Reasoning and Intelligence

David F. Lohman

Joni M. Lakin

The University of Iowa

October, 2009

To appear in

R. J. Sternberg and S. B. Kaufman (Eds.), *Handbook of Intelligence* (2nd ed.).

New York: Cambridge University Press.

Reasoning and Intelligence

The topic of reasoning has always been central to Western philosophy. Early psychological speculations about the nature reasoning (e.g., James, 1890/1950, chap. 22) grew out of these traditions, especially from the work of philosophers such as David Hume and John Locke. Normative standards for good reasoning are fundamental to philosophy. Building on this legacy, some psychologists have studied reasoning on formal logic tasks and the consistent violations of these normative standards that characterize much human reasoning (Stanovich chapter in this book). Researchers in this tradition study logical problem solving using the methods of inquiry developed in cognitive psychology (Leighton & Sternberg, 2004). A related tradition has focused on probabilistic reasoning in knowledge-rich domains such as law or medicine (Ellsworth, 2005; Patel, Arocha, & Zhang, 2005). Other researchers focus instead on individual differences in reasoning and the place of reasoning abilities within the larger domain of human abilities (Carroll, 1993). Typically, these researchers administer batteries of psychological tests to large samples of people and study the patterns of covariation among test scores using latent variable models. Finally, other researchers have attempted to understand individual differences in reasoning by modeling the processes individuals use when solving items on tests that define reasoning abilities in these latent variable models (e.g., Pellegrino, 1985; Sternberg, 1986).

Reasoning is closely allied with other domains of inquiry in psychology. Reasoning, problem solving, and decision-making represent different but overlapping aspects of human intelligence. Although interrelated, research on each of these three aspects of thinking is enormous (e.g., Holyoak & Morrison, 2005). In this chapter, we will survey only a small part of the field. Our emphasis will be on individual differences in reasoning as it is reflected in solving problems taken from or modeled after those used on psychometric tests of reasoning.

Reasoning refers to the process of drawing conclusions or inferences from information. Reasoning always requires going beyond the information that is given (Bruner, 1957). In logic, an inference is called *deductive* if the truth of the initial information (or premises) guarantees the truth of the conclusion. The inference is called *inductive* if the truth of the premises makes the conclusion probable but not certain. Distinctions between deductive and inductive reasoning can be important in understanding logic; but in practice, these distinctions may exist more in the mind of the researcher developing a task than in the performance of examinees on that task. Many researchers have found that performance on deductive and inductive tests is strongly related (Wilhelm, 2005).

Cognitive-Psychological Studies of Reasoning

Those researchers following the *cognitive psychological approach* to the study of reasoning typically study the responses of a small number of participants to logical tasks such as syllogisms or formal logic tasks. Researchers analyze how features of the problem influence the types of errors that participants make and often base their generalizations on the proportion of participants making certain errors (e.g., Stanovich, 1999). One source of debate in the cognitive approach is whether humans are fundamentally rational, as Aristotle assumed, or whether consistent demonstrations of irrational behaviors in the laboratory mean that humans function with pervasive biases that impede or prevent rational decision making. Researchers who conclude that humans operate with biases cite data showing that people are swayed by personal testimony over data and readily accept believable conclusions that are based on unlikely premises. However, critics of this research argue that the abstract structure of the problems can influence how they are solved and participants' misunderstandings of the format may explain some of these apparent failures in logical reasoning (Leighton, 2004). For example, illogical behavior on artificial tasks can disappear when the task is framed in a meaningful way (Evans & Feeney, 2004; Stenning & Monaghan, 2004).

Followers of the cognitive approach have debated how best to explain variation in performance across tasks: Although some have argued that failures of logical reasoning are caused by random errors, others have shown that these errors are correlated across tasks. That some people make more errors than others suggests computational limitations that vary systematically across individuals (Stanovich, 1999). That such a finding could be controversial would astonish most researchers coming from the psychometric approach.

Mental Rules or Mental Models?

Two theories have dominated psychological theorizing about reasoning: mental rules and mental models. Both theories were first applied to the study of deductive reasoning tasks such as syllogisms and then later applied to a broader range of reasoning tasks. The mental rules theory of deductive reasoning (Rips, 1994) posits mental processes common to all normally developed adults that operate directly on the representations of the premises. Humans are assumed to be natural logicians who are sometimes fallible because of errors in processing or because of limitations of the human cognitive system. According to mental rules theory, the basic processes involved in solving deductive reasoning problems are (a) encoding the premises into representations stored in working memory, (b) applying abstract, rule-based schemas to these representations to derive a conclusion, and (c) applying other rules to check the contents of working memory for incompatibilities. Although the model posits several sources of error, the number of steps to be executed in applying rules is the major source of difficulty. Errors in performance are thus primarily attributable to working memory overload (Gilhooly, 2004).

The mental models theory (Johnson-Laird, 2004) of deductive reasoning posits that the individual first transforms the premises of an argument into another representation (i.e., a mental model) that is consistent with the premises. Importantly, multiple mental models that are consistent with the premises must often be constructed and then compared in order to reach a valid conclusion. Each mental model represents a possible state of affairs that must be evaluated. Bara, Bucciarelli, and Johnson-Laird (1995) identified the following factors that affect syllogistic inference in the mental models approach: (a) assembling a propositional representation of premises; (b) constructing models that integrate information from premises; (c) formulating a conclusion which integrates relationships not expressed in the premises; (d) searching for alternative models to refute conclusions; and (e) recognizing similarities between models. All these processes require working memory resources. Limitations of working memory are considered especially important in understanding individual differences in reasoning in this theory, because working memory limits the number of mental models that can be held in mind at once. Individuals with limited working memory capacity can fail to generate enough models to evaluate the validity of a conclusion (Stanovich, Sá, & West, 2004).

The mental rules and mental models theories of reasoning propose universal but somewhat contradictory mechanisms for deductive reasoning (Roberts, 1993). Furthermore, advocates of both theories have been able to marshal considerable evidence in support of their position. Research that explicitly attempts to account for individual differences in reasoning offers a possible explanation for this paradox: On some problems, the behavior of some reasoners is more consistent with the mental models theory, whereas the behavior of other reasoners is more consistent with the predictions of a mental rules theory (Stanovich et al., 2004). In addition to stable individual differences in propensity to solve reasoning problems in one way or another, how the problem is presented can encourage individuals to change their strategies across items (Galotti, Baron, & Sabini, 1986). Therefore, what a task measures cannot be determined by simple inspection. Rather, what is measured depends on a complex interaction between the characteristics of the examinee, the task, and the situation. This does not mean, however, that one cannot know what tasks typically measure when they are attempted by individuals of known characteristics, but that what tasks measure and for whom and under what circumstances are inferences that must be supported by other data—not merely presumed to be the case.

Tacit and Explicit Processes

Human reasoning occurs at different levels of awareness. Most cognitive scientists distinguish between tacit and intentional (or explicit) reasoning processes (Evans & Over, 1996; Stanovich, 1999). *Tacit* processes that facilitate reasoning occur without conscious intervention and outside 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 (James, 1890/1950). Tacit processes are used when we make a decision in a quick or intuitive way, often 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 an initial 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 seek and know what to ignore (Horn & Masunaga, 2006). In part, this is due to greater experience; in part, to better use of past experiences. Other researchers describe this automatic attention as the extent to which the person is attuned to certain aspects of a situation and not others (Gobet & Waters, 2003). By temperament or training, some people are more attuned to the distress of others, the beauty in a painting, the mathematical properties of objects, or the alliteration in a poem. Tacit processes are also importantly linked to feelings, which seem essential for solving ill-structured problems of all sorts. 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 our conscious awareness. We are aware not only of the outcome of our thinking (as with tacit processes), but also with the processes themselves. This is the type of reasoning that is most distinctly human. Such thinking is often described as *strategic* or rule based. It typically requires effort, and it allows us to bypass the relatively slow accumulation of experiences that underlie tacit learning. We can thereby transfer principles (e.g., *always capitalize proper nouns*) rather than an accumulation of varied experiences (e.g., *I always capitalize this 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 comparatively slow and effortful, but flexible.

Thus, reasoning involves both conscious (or explicit) and unconscious (or tacit) processes. Although some refer to both explicit and tacit reasoning processes, other psychologists argue that tasks elicit reasoning only to the extent that they require conscious application of particular mental processes (Elshout, 1985; Sternberg, 1986).

The Role of Knowledge in Reasoning

Reasoning well in domains of nontrivial complexity depends importantly on knowledge. Expertise is rooted in knowledge, and experts reason differently about problems than do novices (Feltovich, Prietula, & Ericsson, 2006). 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. Everyday reasoning depends heavily on the efficacy of past reasoning processes (stored as knowledge) as well as the efficacy of present reasoning processes. An increasingly sophisticated knowledge base supports increasingly sophisticated forms of reasoning. A more sophisticated knowledge base has richer, more abstract associative connections between concepts and more metacognitive knowledge that links strategies to goals. This frees working memory resources for problem solving (Gobet & Waters, 2003; Feltovich et al., 2006; Horn & Masunaga, 2006; Proctor & Vu, 2006).

Successful problem solvers form problem representations that are not only more abstract than those of novices, but also more finely tuned to the problem at hand. Markman and Gentner (2001) argue that the formation of moderately abstract conceptual relations may be a precursor to the detection of coherent patterns that help successful problem solvers make connections to similar problems with known solutions. 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

(a) missing grounds (e.g., begging the question); (b) irrelevant grounds (e.g., red herring); (c) defective grounds (e.g., hasty generalization); (d) unwarranted assumptions; and (e) 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 especially helpful here. *Strong methods* of reasoning rely heavily on knowledge within a particular domain, whereas *weak methods* depend less on content and context. That is, strong (or domain-specific) methods describe what people do when they *do know* what to do; weak (or domain-general) methods describe what people do when they *do not know* what to do. Therefore, children and novices are more likely to use domain-general methods. The distinction is sometimes overblown; Markman and Gentner (2001) observed that many instances of domain-specific thinking result from domain-general processes operating on domain-specific representations and 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. 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.

A Classification Scheme for Reasoning Processes

Sternberg (1986) offered a helpful way to categorize the kinds of mental processes used on commonly investigated reasoning tasks: He calls them *selective encoding*, *selective comparison*, and s*elective combination*. We will alter these labels somewhat in the discussion that follows. Recall from the discussion of mental models that although a test item or experimental task may elicit these processes for some or even most people, it may elicit other (non-reasoning) processes for any particular person or item. As Sternberg puts it, “the extent to which a task elicits reasoning is a function of the interaction between person and task, rather than merely a function of the task” (p. 287).

Selective encoding refers to the process of distinguishing relevant from irrelevant information. Such encoding can be effortful and deliberate, in which case it is clearly a reasoning process, or automatic, in which case it is at best a process that facilitates reasoning. For example, expert problem solvers generally attend to the deep structure of a problem and notice features and tasks similarities invisible to the untrained eye, whereas novices attend to its surface features. For the expert, then, encoding processes facilitate problem solution but are automatized and not truly part of reasoning on the task; for the novice, however, attempting to encode the most important features is an effortful and multi-step process that can impede problem solution. Learning what to notice and what to ignore is the essential first step in reasoning about any problem.

Whereas selective encoding means attending only to a subset of the information in a situation, *selective comparison* means retrieving and then comparing only a subset of the potentially relevant information about these concepts from long-term memory. We know a lot about many things that we think we do not know very well, and vastly more about things we know intimately; choosing what knowledge to apply to a new problem is a nontrivial source of reasoning complexity. Developmental psychologists have long known that children reveal much about the sophistication of their reasoning by how they classify or sort objects: on the basis of an arbitrary association, or by using perceptual characteristics, or, at the highest level, by using several different abstract concepts (e.g., Piaget, 1963). Therefore, deciding how best to describe the relationships among two or more concepts is the critical second step in reasoning. For example, consider the analogy:

teacher : student :: coach : (a) athlete (b) child

There are many things a particular examinee knows about teachers: that teachers are people, that her English teacher is Mrs. Smith, that teachers are adults, that teachers have college degrees, and so on. Solving the analogy requires that the student focus on that small subset of features of the concept *teacher* that overlaps with the concept *student.* Comparison refers to the inference process—that is, the process of finding relationships between the two concepts and then selecting one that best characterizes the type of association between them given other contextual clues. For example, a vague relationship would be that teachers and students are both people, but this will not lead to a unique answer in this problem. One of the critical differences between good and poor reasoners is that poor reasoners often settle for a vague relationship or rule rather for a more exact one (Sternberg, 1985). This could be because they terminate the search for a rule or relationship too quickly, or because they do not critically examine how well candidate rules or relationships describe the data, or because they simply do not see or know the rule. Thus, what is called the comparison phase of reasoning actually has two parts: (a) the generation of plausible rules or relationships and (b) the evaluation of these rules or relationships. Oftentimes the problem itself provides the context for at least a partial evaluation of the rule. In an analogy, the relationship between the first two terms (A and B) must also be applicable to the third term (C) and one of the options (D_1, D_2, ...). If the A-B relationship cannot be mapped on to one of the C-D pairs, then one must try to generate other possible relationships. When inferring the meaning of a word or phrase in a text, the surrounding text provides the context for evaluation.

Finally, the third category of reasoning processes may be called orderly, strategic, or planful combination of information in working memory. *Strategic combination* is often required on tasks that require deductive reasoning, such as formulating a logical argument or a mathematical proof.

Syllogisms capture key aspects of this type of reasoning, albeit in an artificial format. Consider the following syllogism:

All A are B.

Some B are C.

Some C are A. (True or False?)

The difficulty in such problems lies not in discovering relationships or in understanding the meaning of concepts such as *all* or *some*. Rather, the difficulty lies in keeping track of all the ways in which the three terms (A, B, and C) can be combined. This quickly taxes working memory and can lead to a failure to consider combinations that disprove the rule (Stanovich et al., 2004). Memory burdens (and thus errors) are reduced if one has or can assemble a systematic method for solving the problem. For example, abstract syllogisms can be made more understandable by replacing abstractions (A, B, and C) with concrete nouns:

All dogs are animals.

Some animals are cats.

Some cats are dogs. (True or False?)

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. Because of the importance of strategy use in deductive reasoning, many investigators have noted that such tasks are particularly susceptible to training. This also means that deductive reasoning tests can measure different abilities in examinees who have learned strategies for solving problems like those used on the test than for examinees who must invent a strategy on the spot.

There are several other processes that, while not reasoning processes, are often essential. All are routinely used to regulate processing in working memory. Particularly important are the executive functions of self-monitoring and coordination. In order to be strategic or planful in working out the ways in which concepts can be combined or rules can be applied, one must monitor the success of one’s efforts. Thoughtful adaptation of old strategies, the invention of new strategies, or the ability to learn from each problem attempted, all depend on the ability to monitor the success of one’s efforts. Thus, self-monitoring is a critical skill. Similarly, when solving reasoning problems, one must frequently coordinate different types of mental models. Understanding a text, for example, requires that one coordinate what Kintsch and Greeno (1985) calls a text-based model (i.e., the network of ideas) with a situation model (often an envisioning of the situation being described).

Working Memory

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 (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Süß, Oberauer, Wittman, Wilhelm, & Schulze, 2002). However, critics complained that some tasks used to estimate working memory in these studies were indistinguishable from tasks used to estimate reasoning. 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, a meta-analysis of the existing data yielded a true-score correlation of $r = .48$ between working memory and *g*, far below the unity some claim (Ackerman, Beier, & Boyle, 2005).

In part, this is a problem of words. The term *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 usually underrepresented.

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; to assemble and then revise a strategy for performing a difficult, attention-demanding task; to maintain a high level of effort across a substantial number of trials; and then to repeat the process for a new task with a new set of directions. In addition, many working memory tasks require individuals to process simultaneously one set of ideas while remembering another set. Although the individual tasks are generally thought to be easy, they are certainly not trivial, especially when performed under memory load. These tasks 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. Such executive functions clearly overlap with many researchers' conception reasoning or even of general intelligence. This heated debate may boil down to a difference in branding caused by the parallel development of closely related constructs in both psychometric and cognitive traditions.

Measuring Reasoning Abilities

Performance on one item provides little information about individual differences that would generalize to a test composed of similar items, and even less information about the broader ability construct defined by performance on several tests. Research on reasoning requires a method for measuring reasoning abilities. Although a single test-task is often used in experimental research, the term ability implies consistency in performance across some defined class of tasks. Indeed, some of the confusions and controversies in the field stem from equating performance on a particular task with the broader psychological construct. Psychological tests are simply organized collections of such tasks. However, typically less than half of the variation on well constructed, reliable tests is shared with other tests that measure the same construct using somewhat different kinds of test tasks. An early but still reasonable rule in psychological measurement is that when measuring any ability, one should combine performance across at least three different measures that use different formats to reduce the specific effects of individual tasks (Süß & Beauducel, 2005).

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.

Examples of the nine reasoning tasks used in the most recent revision of Thorndike and Hagen's Cognitive Abilities Test (CogAT, Lohman, in press) are presented in Figure 1. Although unfamiliar to most researchers, the CogAT is the most widely used group ability test in the United States and the United Kingdom. The three reasoning abilities measured by the test correspond with the three aspects of fluid reasoning ability identified in Carroll's (1993) compendium. Each reasoning ability is estimated by three subtests that require somewhat different processing.

Uses of Reasoning Tests

Traditionally, tests such as the CogAT or the SAT have been used (a) to predict achievement, (b) to provide a measure of cognitive development that supplements or can be contrasted with other measures of a student's cognitive development, and (c) to guide efforts to adapt instruction to the abilities of students. One need not have much of a theory of reasoning abilities in order to use a test such as the SAT Reasoning Test to predict college grade-point average (GPA). Indeed, the primary contribution of a theory of reasoning in such cases would be to avoid misinterpretation of predictions. Naive interpreters will see causal arrows running only from reasoning ability to achievement (or GPA), rather than seeing both as outcomes of education and experience (Snow, 1991). Understanding of the nature of reasoning abilities is also required when scores on an ability test are used as a measure of a student's level of cognitive development. For example, SAT scores can provide new information on a student's cognitive development only if the interpreter has some understanding of what reasoning abilities are and how they develop. Diagnostic interpretations of test scores attempt to provide this information at a more skill-based level (see Mislevy, 2006).

The third use of reasoning tests—to guide instructional adaptations—often requires the most sophisticated understanding of reasoning abilities. Every effort to make instructional adaptations on the basis of student performance on an ability test makes some implicit or explicit assumption about what those measured abilities are. For example, if ability is primarily a matter of speed of processing, then slowing the pace of instruction may be the most effective adaptation for students with relatively poorly developed reasoning abilities. If, on the other hand, reasoning has more to do with the type of thinking one uses to solve problems than the speed of processing, then slowing the pace of instruction may not be the most effective adaptation. Knowing what elements of a task elicit or circumvent reasoning helps us better understand what those abilities are and how instruction might be modified to require or circumvent the need for those abilities.

One really does not know what abilities are unless one knows how they develop. Reasoning abilities are not only critical aptitudes for learning, they are also among its most important outcomes. Instructional interventions that explicitly require and succeed in developing students' reasoning abilities comprise one of the best sources of evidence on the construct validity of reasoning tests (Snow & Lohman, 1989).

The Construct Validity of Reasoning Tests

Inferences about the psychological constructs that a test measures in any particular application require multiple sources of evidence. The two major aspects of construct validation 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 individuals typically use in responding to items on a test. The cognitive psychological research on families of reasoning tests or tasks summarized in previous sections of this chapter provides the foundation for this aspect of construct validation.

However, inferences about processes do not depend on or explain individual differences on a task. 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 major sources of individual differences across several tasks. And only a part of these common individual differences will be attributed to the latent variable that best represents the reasoning construct. 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.

Nomothetic span, on the other hand, concerns evidence on the nature of a construct that derive from its relationships with other constructs. For constructs that are grounded in individual differences, these inferences are based on the complex web of relationships among scores on tests that are designed to measure different constructs. Since the patterns of individual differences on a test depend on the characteristics of both the sample of test takers and of the number and nature of other tests included in the study, inferences about the nomothetic span of a test gain credence only after the test has been used in many different studies. The aspect of construct validation captured by nomothetic span affirms the importance of understanding individual differences on families of reasoning tasks, not simply on one or two tasks that have sparked interest among researchers. It follows that using a test in which all items follow the same format to define *reasoning* (or even worse, to define *intelligence*) reflects a fundamental misunderstanding of psychological measurement.

Nomothetic Span of Reasoning Tests

Psychologists have been investigating the number and organization of cognitive abilities for over a century now. Carroll (1993) reanalyzed and then summarized much of this work. His conclusions generally conform with those of other researchers in the field (McGrew, 2005). The first important finding is that human abilities are organized hierarchically. This means that some cognitive competencies are more broadly useful than others. It also means that theories that postulate an independent set of abilities (Gardner, 1983; Thurstone, 1938) or that only one ability of any consequence (Jensen, 1998) are fundamentally flawed. The hierarchy that Carroll proposes starts with *g* (general mental ability) at the topmost level: Although the broadest factor in the model, *g* is also the least psychologically transparent. Eight broad group factors that are somewhat more psychologically transparent define the second level. These factors vary in their closeness or association with *g*. The closest is an ability factor that Cattell (1963) called *Gf* (general fluid ability). Other broad factors closely related to *g* at this level include *Gc* (General verbal crystallized ability), *Gv* (general spatial visualization ability), and *Gm* (general memory ability). Finally, a longer list of primary factors that are even more psychologically transparent defines the third level. These factors include such abilities as verbal comprehension, verbal fluency, inductive reasoning, spatial visualization, perceptual speed, and number facility. Most of these specific abilities have quite narrow predictive ranges.

The second critical finding in the literature on human abilities is that the general reasoning factor (*Gf*) may be decomposed into subfactors: (a) sequential reasoning (verbal logical or deductive reasoning), (b) quantitative reasoning (inductive or deductive reasoning with quantitative concepts), and (3) inductive reasoning (often measured with figural tasks). A good reasoning test, then, should probably measure all three of these reasoning factors—or at least not be strongly biased towards one (Wilhelm, 2005). This fact is commonly overlooked in studies that represent fluid reasoning abilities with a single figural reasoning test such as the Progressive Matrices test (Raven et al., 1977).

The third critical finding is that the topmost factor in the hierarchy (*g*) is virtually synonymous with the factor called *Gf* (general fluid ability) at the second level. And *Gf* is in turn virtually synonymous with the primary factor called inductive reasoning (IR). Gustafsson (1988; Kvist & Gustafsson, 2008) claims that the three factors are in fact identical (i.e., $g = Gf =$ IR). Others would describe the relationship between *g* and *Gf* as more of an approximation than an identity (Carroll, 1993; Horn & Blankson, 2005). In either case, however, we are left with the important insight that reasoning abilities are at the core of human cognitive competence. In other words, the least psychologically transparent dimension (*g*) is in large measure isomorphic with one of the most psychologically transparent dimensions (IR).

Evidence from School Learning

Information on the nomothetic span of a test also comes from the sorts of criterion behaviors that the test predicts. Measures of general reasoning ability (or *Gf*) are good predictors of success in learning a broad range of tasks. Correlations are generally highest for the early phases of learning new, especially open-ended skills (Ackerman, 1988) and for learning the sorts of organized systems of meaningful concepts that are commonly required in formal schooling. Population correlations with measures of school success range from $r = .4$ to .8, depending on the criterion measure (e.g., grades, achievement tests) and of content of reasoning test (e.g. verbal, quantitative, or figural reasoning). Predictive and concurrent correlations based representative samples of U.S. school children are commonly reported in technical manuals for group ability and achievement tests, most of which are updated and renormed every six to 10 years (e.g., Lohman & Hagen, 2002).

Reasoning tests correlate with academic success because school learning requires reasoning abilities. Understanding a story, inferring the meaning of an unfamiliar word, detecting patterns and regularities in information, abstracting the information given to form more general rules or principles, applying mathematical concepts to solve a problem...in these ways and in a hundred other ways, successful learning requires reasoning strategies. Indeed, the best way to develop reasoning abilities is through challenging instruction that requires students to exercise old reasoning strategies and to invent or learn new ones (Martinez, 2000; Nickerson, 2004).

These important reasoning skills are captured even by what some would consider narrow measures of achievement like vocabulary tests. Individual differences on vocabulary tests may arise from variance in how well learners use certain metacognitive or performance processes when learning—such as systematically testing alternative interpretations of a word when it is used in unfamiliar contexts—that then lead to a richer and more usefully organized knowledge base to guide new learning (e.g., Robinson & Hayes, 1978). Marshalek (1981) concludes that the ability to infer word meanings from the contexts in which they occur is the cause of high correlation typically observed between vocabulary and reasoning tests. But there is also a synergism in that vocabulary knowledge allows comprehension and expression of a broader array of ideas, which in turn facilitate the task of learning new words and concepts. Thus, language functions as a vehicle for the expression, refinement, and acquisition of thought, and the humble vocabulary test masks an enormous amount of reasoning and remembering.

Aptitude-Treatment Interaction Research

One of the best sorts of evidence for construct validity via nomothetic span comes from experiments in which the treatment conditions are designed to vary in their demands for the construct presumably measured by a test (Messick, 1989). Those who understand that abilities are multiple, not unitary, have always believed that student's profiles on the sort of primary abilities that Thurstone (1938) identified would be the key to effective instructional adaptation. In the 1950s, research on the problem began in earnest (see Cronbach, 1957). The idea is straightforward. First, measure students' abilities. Then, randomly assign them to different instructional treatments, each of which is designed to appeal to students with different patterns of abilities. Finally, measure outcomes to see whether students with a particular ability profile performed better in one instructional treatment than another treatment. Statistically, the goal is to look for interactions between aptitude variables (such as verbal ability or spatial ability) and treatments (such as the use of demonstrations and films versus written texts) or aptitude by treatment interactions (ATI).

Hundreds of studies of ATI studies now have been conducted. Cronbach and Snow (1977) provided an initial summary; more recently, Corno et al. (2002) have updated the record. The most astonishing find in this vast research effort is this: Contrary to the expectations of virtually all, the profile of specific abilities or learning styles generally does not account for much of the variation in outcomes. Indeed, interactions between learning styles (such verbalizer versus visualizer) and instructional methods (such as an emphasis on visual versus verbal media) are usually small and frequently in opposite directions in different studies. Instead, the ability dimensions that routinely interact with instructional methods are *Gc* (general verbal-crystallized achievement) and *Gf* (general fluid reasoning abilities). or Gv (general spatial visualization abilities) This means that it is the students' knowledge and skills in a domain, and their abilities to reason in the symbol system of that domain that matters most when deciding how best to help them learn. For example, it is not the ability to generate visual images that matters, but the ability to reason with and about those images. Similarly, it is not the ability to remember words or to speak with fluency but rather to reason about what the concepts words signify.

The nature of the statistical interaction between instructional treatments and reasoning abilities is straightforward. Instructional methods that place the burden of making inferences and deductions on the student increase the relationship between reasoning abilities and achievement. Instructional methods that scaffold, remove, or otherwise reduce this burden reduce the relationship between reasoning abilities and achievement. The relationship is moderated by other variables, particularly anxiety, but reasoning abilities and prior knowledge in the domain are clearly the most important aptitudes for learning from instruction. Put differently, those who hope to enhance the probability of successful completion of school by offering different instructional opportunities are most likely to succeed if these adaptations are based on the developed broad reasoning abilities of students rather than narrow cognitive styles.

In summary, studies that address the nomothetic span of reasoning tests show that they (a) are at the core of human cognitive abilities, (b) are among the best predictors of success of meaningful learning, and (c) routinely interact with instructional methods that vary in the demands that reasoning tests place on students to think for themselves. Such evidence confirms the important role that reasoning tests play in human abilities. But other information is needed to understand exactly what these tests measure.

Hypotheses about the Construct Representation of Reasoning Tests

Hundreds of studies have estimated relationships between reasoning tests and other kinds of ability tests show that reasoning tests are good measures of the general ability (*g*). But evidence of construct representation is needed to explain *why* reasoning tests are such good measures and what essential processes they tap into that could explain this relationship. Two-dimensional scalings of the correlations among large batteries of tests reveal something that can serve as a useful bridge between the cognitive psychological studies that investigate the construct representation of reasoning tests and the correlational studies that address the nomothetic span of reasoning tests. In these scaling, complex tests that load heavily on *g* (or *Gf*) fall near the center of the plot, whereas simpler tasks are distributed around the periphery. (See Figure 2.) Complex reasoning tasks occupy the spots closest to the center.

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) an increase in the 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 2) without also increasing the number of component processes required (Hypothesis 1). Despite 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 plot shown in Figure 2 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, Kyllonen, and Marshalek's (1984) reanalysis of old learning-task and ability-test correlation matrices showed similar continua. Spilsbury (1992) argued that the crucial manipulation was an increase in the factorial complexity of a task (that is, the number of different abilities required). 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 possibly as a result of more efficient brain structures (Jensen, 1998). Although disattenuated correlations between reaction time (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 $-.4$ range (Deary & Stough, 1996; Jensen, 1982; Roberts & Stankov, 1999; Sternberg, 1985). In principle, processing speed could be estimated on any elementary cognitive task that minimizes the import of learning, motivation, strategy, and other confounding variables. In fact, response latencies on many tasks show a pattern of increasing correlation with an external estimate of *g* as task complexity decreases. 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 across items.

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 of 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 such as reading comprehension. 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 on a complex task 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 Critical Performance Components

If the *g*-loading of a test 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 (1977) 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 require education of relations 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 these reasoning tests. Greeno (1978) refers to such tasks, collectively, as problems of inducing structure. Indeed, the need for learners to induce 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 unequivocally supporting the hypothesis that individual differences in particular component processes correlate strongly with *g* has been surprisingly difficult to obtain. Sternberg's (1977) investigations of analogical reasoning found little generalizability across tasks or scores for the inference component (Spearman's education of relations), 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 intended to require different amounts of a particular component processes will appear in the intercept (the mean score of the individual across items) rather than in the component scores (reflecting factors that vary within the individual). 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 because they omit important variance due to individual differences in reasoning.

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) focused on manipulations that affected the demands placed on his three component processes: selective encoding, selective comparison, and selective combination, described previously. Demands on selective encoding skills are amplified by increasing distractions caused by salient but irrelevant information, or, when solving items on mental tests, by preventing examinees from looking ahead to the alternatives before studying the stem (Bethell-Fox, Lohman, & Snow, 1984). Demands on selective comparison are increased by manipulating the familiarity of concepts. Presenting somewhat unfamiliar concepts or using familiar concepts in unfamiliar ways places heavy demands on the ability to retrieve and compare information. Selective combination can be manipulated by providing algorithms or strategies that reduce working memory burdens. Practice on items that are similar to those used on a test can undermine the *Gf*-loading of a test because the processes and strategies used become increasingly automatized; this is especially apparent on deductive reasoning tasks and their demands on selective combination (Sternberg, 1986).

Attention and Working Memory Capacity

All information-processing models of memory and cognition posit the existence of a limited capacity 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 in terms of attentional resources, and yet 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. 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 greater 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 (Ackerman & Woltz, 1994). Attributing individual differences in reasoning to individual 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 (Engle, Tuholski, Laughlin, & Conway, 1999; Holzman, Pellegrino, & Glaser, 1982). Controlling attention in this way is a critical aspect both of selective encoding and goal management within the constraints of working memory (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 a chosen strategy may be effective, it must be performed within the limits of the working memory system while sharing resources with retrieval, executive, and other processes. 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 to maintain these competing processes.

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 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, Lohman, & Snow (1983), on the other hand, focused more narrowly 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 and complex problem solving. 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.

Several investigators have attempted to manipulate the extent to which items require assembly and control processes 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 requiring maximally flexible adaptation.

Results showed that low-ability students were more adversely affected by mixing items than high-ability students, 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 items that required that students identify or apply difficult rules. Retrospective reports supported the conclusion that high-*g* subjects were better able to adapt their strategies flexibly to meet changing task demands. Swiney (1985) 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 similar studies contrasting blocked versus mixed item presentations and found small relationships consistent with Swiney's (1985) hypotheses that mixed items would show greater *g*-loading. An opposite finding, however, was reported in a study by Carlstedt, Gustafsson, and Ullstadius (2000). Three kinds of inductive reasoning problems were administered to groups of Swedish military recruits. Carlstedt et al. unexpectedly 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: On plots of two-dimensional scalings of test correlations, tests increase in apparent complexity as one moves from periphery to center of the plot. 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 thinking 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 items rules that will be needed in combination to solve hard items. All of these elements may be necessary to explain the relationships among batteries of diverse collections of ability tests.

Defining Reasoning

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:

- 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 (e.g., words, symbols, figures, sounds, movements) 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.

- 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.

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. The 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 both on reasoning tests and in everyday tasks. 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 by experts in a knowledge-rich domain, however, appear to be largely the same as those which enable the novice to infer consistencies or deduce likely consequents in novel problem-solving.

There are many sources of evidence that bear on the construct validity and practical importance of reasoning tests. First is the fact that reasoning is the central or most general cognitive ability in any diverse battery of tests. Second, reasoning tests predict success in academic learning because—as Snow, Greeno, Resnick, Bruner, and others have pointed out—academic learning is at its core one grand game of inference and deduction making. All instruction is incomplete in some respects. Effective learning requires that the student continually go beyond the information given to find similarities and differences between new patterns and concepts already in memory. Third, reasoning abilities are the critical moderator of instructional adaptations. By tracking what increases or decreases the relationship between reasoning ability and learning outcomes, we understand better both what reasoning abilities are and how instruction can be made more effective for more learners. Fourth, there is now a substantial research base in cognitive psychology on the nature of human reasoning (e.g., Evans & Over, 1996; Holyoak & Morrison, 2005; Johnson-Laird, 1999; Leighton & Sternberg, 2004; Rips, 1994; Stanovich, 1999). Especially helpful are studies of individual differences in reasoning measured on test-like tasks modeled after those used on ability tests. Indeed, one would be hard pressed to think of any construct in psychology that is better understood, and whose practical relevance for education at all levels is better demonstrated than reasoning abilities.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288-318.

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.

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., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, volition, and motivation. *Journal of Educational Psychology, 86,* 487-515.

Bara, B. G., Bucciarelli, M., Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *American Journal of Psychology, 108*, 157-193.

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.

Bruner, J. S., (Ed.). (1957). Going beyond the information given. In *Contemporary approaches to cognition: A symposium held at the University of Colorado* (pp. 41-69). 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., & Shell, 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, England: Cambridge University Press.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.

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., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., & Talbert, J. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow.* Hillsdale, NJ: Erlbaum. (Educational psychology series; work completed by the Stanford Aptitude Seminar after the death of R. E. Snow; L. J. Cronbach, Ed.).

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

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671-684.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. New York, NY: Irvington.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York, NY: 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? Contemporary viewpoints on its nature and definition* (pp. 57-61). Norwood, NJ: Ablex.

Ellsworth, P. C. (2005). Legal reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 685-704). New York, NY: Cambridge University Press.

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

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,* 309-331.

Evans, J. St. B. T., & Feeney, A. (2004). The role of prior belief in reasoning. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 78-102). New York, NY: Cambridge University Press.

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

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 41-68). New York, NY: Cambridge University Press.

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

Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General, 115*, 16-25.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York, NY: Basic Books.

Gilhooly, K. J. (2004). Working memory and reasoning. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 49-77). New York, NY: Cambridge University Press.

Gobet, F., & Waters, A. J. (2003). The role of constraints in expert memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 1082-1094.

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.

Holyoak, K. J., & Morrison, R. G. (Eds.). (2005). *The Cambridge handbook of thinking and reasoning.* New York, NY: Cambridge University Press.

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

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test, and issues* (2nd ed., pp.41–68). New York, NY: Guilford Press.

Horn, J., & Masunaga, H. (2006). A merging theory of expertise and intelligence. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 587-612). New York, NY: Cambridge University Press.

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, NY: 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.

James, W. (1950). *Principles of psychology* (Vol. 2). New York, NY: Dover. (Original work published in 1890)

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.

Johnson-Laird, P. N. (2004). Mental models and reasoning. In J. P. Leighton, & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 169-204). New York, NY: Cambridge University Press.

Kintsch, W., & Greeno, J. G., (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92,* 109-129.

Kvist, A.V., & Gustafsson, J.- E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence, 36,* 422-436.

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

Leighton, J. P. (2004). The assessment of logical reasoning. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 291-312). New York, NY: Cambridge University Press.

Leighton, J. P., & Sternberg, R. J. (Eds.). (2004). *The nature of reasoning.* New York, NY: Cambridge 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: Erlbaum.

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

Lohman, D. F. (in press). *Cognitive Abilities Test (Form 7).* Rolling Meadows, IL: Riverside.

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

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

Marshalek, B. (1981). Trait and process aspects of vocabulary knowledge and verbal ability (Tech. Rep. No. 15). Stanford, CA: Stanford University, Aptitude Research Project, School of Education. (NTIS No. AD-A102 757). 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.

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test, and issues* (2nd ed., pp. 136-181). New York, NY: Guilford Press.

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257-353). Westport, CT: American Council on Educational/Praeger.

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

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

Nickerson, R. S. (2004). Teaching reasoning. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 410-442). New York, NY: Cambridge University Press.

Patel, V. L., Arocha, J. F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 727-750). New York, NY: Cambridge University Press.

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

Piaget, J. (1963). *The psychology of intelligence.* New York, NY: International Universities Press.

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

Proctor, R. W., & Vu, K. L. (2006). Laboratory studies of training, skill acquisition, and retention of performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 265-286). New York, NY: Cambridge University Press.

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, NY: Psychological Corporation.

Rips, L. J. (1994). *The Psychology of proof*. Cambridge, MA: MIT Press.

Roberts, M. J. (1993). Human reasoning: Deductive rules or mental models, or both? *Quarterly Journal of Experimental Psychology, 46A*, 569-589.

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. (1990). 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, England: Cambridge University Press.

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

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.

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, NY: Macmillan.

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

Stanovich, K. E. (this volume). Chapter name. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. ?). New York, NY: Cambridge University Press.

Stanovich, K. E., Sá, W. C., & West, R. F. (2004). Individual differences in thinking, reasoning, and decision making. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 375-409). New York, NY: Cambridge University Press.

Stenning, K., & Monaghan, P. (2004). Strategies and knowledge representation. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 129-168). New York, NY: Cambridge University Press.

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, England: Cambridge University Press.

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

Süß, H.- M. & Beauducel, A. (2005). Faceted Models of Intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of measuring and understanding intelligence* (pp. 313-332). Thousand Oaks, CA: Sage Press.

Süß, 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.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs, 1.*

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

Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of measuring and understanding intelligence* (pp. 373-392). Thousand Oaks, CA: Sage Press.

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.

List of Figures

Verbal Reasoning

Verbal Analogy

movie -> watch : book -> ?
A) library B) rent C) read D) write

Verbal Classification

discover create imagine ?
A) start B) think C) invent D) learn

Sentence Completion

Even though I am older than Bob, Bob is ______ than I am.
A) younger B) shorter C) taller D) happier

Quantitative Reasoning

Number Analogy

[11 -> 16] [8 -> 13] [3 -> ?]
A) 6 B) 7 C) 8 D) 9

Number Puzzle

$\square = \lozenge \times 2$

$\lozenge = 4$

$\square = ?$

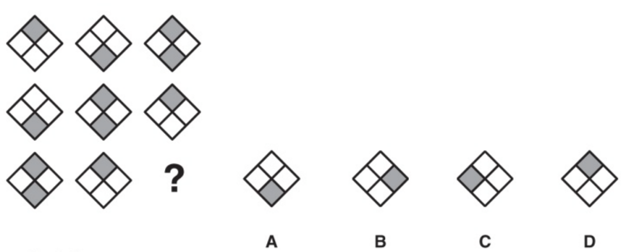
A) 4 B) 6 C) 8 D) 10

Number Series

3 6 9 12 15 ?
A) 15 B) 16 C) 17 D) 18

Nonverbal Reasoning

Figure Matrices



Paper Folding

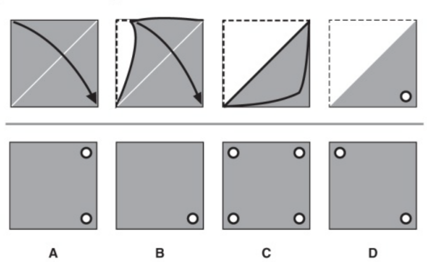
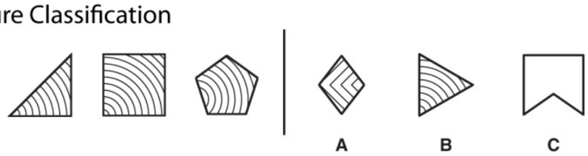
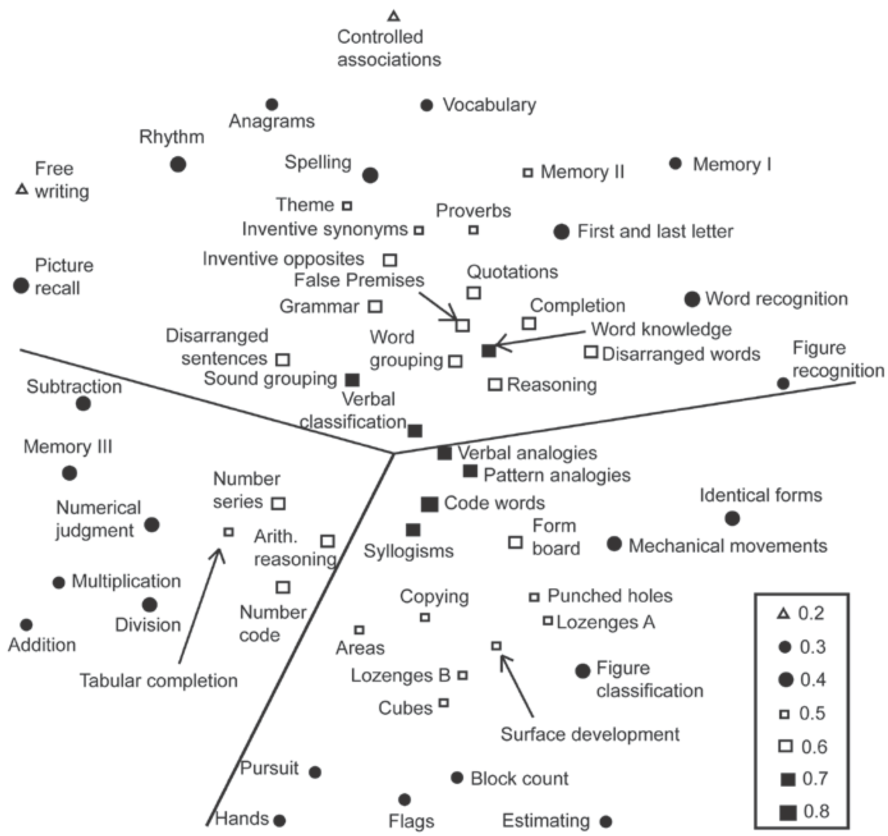


Figure Classification





Controlled associations
Vocabulary
Anagrams
Rhythm
Free writing
Spelling
Memory II
Memory I
Theme
Proverbs
Inventive synonyms
First and last letter
Picture recall
Inventive opposites
Quotations
False Premises
Grammar
Completion
Word recognition
Word knowledge
Disarranged sentences
Word grouping
Disarranged words
Figure recognition
Sound grouping
Reasoning
Subtraction
Verbal classification
Memory III
Verbal analogies
Pattern analogies
Number series
Identical forms
Numerical judgment
Code words
Arith. reasoning
Form board
Syllogisms
Mechanical movements
Multiplication
Division
Number code
Copying
Punched holes
Lozenges A
Addition
Areas
Tabular completion
Lozenges B
Figure classification
Cubes
Surface development
Pursuit
Block count
Hands
Flags
Estimating
0.2
0.3
0.4
0.5
0.6
0.7
0.8