Research Report

Toward a Definition of Verbal Reasoning in Higher Education

Nancy W. Burton, Cynthia Welsh, Irene Kostin, and Thomas VanEssen

November 2009

ETS RR-09-33

Toward a Definition of Verbal Reasoning in Higher Education

Nancy W. Burton, Cynthia Welsh, Irene Kostin, and Thomas van Essen
ETS, Princeton, New Jersey

November 2009

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

This paper briefly summarizes the literatures of reading and reasoning in the last quarter century, focusing mainly on the disciplines of cognitive science, cognitive developmental psychology, linguistics, and educational psychology. These literatures were synthesized to create a framework for defining verbal reasoning in higher education. Eight general cognitive and meta-cognitive operations were identified (including, for example, evaluating discourse, seeking and solving problems, and monitoring one's comprehension). Several dimensions underlying these operations on which individual skills may vary were identified (such as breadth of understanding, precision of understanding, or familiarity and facility). Finally, these ideal descriptions of verbal reasoning are applied to the assessment of verbal reasoning for selection in higher education. Problems in measurement and unanticipated consequences of measurement are discussed.

Key words: Cognitive psychology, critical reading, reading research, reasoning research, verbal reasoning, verbal reasoning tests

Acknowledgments

Carol Dwyer conceived the idea of a set of papers defining the general skills constructs used in ETS admission tests and provided the intellectual leadership and dogged determination it took to complete the task. Drew Gitomer, the former senior vice president of ETS Research & Development, made funding available for the project.

Many people helped us think through the issues in the verbal reasoning paper, but we would especially like to thank David Lohman, who suggested ideas, pointed us to other authors and other viewpoints, and read many drafts with critical precision. We would also like to thank the GRE® verbal redesign team, which is chaired by Ed Shea and includes Nancy Burton, Marna Golub-Smith, John Hawthorn, James Hessinger, and Karen Riedeburg, who applied this general construct to a specific test and created the summary set of eight cognitive operations. Finally, we would like to thank Kathy O'Neill, who arranged for extensive technical reviews by external scholars.

The views expressed in this report are those of the authors and not necessarily those of ETS.

Table of Contents

> Language is the most powerful, most readily available tool we have for representing the world to ourselves and ourselves to the world. Language is not only a means of communication, it is a primary instrument of thought, a defining feature of culture, and an unmistakable mark of personal identity.
> (National Council of Teachers of English [NCTE] & International Reading Association [IRA], 1996, p. 12)

The purpose of this paper is to develop a framework for thinking about verbal reasoning in higher education. The framework is based on current theories about and research on reading and reasoning. It is applied to the verbal skills needed to succeed in higher education, specifically those that can be measured before admission. It does not attempt to define a single model of verbal reasoning, since there is not a scientific consensus despite major advances in theory and research in the past quarter century. The discussion is largely theoretical, although from time to time we use practical examples from the verbal reasoning tests we have worked on—the PSAT/NMSQT®, the SAT® I: Reasoning Test, the Graduate Management Admission Test (GMAT), and the Graduate Record Examinations® (GRE®). The framework we develop is defined broadly in order to encompass the function of verbal reasoning in higher education: to help define the outlines that a complete model would have, whether or not all aspects of it are measurable in an admission test. Some of the important limiting conditions of existing admission tests include testing of widely diverse populations of domestic and international students who are applying to institutions with quite different missions in an expanding array of disciplines; the decisions based on the assessments are considered *high stakes* in that they involve admission or non-admission to a degree program, or the provision of financial aid that would make attendance possible. Concerns have arisen in recent years about the effect of admission tests on instruction in high school and college, their fairness to test takers of diverse backgrounds, and the validity of scores in the face of coaching and cheating. These conditions of admission testing tend to limit, to some extent, the skills that can be measured and how they are measured.

The assumption of general admission tests is that there are fundamental academic skills that apply across a wide array of disciplines. Language skills in particular apply to all areas of instruction: Learning requires listening and reading, and demonstrating one's learning requires writing and speaking. The emphasis in verbal measures on higher level skills of critical reading is appropriate in higher education, where instruction is intended to produce independent students and practitioners. General admission test scores are particularly important in supplying a common standard across students.

The following two sections, Concepts of Reading and Reasoning and A Framework for Verbal Reasoning, discusses how current research and thought characterize verbal reasoning, from relatively straightforward comprehension to high levels of expertise in verbal reasoning. The third section, Measuring Verbal Reasoning, discusses some of the difficult theoretical and practical issues in measuring verbal reasoning.

Concepts of Reading and Reasoning

The overview of the cognitive literatures of reading comprehension, critical reading, expertise, and reasoning that follows yields a collection of related conceptions of verbal reasoning, rather than a single coherent cognitive theory. Verbal reasoning appears to involve a number of logically distinct cognitive operations and multiple dimensions on which individual performance may vary. We will attempt to formulate both the cognitive operations and the underlying dimensions at the end of this section. Although we were not able to develop a succinct definition of verbal reasoning, we have adopted a few assumptions about verbal reasoning that may help frame the discussion:

- Comprehending discourse is a key element in verbal reasoning. It involves constructing meaning using information given in discourse, inferences the reader makes about discourse, and the reader's own prior knowledge. (See Graesser, Millis, & Zwaan, 1997, for a review of comprehension models.) Thorndike (1917) only slightly overstated the case nearly a century ago when he said that reading *is* reasoning.

- Reasoning involves going beyond the information given (Bruner, Goodnow, & Austin, 1956) to a more structured and precise understanding.

- Lifelong learning of cognitive skills and knowledge requires a continuing ability to apply general reading and reasoning skills to relatively unfamiliar material.

The following discussion of various concepts of verbal reasoning focuses on ideal conceptions, constrained mainly by limits in current models and data. We start with a discussion of critical reading, which, as stated above, we consider to be the central or defining skill in verbal reasoning. The second part of the discussion, Reasoning, goes back to earlier conceptions and literatures about reasoning and finds a link between the reading and reasoning conceptions in the work that has been done on expertise and on the self-regulated use of knowledge.

Critical Reading

Definition. Verbal reasoning and reading are not the same thing, but adept critical reading is one of the most useful aspects of verbal reasoning. Nist and Simpson (2000) cited research showing that "approximately 85% of all college learning involves reading, and … texts are central to learning at all levels of education" (p. 648). Wagner and Stanovich (1996) emphasized the importance of reading to learning in a variety of domains:

> In certain domains, reading is especially likely to be a substantial contributor to cognitive growth. For example, as a mechanism for building content knowledge structures (Glaser, 1984), reading seems to be unparalleled (Goody, 1977). The world's storehouse of knowledge is readily available for those who read, and much of this information is not usually obtained from other media. (p. 208)

The consensus of a quarter century of cognitive research on reading characterizes it as an active process that involves building a mental representation of the text ("constructing meaning"), calling up relevant knowledge from memory, evaluating differences between text and the reader's existing knowledge and beliefs, making inferences needed to fill gaps in understanding or clarify meaning, integrating pertinent new information into the reader's knowledge base, and thinking about what are the important and unimportant points in the text and how the information can be used (Chapman, 1993; Sweet, 1993). For an expert reader, a number of these processes become automatic, while others may remain conscious and require effort (Graesser, Singer, & Trabasso, 1994; Kintsch, 1998, chapter 4, pp. 93-120; Sternberg, 1986). In general, reasoning is always required when the reader is first learning to read or is confronting new content. As the reader becomes more proficient, and as his or her knowledge of the content grows, reading comprehension becomes more automatic, requiring less reasoning at a conscious level.

Much of the following research is based on theories or models of comprehension that differ from each other, yet many of them share similar ideas of the basic cognitive components and processes involved in comprehension. Some of the key ideas, adapted from Graesser et al. (1997, pp. 174-175), are as follows:

1. The mental representation of the text, as well as the reader's knowledge base, is thought of as containing nodes interconnected by relational ties or arcs. The nodes may be such things as concepts or objects (Graesser & Clark, 1985; van Dijk & Kintsch, 1983).

2. Nodes in the reader's knowledge base are activated when they appear in text; the activation spreads to closely related nodes in the knowledge base by way of relational arcs (Anderson,1983). Continued reading may activate other nodes, increase the activation level of nodes previously activated, or inhibit or suppress nodes. For example, mention of bridge in text may activate a node for a structure over a river and another node for a card game; as reading proceeds, one of these nodes will be suppressed (Gernsbacher, 1990; Kintsch, 1988, 1998).

3. Various memory stores are involved in most reading models: short-term memory, working memory, and long-term memory. Short-term memory and working memory are thought of as having strictly limited capacity, only holding the most recent information being processed. Some models only have one of these two; in models that have both, short-term memory has a smaller capacity and working memory has some processing capacity, such as recycling important information (Fletcher & Bloom, 1988; Kintsch & van Dijk, 1978; Trabasso & Magliano, 1996).

4. A knowledge structure is strengthened (accessed faster, remembered better) when

    - it is consistent with other knowledge structures (a text can easily be integrated in the reader's knowledge base if it fits the constraints of the existing web of nodes and relations [Graesser & Clark, 1985; Kintsch, 1988; MacDonald, Pearlmutter, & Seidenberg, 1994]),

- the reader constructs causal explanations for the content or presentation of the text; for example, what may have caused the events in a narrative or what the writer intends (Chi, de Leeuw, Chiu, & LaVancher, 1994; Graesser et al., 1994; Pressley, Symons, McDaniel, Snyder, & Turnure, 1988; Trabasso & Magliano, 1996; Zwaan & Brown, 1996), and
- it is repeatedly accessed.

An enlightening and different way to understand critical reading is to consider the behaviors of *less competent readers*, who

- have restricted vocabularies,
- have derived less knowledge from previous reading and may have less personal experience about the subjects read,
- stay almost totally within the literal meaning of the text,
- focus on individual words and sentences, details, and separate pieces of information,
- add isolated facts to their preexisting knowledge base,
- focus almost exclusively on content rather than context, structure, rhetorical devices, or author's intentions,
- organize recall in the form of lists,
- summarize in a linear fashion, not relating concepts in different parts of the text,
- assess their comprehension in terms of how many facts they can recall,
- make little attempt to monitor their own understanding,
- rely on one or two critical reading strategies, not necessarily using them consistently,
- do not differentiate between the demands of different kinds of texts and different purposes for reading, and
- give up when unable to comprehend text (adapted from Chapman, 1993, pp. 16-17).

Such readers may succeed in reading without the active participation in reading described earlier, and without making any but the simplest inferences about what they have read, but they are unlikely to learn much from what they read, or to remember it. Myers (1996) defines this as decoding literacy. Critical reading requires more than decoding literacy.

Critical reading includes both cognitive and metacognitive components. Pressley and Afflerbach (1995) did a comprehensive review of cognitive and metacognitive reading strategies, including preparing to read (such as overviewing or determining the purposes for reading), reading (constructing and revising meaning, activating prior knowledge, or monitoring one's comprehension), and processes that may follow reading (planning for use of new information, monitoring memory). Metacognitive processes are those that involve consideration of the reading process, as opposed to the content of reading, and include such acts as monitoring comprehension and changing reading strategies to improve comprehension. As mentioned earlier, the cognitive processes may be conscious or unconscious. Such cognitive processes as activating prior knowledge are treated as automatic neurological responses in some models of reading. Others, such as inferring pronoun referents, start as conscious for beginning readers but become automatic. Still others, such as deciding whether information will change one's knowledge, beliefs, or actions, will most likely always be conscious.

Critical reading requires background knowledge. Researchers have concluded that background knowledge plays an essential role in the process of comprehending text. Johnston (1984) considers prior knowledge an integral part of reading comprehension and argues that a task that does not require prior knowledge cannot therefore involve reading. Pearson and Johnson (1978) make a similar point when they portray comprehension as building bridges between the new and the known. A review of reading research (Roller, 1990) concludes that there is extensive support in the literature for broad general knowledge as one of the major determinants of fluent reading comprehension. This research implies that assessments of verbal reasoning can only measure reasoning in test takers who possess a given level of background knowledge. Otherwise, a low score can be due either to lack of knowledge or lack of reasoning skill—there is no way to tell.

One specific kind of background knowledge needed for fluent reading is *vocabulary.* Many researchers have reported a strong relationship between vocabulary and measures of reasoning or intelligence (Carroll, 1993; Sternberg, 1986). Lohman (2000) went on to say,

> … the high correlation between vocabulary knowledge and reasoning seems to reflect primarily the fact that word meanings are generally inferred from the contexts in which they are embedded. But there is a synergism here 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. (p. 319)

Indeed, the cognitive research literature is quite consistent on the importance of vocabulary in reading. Miller (1999) pointed out that reading is an important source for vocabulary growth beyond the early years. Miller cited Anglin's (1993) estimate that school children learn about 20 words per day, more than anyone could teach them, which suggests that children learn much of their vocabulary by observing the use of words in text. A summary of the literature by the RAND Reading Study Group (2002) said that "the large body of research on vocabulary has consistently shown the importance of vocabulary knowledge for comprehension. This relationship is remarkably robust across ages and populations" (pp.10-11).

Vocabulary indicates that critical reading and reasoning have occurred in the past, and it is also an important tool for facilitating future comprehension and expression. A broad vocabulary suggests that a broad array of texts were read in the past and constitutes evidence that a range of subject matter can be learned and understood in the future.

Critical reading goes beyond reading comprehension. Successful students need to do more than decode and form a basic comprehension of what they read. To participate effectively in higher education, they also need to be able to reason about what they read and hear. In a study that included observations of undergraduate history classes at two universities, Rohwer and Thomas (1989) found that 99% of the exam questions given in these courses required that students go beyond basic comprehension of the course material and use integrative reasoning processes such as elaborating, reorganizing, contrasting, integrating, or summarizing (pp. 119, 122). Similarly, Chase, Gibson, and White (1994), who studied reading demands in four college courses (history, political science, biology, and English), found that the exam questions in these courses required students to make critical judgments and synthesize material from texts and lectures (p. 12). Rosenfeld, Leung, and Oltman (2001) queried both undergraduate and graduate faculty regarding the skills needed for successful academic performance, and reported high faculty ratings for comparing and contrasting ideas in a single text and/or across texts, and for synthesizing ideas in a single text and/or across texts. 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 reasoning skills that were most important for successful performance in graduate school. Two sets of general skills crossed disciplines: critical thinking related to argumentation (e.g., being able to understand, analyze, and evaluate arguments) and critical thinking related to drawing conclusions (e.g., generating valid explanations to account for observations).

Critical reading is important for all students. Critical reading assessments developed for higher education have traditionally attempted to determine whether prospective college and graduate students can evaluate, reflect on, and apply what they have read. Current language arts standards emphasize critical reading and critical thinking for *all* students (Myers, 1996; NCTE & IRA, 1996;), and language arts curricula designed to teach critical thinking skills to *all* students are spreading (Klooster, Steele, & Bloem, 2001; Morgan, 1990; Pearson & Fielding, 1991; Wolf, 1995). Assessments such as the New Standards project and the National Assessment of Educational Progress (NAEP) seek to publish data on the prevalence of higher level skills across *all* students in national or state school populations (Lewis, 1995; Loomis & Bourque, 2001).

Braunger & Lewis (1998) describe the various higher levels of performance beyond basic competency sought in the National Assessments of Reading. To attain a *proficient* level, readers "should be able to extend the ideas of the text by making inferences, drawing conclusions, … making connections to their own personal experiences and other readings, … [and analyzing] the author's use of literary devices." (Mullis, Campbell, & Farstrup, 1993, pp. 12-17). At the *advanced* level, readers construct new understandings within and across texts. Advanced readers must be able to use what they read for creative and critical thinking and for problem solving (Pace, 1993). Advanced readers construct meaning from different perspectives and understand how their meaning may differ from those of others (Hiebert, 1991, p. 2). These descriptions help define a range of expertise in reading, going beyond the minimal or average performance levels usually reported. However, their real importance is that they were developed as standards for *all* secondary school students. Critical reading and reasoning are no longer expected of college-bound students only.

Reasoning

The early psychological literature on reasoning was influenced by Darwin's (1859) conception that biological variations among individuals allow natural selection of advantageous traits and thus the evolutionary change in species. Galton (1869) applied this concept to mental abilities in *Hereditary Genius.* Terman (1916, 1925) used Binet's early work on intelligence testing (Binet & Simon, 1905) to examine the concept of intelligence as a biologically determined trait. Intelligence testing was greatly influenced by this conceptualization. At about the same time, however, the behaviorist movement (Watson, 1913) was very influential in discouraging psychologists from hypothesizing causes that could not be scientifically confirmed or disconfirmed.

In the early twentieth century, a series of important developments in mathematical systems of reasoning also took place. Whitehead and Russell's *Principia Mathematica* (1910-1913), a major advance in symbolic logic, was made possible by such mid-nineteenth-century mathematical advances as Boolean algebra. The Russell-Whitehead system could derive large parts of mathematics. Probability, originally developed by seventeenth- and eighteenth-century mathematicians, was found useful in explaining subatomic physics, further developing methods of probabilistic reasoning. Factor analysis, a statistical method of reducing large amounts of quantified observations of behavior into general "factors," was developed from an intuitive application by Spearman (1904a). Factor analysis, like the probabilistic methods of physics, allowed scientists to classify behavior and learn about its regularities without having to describe underlying, unknown, causal mechanisms. This allowed a great deal of good empirical work to be done on the organization of human behavior, despite the intractability of questions of cause.

Factor analytic conceptions of reasoning. In a massive reanalysis covering much of the history of factor analytic research on human abilities, Carroll (1993) concluded that there is evidence for three hierarchical levels of ability, with overall intelligence at the most general level. The next level of Carroll's hierarchy includes eight broad abilities that are in turn related to abilities at the third, most specific, level of the hierarchy. An alternative hierarchical model has been proposed by Cattell (1957, 1987) and extended by Horn (e.g., Horn & Noll, 1997). This alternative model consists of two levels. The top level in the Cattell-Horn model is almost identical to the middle level in Carroll's model; a consensus in the field of individual differences research is developing at this level.

The eight broad abilities are: reasoning, knowledge, visual processing, auditory processing, working memory, retrieval from long-term memory, and two speed abilities (general cognitive speediness and correct decision speed). One difference between the Cattell-Horn model and the Carroll model is that Cattell and Horn add quantitative reasoning as a ninth broad ability, whereas Carroll considers this ability dimension to be one of the third-level "primary" abilities under reasoning. Carroll's three primary reasoning abilities are sequential reasoning (deductive and other rule-based reasoning), inductive reasoning, and quantitative reasoning (which may include sequential, inductive, and other methods applied to quantitative material). The more important distinction between these two models is that the Cattell-Horn model does not include a general intelligence factor. Unlike the agreement that seems to be forming about the eight or so general cognitive abilities, there is little agreement in the field about the necessity for a single general intelligence factor (the G factor proposed by Spearman, 1904b) governing all of the midlevel factors.

The factor analytic literature tells much about how specific behaviors observed in a testing situation are related and about how many distinct causal mechanisms might be needed to explain these behaviors, but very little about how such mechanisms might work. Reasoning, in this formulation, is somewhat narrowly defined in that it is generally confined to the more formalized and rule-driven sequential systems of reasoning and simplified, puzzle-like measures of inductive reasoning. The limitation arises from the researchers' decision to abstain from postulating theories. As a result, meaning must be derived from empirical observations—in this case, tests, test items, and their interrelations.

Philosophic conceptions of informal reasoning. Discussions of less formalized reasoning among philosophers (van Eemeren, Grootendorst, & Henkemans, 1996; Ennis, 1986; Toulmin, 1958) emphasize the forms of reasoning found in academic disciplines, professions, and normal discourse, rather in than highly structured and relatively content-free methods such as deductive logic. Miller-Jones (1991) introduced a distinction between "structure *seeking*" and "structure *using*" reasoning. Johnson and Blair (1991), in discussing Miller-Jones's paper, suggest that informal reasoning might be characterized as structure seeking, while formal reasoning systems are structure using.

Perkins, Farady, and Bushey (1991) hold that a successful nonformal (structure seeking) argument must include all relevant factors and must not be one-sided. This standard, calling for all sides of an argument to be presented fairly, is not necessarily consistent with current educational practice, since students are more likely to have training in presenting their own views on a subject effectively. However, Perkins et al. concluded that “such know-how can be taught, and such attitudes can be fostered” (p. 103). The psychological literature on expert performance also emphasizes the importance of teaching and learning. Ericsson and his colleagues (Ericsson & Lehmann, 1996) believed that expert performance is based on deliberate practice that involves activities designed by a coach or teacher to improve specific aspects of performance through repetition and refinement. The individual needs to spend at least 10 years at such work, monitoring his or her performance with full concentration, to attain expert levels.

Reading as expert performance. Wagner and Stanovich (1996) applied the concept of expert performance to expertise in reading. They point out that it is difficult to determine the effects of such variables as talent or intelligence on the usual kind of expert studied, who is a product of several quite severe levels of selection and may be rarer than one in a million. Reading instruction and practice, in contrast, is something that virtually all children in many countries are exposed to for 13 or more years. Wagner and Stanovich held that “… one can view schooling as providing [a] natural developmental study …in which millions of children undergo years of preparation to achieve fluent levels of reading” (p. 191). Unlike Ericsson and colleagues, who emphasized the importance of “deliberate practice,” Wagner and Stanovich suggested that simple *print exposure* (how many hundreds of thousands or millions of words a child reads over a number of years) has a significant unique effect on level of reading comprehension, even after controlling for prior reading comprehension skill, verbal ability, and nonverbal ability (all of which are also significantly related to reading comprehension test scores). They concluded that a large number of people can and do achieve expertise in reading.

Thus it appears that at least in the area of reading (and, one might argue, in the areas of mathematics and writing) students have been learning with guidance and feedback from teachers for many years by the time they go to college or graduate school. It is reasonable to consider that many have had an opportunity to attain some level of expertise. Verbal reasoning tests, then, can be seen as measures of skill areas in which a large number of students have had the chance to develop very high levels of performance. It is necessary to acknowledge, however, that not all students have had the opportunity to develop expertise. Curriculum reforms meant to introduce critical thinking to all students have not been fully implemented. NAEP reading assessments

(National Center for Education Statistics, 1999) show that about 40% of high school seniors reach proficient or higher levels of reading; only 6% read at the advanced level.

In the case of expertise, *opportunity* has at least two aspects: Students must have access to appropriate instruction and must supply long-term sustained effort themselves. Both are needed. In a discussion of *self-regulated learners* that emphasizes the importance of the students' own efforts, Paris and Byrnes (1989; cited in Winne, 1995) developed a picture of students

> … who thirst for learning, who … seek challenges and overcome obstacles sometimes with persistence and sometimes with inventive problem solving. They set realistic goals and utilize a battery of resources. They approach academic tasks with confidence and purpose. The combination of positive expectations, motivations, and diverse strategies for problem solving are virtues of self-regulated learners. (p. 169)

Pressley (1995, p. 209), emphasizing the importance of long familiarity and effort before self-regulation can develop, noted that students do not transfer strategies they have just been taught. "This phenomenon is so striking that it recently inspired an entire volume questioning the assumption that transferable competence can be developed via instruction (Detterman & Sternberg, 1993)." Pressley went on to note that students do not apply newly acquired conceptual knowledge routinely, even when it is learned to some level of mastery, citing common instances of scientific misconceptions. The reasons Pressley adduced for such failures of self-regulated use of new knowledge include the fact that old strategies and concepts, even if known to be incorrect, are familiar, easy, and widely connected in the individual's knowledge structures. The individual not only knows *how* to use the old procedure, but also has learned *when* and *where* to use it, and how to adapt it to local circumstances. Procedural and conditional knowledge will be missing or incomplete for newly acquired knowledge. In other words, not even an expert can routinely apply new knowledge until it becomes assimilated in a variety of content and strategic knowledge systems.

These insights into expertise and the self-regulated use of cognitive knowledge help to define the desirable content of large-scale reasoning assessments. Since it is not yet feasible to tailor the content of these tests to the knowledge and interests of each test taker, the other alternative is to limit test content to widely available general world knowledge, mathematical concepts that have been taught and practiced in mathematics classes since middle school, and adaptable writing topics that students can respond to using their own knowledge and interests. Tasks should call on knowledge that the test population can be assumed to have learned long ago, and to have had chances to apply over and over again, and assessment descriptions should make clear to test takers and test users the knowledge that is assumed.

Prospective undergraduate, graduate, or professional students in admission testing populations should have 10 to 20 years of increasingly adept reading experience. Some have progressed far enough on the continuum of expertise to be considered promising candidates for further academic work. The best of them are active, engaged, critical readers. They are likely to know how to learn new content (Ackerman, 1987); to evaluate, analyze, apply, and expand upon what they read (Chapman, 1993; Cioffi, 1992); and some may be able to reason in an unbiased manner (Perkins et al., 1991).

This brief review of current research and thinking about reading and reasoning has dealt with how *verbal reasoning* is conceptualized in current models of cognition. While there are differences among the conceptions, there are also many common themes. The next section of this paper summarizes the dimensions underlying verbal reasoning in this review and proposes a list of eight cognitive and metacognitive verbal operations that are important, if not necessary, for success in higher education.

A Framework for Verbal Reasoning

This section presents a list of cognitive operations that comprise a proposed framework for verbal reasoning in higher education. The list includes operations that can be measured in national, high-stakes assessments and those that probably cannot be measured. Some fall between these extremes and could possibly be measured in a national high-stakes assessment with more flexible, more complex item types than those currently used. The framework includes this range of operations in order to indicate the full extent of verbal reasoning. It should be emphasized that this list is not a unitary, scientific model of reasoning, but is a consensus drawn from related but by no means completely consistent literatures. The review of literature has revealed a number of important dimensions that underlie specific verbal reading and reasoning skills. These dimensions introduce the structure behind the list of cognitive operations that will follow.

Dimensions Underlying Verbal Reasoning

Breadth of understanding. One can think about verbal reasoning as spanning a dimension from understanding words to sentences to units of text to multiple texts and finally to whole systems of discourse. The concept of size of units of meaning is especially important in the discussion of expertise, where the ability to deal with complex problems is related to the ability to organize one's background knowledge into larger meaningful chunks. The ability to chunk relieves the constraints of working memory that can prevent one from attending to a complicated problem as a whole. Discussions of expertise also emphasize acquiring an extensive knowledge base.

Depth of understanding. Alternatively, one can characterize increasing verbal reasoning skill as a matter of depth or sensitivity or precision of comprehension, rather than as an increase in the amount of information that is processed. These first two dimensions correspond roughly to the distinction between depth and breadth of understanding.

One can also focus on the dimension of *familiarity of content and facility of performance*. Characterizations of the growth of expertise, for example, emphasize years of deliberate practice to improve and refine performance. Self-regulated use of knowledge, including transfer of learning, requires familiarity, facility, and a support system of procedural and situational knowledge.

Focus. The development of verbal reasoning also implies differences in the agent's focus on discourse. In comprehending discourse one goes from surface features to deeper structures of discourse, but in both cases the agent is closely focused on the discourse. The agent steps back somewhat from discourse in order to monitor comprehension, to compare it to existing knowledge and beliefs, and to judge its value. Converting discourse from a separate entity to part of the learner's knowledge and belief systems involves a partial or complete transformation of the discourse, and also a change in the learner's knowledge base. When the transformed discourse is used, for example, to solve problems, the focus has shifted away from discourse toward the agent's goals and actions.

Receptive and productive modes. In discussing verbal skills, it is conventional to distinguish between receptive modes (reading and listening) and productive modes (writing and speaking). The term *receptive* is not ideal in its implication of passive reception, since cognitive models make clear that the reader is always an agent. This distinction is perhaps better thought of as an aspect of the agent's purpose. In reading and listening, the agent's purpose is centered on the discourse (to understand it, to learn it, to evaluate it); whereas in writing and speaking, the purpose is to communicate a particular message to a particular audience.

Finally, one can vary the *category of reasoning* required. Although this paper does not attempt to engage the large philosophical and psychological literatures on reasoning, we have tried to distinguish two general kinds of reasoning. One category is reasoning that uses an existing structure or set of principles to warrant its conclusions; this can include logical deductions, formal applications of probabilistic systems, or sequential systems of rules. We have called this first category *structure using*. The second category, called *structure seeking*, includes arguments based on likeness, such as analogy, inference, or induction.

Eight Verbal Cognitive Operations Important in Higher Education

The framework of cognitive operations is based on several key resources. The first was a list of approximately 70 cognitive operations extracted from the literature reviewed in the Reading and Reasoning sections above. These operations were supplemented by consulting several other references pertinent to higher education. Research on success in higher education was reviewed (Blackburn, 1990; Bowen & Bok, 1998; Campbell, Kuncel, & Oswald, 1998; Hartnett & Willingham, 1980; Klitgaard, 1985; Willingham & Breland, 1982). Further sources included surveys of important reasoning skills at the graduate level that were done when the GRE Board was considering revisions to its Analytical measure (Powers & Enright, 1987; Tucker, 1985). Another source was a summary of key reasoning skills developed as part of the work of a committee to consider measuring reasoning in context for the GRE (Linn, 1992). A survey of language skills required of graduate and undergraduate students (Rosenfeld et al., 2001), sponsored by the TOEFL® Board, was another source. The final source was a set of definitions of success in graduate school based on interviews with faculty and graduate school personnel to plan future GRE assessments and services (Walpole, Burton, Kanyi, & Jackenthal, 2002).

These sources were used to create lists of specific verbal cognitive and metacognitive operations that were synthesized to form the following eight general categories. Operations 1 through 5 are sequential in that all prior operations must be completed before the next can take place. Operations 1 through 5 are also prerequisites for Operations 6 and 7. Operation 8, monitoring, can occur with all of the other operations. Note that we use the general term *discourse* to refer to written text in current admission assessments, but it can also be interpreted to include spoken, symbolic, or graphical representations.

1. *Understand discourse.* Understand the meanings of words, sentences, and entire texts. Understand relationships among words and among concepts, and understand the structure of text. Reason from incomplete data, inferring missing information or connections. Select important points, distinguish major from minor or irrelevant points, summarize. Use different reading strategies, depending on the text and one's purpose in reading; use multiple strategies for remembering.

2. *Interpret discourse.* Analyze and draw conclusions from and about discourse. Identify author's/speaker's perspective and assumptions. Understand multiple levels of meaning (such as literal, figurative, author's intent, etc.).

3. *Evaluate discourse.* Identify strengths and weaknesses. Raise questions about the implications of discourse. Consider alternative explanations. Understand and balance multiple perspectives. Appraise author's definitions and assumptions, evaluating sources for bias, self-interest, and lack of expertise. Recognize fallacies in argument.

4. *Incorporate discourse with knowledge base and beliefs.* Evaluate differences between one's knowledge base and beliefs and discourse; integrate new information into one's knowledge base; revise/reorganize prior knowledge and beliefs based on discourse. This might include discarding ideas that one judges to be wrong, but it might also include retaining knowledge and beliefs that are in conflict in recognition of the value of alternative viewpoints.

5. *Create new understandings*. Move beyond the reception of knowledge to the use and application of knowledge. Build upon discourse by integrating, elaborating, and transforming the content. Synthesize information from a variety of sources. Incorporate understanding in a larger framework. Compare, contrast, and integrate perspectives.

6. *Seek and solve problems.* Identify areas that require further thought and research. Develop possible explanations, and test them. Apply knowledge and verbal reasoning strategies to new problem situations. Use verbal reasoning and self-monitoring skills to set goals, plan, and overcome obstacles in the course of problem solving.

7. *Communicate.* Write, present, explain, define, persuade, teach, provide feedback to, and interact with people from a variety of communities of discourse. Become fluent in the language and conventions of one's own discipline.

8. *Monitor one's own comprehension, reasoning, and habits of mind*. Use multiple criteria to monitor comprehension while reading; change strategies when comprehension is unsatisfactory. Use multiple strategies for overcoming obstacles in problem solving. Strive to be well-informed, open-minded, flexible, creative; to maintain personal and professional integrity; and to maintain a broad perspective.

These eight processes seem to be essential verbal reasoning skills with broad application across disciplines. However, some may not be completely measurable in an operational admission assessment. For example, there is little attention to communication (Operation 7) in such verbal reasoning assessments as PSAT, SAT, GRE, or GMAT, except for the GMAT sentence-correction items that measure recognition of correct and effective expression. More significant aspects of communication are measured, however, by the writing assessments associated with each of these assessments.

Other cognitive operations may not be measurable at all in a high-stakes, large-scale assessment—those that have to do with internal knowledge structures, beliefs, attitudes, and actions. For example, it is difficult to observe changes in a learner's internal conceptual structure (Operation 4), even in a carefully designed experiment, and is quite unlikely to be done in a large-scale, impersonal testing environment. An objective measure of such an internal operation would require a great deal of knowledge about the examinee and carefully constructed exercises to capture the examinee's actual current cognitive state. A subjective measure—just asking the examinee—would be very tempting to fake in a high-stakes situation. The solution to a problem (Operation 6) can be observed, but not how the solution was reached. Did the solver apply a well-known procedure? Or did the solution involve reasoning in a novel situation? The

Analytical Writing argument task used by both GRE and GMAT, however, presents evidence that the student can develop alternate explanations for a situation and judge their reasonableness, and, in addition, communicate the results of this reasoning process. Monitoring (Operation 8) is another operation that is difficult to observe.

Mapping the Cognitive Operations Into the Underlying Dimensions

The following paragraphs discuss how the eight cognitive operations relate to the underlying dimensions discussed above.

Breadth of understanding. The first three operations in the list deal with words, sentences, and single units of text; the fourth operation (balancing and integrating discourse with one's knowledge and belief systems) applies to all levels of discourse and makes it possible for one to comprehend larger collections of discourse. Operations 4 through 7 at least implicitly deal with larger systems of discourse. Individual differences in performance of all eight operations implicitly depend on the extent of knowledge and the extent of vocabulary.

Depth of understanding. Precision is only implicit in the proposed list—for example, it can be assumed that as learners continually evaluate and modify their knowledge and belief systems (Operation 4), knowledge becomes more precise as well as more extensive. An actual measure of individual differences in verbal reasoning should explicitly cover different levels of precision and refinement.

Facility and familiarity. Here, amount of practice and extent of reading lead to individual differences in performance. Most theorists also believe that good coaching and the learner's dedication are also key to performance. Like precision, this dimension is only implicit in the operation of evaluating and modifying the agent's knowledge and beliefs based on learning (Operation 4). Familiarity or extent of knowledge is generally measured best in achievement tests such as the SAT Subject Tests or GRE Subject Tests, since expertise is most likely subject-specific. At the expert level, both knowledge and high-level verbal (and quantitative) reasoning skills tend to be subject-specific, and high-level achievement tests focus at least as much on subject-area applications of reasoning skills as they do on knowledge.

Focus. Focus is important in all eight cognitive operations. By understanding and interpreting discourse (Operations 1 and 2), one moves the focus from surface features to deeper structures; in evaluating (Operation 3) and monitoring (Operation 8), the agent must step back from discourse; incorporating (Operation 4) converts the discourse from a separate entity to part of the agent's knowledge and belief systems; while in creating, problem solving, and communicating (Operations 5, 6, and 7), the agent changes the primary focus from discourse to his or her goals and actions.

Reception and production. Although admission tests of verbal reasoning have been built around reading exercises, there is no theoretical reason why skills of listening and speaking could not be measured (writing being currently assessed in a separate measure). In accommodations for students with disabilities, listening and speaking are permitted, since some test takers must demonstrate their comprehension and reasoning in those modes. In the interpretation of this dimension as an aspect of purpose, there is a progression in the agent's purposes over the first seven operations; that is, the agent's purposes become progressively more complex from comprehension (Operation 1) to integration (Operation 4). In communication (Operation 7), creation (Operation 5), and problem solving (Operation 6), discourse becomes a means for achieving the agent's purposes.

Kind of reasoning. It seems clear that structure seeking is involved in all eight cognitive operations. In understanding, for example, the reader may only need to make small inferences having to do with pronoun references, implied but not directly stated information, and so on. These inferences, made within discourse, are called *near inferences* (Royer, Carlo, Dufrense, & Mestre, 1996.) In more complex operations, the framework or structure sought may become more important than the discourse, as in solving problems. Structure using reasoning is a bit different, since it involves applying a specific set of rules to accomplish a task. Basically, it can appear in all of the listed operations, either as a technique that the reader must understand, interpret, or evaluate, or as a technique the reader may use to create new understandings, solve problems, communicate, and monitor her or his own cognitive reading and reasoning operations.

Measuring Verbal Reasoning

What Verbal Reasoning Is and Is Not

In ordinary parlance, verbal skills include at least reading, writing, speaking, and listening. Some current definitions of literacy include graphic and symbolic material, as well as written material. However, the verbal reasoning assessments for PSAT, SAT, GMAT, and GRE have tended to use a narrower definition, organized around the concept of critical reading and related skills. Critical reading is distinguishable from simpler stages of reading, such as decoding and basic comprehension, but because the dividing line differs for different readers, a critical reading assessment is likely to include some tasks that are simply routine comprehension tasks for some more sophisticated readers. However, cognitive researchers who have reviewed the critical reading items in the PSAT and SAT say that the actual questions posed will require almost all test takers to reflect and reason about the text (Graesser, Daneman, & Lohman, 1998).

Admission assessments can be distinguished from assessments such as the Test of English as a Foreign Language™ (TOEFL) and the next-generation TOEFL under development. TOEFL program publications describe the current test as a measure of general English proficiency (ETS, 2000), and the next-generation TOEFL is intended as a test of communicative competence (Bachman, 1990; Canale, 1983; Canale & Swain, 1980). The reading section of the current TOEFL requires the ability to decode and comprehend text—to make inferences within the text, and to understand main ideas, vocabulary, factual information stated in the text, and pronoun referents, as well as the organization and purpose of a text. The reading section of the next-generation TOEFL will require all of those skills and, in addition, will require examinees to demonstrate that they can distinguish among major points and minor or supporting points and organize information from the text into a coordinated structure. These tests, however, do not attempt to assess critical thinking about text (M. Enright, personal communication, March 11, 2004).

Verbal reasoning assessments are also distinguishable from assessments of other verbal skills, such as writing, listening, and speaking, or achievement in literature. In general, the reading and language-processing skills measured in verbal reasoning assessments are correlated with other language skills and achievements (Donlon, 1984, p. 21; ETS, 2002, p. 15) and can be used to predict future performance in writing, speaking, listening, and comprehension of literature. Writing is an especially important case to consider. Any writing task that is even minimally complex entails elements of verbal reasoning. Writing to convince, criticize, or describe involves making a reasoned decision about what to include and what not to include and an awareness of the conventions of reasoned discourse. Research has demonstrated the close links between reading and writing (Tierney & Shanahan, 1991), and teachers now generally agree that they should be taught together (Bushman, 1992; Dickson, 1999). However, writing is such an important skill in higher education that all of the major admission assessments—SAT, ACT, GRE, GMAT, LSAT (Law School Admission Test), and MCAT (Medical College Admission Test)—now include or plan to include a writing measure.

The example verbal reasoning assessments all measure verbal reasoning through written text, partly because it is an efficient way of testing verbal reasoning and partly because of the importance of critical reading to higher education. The concept of verbal reasoning does not require that the test be written. Students also use verbal reasoning in processing spoken discourse. Some students with visual disabilities may necessarily deal mostly with spoken discourse. Although one might argue that a reading task needs to involve written text, verbal reasoning tasks are most appropriately presented in the student's normal mode of reasoning about discourse.

Finally, it is important to distinguish verbal reasoning from a very common conception of intelligence. The view of verbal reasoning developed here is that it is based on a set of cognitive and metacognitive skills that can and must be taught, and years of disciplined practice that require both individual effort by the student and careful guidance by a teacher. Intelligence, on the other hand, is often spoken of as something innate, a "gift" that cannot be taught or achieved through work. Many psychologists continue to use the term *intelligence* with a technical definition different from the above more common usage; however, we have decided that this too often leads to misunderstandings and we have avoided the use of *intelligence* and related words, such as *aptitude* and *ability*.

How Should Verbal Reasoning Be Measured?

Resnick (1987) listed attributes of higher level thinking—nonalgorithmic; complex; having multiple, probabilistic solutions; involving nuanced judgments—that make it clear such skills are not easy to measure. It is also important to recognize that "there is no such thing, strictly speaking, as a 'reasoning task' independent of the persons who are to solve that task" (Sternberg, 1986, p. 287). For one person, a given task may be relatively new and require careful, conscious reasoning to solve; another person, with a great deal of experience in the area, may recognize the problem, know the proper procedure to solve it, and proceed in a mostly routine manner.

There are a number of ways of ameliorating the problem of finding appropriate reasoning tasks for a widely differing group of test takers. The problem cannot be definitively solved. Its solution requires a detailed knowledge of what each test taker knows and can do, but usually the assessment is being given to determine that information. One way to address this situation, however, is to present items at varying levels of difficulty, so that at least some are at the appropriate difficulty level for each test taker. This is what many standardized paper-and-pencil assessments do. Another solution is to tailor item difficulty to the test taker, as computer-adaptive tests do. One can address the situation in another dimension by presenting items in a variety of contexts and subject areas, since test takers have different areas of interest and experience. The higher education assessments we have worked on—PSAT, SAT, GRE, and GMAT—all specify three or four broad areas of content to cover, and ensure that each test administration includes all content areas.

The research cited in this paper suggests that having a broad knowledge base is a strong contributor to fluent reading. Reading allows the student to develop vocabulary, also key to fluent reading. And, finally, since education often involves learning new content, experience in reading, comprehending, and using knowledge in a wide range of fields is excellent preparation for future learning. Broad coverage in an assessment, however, can only be achieved with some loss of depth. The usual compromise involves using many relatively brief, simple assessment tasks to achieve broad coverage.

Another efficient way to measure reasoning involves looking for the products of past reasoning, rather than trying to capture reasoning at the moment of testing. One can visualize information as divided into two areas: the large area of things already known to a given person, and the larger area of everything unknown to that person. At the boundary of those two areas is the fairly narrow line of things not yet fully known, but attainable by a reasoning effort. This is a narrow, idiosyncratic, shifting target of assessment that varies for each individual. It is far easier to mine the person's area of knowledge for evidence of past reasoning, and such assessment tasks tend to be somewhat simpler and briefer than direct measures of reasoning. This technique is used to some extent by all four of the admission assessments under discussion. The ability to answer decontextualized vocabulary items is probably the most obvious example, since little *present* reasoning is possible. The ability to do abbreviated analogy items without text is both a current instance of reasoning and also pretty good evidence that the student has reasoned about relationships in past experiences with text. Similarly, fluent reading in a variety of subject areas not only requires reasoning in the test situation, but also stands on the shoulders of many years of the kind of reasoning necessary to develop good reading skill.

Unanticipated Effects of Assessment

Assessment items that measure past rather than present reasoning, or that use unrealistic, puzzle-like formats, won't do for every purpose. For assessments meant to predict the likelihood of future educational success, evidence of past reasoning can supplement direct measures of reasoning. Artificial items can also be useful, so long as performance on them predicts performance in realistic settings. For a teacher who wants to find out if a student has learned the critical reading skills just taught, evidence of reasoning that occurred before instruction will *not* do. The problem is similar for a cognitive psychologist running experiments on reasoning: The psychologist, too, must capture current reasoning, and usually in a way that helps to illuminate the actual reasoning process.

The authentic assessment movement has raised new concerns about indirect or unrealistic assessment items. Even items that meet the purpose of the assessment may cause unanticipated harm, because teachers will prepare students for assessments that have strong consequences for the students or the teachers themselves. But the obvious preparation for indirect or unrealistic assessment items may be inappropriate. For example, most coaches prepare students for vocabulary tests by having them memorize definitions, although the items are meant as an indirect measure of a history of broad reading. Messick (1989, 1996, 1998) defined studies of the unintended effects of assessments as *consequential validity* studies. The president of the University of California, Richard Atkinson (2001), has raised the concern that high school students are drilling on vocabulary and studying analogies flash cards to prepare for the SAT. It is possible that they may learn something about analogical reasoning by practicing flash cards (although it is unlikely if they are just memorizing answers), but there is good evidence that memorizing definitions is not an effective way to learn vocabulary. One recent review of the literature concluded that memorizing definitions does not appear to increase vocabulary or improve comprehension (RAND Reading Study Group, 2002, pp. 10-11). Sternberg (1988, p. 199) remarked that "vocabulary [that] is directly taught seems most likely to be forgotten later."

Unanticipated consequences are among the principal reasons for recent or in-process revisions to the PSAT, SAT, GRE, and GMAT. Antonyms were dropped in the 1993/94 revision of the PSAT and SAT, not because they failed to perform their measurement function, but to discourage an ineffective teaching practice. Analogies were dropped in the 2005 revision of the PSAT and SAT for the same reason. GRE is currently investigating new item types as possible replacements for antonyms and analogies; GMAT has not used antonyms and analogies since 1976. In the 1993/94 revision of the PSAT and SAT, reading passages were also changed based on reading research. Opening descriptions of where the text appeared and its purpose were added, as were other context-providing cues such as headings. The passages were only lightly edited, no longer severely compressed to save testing time. Assessments that reflect current thought about reading are considered valuable both because such assessments give pertinent information to teachers and because such assessments encourage good teaching practice (Sweet, 1993). In similar responses to both research findings and score users' needs, GMAT added a writing measure in 1994, PSAT added an all-multiple-choice writing skills measure in 1997, GRE moved its Writing Subject Test into the General Test in the fall of 2002, and the College Board moved its SAT Writing Subject Test into the SAT in 2005.

Validity and Fairness of Assessments

Admission assessments are used by students to select institutions of higher education and by institutions to select students. Among the most important of their many missions, institutions of higher education seek to admit students who are able to succeed academically and benefit from the instruction offered (Blackburn, 1990; Bowen & Bok, 1998; Campbell, Kuncel, & Oswald, 1998, Hartnett & Willingham, 1980; Klitgaard, 1985; Walpole et al., 2002; Willingham & Breland, 1982).

Predictive validity of admission assessments. Admission test scores and previous academic record are generally used together, sometimes with extensive additional information, to provide evidence about the academic skills of applicants and to predict their likelihood of success in higher education. Verbal reasoning tests are one part of that evidence. Predictive validity studies gather pre-admission information, which institutions use to decide which applicants to admit, and correlate these predictors with subsequent performance in higher education. The most common studies use first-year average (FYA) as a measure of successful academic performance, but others use cumulative grade point average (GPA), attainment of degree, ratings or measures of honors, achievements, leadership, further participation in higher education, employment, income, and other measures of success or satisfaction in later life. (See, for example, predictive validity studies reported by Bowen & Bok, 1998; Braun & Jones, 1984; Braun, Ragosta, & Kaplan, 1986; McCamley-, Jenkins, & Ervin, 2000; Briel et al., 1993; Burton & Ramist, 2001; Hecht & Schrader, 1986; Kuncel, Hezlett, & Ones, 2001; Ones et al., 2001; Pennock-Roman, 1990; Ragosta, Braun, & Kaplan, 1991; Ramist, Lewis, & McCamley-Jenkins, 1994; Swinton, 1987; Willingham, 1974, 1985; Wilson, 1979, 1986.)

Table 1 summarizes predictive validity study results for three of the four example assessments. (Because PSAT is intended to prepare students for the SAT, the appropriate analysis determines the extent to which the PSAT scores accurately predict SAT scores.) The students who attend a particular institution are likely to have a limited range of scores because of the institution's mission, its intended population of students, and because admission tests are used to select students. This phenomenon is known as *restriction of range* and can be shown mathematically to reduce the size of the correlation between predictors and measures of success in the institution. For two of the three assessments, a statistical correction for restriction of range was used in order to make validity estimates for different institutions more comparable. Both corrected and uncorrected correlations are presented.

Table 1

Predictive Validity of Three Assessments

	Uncorrected correlations		Correlations corrected for restriction of range	
Assessment	Verbal score	V, Q (or M), and previous GPA	Verbal score	V, Q (or M), and previous GPA
SAT	.30	.48	.47	.61
GRE	.30	.52	.42	.65
GMAT	.25	.45	-	-

Note. The data for this table were taken from Bridgeman et al. (2000) for SAT; Wang, (2002) for GRE; and Zhao et al. (2000) for GMAT.

It can be seen that, for all three assessments, unadjusted correlations for the verbal assessment taken alone are in general of moderate size (.3 according to Cohen, 1997), and the unadjusted multiple correlations approach the large category (.5 according to Cohen). The correlations adjusted for restriction in range are, in general, large. The verbal test almost always makes a significant contribution to the multiple correlation; the size of its contribution varies depending on discipline—smaller for students primarily taking mathematics and science courses, larger for students in humanities and social sciences courses.

Validity of decisions made about diverse groups of students. There has been a great deal of research on test fairness, and studies have been published that show admission tests to be roughly equally valid for ethnic minority students compared to White students and for women compared to men (Bowen & Bok, 1998; Bridgeman et al., 2000; Burton & Ramist, 2001; Ones et al., 2001; Pennock-Roman, 1990; Ramist et al., 1994), and for students with disabilities (Braun et al., 1986; Ragosta et al., 1991). SAT data on how well predicted grades match actual grades are also available (Burton & Cline, in press). For women, actual grades are slightly higher than predicted by the verbal test (for a predicted FYA of 3.00, for example, actual FYA averaged 3.02); African American and Hispanic students, on the other hand, tend to get lower grades than predicted by the verbal test (for a predicted FYA of 3.0, African American students actually attained an average of 2.92, while Hispanic students attained an average of 2.96). Research has shown that these small but consistent differences are partly explained by differences in course-taking patterns among gender and ethnic groups (Ramist et al., 1994).

Data on predictive validity for students with disabilities are available for the SAT and the GRE. The correlation of SAT scores and high school GPA with four-year cumulative college GPA is .70 for students without disabilities; the correlations are .62 for students with learning disabilities, .62 for students with physical disabilities, .63 for students with visual disabilities, and .45 for students with hearing disabilities (Ragosta et al., 1991). The correlation of GRE scores and undergraduate GPA with FYA is .63 for students without disabilities, but for 216 students with disabilities it is .27 (Braun et al., 1986). Predictive validity is very hard to estimate for GRE, since most graduate departments are small and have few students with disabilities. In both studies there was a slight tendency for test scores and previous grades to overpredict actual grades for students with disabilities.

Students for whom English either is not the native language or is not the best language are a group of particular concern for assessments of verbal reasoning in English. The current review of research suggests that nonnative speakers would not be able to display their true level of skill in verbal reasoning on a test presented in English. However, the pertinent question in a selection decision is whether a test presented in English gives a valid prediction of academic performance in an English-speaking university for nonnative speakers. This depends on whether reasoning *in English* is crucial to performance in higher education. Some data on this question have recently become available and are displayed in Table 2.

Table 2

Predictive Validity for Students Who Are Not Native Speakers of English (Correlations Not Corrected For Restriction of Range)

Assessment	Verbal score	Multiple correlation
SAT	.29	.53
GRE	.27	.50

Note: Data for this table were taken from Burton & Cline (in press) for the SAT and from Wang (2002) for the GRE. For SAT, the predictors were verbal scores, mathematical scores, and high school GPA; for GRE the predictors were verbal scores, quantitative scores, and analytical scores, but not undergraduate GPA since most of these students attended undergraduate school outside the United States.

Comparing these results to the uncorrected correlations in Table 1, we can see that the correlations for the verbal test alone and in combination with other predictors are just as high for nonnative speakers as for the entire test-taking populations for both SAT and GRE. SAT data on how well predicted grades match actual grades are also available (Burton & Cline, in press). For native speakers, actual grades are barely lower than predicted (students with a predicted GPA of 3.00 on average attained a GPA of 2.99); for nonnative speakers, actual GPAs were slightly higher (predicted = 3.00, actual = 3.03). This difference is of no practical significance.

In summary, predictive validity research provides extensive support for using current admission assessments of verbal reasoning, in conjunction with measures of mathematical reasoning, writing, and previous grades, to make admission decisions. Support exists for decisions about women and men, White students and ethnic minority students, students with and without disabilities, and students for whom English is and is not the best or native language. While there are some variations in the validity coefficients observed for these groups, in general the correlation coefficients are large, and current verbal reasoning measures make substantial contributions to predicting success in higher education for all groups studied. These findings are important steps in establishing the validity and fairness of verbal reasoning assessments as part of the information used in making admission decisions, although more evidence is also needed.

Researchers need to develop assessments that more fully measure the construct of verbal reasoning, and measures that capture some of the other elements of success in higher education (such as practical judgment or persistence). In particular, the search needs to continue for measures that reveal alternative strengths of a diverse population of applicants.

Using the Framework to Develop and Improve Assessments

The eight cognitive operations and underlying dimensions are intended to form a possible framework for developing assessments of verbal reasoning for young adults. Despite the fact that the framework was developed in the specific context of admission to higher education, the goal was to make the structure broad and general enough to include operations likely to be measured in a high-stakes, national test. Therefore, we believe that the framework may be useful for other assessments of young adults, and that, in the future, it might also be extended to younger students, although the formidable literatures on reading and reasoning for younger students would need to be reviewed and integrated.

We developed the framework because we believed that it could be useful to existing assessment sponsors and developers. The framework will need to be adapted to fit a specific assessment's purposes and intended testing population, but that in itself should help the assessment sponsors and developers clarify and update goals that may have been set decades ago.

Perhaps most obviously, the framework can be used to explain an existing assessment to test takers and to institutional score recipients. It provides a structure for discussing the purposes and uses of an assessment and for describing the meaning of scores.

The framework can also assist in planning revisions or expansions to existing assessments. It will illuminate missing elements in existing test coverage, and provide a theoretical basis for developing and evaluating new measures. The cognitive models behind the framework should assist in identifying and solving measurement problems. They may suggest possible solutions to known weaknesses in existing assessment items, and guide the refinement of new assessment items as they are tried out.

The framework can be used to guide a research agenda. It can provide a context for setting priorities on individual studies. It can help researchers form coherent programs of research, with a series of studies building upon earlier work. In particular, the framework is an essential basis for test validation. It can be used to determine areas where validation evidence needs to be improved for an existing assessment. It can be used to generate hypotheses about verbal reasoning behavior that, if confirmed, provide evidence toward the construct validity of the assessment.

Finally, the framework can be use to generate new assessments on a theoretical basis. The cognitive models behind the framework not only specify what needs to be measured and suggest what kinds of tasks are most likely to be good measures, but they also have implications for the kinds of evidence needed to diagnose test takers' specific cognitive problems, and for the kind of instruction needed to solve those cognitive problems.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102,* 3-27.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge MA: Harvard University Press.

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58* (10, Serial No. 238), 1-166.

Atkinson, R. (2001, Spring). Rethinking the SAT. *The Presidency, 4,* 20-27.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

Binet, A., & Simon, T. (1905). New methods for diagnosing the intellectual level of abnormals. *Annee Psychologique, 11,* 191-336.

Blackburn, J. (1990). *Assessment and education in admission*. New York: The College Board.

Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions.* Princeton, NJ: Princeton University Press.

Braun, H. I., & Jones, D. H. (1984). *Use of empirical Bayes methods in the study of the validity of academic predictors of graduate school performance* (GRE Board Research Rep. No. 79-13R, ETS RR-84-34). Princeton, NJ: ETS.

Braun, H., Ragosta, M., & Kaplan, B. (1986). *The predictive validity of the GRE General Test for disabled students* (ETS RR-86-42). Princeton, NJ: ETS.

Braunger, J., & Lewis, J. P. (1998). *Building a knowledge base in reading* (2nd ed.). Portland, OR: Northwest Regional Educational Laboratories, Curriculum and Instruction Services.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade point average from the revised and recentered SAT I: Reasoning Test* (College Board Research Rep. No. 2000-1, ETS RR-00-01). New York: The College Board.

Briel, J. B., Kubiak, A. T., Marr, D. B., O'Neill, K. A., Powers, D. E., Scheuneman, J. D., & Schrader, W. B. (1993). *Graduate Record Examinations technical manual.* Princeton, NJ: ETS.

Bruner, J. S., Goodnow, J. J., & Austin, G. (1956). *A study of thinking.* Oxford, England: Wiley.

Burton, N. W., & Cline, F. (in press). *The predictive validity of SAT item types for gender and ethnic groups: two studies*. Princeton, NJ: ETS.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (College Board Research Rep. No. 2001-2). New York: The College Board.

Bushman, J. H. (1992). The role of young adult literature: The reading/writing connection. *ALAN Review*, *20*(1), 42-46.

Campbell, J. P., Kuncel, N. R., & Oswald, F. L. (1998, April). Predicting performance in graduate school: The criterion problem. In J. P. Campbell & D. S. Ones, *Selection into I-O programs: Focus on GRE validity*. Symposium presented at the 13th annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Canale, M. (1983). Language assessment: The method is the message. In D. Tannen & J. E. Alatis (Eds.), *The interdependency of theory, data, and application* (pp. 249-262). Washington, DC: Georgetown University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1,* 1-47.

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

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York: World Books.

Cattell, R. B. (1987) *Intelligence: Its structure, growth and action*. Amsterdam: North-Holland

Chapman, A. (Ed.). (1993). *Making sense: Teaching critical reading across the curriculum.* New York: The College Board.

Chase, N. D., Gibson, S. U., & White, J. G. (1994). An examination of reading demands across four college courses. *Journal of Developmental Education, 18,* 10-16.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Cioffi, G. (1992). Perspective and experience: Developing critical reading abilities. *Journal of Reading, 33,* 48-52.

Cohen, J. (1997). *Statistical power analyses for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: John Murray.

Detterman, D. K., & Sternberg, R. J. (Eds.). (1993). *Transfer on trial: Intelligence, cognition, and instruction.* Norwood, NJ: Ablex.

Dickson, S. (1999). Integrating reading and writing to teach compare-contrast text structure. *Reading and Writing Quarterly*, *15*, 49-79.

Dijk, T. A. van, & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Donlon, T. F. (Ed.). (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: The College Board.

Eemeren, F. van, Grootendorst, R., & Henkemans, F. (1996). *Fundamentals of argumentation theory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Ennis, R. H. (1986). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. S. Sternberg (Eds.), *Teaching thinking skills: Theory and practice.* New York: W. H. Freeman.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), *Annual Review of Psychology, 47*, 274-305.

ETS. (2000). *Computer-based TOEFL score user guide* (2000-2001 ed.). Princeton, NJ: Author.

ETS. (2002). *LanguEdge courseware score interpretation guide.* Princeton, NJ: Author.

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language, 19*, 70-80.

Galton, F. (1869). *Hereditary genius: An enquiry into its laws and consequences.* London: Macmillan.

Gernsnbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Erlbaum.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 65-108.

Goody, J. (1977). *The domestication of the savage mind.* New York: Cambridge University Press.

Graesser, A. C., & Clark, L. F. (1985). *Structure and procedures of implicit knowledge.* Norwood NJ: Ablex.

Graesser, A. C., Daneman, M., & Lohman, D. (1998, Aug. 19-20). *Exploring the SAT verbal construct in terms of basic psychological processes: Critical reading*. Presentation at meeting at ETS, Princeton, NJ.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163-189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395.

Hartnett, R. T., & Willingham, W. W. (1980). The criterion problem: What measures of success in graduate education? *Applied Psychological Measurement 4*(3), 281-291.

Hecht, L. W., & Schrader, W. B. (1986). *Graduate Management Admission Test: Technical report on test development and score interpretation for GMAT users.* Princeton, NJ: ETS.

Hiebert, E. (Ed.). (1991). *Literacy for a diverse society: Perspectives, practices, and policies.* New York: Teacher's College Press.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. Flanagan & J. L. Genshaft (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues.* New York: Guilford Press.

Johnson, R. H., & Blair, J. A. (1991). Contexts of informal reasoning: Commentary. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnston, P. (1984). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly*, *19*, 219-239.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A constructive-integration model. *Psychological Review, 85,* 363-394.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

Kintsch, W. & Dijk, T. A. van. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-394.

Klitgaard, R. (1985). *Choosing elites.* New York: Basic Books.

Klooster, D. J., Steele, J. L., & Bloem, P. L. (Eds.). (2001). *Ideas without boundaries: International education reform through reading and writing for critical thinking.* Newark, DE: International Reading Association.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*(1), 162-181.

Lewis, A. C. (1995). An overview of the standards movement. *Phi Delta Kappan*, *76*(10), 744-750.

Linn, M. (1992, May). *Meeting summary of GRE Reasoning in Context Committee Meeting*. Unpublished manuscript.

Lohman, D. F. (2000). Complex information processing and intelligence. In R. S. Sternberg (Ed.), *Handbook of intelligence.* Cambridge, England: Cambridge University Press.

Loomis, S., & Bourque, M. (2001). From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In J. Cizek (Ed.), *Setting performance standards: Concept, methods and perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676-703.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement,* (3rd ed.). New York: Macmillan.

Messick, S. (1996). *Validity and washback in language testing* (ETS RR-96-17). Princeton, NJ: ETS.

Messick, S. (1998). *Consequences of test interpretation and use: The fusion of validity and values in psychological assessment* (ETS RR-98-48). Princeton, NJ: ETS.

Miller, G. A. (1999). On knowing a word. *Annual Review of Psychology, 50,* 1-19.

Miller-Jones, D. (1991). Informal reasoning in inner-city children. In J. F. Voss, , D. N. Perkins., & J. W. Segal (Eds.), *Informal reasoning and education.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Morgan, M. (1990). Encouraging critical thinking in the language arts. *Language Arts, 67*(7), 780-782.

Mullis, I. V., Campell, J., & Farstrup, A. (1993). *Executive summary of the NAEP 1992 reading report card for the nation and the states: Data from the national and trial state assessments.* Washington, DC: National Center for Education Statistics.

Myers, M. (1996). *Changing our minds: Negotiating English and literacy.* Urbana, IL: National Council of Teachers of English.

National Council of Teachers of English, & International Reading Association. (1996). *Standards for the English language arts.* Urbana, IL: Author.

National Center for Education Statistics. (1999). *The NAEP 1998 reading report card: National and state highlights*. Washington, DC: Author.

Nist, S. T., & Simpson, M. L. (2000). College studying. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III). Mahwah, NJ: Lawrence Erlbaum Associates.

Ones, D. S., Hezlett, S. A., Kuncel, N. R., Vey, M., Ahart, A., Campbell, J. P., et al. (2001, April). *The predictive validity of academic performance in high school.* Paper presented at the annual meeting of the American Educational Research Association.

Pace, G. (1993). *Making decisions about grouping in language arts* (Literacy Improvement Series for Elementary Educators). Portland, OR: Northwest Regional Educational Laboratory, Literacy, Language and Communication Program.

Paris, S. G., & Byrnes, J. P. (1989). The constructivist approach to self-regulation and learning in the classroom. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theory, research, and practice* (pp. 169-200). New York: Springer-Verlag.

Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson, (Eds.), *Handbook of reading research* (Vol. II). White Plains, NY: Longman.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart & Winston.

Pennock-Roman, M. (1990). *Test validity and language background: A study of Hispanic American students at six universities*. New York: The College Board.

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Powers, Donald E.; Roever, Carsten; Huff, Kristin L.; & Trapani, C. (2003). *Validating LanguEdge Courseware scores against faculty ratings and student self-assessments* (ETS RR-03-11). Princeton NJ: ETS.

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.

Pressley, M. (1995). More about the development of self-regulation: Complex, long-term, and thoroughly social. *Educational Psychologist*, *30*(4), 207-212.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Hillside, NJ: Lawrence Erlbaum Associates.

Pressley, M., Symons, S., McDaniel, M. A., Snyder, B. L., & Turnure, J. E. (1988). Elaborative interrogation facilitates acquisition of confusing facts. *Journal of Educational Psychology, 80,* 268-278.

Ragosta, M., Braun, H., & Kaplan, B. (1991). *Performance and persistence: A validity study of the SAT for students with disabilities* (College Board Research Rep. No. 91-3, ETS RR-91-41). New York: The College Board.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic group* (College Board Research Rep. No. 93-1, ETS RR-94-27). New York: The College Board.

RAND Reading Study Group. (2002). *Reading for understanding: Towards an R & D program in reading comprehension.* Santa Monica, CA: RAND Corporation.

Resnick, L. B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Rohwer, W. D., & Thomas, J. W. (1989). Domain-specific knowledge, metacognition, and the promise of instructional reform. In C. B. McCormick, G. Miller, & M. Pressley (Eds.), *Cognitive strategy research: From basic research to educational application.* New York: Springer-Verlag.

Roller, C. (1990). The interaction between knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly*, *25*, 79-89.

Rosenfeld, M., Leung, S., & Oltman, P. K. (2001). *The reading writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels* (TOEFL Monograph No. MS-21, ETS RM-01-03). Princeton, NJ: ETS.

Royer, J. M., Carlo, M. S., Dufrense, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and instruction, 14,* 373-408.

Spearman, C. (1904a). The proof and measurement of association between things. *American Journal of Psychology, 15,* 72-111.

Spearman, C. (1904b). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–295.

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

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence.* New York: Viking.

Sweet, A. P. (1993). *State of the art: Transforming ideas for teaching and learning to read.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement Education Information.

Swinton, S. S. (1987). *The predictive validity of the restructured GRE with particular attention to older students* (GRE Board Professional Rep. No. 83-25P, ETS RR-87-22). Princeton, NJ: ETS.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision of the Binet-Simon intelligence scale.* Boston: Houghton-Mifflin.

Terman, L. M. (1925). *Genetic studies of genius: Mental and physical traits of a thousand gifted children.* Oxford, England: Stanford University Press.

Tierney, R., & Shanahan, T. (1991). Research on the reading writing relationship. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2; pp. 246-280). Hillside, NJ: Lawrence Erlbaum Associates.

Thorndike, E. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology, 8,* 323–332.

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

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Process, 21*(3), 255-287.

Tucker, C. (1985). *Delineation of reasoning processes important to the construct validity of the analytical test* (GRE Board Research Rep. No. 84-17). Princeton, NJ: ETS.

Wagner, R. K., & Stanovich, K. E. (1996). Expertise in reading. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games.* Mahwah, NJ: Lawrence Erlbaum Associates.

Walpole, M. B., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). *Selecting successful graduate students: In-depth interviews with GRE users* (GRE Research Rep. No 99-11R, ETS RR-02-08). Princeton, NJ: ETS.

Wang, M-m. (2002). [Reanalysis of GRE validity data from 1986-1991]. Unpublished raw data.

Watson, J. (1913). Psychology as the behaviorist views it. *Psychological Review, 20,* 158-177.

Whitehead, A. N., & Russell, B. (1910-1913). *Principia mathematica* (Vols. 1–3). Cambridge, England: Cambridge University Press.

Willingham, W. W. (1974). Predicting success in graduate education. *Science, 183,* 273-278.

Willingham, W. W. (1985). *Success in college.* New York: The College Board.

Willingham, W. W., & Breland, H. M. (1982). *Personal qualities and college admissions.* New York: The College Board.

Wilson, K. M. (1979). *The validation of GRE scores as predictors of first-year performance in graduate study: Report of the GRE cooperative validity studies project* (GRE Board Research Rep. No. 75-8R). Princeton, NJ: ETS.

Wilson, K. M. (1986). *The relationship of GRE General Test scores to first-year grades for foreign graduate students: Report of a cooperative study* (GRE Board Professional Rep. No. 82-11P, ETS RR-87-4). Princeton, NJ: ETS.

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist, 30*(4), 173-187.

Wolf, D. P. (1995). Of courses: The pacesetter initiative and the need for curriculum-based school reform. *English Journal, 84*(1), 60-68.

Zhao, J. C., Patsula, L., Stone, E., Paez, R., Babula, J., Tye, E., et al. (2000). *GMAT validity study service: A three-year summary (1996-97 through 1998-99).* Princeton, NJ: ETS.

Zwaan, R. A., & Brown, C. M. (1996). The influence of language proficiency and comprehension skill on situation model construction. *Discourse Process, 21,* 289-327.