources of difficulty on the logical reasoning problems from the Analytic subtest of the Graduate Record Examination. They identified three sources of difficulty: the nature of the task (it is easier to identify which conclusion a text implies rather than a missing premise), the nature of the foils (it is easier to reject foils that are inconsistent with the text than foils that are consistent with it), and the nature of the conclusions (it is easier to accept a conclusion that is consistent with the text than one that is inconsistent with it). The second and third sources of difficulty stem from the principle of truth: Individuals minimize the load on working memory by tending to construct mental models that represent explicitly only what is true, and not what is false (Johnson-Laird, 1999). Given the truth of the premises, the probability of a conclusion depends on the proportion of models in which it holds. It is considered possible if it holds in at least one model of the premises and necessary of it hold in all models.

Nomothetic Span of Reasoning Tests

Understanding the common processing demands of tasks is one way to understand the construct they help define. The emphasis is on explaining what makes tasks difficult. Another route is to examine those features of tasks that seem to moderate their relationships with the target construct -- here Gf or G. One of the primary uses of visual models of test correlations (such as a two-dimensional radex) is to make these general themes more apparent. Tests that load heavily on the G or Gf typically fall near the center of the radex, whereas seemingly simpler tasks are distributed around the periphery.

Several hypotheses have been advanced to explain how processing complexity increases along the various spokes that run from the periphery to G: (1) an increase in the number of component processes; (2) an accumulation of differences in speed of component processing; (3) increasing involvement of one or more critically important performance components, such as the inference process; (4) an increase in demands on limited working memory or attention; and (5) an increase in demands on adaptive functions, including assembly, control, and monitor functions. Clearly these explanations are not independent. For example, it is impossible to get an accumulation of speed differences over components (Hypothesis 1) without also increasing the number of component processes required (Hypothesis 1). In spite of this overlap, these hypotheses provide a useful way to organize the discussion.

More Component Processes. Even the most superficial examination of tasks that fall along one of the spokes of the radex reveals that more central or G-loaded tasks require subjects to do more than the more peripheral tests. Many years ago, Zimmerman (1954) demonstrated that a form-board test could be made to load more on perceptual speed, spatial relations, visualization, and reasoning factors, in that order, by increasing the complexity of the items. Snow et al.'s (1984) reanalyses of old learning-task and ability-test correlation matrices showed similar continua. Spilsbury (1992) argues that the crucial manipulation here is an increase in the factorial complexity of a task. However, increases in the number or difficulty of task steps beyond a certain point can decrease the correlation with G (Crawford, 1988; Raaheim, 1988; Swiney, 1985). Thus, one does not automatically increase the relationship with G simply by making problems harder, or even by increasing the factorial complexity of a task. Indeed, there are many hard problems (e.g., memorizing lists of randomly chosen numbers or words) that are not particularly good measures of G. Furthermore, even for problems that do require the type of processing that causes the test to measure G, problems must be of the appropriate level of difficulty for subjects.

Speed or Efficiency of Elementary Processing. This hypothesis has taken several forms. In its strongest form, the assertion has been that individuals differ in the general speed or efficiency with which they process information (Jensen, 1998). In principle, processing speed could be estimated on any elementary cognitive task that minimizes the import of learning, motivation, strategy, and other confounding variables. Although disattenuated correlations between RT and G can be substantial when samples vary widely in ability (even, for example, including mentally retarded participants), samples more typical of those used in other research on abilities yield correlations between RT and G in the $r = -.1$ to $r = -.4$ range (Jensen, 1982; Roberts & Stankov, 1999; Sternberg, 1985; Deary & Stough, 1996). Furthermore, response latencies on many tasks show a pattern of increasing correlation with an external estimate of G as task complexity is decreased. In other words, response latencies for simpler tasks typically show higher correlations with G than do response latencies for more complex tasks. But this is unsurprising. The more complex the task, the more room there is for subjects to use different strategies or even to be inconsistent in the execution of different components.

In its weak form, the hypothesis has been that although speed of processing on any one task may be only weakly correlated with more complex performances, such small differences cumulate over time and tasks. Thus, Hunt, Frost and Lunneborg (1973) noted that although latency differences in the retrieval overlearned name codes correlated only $r = .3$ with verbal ability, such small differences on individual words cumulate to substantial differences in the course of a more extended activity. Detterman (1986) emphasized the cumulation across different component processes rather than across time. He showed that although individual component processes were only weakly correlated with G, their combined effect was more substantial.

Although individual differences in speed of processing are an important aspect of G, G is more than rapid or efficient information processing. Furthermore, the strength of the relationship between speed of processing and G varies considerably across domains, being strongest ($r \approx -.4$) in verbal domain and weakest ($r \approx -.2$) in the spatial domain. Indeed, for complex spatial tasks, the speed with which individuals perform different spatial operations is usually much less predictive of overall performance than the richness or quality of the mental representations they create (Lohman, 1988; Salthouse, Babcock, Mitchell, Palmon, & Skovronek, 1990).

More Involvement of Central Components. If G is not simply a reflection of more or faster processing, might it be the case that G really reflects the action of particular mental processes? Spearman (1927) was one of the first to argue for this alternative. For him, the essential processes were the "eduction of relations," which Sternberg calls *inference*, and the "eduction of correlates," which Sternberg calls *mapping* and *application*. Evidence favoring this hypothesis is substantial. A common characteristic of tests that are good measures of Gf--such as the matrices, letter/number series, analogies, classification, and various quantitative reasoning tests--is that they are all measures of reasoning, particularly inductive reasoning. Many school learning tasks, particularly in science and mathematics, bear formal similarity to Gf tests. Greeno (1978) refers to such tasks, collectively, as problems of inducing structure. Indeed, the problem of inducing structure in instruction is probably why reasoning tests correlate with achievement tests (Snow, 1980). But to describe the overlap in this way is not to explain it.

Evidence supporting the hypothesis that particular component processes are central to G has been surprisingly difficult to obtain. Sternberg's (1977) investigations of analogical reasoning found little generalizability across tasks of scores for the inference component, and at best inconsistent correlations of these scores with reference reasoning tests. Rather, it was the intercept (or "wastebasket" parameter) that showed more consistent correlations with reference abilities. We now know that this was in large measure an inevitable consequence of the way component scores are estimated (Lohman, 1994). Individual differences that are consistent across items that require different amounts of a particular component processes will appear in the intercept rather than in the component scores. Therefore, low or inconsistent correlations between scores for particular component processes and other variables do not provide much evidence against the hypothesis that these processes are important.

A second line of evidence on the centrality of particular component processes comes from demonstrations that certain types of task manipulations are more likely than others to increase the Gf loading of a task (Pellegrino, 1985; Sternberg, 1986). Sternberg (1986) calls these selective encoding, i.e., the requirement to attend selectively to information and to encode only that subset that is likely to be needed for solving a problem; selective comparison, i.e., to retrieve only information that is relevant to a problem, especially when the set of potentially relevant information in memory is vast; and selective combination, i.e., to assemble in working memory information already selected as relevant. Selective encoding depends heavily on the individual's store of prior knowledge (schema) and its attunement to the affordances of the situation. It also means the ability to resist the distractions of salient but irrelevant information, or, when solving items on mental tests, looking ahead to the alternatives before studying the stem (Bethell-Fox et al., 1984). Selective comparison also depends heavily on the store of knowledge, but also on its organization and accessibility, especially the ability to search rapidly through memory for overlap between two concepts. This is the essential feature of inference or abstraction problems: finding ways in which concepts A and B are not merely associated with each other, but rather finding the rules or relations that most specifically characterize their association. Problems in inductive reasoning emphasize selective encoding and comparison. Problems in deductive reasoning, on the other hand, emphasize selective combination. For example, syllogistic reasoning problems are difficult not because it is difficult to discern the relevant information in statements such as “all A are B” or in the understanding of the relations between words such as “all” and “some” (although this is a source of confusion for some), rather, the main difficulty in keeping track of all of the ways in which the premises can be combined. This taxes both working memory and the ability to manipulate symbols. Thus, although certain processes may be central to intelligent thinking, individual differences in those processes may be in part due to other system limitations--such as working-memory resources.

Attention and Working Memory Capacity. All information-processing models of memory and cognition posit the existence of a limited capacity short term or working memory that functions not only as a central processor, but as a bottleneck in the system. Some see this in terms of structure or capacity limitations; others view it in terms of attentional resources, and others in terms of differences in knowledge or experience (see Miyake & Shah, 1999). Hunt and Lansman (1982) and Ackerman (1988) argue that tasks that show higher correlations with G require more attentional resources. Attempts to manipulate the attentional demands of tasks often use a dual-task paradigm. Here, participants are required to do two things simultaneously, such as searching for a particular stimulus in a visual display while simultaneously listening for a specified auditory stimulus. Although the effect is often not observed, differences between more and less able subjects are typically greater in the dual task than in the single task condition. However, interpretation of this finding is problematic. For example, in one study, Stankov (1988) found that correlations with both Gc and Gf, but especially Gf, were higher for dual tasks than for single tasks. However, high levels of performance in the dual task situation were due to a strategy of momentarily ignoring one task while attending to the other. Thus, what on the surface seemed to implicate attentional resources on closer inspection implicated self-monitoring and the shifting of attentional resources.

Attentional requirements of tasks vary according to an individual's familiarity with the task and to the susceptibility of the task to automatization. Tasks--or task components--in which there is a consistent mapping between stimulus and response can be automatized in this way. Individuals who recognize the consistencies thus automatize task components more rapidly than those who are not so attuned. Put differently, knowledge guides attention, and thus constrains the number of features that must be considered in understanding the problem.

The explanation of differences in reasoning as reflecting differences in working memory capacity parallels the attentional explanation. Many researchers have claimed that a major source of individual differences on reasoning tasks lies in how much information one must maintain in working memory, especially while effecting some transformation of that information (Holzman, Pellegrino, & Glaser, 1982). Some argue that the critical factor is the ability to maintain a representation in the face of interference from automatically activated but distracting representations (Engle, Tuholski, Laughlin, & Conway, 1999). Controlling attention in this way is a critical aspect both of selective encoding and goal management (Primi, 2001). Furthermore, as Kyllonen and Christal (1990) noted, most of the performance processes (such as encoding and inference) and executive processes (such as goal setting, goal management, and monitoring) are presumed to occur in working memory. Thus, even though, say, the inference process may be effective, it must be performed within the limits of the working memory system. Therefore, although many different processes may be executed in the solution of a task, individual differences in them may primarily reflect individual differences in working memory resources.

Adaptive Processing. While acknowledging that individual differences in G reflected differences in all of these levels--in the speed and efficacy of elementary processes, in attentional or working memory resources, in the action of processes responsible for inference and abstraction (which includes knowledge, skill, and attunement to affordances in the task situation)--several theorists have argued that more is needed. Sternberg (1985) argued that intelligent action requires the application of metacomponents--i.e., control processes that decide what the problem is, select lower-order components and organize them into a strategy, select a mode for representing or organizing information, allocate attentional resources, monitor the solution process, and attend to external feedback. Marshalek et al. (1983), on the other hand, focused on assembly and control processes. They hypothesized that "more complex tasks may require more involvement of executive assembly and control processes that structure and analyze the problem, assemble a strategy of attack on it, monitor the performance process, and adapt these strategies as performance proceeds, within as well as between items in a task, and between tasks" (Marshalek et al., 1983, p. 124). The Carpenter, Just, and Shell (1990) analysis of the Raven test supports this hypothesis. In their simulation, the crucial executive functions were (a) the ability to decompose a complex problem into simpler problems and (b) the ability to manage the hierarchy of goals and subgoals generated by this decomposition.

In general, assembly processes are reflected in activities in which an individual must organize a series of overt acts or covert cognitive processes into a sequence. They are thus essential for all high-level thinking. These processes are greatly facilitated by the ability to envision future states (i.e., goals) that differ from present states (i.e., what is currently in mind or in view). This is an especially important activity when attempting novel or ill-structured tasks. Control processes are more diverse, although all involve the ability to monitor the effects of one's cognitions and actions, and adjust them according to feedback from the environment or one's body. Both types of processing depend heavily on the ability to maintain ideas or images in an active state in working memory, especially when several ideas must be considered simultaneously or when goal images differ from images activated by perceptions.

More able problem solvers are not always more strategic or flexible or reflective in their problem solving (cf. Alderton & Larson, 1994). Indeed, subjects who are most able often show little evidence of shifting strategies across items on a test. For example, in the Kyllonen, Lohman, and Woltz (1984) study of a spatial synthesis task, subjects very high in spatial ability (but low in verbal ability) were best described by a model that said that they always mentally synthesized stimuli. These subjects probably did not have to resort to other strategies. Rather, it was the subjects who had less extreme profiles but relatively high scores on G that showed the most strategy shifting.

Several investigators have attempted to manipulate the extent to which items require flexible adaptation and thereby alter their relationship with G. For example, Swiney (1985) sought to test the hypothesis that correlations between performance on geometric analogies and G would increase as more flexible adaptation was required, at least for easy and moderately difficult problems. Correlations with G were expected to decline if task difficulty was too great. Adaptation was manipulated by grouping items in different ways. In the blocked condition, inter-item variation was minimized by grouping items with similar processing requirements (estimated by the number of elements, and the number and type of transformations). In the mixed condition, items were grouped to be as dissimilar as possible.

Results showed that low-ability subjects were more adversely affected by mixing items than high ability subjects, regardless of treatment order. Relationships between task accuracy and G varied systematically as a function of item difficulty and task requirements. Strongest relationships were observed for identifying (i.e., inferring) and applying difficult rules. Weakest relationships were observed for applying easy rules or discovering difficult rules, especially in the mixed condition. Retrospective reports supported the conclusion that high-G subjects were better able to adapt their strategies flexibly to meet changing task demands. Swiney also found that low-G subjects overestimated their performance on highly difficult items; they also consistently underestimated the difficulty of problems. This suggests differences in monitoring and evaluation processes.

Chastain (1992) reported three additional studies contrasting blocked versus mixed item presentations. Experiments 1 and 2 used items from the Wonderlic Personnel Test, a 50-item test that samples a broad range of item formats. The third experiment used a figural encoding task and a dynamic spatial task. In all studies, flexible adaptation was estimated by a simple difference score (mixed minus blocked) and by a residual score (regression of mixed on blocked). Correlations between these two scores, reference tests, and performance on a logic-gates learning task were small, but generally in the expected direction.

A study by Carlstedt, Gustafsson , and Ullstadius (2000) challenges this interpretation of the blocked-mixed contrast. Carlstedt et al. administered three kinds of inductive reasoning problems to groups of Swedish military recruits. Unexpectedly, they found that G loadings were higher in the blocked condition than in the mixed condition. They argue that the homogeneous arrangement affords better possibilities for learning and transfer across items. However, items were extremely difficult, and so generalization is difficult.

To summarize: as one moves from periphery to center in a two (or even three) dimensional radex, tasks increase in apparent complexity. Tasks near the center typically require more steps or component processes, and emphasize accuracy rather than speed of response. But this does not mean that speed of processing is unimportant or that the addition of any type of process will increase the correlation with G. Increasing the demand on certain types processing, which Sternberg describes as selective encoding, comparison, and combination, also increases the correlation with G. Importantly, though, such processes require controlled, effortful processing and place heavy demands on working memory resources. They also require subjects to be more strategic or flexible or adaptive in their problem solving, or to learn from easy item rules that will be needed in combination to solve hard items

Limitations of the information-processing paradigm

The information-processing paradigm has enormously enriched our understanding of cognitive tests and the ability constructs they estimate. We have moved from trait labels and vague notions of "process" to detailed models of thinking. However, all paradigms are inadequate in some respects. Two shortcomings of the information-processing approach are particularly salient: (a) the neglect of affect and conation, and (b) the failure to understand the contextual specificity of abilities.

Affect and conation. Although theorizing about the influence of affect (or feeling) and conation (or willing) on cognition dates back to the Greek philosophers, it is only recently that investigators have attempted to study the complex and reciprocal influences on each other. Many promising leads have been identified (see Snow, Corno, & Jackson, 1996; Boekaerts, 1995). It is clear that persons who do well on ability tests expend effort differently from persons who score poorly. The difference is most striking in comparisons of experts and novices in skill domains such as reading. Experts expend their efforts on high-level processes (that include, but go beyond comprehension), whereas novices struggle to identify words and the sentences they comprise. Affect enters not only as anxiety or frustration, which further constricts cognition, but also as interest and surprise, which enhance and direct cognition. In particular, those who adopt a constructive motivational orientation towards a task will tent to exhibit more and better self-regulation than individuals who adopt a less constructive or even defensive orientation. Situations differentially elicit these conative and affective resources. Indeed, understanding the role of affect in cognition seems to demand a mode of theorizing and experimentation that attends not only to persons or to situations, but also to the attunement of particular individuals to particular aspects of situations.

Including situations and their affordances. A theory of G must explain individual differences in problem solving not only on tests, but in school and other everyday contexts. Although occasionally nodding to the role of culture, cognitive theories of abilities have not yet found ways to incorporate the fact that cognition is, to a greater or lesser degree, situated. Theories that would explain how abilities facilitate goal attainment need to start with the proposition that such action is always situated. Situations evoke or afford the use of some concepts or ways of thinking, but only for those tuned to perceive them. Some tunings reflect biological adaptations; but most are mediated by experience. In the language of Corno et al. (2002), abilities that are actually elicited in a particular situation function as aptitudes. Aptitudes are any characteristics (including affect and motivation, for example) that aid goal attainment in a particular situation. For example, inductive reasoning abilities may be elicited when situations require the identification of pattern or rule and the person has no ready-made solution. The perception that evokes structure-mapping processes is trivial when someone asks “What do these situations have in common?” More often, it occurs because the individual is actively engaged in making sense of the world. Making sense means finding commonalities. Thus, as Snow (1994) put it, aptitudes are reflected in the tuning of particular persons to the particular demands and opportunities of a situation, and thus reside in the union of person in situation, not "in the mind" alone.

Toward a definition of reasoning

In his summary of correlational studies of reasoning abilities, Carroll (1993) suggests that the general reasoning factor can be decomposed into three subfactors: sequential reasoning, inductive reasoning, and quantitative reasoning. Sequential reasoning is most commonly measured by tasks that require deductive or logical reasoning. Tasks are often (but not always) verbal. Inductive reasoning is commonly measured by tasks that require identification of a pattern or rule in a stimulus set. Tasks are often (but not always) figural. Quantitative reasoning is measured by tasks that require either inductive or deductive reasoning on quantitative concepts. Setting aside task content, then, the critical reasoning processes are sequential (or deductive) and inferential.

When people reason, they must, in Bruner’s (1957) helpful phrase, go “beyond the information given.” They do this in one or both of the following ways:

(a) They attempt to infer (either automatically or deliberately) concepts, patterns, or rules that best (i.e., most uniquely) characterize the relationships or patterns they perceive among all the elements (words, symbols, figures, sounds, movements, etc.) in a stimulus set. Better reasoning is characterized by the use of concepts or rules that simultaneously satisfy the opposing needs for abstraction (or generalization) and specificity. Such concepts or rules tend to be at least moderately abstract yet precisely tuned. Put differently, a poor inference is often vague and captures only a subset of the relationships among the elements in the set.

The judgment of what constitutes better reasoning is in part dictated by the shared knowledge and conventions of particular communities of discourse and in part by the precision and generality of the inference.

(b) They attempt to deduce the consequences or implications of a rule, set of premises, or statements using warrants that are rendered plausible by logic or by information that is either given in the problem or assumed to be true within the community of discourse. They often seem to do this by creating and manipulating mental models of the situation. Such models tend to represent explicitly only what is assumed to be true about the situation. Better reasoning involves providing warrants that are more plausible or consistent with the rules of logic or the conditions embodied in a comprehensive mental model. More advanced deductive reasoning involves providing either multiple (possibly divergent) warrants for a single claim or an increasingly sophisticated chain of logically connected and separately warranted assertions.

Clearly, then, reasoning abilities are not static. They are developed through experience and rendered easier to perform through exercise. Recall that individual differences in reasoning are substantially correlated with the amount of information individuals can hold in working memory while performing some transformation on it. Their ability to do this depends in large measure on the attentional resources individuals bring to a task, their familiarity with the to-be-remembered information, and their skill in performing the required transformations. Thus, prior knowledge and skill are critical determiners of the level of reasoning that one can exhibit. The dependence on prior knowledge is most pronounced on tasks that require deductive reasoning with authentic stimulus materials, and is least pronounced on tasks that require inferential reasoning with simple geometric or alphanumeric stimuli. The processes that support sophisticated reasoning in a knowledge-rich domain, however, appear to be largely the same as those which enable the neophyte to infer consistencies or deduce likely consequents.

One of the most important uses of tests of reasoning abilities is as an indicator of readiness to discover what to do in situations in which the person cannot rely on stored routines to solve problems. Reasoning tests have long been used in this way to inform decisions about college admission for students who come from impoverished backgrounds. Indeed, good reasoning tests shows smaller differences between majority and minority students than good achievement tests (Lohman, in press). Measures of general reasoning abilities also routinely interact with instructional methods. In particular, they predict academic success better when instructional methods require that students discover concepts and relationships for themselves than when instruction is more didactic (Snow & Lohman, 1989). Because of this, one can improve the likelihood that students with poorly developed reasoning abilities will succeed -- by reducing either the need for prior knowledge or the working memory demands of ancillary processes. In other words, understanding why individuals differ in their reasoning abilities allows one to alter the prediction of academic success. For this reason alone, educators should pay more attention to students' current levels of reasoning abilities. Indeed, one would be hard pressed to think of any ability construct that is better understood or has more practical relevance to education at all levels than reasoning abilities.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: A theory of cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 299-329.

Ackerman, P. L., Beier, M., & 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.

Alderton, D. L., & Larson, G. E. (1994). Cross-task consistency in strategy use and the relationship with intelligence. *Intelligence, 18*, 47-76.

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence, 8*, 205-238.

Boekaerts, M. (1995). The interface between intelligence and personality as determinants of classroom learning. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 161-183). New York: Plenum Press.

Brody, N. (1992). *Intelligence.* San Diego: Academic Press.

Bruner, J. (1957). Going beyond the information given. In J. S. Bruner (Ed.), *Contemporary approaches to cognition* (pp. _____). Cambridge, MA: Harvard University Press.

Carlstedt, B., Gustafsson, J.- E., & Ullstadius, E. (2000). Item sequencing effects on the measurement of fluid intelligence. *Intelligence, 28,* 145-160.

Carpenter, P. A., Just, M. A., & Schell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review, 97*, 404-431.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Chastain, R. L. (1992). Adaptive processing in complex learning and cognitive performance. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Corno, L., Cronbach, L. J., Lohman, D. F., Kupermintz, H., & Talbert, J. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow.* Hillsdale, NJ: Erlbaum.

Crawford, J. (1988). Intelligence, task complexity and tests of sustained attention. Unpublished doctoral dissertation, University of New South Wales, Sydney, Australia.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist, 51*, 599-608.

Detterman, D. K. (1986). Human intelligence is a complex system of separate processes. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?* (pp. 57-61). Norwood, NJ: Ablex.

Elshout, J. J. (1985, June). *Problem solving and education*. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179-197.

Engle, R. W. , Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent variable approach. *Journal of Experimental Psychology: General, 128* (3), 309-331.

Estes, W. K. (1974). Learning theory and intelligence. *American Psychologist, 29*, 740-749.

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning.* Hove, UK: Psychology Press.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: evidence for a developmental cascade. *Psychological Science, 7*, 237-241.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review, 109,* 116-136.

Goldman, S. R., & Pellegrino, J. W. (1984). Deductions about induction: Analyses of developmental and individual differences. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 149-197). Hillsdale, NJ: Erlbaum.

Greeno, J. G. (1978). A study of problem solving. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 13-75). Hillsdale, NJ: Erlbaum.

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: Erlbaum.

Gustafsson, J.- E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186-242.) New York: Simon & Schuster Macmillan.

Holzman, T. G., Pellegrino, J. W., & Glaser, R. (1982). Cognitive dimensions of numerical rule induction. *Journal of Educational Psychology, 74*, 360-373.

Hunt, E. B., Frost, N., & Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *The psychology of learning and motivation: Vol. 7* (pp. 87-122). New York: Academic Press.

Hunt, E., & Lansman, M. (1982). Individual differences in attention. In R. J. Sternberg (Ed.), *Advances in the psychology of human abilities* (Vol. 1, pp. 207-254). Hillsdale, NJ: Erlbaum.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 3/4,* 447-472.

Jensen, A. R. (1982). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 255-310). Hillsdale, NJ: Erlbaum.

Jensen, A. R. (1998). *The* g *factor: The science of mental ability*. Westport, CT: Praeger.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50*, 109-135.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*, 389-433.

Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology, 76*, 1325-1345.

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: Erlbaum.

Lohman, D. F. (1994). Component scores as residual variation (or why the intercept correlates best). *Intelligence, 19,* 1-11.

Lohman, D. F. (2000). Complex information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 285-340). Cambridge, UK: Cambridge University Press.

Lohman, D. F. (in press). Aptitude for college: The importance of reasoning tests for minority admissions. Paper presented at the conference "Rethinking the SAT: The future of standardized testing in university admissions," University of California, Santa Barbara. Retrieved October 29, 2002, from The University of Iowa, College of Education Web site: http://faculty.education.uiowa.edu/dlohman/

Lohman, D. F., & Hagen, E. (2001a). *Cognitive Abilities Test (Form 6): Interpretive guide for teachers and counselors.* Itasca, IL: Riverside.

Lohman, D. F., & Hagen, E. (2001b). *Cognitive Abilities Test (Form 6): Interpretive guide for school administrators.* Itasca, IL: Riverside.

Lohman, D. F., & Hagen, E. (2002). *Cognitive Abilities Test (Form 6): Research handbook.* Itasca, IL: Riverside.

Markman, A. B., & Genter, D. (2001). Thinking. *Annual Review of Psychology, 52,* 223-247.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*, 107-128.

Martinez, M. E. (2000). *Education as the cultivation of intelligence.* Mahwah, NJ: Erlbaum.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge, UK: Cambridge University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Pellegrino, J. W. (1985). Inductive reasoning ability. In R. J. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 195-225). New York: W. H. Freeman.

Pellegrino, J. W., & Glaser, R. (1980). Components of inductive reasoning. In R. E. Snow, P.- A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Vol. 1. Cognitive process analyses of aptitude* (pp. 177-218). Hillsdale, NJ: Erlbaum.

Primi, R. (2001). Complexity of geometric inductive reasoning tasks contribution to the understanding of fluid intelligence. *Intelligence, 30*, 41-70.

Raaheim, K. (1988). Intelligence and task novelty. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 73-97). Hillsdale, NJ: Erlbaum.

Raven, J. C., Court, J. H., & Raven, J. (1977). *Raven's Progressive Matrices and Vocabulary Scales.* New York: Psychological Corporation.

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.

Salthouse, T. A., Babcock, R. L, Mitchell, D. R. D., Palmon, R., & Skovronek, E. Sources of individual differences in spatial visualization ability. *Intelligence, 14,* 187-230.

Snow, R. E. (1980). Aptitude and achievement. *New Directions for Testing and Measurement, 5*, 39-59.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3-37). Cambridge, UK: Cambridge University Press.

Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law, 2,* 536-560.

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., Corno, L., & Jackson, D., III. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243-310). New York: Simon & Schuster Macmillan.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 47-104). Hillsdale, NJ: Erlbaum.

Spearman, C. E. (1927). *The abilities of man*. London: Macmillan.

Spilsbury, G. (1992). Complexity as a reflection of the dimensionality of a task. *Intelligence, 16*, 31-45.

Stankov, L. (1988). Single tests, competing tasks and their relationship to broad factors of intelligence. *Personality and Individual Differences, 9*, 25-33.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge, UK: Cambridge University Press.

Sternberg, R. J. (1986). Toward a unified theory of human reasoning. *Intelligence, 10*, 281-314.

Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112*, 80-116.

Süβ, H.- M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261-288.

Swiney, J. F. (1985). A study of executive processes in intelligence. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning* (2nd ed). New York: Macmillan.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Yanng, Y. & Johnson-Laird, P. N. (2001). Mental models and logical reasoning problems in the GRE. *Journal of Experimental Psychology: Applied, 7*(4), 308-316.

Zimmerman, W. S. (1954). The influence of item complexity upon the factor composition of a spatial visualization test. *Educational and Psychological Measurement, 14*, 106-119.

Author Notes

[1] Portions of this chapter appear in Lohman, D. F. (2002) "Complex information processing and intelligence" in R.J. Sternberg (Ed.) Handbook of intelligence. Cambridge, MA: Cambridge University Press.