RESEARCH MEMORANDUM

May 2003
RM-03-01

Toward Specifying a Construct of Reasoning

Donald E. Powers
Carol A. Dwyer

Educational Testing Service

Research & Development Division
Princeton, NJ 08541

Toward Specifying a Construct of Reasoning

Donald E. Powers and Carol A. Dwyer

Educational Testing Service, Princeton, NJ

May 2003

Research Memoranda provide preliminary and limited dissemination of ETS research prior to publication. They are available without charge from:

Research Publications Office
Mail Stop 10-R
Educational Testing Service
Princeton, NJ 08541

Abstract

The main purpose of this paper was to work, within the framework of construct validity theory, toward the delineation of a construct of *reasoning*. In so doing, we address *reasoning*, *broadly conceived*. We make no claim to have considered every aspect of this complex construct.

The outcome of our deliberations is a list of elements that, we believe, are useful to consider in any attempt to specify the facets of a construct of reasoning ability. As we note, befitting such an important construct, there have been many attempts to articulate its (multifaceted) nature, some of which are discussed here.

Given the variety of ways in which reasoning has been characterized, however, we offer no single definition or comprehensive vision of reasoning. Rather, we attempt only to set forth some of the salient features of reasoning that, to us, seem worthy of consideration and possibly of incorporation in any construct statement of reasoning.

Our main conclusion is that there is no *single* construct of reasoning. Instead, any of several formulations may be equally useful and informative depending on the particular context and the user's purpose. We hope, though, that our reflections provide some of the "grist" that users may need in order to specify a construct of reasoning that meets their requirements.

Key words: Reasoning, thinking, analytical ability, construct validity, test development, test validation

Acknowledgements

The authors would like to thank Matthew Bell, Linda Cook, Dan Eignor, Bill Eilfort, David Lohman, and Carol Tucker for helpful comments on earlier drafts of this paper.

Table of Contents

The purpose of this paper is primarily to delineate, within the framework of construct validity theory, a construct of *reasoning*, and secondarily to describe how selected testing programs have operationalized this construct in the assessments that they offer. Our aim, as well as that of the authors of other papers in this series (on verbal ability, quantitative ability, and writing ability), is to create a shared understanding of the nature of selected constructs that will help guide future discussion and decision-making for research and operational purposes. In this paper we address *reasoning*, broadly conceived. We make no claim to have considered every aspect of this complex construct. Nor do we attempt to consider all of the many issues involved in designing assessments, such as the various policy issues that are beyond the scope of this paper.

The Nature of Construct Description

The measurement framework for this description is based on construct validation theory as articulated by Messick (e.g., Messick, 1975, 1980, 1988, 1989, 1994, 1995) and as represented in the most widely accepted standards for educational and psychological testing, the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999). An essential element of Messick's framework is that it requires consideration of both the nature of the construct itself and of related but distinguishable constructs. It also requires consideration of both the adequacy with which a construct is represented (how much of what is important has been included in a particular test) and the possible threats to validity that may be introduced by the inclusion of material unrelated to the construct (how much of what is included is irrelevant).

Throughout this paper, we have tried to adhere to the meaning of "construct" as defined in the *Standards for Educational and Psychological Testing* as: "the concept or the characteristic that a test is designed to measure" (AERA, APA, & NCME, 1999, p. 173).

A construct provides a target for a particular assessment; it is not synonymous with the test itself. As Messick made clear, "The measure is not equated with the construct nor considered to define the construct" (1975, p. 955). There are many different ways that the same construct could be measured, and the validation process is aimed at establishing how well a particular test has succeeded in measuring its target construct. This point is central to distinguishing the modern view of construct validation from earlier operationalist views, which viewed a test as an entity having no meaning beyond its own characteristics. According to modern views, a clearly specified construct can serve as the fundamental basis for judging the quality of the inferences that can legitimately be drawn about test takers from scores on an assessment. A comprehensive articulation of a construct is thus necessary in order to provide a reference point for addressing a range of important validity and fairness issues about specific assessments. Answers to some of the following questions should be more readily apparent with the aid of such a document:

- Which aspects of the target construct are addressed in a particular assessment and which are not?
- How are similar but distinct constructs assessed?
- How might any changes in an assessment be expected to impact construct coverage and construct-irrelevant factors?
- And, perhaps most importantly, what inferences about an individual can legitimately be drawn from a particular assessment?

Our aim here is to describe a concept of reasoning that might serve as a basis for answering such questions. This specification could in turn be used to guide the development of measure(s) of reasoning as well as research on the extent to which any such measure adequately reflects the concept we have described.

The general strategy for accomplishing our objective has several prongs. First, we attempt, in a very general manner, to situate reasoning within the context of individual abilities. Next, we attempt to explicate the various facets of reasoning by considering (a) general notions/definitions of reasoning and thinking, (b) various theories of individual abilities, (c) surveys of the role of reasoning skills in academic contexts, and (d) ways in which reasoning has been measured. The end product of our search and deliberation is a list of observations that suggest, we believe, possible aspects or facets of a construct of reasoning.

General Definitions

A potentially useful first step in circumscribing a construct of reasoning is to consider various definitions that have been offered for thinking and, especially, for reasoning. Reasoning has been broadly defined by educators, philosophers, and psychologists as, for example:

- A logical thinking process that results in a decision or conclusion (Northrop, 1977, p. 1)
- A process of thought that yields a conclusion from precepts, thoughts, or assertions (Johnson-Laird, 1999, p. 110)
- A process whose goal is to enable the reasoner to make the best decision about a particular issue, in particular circumstances, within a particular forum and enterprise (Toulmin, Rieke, & Janik, 1984, p. 19)
- A way of testing and sifting ideas critically (Toulmin et al., p. 10)
- A means of providing support for our ideas when they are open to challenge and criticism (Toulmin et al., p. 11)
- (More narrowly) as the central activity of presenting the reasons in support of a claim to show how those reasons succeed in giving strength to the claim (Toulmin et al., p. 14)
- A process that enables humans to go beyond the information that is readily available in the environment (Markman & Gentner, 2001, p. 223)
- Inductive reasoning as the process of using the known to predict the unknown (Heit, 2000, p. 569)

By offering these definitions, we do not mean to suggest that the various facets of reasoning (such as induction) are completely distinct and unambiguous. For example, many psychologists tend to view induction as any reasoning that is not deductive. Philosophers, on the other hand, are inclined to view it largely as generalization from cases. Moreover, for philosophers, there are also other kinds of nondeductive reasoning—abduction (or hypothesis generation), for example (C. Tucker, personal communication, September 25, 2001). Some of these distinctions are discussed below.

Connotations

Our next step in identifying the various facets of reasoning will be to consider various adjectives that are often used in conjunction with the terms reasoning and thinking. These have included, for example:

- (For reasoning) formal, informal, everyday, logical, deductive, inductive, abstract, analytical, verbal, quantitative

- (For thinking) higher-order, critical, logical, rational

Consider, for example, the term "abstract." Although preferences for different modes of reasoning may not be universal across all cultures, some (e.g., Stanovich, 1999) have noted a general agreement about one characteristic of good thinking: the importance of decontextualized reasoning styles, i.e., the tendency to evaluate evidence and arguments in a manner that is unaffected by prior beliefs (in contrast to a resistance to reevaluating prior beliefs in light of newly acquired information). This style includes being disposed to reason from starting points with which one may disagree (Norris & Ennis, 1989) and being able to accept statements as true even when they run contrary to one's own beliefs (Zechmeister & Johnson, 1992).

In a related vein, it is also regarded as important to be aware of the fallibility and possibility of bias in one's own opinions (Nickerson, 1987). Stanovich (1999) also notes the importance of skills that enable reasoners to operate independently of world knowledge, prior beliefs, or specific examples that are particularly vivid. His preference is to think of these skills not as abstract problem solving ability, but rather as styles of decontextualizing, abstracting, or disembedding that enable a thinker to decouple content from reasoning processes. Numerous others have also noted the importance of detaching one's own beliefs from the critical evaluation of arguments (e.g., Baron, 1991, 1995; Klaczynski, Gordon, & Fauth, 1997; Voss, Perkins, & Segal, 1991). The tendency to contextualize all information encountered has, in fact, been deemed so noteworthy that it has been called "the fundamental computational bias in human cognition" (Stanovich, 1999, p. 192).

Others (Markman & Gentner, 2001) have stressed the distinction between abstract logical reasoning and newer conceptions of reasoning that emphasize the importance of domain-specific reasoning, i.e., reasoning that is knowledge-based. They characterize reasoning as a mix of domain-specific and domain-general aspects. Markman and Gentner (2001) cite Newell and Simon's (1972) distinction between strong and weak methods of reasoning, with the former making intensive use of knowledge (as in reasoning by cases) and the latter consisting of general strategies (e.g., rules of logic) that can operate without any particular regard for content. Because of their generality, weak methods are often useful in novel or knowledge-poor environments. The stronger methods utilize mental simulations, analogies, or various kinds of induction.

In their discussion of arguments, Toulmin, Rieke, and Janik (1984) also emphasize the difference between universal or field invariant rules of procedure that apply to rational criticism in all fields and those that are field-dependent (law, science, etc.). They note that arguments may vary by field (law, medicine, politics, for example) in terms of the degree of formality, the degree of precision, the modes of resolution, and even the purpose of arguing.

Consistent with this view, Pithers and Soden (2000) cite research (Biggs & Collins, 1982; Laurillard, 1993; Ramsden, 1992) that supports the notion that learning to think involves "learning to use content in successively more sophisticated ways in understanding the world" (p. 241). Others (Bereiter & Scardamalia, 1993; Chi, Glaser, & Farr, 1988) have also noted the inextricable link between content and reasoning ("good knowledge and good thinking").

Several researchers have distinguished between formal and informal (or everyday) reasoning. In their factor analytic studies, Emmerich, Enright, Rock, and Tucker (1991) and Enright, Tucker, and Katz (1995) found evidence of two major subdomains of reasoning, which they termed informal reasoning and formal-deductive reasoning. The first corresponds roughly, in Carroll's (1993) scheme, to induction, and the latter to sequential reasoning. Enright et al. noted that this distinction had been made previously in the field of education (Voss, Perkins, & Segal, 1991), in philosophy (Toulmin, 1958), and in cognitive psychology (Galotti, 1989). A primary distinction between the two is that, for formal reasoning, all of the needed information is usually set forth in the problem; for informal reasoning, a search for information is required initially, or a determination is needed of what is relevant to the problem at hand. Formal reasoning is concerned with the validity and logical consistency of arguments. Informal reasoning, too, is concerned with validity and logical consistency, but also with relevance, cogency, strength of argumentative connections, and consistency with a body of background knowledge.

As with reasoning, thinking is also often preceded by explicit modifiers. For instance, the term critical is often used in conjunction with thinking in order to describe a particular set of skills. Critical thinking has been discussed extensively in the literature, and, in fact, there are several taxonomies of critical thinking skills in the educational and psychological literature. For example, emphasizing the importance of imparting these skills to students, Halpern (1996) has offered the following eight categories of skills:

- Thought and language (comprehending and defending against the persuasive techniques that are embedded in everyday language)
- Deductive reasoning (determining if a conclusion is valid)
- Argument analysis (judging how well reasons and evidence support a conclusion)
- Hypothesis-testing (collecting and using information to confirm or disconfirm hypotheses)
- Likelihood and uncertainty (correctly using objective and subjective estimates of probability)
- Decision-making (framing decisions, generating and evaluating alternatives, and analyzing the outcomes)
- Problem-solving (identifying a problem, stating the goal, and generating and evaluating possible solutions)
- Creative thinking (increasing the flow of relevant ideas, which should increase the likelihood of novel responses)

Ennis (1962) asserted that research on thinking had been concerned mainly with such aspects as associative thinking, concept formation, problem solving, and creative thinking, but that little attention had been devoted to critical thinking, which he characterized as "the correct assessing of statements." According to Ennis, there are 12 overlapping aspects of critical thinking. The first entails grasping the meaning of a statement. The others include judgments regarding whether (a) there is ambiguity in a line of reasoning, (b) certain statements contradict each other, (c) a conclusion necessarily follows, (d) a statement is specific enough, (e) a statement is actually the application of a certain principle, (f) an observation statement is reliable, (g) an inductive conclusion is warranted, (h) the problem has been identified, (i) something is an assumption, (j) a definition is adequate, and (k) a statement made by an alleged authority is acceptable.

Ennis (1987) also offered a taxonomy of critical thinking dispositions and abilities. Ennis characterized the term *critical thinking* as being a more useful term than *reasoning* for the kinds of skills specified by the College Board as needed in order to be prepared for college. His taxonomy of abilities included:

- Focusing on a problem
- Analyzing arguments

- Asking and answering clarifying or challenging questions
- Judging the credibility of a source
- Observing and judging observation reports
- Deducing and judging deductions
- Inducing and judging inductions
- Making value judgments
- Defining terms and judging definitions
- Identifying assumptions
- Deciding on an action

Finally, a word is in order here about two additional adjectives—verbal and quantitative—that are often used to modify reasoning. (Verbal reasoning and quantitative reasoning are the subjects of separate papers in this series.) The view taken here is that, on one level, these are distinct facets of a broader construct of reasoning that we attempt to elaborate in this paper. Verbal and quantitative reasoning are themselves, however, so complex and multifaceted that each can also be treated as a separate construct of its own, as has been done in related papers in this series.

Theoretical and Empirically-based Schemes

Other theories or classification schemes have also described reasoning or thinking in several of the terms mentioned at the outset. For instance, Sternberg (1985) has made a very high-level distinction—a triarchic theory—in which he has posited three broad kinds of abilities: practical, creative, and analytical. For Sternberg, practical abilities involve the application of intelligence to real-world contexts in order to implement solutions. Creative abilities involve the generation of problem-solving options. Analytical abilities are described as involving analysis and evaluation—identifying and defining problems, setting up strategies to solve them, and monitoring solution processes. To Sternberg, all of the abilities (e.g., verbal, quantitative, and analytical) measured by academic admissions tests (such as the GRE® General Test, the SAT® I Reasoning Test, or the GMAT®) are regarded as being of the third kind, i.e., analytical (Sternberg & Williams, 1997).

There are also a variety of finer-grained theories, and there is ample evidence from more than a century of research on individual abilities to support these various theories of human abilities. Much of this research has been summarized and reanalyzed by Carroll (1993), whose conclusions generally agree with those of other prominent researchers in the field.

Lohman and Hagen (2001) have provided a good discussion of some of the key research in this area. They note, for example, that Carroll's first important finding is that individual abilities are organized hierarchically, suggesting that some abilities are useful more broadly than others. Carroll's (1993) hierarchy starts with general mental ability (G) at the topmost level. Eight broad, group factors, which vary in their closeness or association with G, define the second level. The most "G-like" is a factor that Cattell (1963) termed general fluid ability (Gf). Other factors at this level include general verbal crystallized ability (Gc), general spatial visualization ability (Gv), and general memory ability (Gm). A third level is defined by an even greater number of primary factors, such as verbal comprehension, verbal fluency, inductive reasoning, spatial visualization, perceptual speed, and number facility.

Lohman and Hagen (2001) also suggest a second critical finding from the literature—that G is virtually synonymous with Gf at the second level and that Gf, in turn, is virtually synonymous with the primary factor called *reasoning*. Although others might describe these factors as being less than strictly equivalent, there is at least some sense that each of these reasoning abilities is central to cognitive competence.

Based on an extensive survey of factor analytic studies of cognitive abilities, Carroll (1993) specified nine major ability domains: language, memory and learning, visual perception, auditory perception, idea production, cognitive speed, knowledge and achievement, psychomotor abilities, and reasoning. He also listed a number of miscellaneous domains of ability and personal characteristics. With respect to reasoning, Carroll concluded that there is evidence for three higher-order reasoning factors: sequential reasoning, induction, and quantitative reasoning.

For Carroll (1993), *sequential reasoning* is deductive in nature, requiring reasoning, often in a series of two or more sequential steps, from premises, rules, or conditions to reach a conclusion that logically follows. Test stimuli can be of almost any kind—verbal, numerical, figural, or pictorial, for example. Carroll specifies three subtypes of sequential reasoning: syllogistic reasoning that involves class membership, syllogistic reasoning involving comparisons of entities on attributes that can vary continuously, and verbal reasoning that entails making inferences or drawing or evaluating possible conclusions on the basis of conditions or rules that are specified. As Tucker has pointed out (C. Tucker, personal communication, March

12, 2001), this list of sequential reasoning processes is a truncated one that represents, primarily, the interests of psychological researchers and does not include all of the formal-logic processes that have been of interest to symbolic logicians.

Induction involves the discovery, from a class of stimulus materials, a common characteristic—a rule, concept, process, trend, principle, or causal relationship, for example—that underlies the materials. Several subtypes are specified under this category: discovering concepts or rules in stimulus materials, which may involve sets, series, matrices, or analogies, for instance. Carroll points out that analogies may involve both inductive and deductive processes.

Quantitative reasoning involves reasoning based on mathematical properties and relations. It can be either deductive or inductive in nature, or both. Tests measuring this factor have involved, for instance, word problems and number series and problems that require the selection of appropriate mathematical operations. Numerical computation is considered to be largely extraneous to this factor. Carroll suggests that both sequential reasoning and induction can be given a quantitative emphasis by making the mathematical relations or attributes more or less difficult.

Processes Involved in Reasoning

As Lohman and Hagen (2001) have noted, reasoning can occur at several different levels—for example, either tacitly or intentionally/explicitly (Stanovich, 1999). Tacit processes facilitate reasoning without conscious intervention or awareness coming into play when decisions are made quickly or intuitively, rather than on the basis of explicitly articulated reasons. (On the other hand, intentional reasoning processes occur consciously and typically enable reasoners to use principles of reasoning rather than rely solely on accumulating experiences.) Tacit processes help to *focus attention*: Effective problem solvers know what to look for and what to disregard. Moreover, they typically focus on different problem features than do less effective counterparts. Thus, knowing on what to focus one's reasoning skills is an important facet of the reasoning process. Tacit processes are also associated with emotions, which play a key role in solving everyday problems of various sorts.

As Lohman and Hagen (2001) note, the kinds of mental processes used in reasoning tasks have been classified in another way also—as *selective encoding*, *selective comparison*, and *selective combination* (Sternberg, 1985). A test need not, however, elicit these same processes for every test taker.

Selective encoding, i.e., the process of distinguishing relevant from irrelevant information can be either deliberate or relatively automatic. Learning what to notice and what to ignore is an essential step in reasoning about any problem.

S*elective comparison* involves retrieving and comparing information that is potentially relevant for reasoning. Deciding how best to describe the relationships among concepts is another critical aspect of reasoning.

Finally, a third category of reasoning processes involves the *strategic combination* of information in working memory. This process is often required on tasks such as formulating an argument or a mathematical proof.

As Lohman and Hagen (2001) suggest, these are some of the cognitive processes that are implicated in effective reasoning. Different types of reasoning tasks call on these processes in differing degrees. For example, whereas inductive reasoning tasks elicit encoding and comparison, deductive reasoning tasks tend to require strategic combination.

Criteria and Standards for Reasoning

Arguably, any definition of reasoning will be facilitated by standards and criteria for judging the degree to which reasoning is sound, convincing, or otherwise "good." For deductive reasoning, there are logical rules to guide this determination, but deductive provability should not be regarded as being synonymous with objectivity. Even for deductive reasoning, there is less than universal agreement on what constitutes a logical error. Nor is there necessarily a consensus on which axioms or principles, rules of inference, or levels of abstraction should be accepted in our deductive systems (C. Tucker, personal communication, September 25, 2001). For nondeductive reasoning, the criteria for what constitutes good reasoning seem somewhat less clear. Nondeductive reasoning is not, however, merely guessing on the basis of subjective preferences. Objectivity is achieved by marshalling evidence, considering alternatives, and applying the procedures developed for judging arguments in various fields.

There is much evidence that people do not always behave according to normative models of rational behavior (e.g., Tversky & Kahneman ,1988). Stanovich (1999) has summarized a variety of gaps that have been identified between these normative models and how people actually behave. For instance, people:

- Assess probabilities incorrectly
- Tend to look for confirmatory rather than disconfirmatory evidence
- Test hypotheses inefficiently
- Violate axioms of utility theory (i.e., the theory of rational choice)
- Calibrate degrees of belief incorrectly
- Overproject their own opinions onto others
- Allow prior knowledge to become implicated in deductive reasoning
- Systematically underweight information about nonoccurrence when evaluating covariation
- Display numerous other information-processing biases (p. 2)

One can argue, however, that not all of these behaviors are entirely irrational. For example, utility theory may be incorrect or at least incomplete. And the use of prior knowledge may at times be entirely appropriate in deductive reasoning.

With respect to argumentation, Toulmin et al. (1984) note that certain ways of construing arguments are recognized as sound and widely accepted across a range of fields. So, too, certain modes of argumentation have been traditionally recognized as *un*sound, with fallacies falling into five broad types as a result of (a) missing grounds (e.g., begging the question), (b) irrelevant grounds (red herring, straw man), (c) defective grounds (hasty generalization), (d) unwarranted assumptions, and (e) ambiguities.

For informal or inductive reasoning, however, the criteria seem somewhat less clear than for formal or deductive reasoning. They depend on rather more subjective criteria ("inductive strength," for example). The more important instances of induction involve generalization from some but not all possible cases to a general statement. In evaluating this kind of reasoning, the proper focus may be less on the form of the reasoning than on the various factors that may influence the strength of connections. As Heit (2000) has pointed out, some philosophers have argued that there is no concrete basis whatsoever for establishing the validity of inductive methods. This seems true at least to the extent that nothing can ever be known with complete certainty.

Studies of the Importance of Reasoning in Academic Contexts

Success in college and in post-baccalaureate settings depends on many factors: academic knowledge and skill, motivation, perseverance, the ability to manage time and resources, personal skill in developing relationships with other students, study strategies, note-taking strategies, financial support, and so forth. But of the many things that matter, two of the most important, we believe, are: (a) academic knowledge and skill in the domain of study, and (b) the ability to reason well in the symbol systems used to communicate new knowledge. Reasoning tests correlate with academic success because reasoning abilities are very often required in school learning, whether for understanding a story, inferring the meaning of an unfamiliar word, detecting patterns and regularities in information, going beyond the information given to form more general rules or principles, or applying mathematical concepts to solve a problem. In these ways and in hundreds of others, successful learning requires reasoning strategies.

Some research is consistent with this notion. For instance, Enright and Gitomer (1989) attempted to specify the traits that characterized successful graduate students. By consulting with eminent psychologists and distinguished graduate faculty from several fields, they identified seven general competencies that seemed to differentiate more successful graduate students from less successful ones. These qualities were creativity, motivation, communication, planning, professionalism, synthesis, and explanation. Underlying several of these competencies were skills related to reasoning. For instance, communication may involve reasoning from differing viewpoints, and creativity may entail recognizing that facts, concepts, and theories can be subjected to criticism, revision, modification, and reinterpretation. Explanation typically involves giving a reason or suggesting a cause for a phenomenon or finding and evaluating competing hypotheses and developing arguments to support explanations. Planning may involve deciding what evidence is needed to answer a question and anticipating possible problems or criticisms. Finally, synthesis entails the organization and restructuring of knowledge (not simply the accumulation of facts) in order to make inferences.

These kinds of reasoning skills have been explicated elsewhere as well. In an effort to expand the GRE analytical measure, Tucker (1985) composed a list of reasoning processes, which was reviewed by cognitive psychologists, philosophers, and ETS test developers. Reviewers were asked to comment on the list, to suggest additional skills, and to rate the processes in terms of importance. The final list included both the kinds of skills involved in *reasoning in context* in particular as well as the more general processes underlying complex reasoning more broadly. The major procedures involved in reasoning in context were:

- Constructing theories, explanations, and interpretations (e.g., noticing what is relevant and important, deriving relevant concepts, formulating alternative possibilities or explanations, incorporating isolated instances into a wider framework, and reducing a complex situation to a simpler model)
- Evaluating theories, explanations, or interpretations (e.g., relating a theory to its implications or applications, determining which parts of a theory are weakest or most vulnerable to objection, and revising flawed theories, concepts, or previously held frameworks)
- Generating and transforming problems (e.g., formulating a problem that expresses a key issue in the field, dividing a vaguely articulated problem into subproblems, deepening a problem statement by reexamining the presuppositions on which it rests, relating problem statements and methods in one area to those in another, and identifying a new method to approach an old problem)
- Following good methods (e.g., developing a comprehensive plan taking into account a variety of factors and monitoring one's own adherence to generally accepted methods in a field)

Carrying out these relatively complex processes can involve processes that are both more general and smaller in scale. Examples of these are:

- Analyzing and criticizing arguments
- Reasoning deductively, i.e., drawing logically necessary conclusions
- Analyzing the relationship of means to goals

In a related attempt to specify the reasoning skills required for graduate study, Powers and Enright (1987) surveyed graduate faculty in six disciplines (education, English, engineering, chemistry, computer science, and psychology) to suggest (and later to rate) the analytical or reasoning skills that were most important for successful performance in graduate school. A factor analysis of the ratings provided for a relatively large number of skills revealed five reasoning factors:

- Critical thinking related to argumentation (e.g., being able to understand, analyze, and evaluate arguments)

- Critical thinking related to the drawing of conclusions (e.g., generating valid explanations to account for observations)
- Defining, setting up, and analyzing problems (e.g., breaking complex problems into simpler ones)
- Inductive reasoning (e.g., recognizing structural similarities between one type of problem and another)
- Generating alternatives (e.g., finding alternative explanations)

Powers and Enright (1987) found some differences among disciplines in the importance that they attached to various individual skills, but some skills were rated consistently as at least moderately important in each discipline. They were:

- Reasoning or problem solving in situations in which all of the needed information is not known
- Detecting fallacies and logical contradictions in arguments
- Deducing new information from relationships
- Recognizing structural similarities between one type of problem or theory and another
- Taking principles from one area and applying them to another
- Monitoring one's own progress in solving problems
- Deriving from the study of single cases structural features or principles that can be applied to other cases
- Making explicit all relevant components in a chain of logical reasoning
- Testing the validity of an argument by searching for counterexamples

Related to standards or flaws, faculty also rated a variety of critical incidents (i.e., specific incidents that caused faculty to either raise or lower their estimation of a student's analytical ability). The incidents rated consistently as at least moderately serious were:

- Accepting the central assumptions in an argument without questioning them
- Being unable to integrate and synthesize ideas from various sources
- Being unable to generate hypotheses independently
- Being unable to see a pattern in results or to generalize when appropriate
- Ignoring details that contradict an expected or desired result

- Submitting a paper that failed to address the assigned issues
- Basing conclusions on analysis of only part of a text or data set

From the preceding review, potential facets of a construct of reasoning begin to emerge. The focus will become even clearer in the next section, we hope, as we selectively review some of the ways in which reasoning has been measured.

Assessing Reasoning

On one level, all tests are, in some sense, measures of reasoning, as they require test takers to make decisions, draw inferences, or formulate conclusions in order to select or produce answers to test questions. Postsecondary assessments such as the SAT I: Reasoning Test, the Graduate Record Examinations (GRE) General Test, the Graduate Management Admission Test (GMAT), and the Law School Admission Test (LSAT), for example, are all described by their sponsors as measures of verbal and/or mathematical reasoning abilities. Each of these assessments employs a variety of test item types to measure reasoning, among them reading comprehension questions that require inferences about reading passages, mathematical problems that require, for example, determining whether or not data are sufficient for solving a problem, and items that require the analysis of arguments and the adequacy of assumptions on which an argument depends. The GRE General Test and the GMAT also both include a measure of analytical writing.

The history of one particular test (the GRE General Test) is especially revealing and will be discussed in more detail below. It is, to our knowledge, the only major national test that has offered measures of verbal and quantitative reasoning, as well as a separate measure of *analytical ability*. It is interesting to note that both the LSAT and the GMAT have included some of the same analytical test item types that have constituted the GRE analytical measure, but they have been included with other verbal and/or quantitative item types. This suggests, perhaps, the close relationship among reasoning in verbal, quantitative, and other domains. Because, however, the GRE program has regarded it appropriate to distinguish analytical ability from verbal and quantitative ability, it is informative to consider the development and history of the GRE General Test's analytical measure.

The GRE General Test

Until the fall of 2002, the GRE General Test included, besides indicators of verbal and quantitative abilities, a measure of analytical ability. (The analytical ability measure was replaced by a measure of analytical writing.) The term *analytical* (instead of *reasoning*, for example) was chosen to describe the measure mainly so as not to imply that the verbal and quantitative portions of the test were not also measures of reasoning (C. Tucker, personal communication, September 25, 2001). The content of the measure, however, was generally consistent with traditional connotations of the word.

Analysis has been described as "one of the most incessantly performed of all our mental processes" (W. James, *Principles of Psychology*, as quoted in the *Oxford English Dictionary*, 1989.) The *Oxford English Dictionary* (1989) discusses *analysis* and its derivatives (in contrast to *synthesis*) as breaking up anything complex into its various simple elements; as judgments that "unfold the contents" of a concept; as dissecting, decomposing, or separating into parts; as determining the elements or components of anything complex.

Analysis is also described in various disciplines as (a) resolving into elements or constituent parts (chemistry and optics), (b) proving propositions by rendering them into simpler, already proven ones (mathematics), (c) clarifying (by reformulating) existing concepts and knowledge (philosophy), (d) critically describing so as to make clear a structure (music), (e) critically examining so as to bring out essential elements (literature), and (f) tracing things to their source, proceeding from the more to the less determinate, and discovering general principles underlying concrete phenomena (logic). While these descriptions are germane to describing a construct of reasoning, we believe they are also insufficient: It is clear that reasoning may also involve a great deal of synthesis. As Carol Tucker (personal communication, September 25, 2001) has pointed out to us, synthesis occurs, for instance, when we draw conclusions, generate explanations, or incorporate alternative considerations into a point of view. Moreover, analysis may involve creative processes as well, as when we envision other relevant possibilities during the evaluation of an argument.

To further zero in on what was meant by *analytical ability* in the GRE General Test, it may be useful to consider briefly the history of the measure. The development of the test began with an attempt to broaden the definition of academic talent beyond the verbal and quantitative domains. Thus, the primary contrast considered early on was the distinction among verbal, quantitative, and analytical abilities. The decision to develop a measure of analytical ability, instead of a measure of some other important trait, followed from a survey of graduate faculty, which revealed that, of several possible areas, the measurement of *abstract reasoning* or *critical thinking* was of most interest to graduate faculty. Essentially then, the analytical ability measure arose from the acknowledgement that verbal and quantitative abilities—the abilities assessed most often in academic admissions—are not the whole of the abilities (or kinds of reasoning) that are implicated in successful performance in graduate school. As mentioned earlier, the eventual naming of the test section as *analytical ability* appears to have been somewhat arbitrary, as various alternative names were often used when the test was being conceptualized. Among them were *abstract reasoning*, *logical reasoning*, and *critical reasoning*. For the purpose of test development, logical and critical thinking were viewed as two aspects of the same ability, with both widely viewed as being concomitant to good scholarship. In unpublished GRE program documents deliberating the development of the analytical measure, the ability to reason logically was defined generally as "the orderliness of thinking which makes it possible to recognize relationships between statements, to organize statements according to their interdependencies, to infer probable implications of certain statements, and to discern analogies between arguments." The ability to reason critically was defined as "the ability to approach claims, evidence, and persuasive appeals with a skepticism that is sufficient to detect flaws in reasoning, unsound use of evidence, and unfair arguments."

The original version of the analytical ability measure employed three item types—analysis of explanations, logical diagrams, and analytical reasoning. The analysis of explanations item type was seen as tapping the kinds of analysis used in the humanities and social sciences, and the logical diagrams and analytical reasoning types (15 of each type) were regarded as relating more closely to the kinds of analysis required in the sciences. In total, these questions were designed to tap into students' abilities to recognize logical relationships (for example, between evidence and a hypothesis, between premises and a conclusion, or between stated facts and explanations); to judge the consistency of interrelated statements; to draw conclusions from a complex series of statements; to use a sequential procedure to eliminate incorrect choices in order to reach a solution; to make inferences from statements expressing relationships among abstract entities such as nonverbal or nonnumerical symbols; and to determine relationships between independent or interdependent categories or groups. The directions and descriptions that accompanied the item types suggested that "the skills are likely to have been learned in a variety of contexts and in academic study of most kinds" and that common sense, rather than knowledge of formal systems of logic, is presupposed (Conrad, Trismen, & Miller, 1977).

The stated rationale for the analysis of explanations item type was that it tested reasoning in context. It was based on C. S. Peirce's (1901/1955) notion that a person begins to think actively when he/she confronts a perceived problem or difficulty. Because an event is usually not perceived as being problematical if it is seen as being quite natural and expected, this item type presents a result that seems somewhat surprising in view of the situation. The result is related to the situation, but it is not entirely explicable in terms of the causal network presented in the situation. The situation is constructed so as to allow more than one causal explanation. An examinee must explore the implications of the facts presented, construct possible explanations that are consistent with those facts, and search for new facts that confirm or disconfirm those explanations. The item type therefore required the test taker to make judgments about the logical consistency of statements, as well as the relevance and explanatory power of various statements. Each statement was to be evaluated in relation to the situation and the (unexpected) result, i.e., the extent to which each statement fills the gap between the given situation and the result. In order to test reasoning skills primarily, instead of verbal skills, the situations were described relatively succinctly with fairly simple language. The main intention was to pose a task that tested the "controlled generation" of hypotheses in an investigative setting (rather than testing fluency without control). The task was thought to be especially relevant to graduate education because it was situated at the investigative stage, at which alternative hypotheses are being considered and factual evidence is being sought and sifted (rather than at the stage of drawing conclusions).

The remaining item type, logical diagrams, required examinees to select a Venn diagram that best represented the logical relationship among independent or interdependent categories or groups (e.g., cows, birds, and mammals). Thus, this item type was thought to tap another aspect of reasoning ability that involves classification or categorization.

In 1980, the analysis of explanations and the logical diagrams item types were removed from the test because they were deemed to be unduly susceptible to improvement by short-term preparation and practice. The most recent version of the analytical ability measure, which was composed of two item types (analytical reasoning and logical reasoning), was described as testing "the ability to understand structured sets of relationships, deduce new information from sets of relationships, analyze and evaluate arguments, identify central issues and hypotheses, draw sound inferences, and identify plausible causal explanations" (Educational Testing Service, 2000, p.5). The last version of the analytical measure was narrower than the original version insofar as it contained only two item types: logical reasoning and analytical reasoning. The latter item type, the more prominent of the two types, measures deductive reasoning skills, i.e., the ability to draw logically necessary conclusions. But as Tucker (1985) has pointed out, reasoning deductively is only a part of what graduate students are called upon to do. This kind of deductive reasoning ability, along with the kind of analysis and criticism involved in small-scale argumentation, are seen as subordinate skills that come into play when thinkers readjust the ideas they hold when considering theories or other larger-scale systems.

The logical reasoning item type was viewed as having much in common with the informal reasoning that occurs in science, law, and daily life. Logical reasoning items entail four distinctly different inferential tasks:

- Determining what conclusion or explanatory principle follows from the argument in a text (e.g., "If the given statements are true, which of the following must also be true?")
- Determining a *missing* premise of an argument (e.g., "To evaluate the conclusion as drawn, it would be most important to know the answer to which of the following questions?")
- Identifying weaknesses in an argument (e.g., "Which of the following, if true, most seriously weakens support for the conclusion?")
- Interrelating one or more parts of an argument

Assessing Analytical Writing at the Postsecondary Level

Thinking or reasoning skills can also be tapped by constructed-response type questions. For example, both the GMAT and GRE programs offer highly similar measures of analytical writing skills that assess both thinking and writing skills. An examination of these measures is instructive regarding these two programs' intentions regarding the assessment of reasoning ability. Although a main focus is on assessing writing skill, both the GRE and GMAT Writing

Assessments seem as much measures of analytical thinking/reasoning as they do tests of the ability to write well. There is not necessarily a consensus, however, regarding what kind of reasoning is measured by these writing assessments and to what degree they measure it (M. Bell, personal communication, February 6, 2003).

Both assessments require examinees to complete two writing tasks—one called Issue and the other Argument. For the Issue task, test takers are asked, in response to an opinion on an issue of broad interest, to discuss the issue from any perspective they wish and to provide relevant reasons and examples to explain and support their views. According to the GRE Web site, for example, the Issue task is "an exercise in critical thinking and persuasive writing" that is designed to see how well-equipped examinees are to develop a compelling argument supporting their own perspectives on an issue and to effectively communicate that argument in writing to an academic audience. Test takers are advised that when writing their Issue essay they might employ any or all of the following strategies:

- Agree or disagree, either in whole or in part, with the claim being made
- Question the assumptions that the claim or opinion seems to be making
- Qualify any of its terms, especially if the way they define terms is important to developing their perspective on the issue
- Point out why the claim may be valid in some situations but not others
- Evaluate points of view that contrast with their perspective
- Develop their own positions with reasons that are supported by one or more relevant examples

Prospective test takers are also told that when they prepare for the Issue task, they should (a) read the claim carefully, making certain that they understand the issue involved and (b) think about the issue in relation to their own ideas and experiences, to events they have read about or observed, and to people they have known, as this is the knowledge base from which they will develop reasons and examples for their argument in order to reinforce, negate, or qualify the claim. Examinees' responses are evaluated, in part, on the extent to which they adequately develop a position and support it with reasons and examples and on the degree to which their discussion is focused and organized. Because test takers are allowed some latitude with respect to the strategies they employ in crafting their essays (for example, they are not *required* to consider opposing points of view), the GRE issue task may require advocating a position more so than it does critical thinking, at least in the sense that some philosophers have defined the term.

In contrast, the Argument task does require examinees to critique a given argument by discussing how well-reasoned they find it. The task is to comment on the thinking that underlies the position that is expressed or the argument that is made by another writer. In doing so, the test taker can demonstrate the perceptive reading, critical thinking, and analytical writing skills that university faculty consider important for success in graduate school. Examinees are told that the task ("primarily a critical thinking task presented in an essay format") is designed to assess "your ability to understand, analyze, and evaluate an argument and to clearly convey your critique in writing." They are advised that, when reading the argument, they should pay special attention to what is offered as evidence, support, or proof, to what is explicitly stated, claimed, or concluded, to what is assumed or supposed, and to what is not stated, but which necessarily follows from what is stated. Furthermore, test takers are encouraged to consider the structure of the argument or the ways in which its elements are linked to provide a line of reasoning and also whether the linkages are logically sound. They are reminded, however, that they are not being asked to judge the truth or accuracy of the statements in the argument, but rather whether the inferences and conclusions are validly drawn. Examinees are told that, as they prepare for the test, they might use the following strategies:

- Carefully read the argument, identify as many of its claims, conclusions, and assumptions as possible
- Think of alternative explanations and counterexamples
- Think of additional evidence that might either weaken or lend support to the claims
- Think what changes in the argument would make the reasoning more sound

Responses are evaluated, in part, with respect to how well the features of the argument are identified and analyzed and how well ideas are developed, organized, and connected. The two tasks are thought to complement each other, insofar as the Issue task requires test takers to take a position and provide supporting evidence, thus constructing their own arguments. The Argument task requires test takers to evaluate *someone else's* argument by assessing its claims and the evidence it provides.

Summary

How does all of the preceding help us specify a construct of reasoning? First, it will be helpful to keep in mind the target population for the assessment. Our discussion has focused mainly on postsecondary assessments. For the GRE program, for instance, the target is prospective graduate students coming from a wide variety of undergraduate disciplines and intending to enter at least as diverse a set of graduate disciplines. This suggests, on one hand, that a test of analytical abilities should be "general" in nature, emphasizing reasoning abilities that are not specific to particular areas of study, but rather are important in many fields. On the other hand, as pointed out by Ver Steeg (1984), for instance, a defining characteristic of graduate education is the following: Whereas undergraduate students are required primarily to master a body of knowledge, graduate students must typically take an extra step by making an original contribution to the field. Besides mastering the knowledge, modes of thought, and techniques of inquiry that are appropriate to one's field, graduate students are typically required to apply these skills to a research problem that advances the discipline. It is during this application that many of the reasoning skills discussed above come into play.

In any event, it seems that the following elements should be considered in any attempt to specify the facets of a construct of reasoning ability:

General Distinctions

- An emphasis on critical facility, judgment, and evaluation
- A recognition that the concern is, primarily, with *analysis*, that is the breaking down into components, etc., not on synthesizing or creating (but see next point)
- An acknowledgement that even though analysis may be a focus, reasoning may also involve an element of creativity, as evidenced, for example, by the ability to construct arguments, to specify counterexamples, and to formulate alternatives
- A differentiation between (a) formal, deductive reasoning based on rules of logic and (b) informal, everyday reasoning with verbal or quantitative material

Processes

- An emphasis on argumentation—either on constructing arguments or on evaluating them

- An awareness of the role of perception—being able to notice, to attend to what is relevant, to see commonalities, similarities, and differences
- A recognition of the ability to *abstract*, i.e., (when appropriate) to decouple knowledge and prior beliefs from the reasoning process
- The ability to *construct* sound arguments vs. the ability to *evaluate* or recognize the weaknesses in the arguments of others

Context

- A distinction between modes or methods of reasoning that are relevant in most or all disciplines vs. those that are discipline-specific
- A sensitivity to the notion that reasoning is manifested in a wide variety of different kinds of contexts, content, and tasks (e.g., verbal reasoning as manifested in comprehending text, quantitative reasoning as manifested in solving mathematical problems)
- An acknowledgement that reasoning underlies a number of the skills or abilities that are deemed important for success in various academic contexts and that it manifests itself in various situations within these contexts

Content

- A recognition of the difference between reasoning when virtually all of the required information is known vs. reasoning when all the needed information and conditions have *not* been specified
- A distinction between reasoning that is heavily dependent on content or prior knowledge and reasoning that is relatively independent of familiarity with specific content
- A differentiation between reasoning with *familiar* content or material vs. reasoning with material that is *novel* or unfamiliar

Standards

- A recognition that for formal reasoning processes there are exact, agreed-upon rules by which conclusions follow necessarily from premises

- An acknowledgment that the standards for informal reasoning pertain to the quality of evidence, the consideration of alternative explanations, and how cogently arguments and their connections are presented
- A distinction between degrees of uncertainty and logical necessity as facets of reasoning
- An awareness of alternative criteria for evaluating reasoning—consistency, contradiction, logical necessity, relevance, cogency, reasonableness, explanatory power, etc.

Outcomes

- A recognition of the *product* of reasoning as a conclusion, decision, solution, diagnosis, determination, judgment, point of view, explanation, etc.

Potential Sources of Extraneous Variance

In attempts to operationalize the measurement of these facets of reasoning, it will be necessary to exercise due diligence to ensure that extraneous factors do not impede measurement of the construct of interest. For instance, in contrast to the measurement of some other kinds of knowledge and abilities (reading comprehension and quantitative problem solving, for example), the measurement of reasoning ability may sometimes require novel, relatively complex, and therefore unfamiliar test item formats. Unless a significant effort is taken to ensure that all test takers are well-versed with item formats and directions, a potential exists for test performance to be unduly influenced by lack of familiarity with the test.

Similarly, some aspects of reasoning (those having to do with the analysis and evaluation of arguments, for example) will necessarily be tested in a verbal context. Care may be needed therefore to ensure that performance does not depend unduly on examinees' previous familiarity with the same kinds of content or on their ability to "merely" comprehend the stimulus materials.

Finally, careful thought will be needed with respect to choosing the facets of reasoning to be measured in order to minimize the chances that some examinees are not unfairly disadvantaged because the test unduly privileges modes of reasoning that are less prevalent in some cultures than in others. For example, some research has shown that Asians may be more inclined to think holistically instead of analytically when compared to Westerners (Nisbett, Peng, Choi, & Norenzayan, 2001). And cultures that differ with respect to their emphasis on context may also exhibit differences in their preferred modes of thinking (e.g., inductive vs. deductive) (Ibarra & Cohen, 1999).

Concluding Statement

However defined, reasoning is perhaps the most general of all cognitive abilities and, arguably, the most central to academic learning (D. Lohman, personal communication, August 2001). As befitting such an important construct, there have been many attempts to articulate its (multifaceted) nature, some of which have been discussed here. However, given the variety of ways in which reasoning has been characterized, we are unable to offer a single definition or comprehensive vision of a construct of reasoning. Rather, we have attempted only to set forth some of the salient features of reasoning that, to us, seem worthy of consideration, and possibly of incorporation, in any construct statement of reasoning. Our main conclusion then is that there is no *single* construct of reasoning. Instead, any of several formulations may be equally useful and informative depending on the context and purpose. We hope though that our reflections have provided the interested reader with some of the "grist" that may be needed for developing a construct statement that is tailored to his or her own particular needs.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington DC: American Educational Research Association.

Baron, J. (1991). Beliefs about thinking. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp.169-186). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and reasoning, 1,* 221-235.

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

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

Chi, M., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Conrad, L., Trismen, D., & Miller, R. (Eds.). (1977). *Graduate Record Examinations technical manual.* Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2000). *GRE 2000-2001. Guide to the use of scores.* Princeton, NJ: Author.

Emmerich, W., Enright, M. K., Rock, D. A., & Tucker, C. (1991). The development, investigation, and evaluation of new item types for the GRE analytical measure (GRE Board Professional Report No. 87-09P, ETS RR-91-16). Princeton, NJ: Educational Testing Service.

Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review, 32,* 81-111.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). New York: W. H. Freeman and Company.

Enright, M. K., & Gitomer, D. (1989). *Toward a description of successful graduate students* (GRE Board Research Report No. 85-17R, ETS RR-89-09). Princeton, NJ: Educational Testing Service.

Enright, M. K., Tucker, C. B., & Katz, I. R. (1995). *A cognitive analysis of solutions for verbal, informal, and formal-deductive reasoning problems* (GRE Board Professional Report No. 90-04P, ETS RR-95-06). Princeton, NJ: Educational Testing Service.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105,* 331-351.

Halpern, D. F. (1996). *Thinking critically about critical thinking.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7,* 569-592.

Ibarra, R. A., & Cohen, A. S. (1999). Multicontextuality: A hidden dimension in testing and assessment. In *GRE, FAME Report Series: Vol 3. New directions in assessment for higher education: Fairness, access, multiculturalism, & equity* (pp. 16-30). Princeton, NJ: Educational Testing Service.

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

Klaczynski, P. A., Gordon, D. H., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology, 89,* 470-485.

Lohman, D. F., & Hagen, E. (2001). *The cognitive abilities test: Form 6. Interpretive guide for teachers and counselors.* Itasca, IL: Riverside Publishing.

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

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30*, 955-966.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35,* 1012-1027.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & I. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Erlbaum.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher, 23*(2), 13-23.

Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Nickerson, R. (1987). Why teach thinking? In J. Baron, & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 27-40). New York: Freeman.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291-310.

Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.

Northrop, L. C. (1977). *The definition and measurement of reasoning* (Technical Memorandum 77-10). Washington, DC: U.S. Civil Service Commission.

Oxford English Dictionary (2nd ed.). (1989). New York: Oxford University Press. Retrieved April 28, 2003, from http://dictionary.oed.com/

Peirce, C. S. (1955). Abduction and induction. In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 302-309). New York: Dover. (Original work published 1901)

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research, 42*, 237-249.

Powers, D. E., & Enright, M. K. (1987). Analytical reasoning skills in graduate study: Perceptions of faculty in six fields. *Journal of Higher Education, 58*, 658-682.

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

Sternberg, R. J. (1985). *Beyond IQ*. New York: Cambridge University Press.

Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychology? A case study. *American Psychologist, 52,* 630-641.

Toulmin, S. E. (1958). *The uses of argument.* New York: Cambridge University Press.

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

Tucker, C. (1985). *Delineation of reasoning processes important to the construct validity of the analytical test*. Unpublished manuscript.

Tversky, A., & Kahneman, D. (1988). Rational choice and the framing of decisions. In D. E. Bell & H. Raiffa (Eds.). *Decision making: Descriptive, normative, and prescriptive interactions* (pp. 167-192). New York: Cambridge University Press

Ver Steeg, C. L. (1984). Who will help the graduate student? *Change, 16*, 22-23.

Voss, J. F., Perkins, D. N., & Segal, J. W. (Eds.). (1991). *Informal reasoning and education.* Hillsdale, NJ: Erlbaum.

Zechmeister, E. B., & Johnson, J. (1992). *Critical thinking: A functional approach*. Pacific Grove, CA: Brooks/Cole.