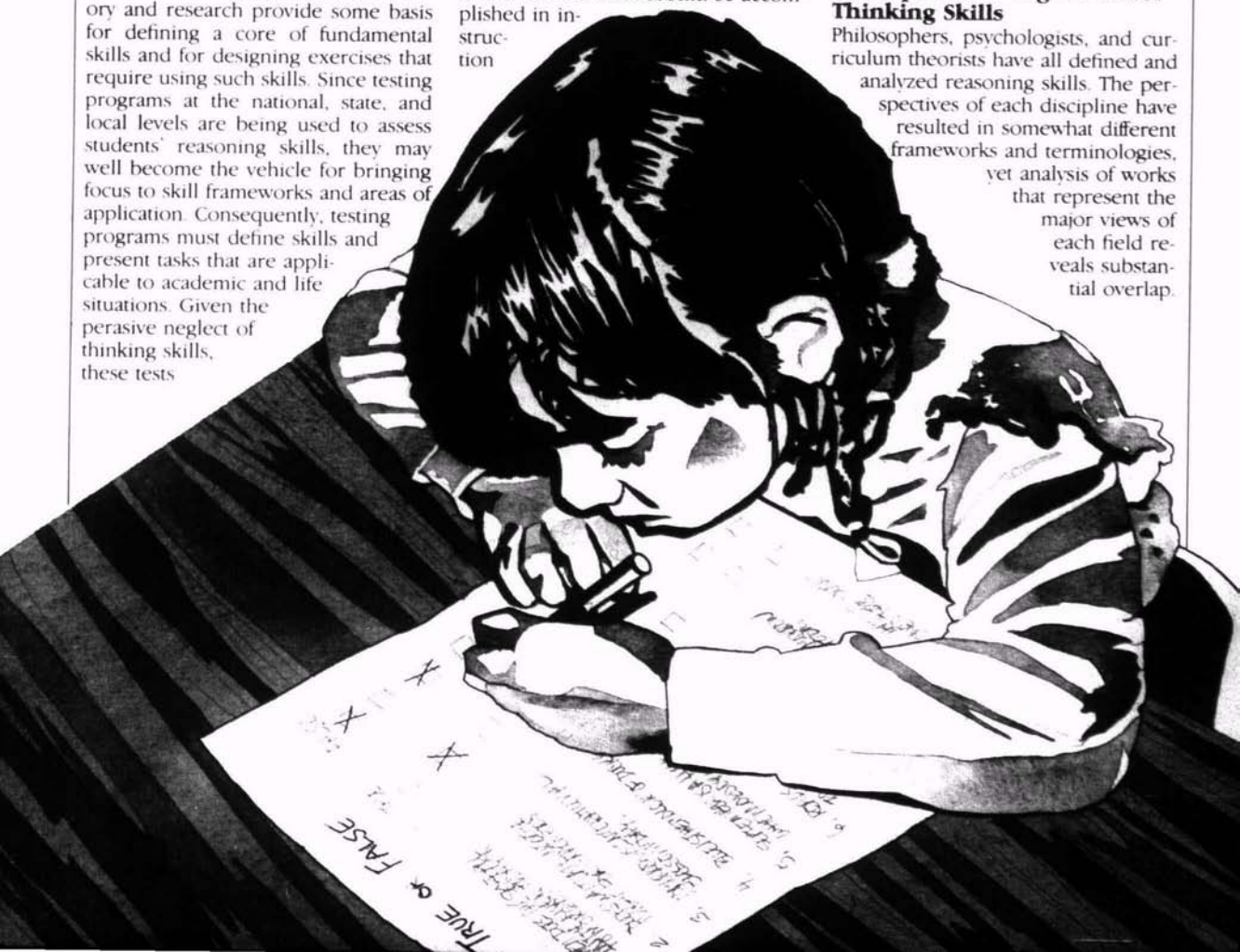
EDYS S. QUELLMALZ

Needed: Better Methods for Testing Higher-Order Thinking Skills

Assessments of student thinking should be based on a core of essential skills that apply to academic, everyday, and novel situations, and should include a variety of test items that require sustained reasoning.

The task of designing a systematic program for testing and teaching higher-order thinking skills is formidable, yet recent advances in theory and research provide some basis for defining a core of fundamental skills and for designing exercises that require using such skills. Since testing programs at the national, state, and local levels are being used to assess students' reasoning skills, they may well become the vehicle for bringing focus to skill frameworks and areas of application. Consequently, testing programs must define skills and present tasks that are applicable to academic and life situations. Given the perasive neglect of thinking skills, these tests will clearly serve a crucial policy role—perhaps the most critical shift in emphasis of the decade. They will define and test *what should be* accomplished in instruction rather than *what is* currently addressed.



Conceptions of Higher-Order Thinking Skills

Philosophers, psychologists, and curriculum theorists have all defined and analyzed reasoning skills. The perspectives of each discipline have resulted in somewhat different frameworks and terminologies, yet analysis of works that represent the major views of each field reveals substantial overlap.

"We need to develop a broader range of tasks that present significant problems, require sustained reasoning, and require component information to be integrated to form a conclusion."

Within the field of *philosophy*, Dewey (1933) defined reflective thought as the careful, persistent examination of an action, proposal, or belief and the analysis or use of knowledge in light of grounds that justify it and its probable consequences. Smith (1953) also emphasized the judgmental aspect of thinking. He defined critical thinking as "what a statement means and whether to accept or reject it." In his landmark paper, "A Concept of Critical Thinking," Ennis (1962) elaborated on Smith's definition of critical thinking by delineating skills that called for the application of formal and informal logic. Ennis has since considerably expanded his concept of critical thinking. His most recent, expanded skill clusters (1984) include clarifying issues and terms, identifying components of arguments, judging the credibility of evidence, using inductive and deductive reasoning, handling argument fallacies, and making value judgments. To date, the Ennis compendium of critical thinking skills both encompasses those forwarded by other philosophers and represents skills appearing on such extant critical thinking tests as his own *Cornell Critical Thinking Test*, the *Watson-Glaser Critical Thinking Test*, and the *New Jersey Test of Reasoning*.

Within the field of *psychology*, definitions of higher-order skills tend to place the reasoning skills proposed by philosophers within broader frameworks. Psychologists studying general intelligence such as Piaget, Guilford, and Sternberg represent one research focus. Piaget's stages of development, particularly the distinction between formal and operational thought, are often used to differentiate among problems requiring logical reasoning. However, Piaget's framework of discrete, hierarchical stages of mental development has been strongly criticized on definitional and empirical grounds by philosophers and psychologists (Ennis, 1976; Linn, 1982; Phillips, 1975). Guilford's (1956) Structure of the Intellect model, based on correlational studies of performance on intelligence test items, has also been criticized on statistical and theoretical grounds. More recently, Sternberg placed the components of intelligence test items into a problem-solving framework. His triarchic model of intelligence includes skills involved in knowledge acquisition, performance, and metacognitive, self-monitoring skills. Sternberg's theory identifies analogical, inductive, and deductive reasoning skills required to perform both novel and familiar tasks (Sternberg, 1983).

Studies of problem solving have identified strategies that generalize across many types of tasks and also sets of distinct steps and strategies for separate subject-matter disciplines and for problem types within them. Studies of how experts play chess and solve physics and mathematical problems have shown us that they call on the logical reasoning skills identified by philosophers. From the work of such psychologists as Bruner (1966), Gagne (1977), and Newell and Simon (1976), we find that people resort to generalized problem-solving strategies when faced with novel tasks, but that expertise depends heavily on extensive knowledge within a particular discipline. We also know that the metacognitive, self-monitoring activities of planning, monitoring, and revising are essential to skillful problem solving.

A final source of information for formulating the thinking skill domain is *curriculum*. In general, thinking skill clusters proposed in curriculum projects (for example, Bruner, 1966; Taba, 1963; Suchman, 1965; Covington, 1968) are the reasoning skills identified by philosophers and psychologists. Bloom's Taxonomy (1956), for example, continues to be the most frequent guide for classifying types of higher-order skills. Curriculum experts also try to map the significant problem types and methods of inquiry within subject matter.

Looking across these three disciplines, we see that:

- Philosophy provides definitions of

Figure 1. Definition of Higher-Order Thinking

Students engage in purposeful, extended lines of thought during which they:
- Identify the task or problem type.
- Define and clarify essential elements and terms.
- Judge and connect relevant information.
- Evaluate the adequacy of information and procedures for drawing conclusions and/or solving problems.

In addition, students become self-conscious about their thinking and develop self-monitoring problem-solving strategies.
Commonly specified higher-order reasoning processes are:

1. **Cognitive**	2. **Metacognitive**
Analyze	Plan
Compare	Monitor
Infer/Interpret	Review/Revise
Evaluate	

reasoning skills and criteria for judging when reasoning is done well.

- Psychology also identifies reasoning skills and their underlying cognitive processes, and sketches a process for how reasoning is used to address purposeful tasks.
- Curriculum theory proposes significant classes of tasks and methods of inquiry.

To force a framework for assessing higher-order reasoning, then, we need to (1) identify the salient, common reasoning skills proposed in psychology and philosophy, (2) track where skills seem to be required for addressing significant tasks, and (3) consider criteria for designing tasks.

Figure 1's definition of higher-order thinking attempts to merge the goal-directed problem-solving paradigm with the inferential, evaluative emphasis of the critical thinking paradigm.

One goal of higher-order thinking instruction proposes that students engage in purposeful, extended lines of thought in which they identify and analyze a problem, identify and relate information necessary to address the task, and evaluate the adequacy of conclusions or solutions. Further, students should be critical of the strategies they use. The cognitive processes of analysis, comparison, inference, and evaluation seem to be involved in various combinations in reasoning tasks, as do the three metacognitive components—planning, monitoring, and reviewing/revising.

Figure 2. Relationship Among Reasoning Skills Proposed by Psychologists and Philosophers

Problem-Solving Strategies (Psychology)	Critical Thinking Skills (Philosophy)	Probable Dominant Cognitive Processes (Psychology)
1. Identify the problem (essential elements and terms)	1. Clarification • Identify or formulate a question • Analyze major components • Define important terms	1. Analogical • Analysis • Comparison
2. Identify appropriate information, content, and procedural schemata	2. Judge credibility of support, the source, and observations	2. Analogical • Analysis • Comparison Evaluate components
3. Connect and use information to solve the problem	3. Inference • Deduction • Induction • Value judgments • Fallacies	3. Inferential—infer/interpret relationships among components
4. Evaluate success of the solution	4. Use criteria to judge adequacy of solution	4. Evaluative—evaluate effectiveness of specific and general strategies

Figure 2 illustrates the relationship among problem-solving, critical thinking, and cognitive strategies. Thus, the problem-solving skill of identifying and analyzing tasks involves the clarification skills Ennis identified. The more basic cognitive strategies primarily involve the analogical reasoning skills of analysis and comparison. The problem-solving skills of accessing appropriate content or procedural schemata and seeing how new information can be used involve the critical thinking skills needed for recognizing the type of information or evidence (analogical reasoning) and judging the adequacy of the information (evaluation). Perhaps the closest correspondence among descriptions of reasoning occurs when learners employ inductive and deductive skills to connect and use information to solve problems and evaluate conclusions. Throughout problem solving, learners use metacognitive, self-monitoring strategies to evaluate and re-check progress. Analytical, comparative, inferential, and evaluative processes are significant, dominant task types or activities as well as processing components of complex tasks. Figure 3 proposes significant tasks that require these processes within three subject domains.

Figure 3. Examples of Higher-Order Reasoning Skills in Three Subject Domains

	Science	Social Science	Literature
1. Analyze	Identify the components of process and the features of animate and inanimate objects	Analyze components or elements of an event	Identify components of literary, expository, and persuasive discourse
2. Compare	Compare the properties of objects or events	Compare causes and effects of separate events; compare social, political, economic, cultural, and geographic features	Compare meanings, themes, plots, characters, settings, and reasons
3. Infer	Draw conclusions; make predictions; pose hypotheses, tests, and explanations	Predict, hypothesize, and conclude	Infer characters' motivation; infer cause and effect
4. Evaluate	Evaluate soundness and significance of findings	Evaluate credibility of arguments, decisions, and reports; evaluate significance	Evaluate form, believability, significance, completeness, and clarity

For example, *analyses* of a process or a literary piece are familiar activities in science, history, and literature. Similarly, *inferring* the consequences of an experiment or interpreting the motives of a character are significant tasks. Further, shoppers use analogical reasoning to *compare* products, much as inventors do when they create inventions. To *evaluate* how well problems have been solved and conclusions drawn, we rely on all the basic processes.

Recommendations for Designing Higher-Order Thinking Skills Tests

Higher-order thinking skills tend to be measured either on specialized tests or within subject-matter tests, most often in a multiple-choice format with a focus on specific, isolated skills. Although many significant interpretive and evaluative problems permit multiple-solution paths and require orchestration of a set of component judgments to arrive at a conclusion, the one-right-answer format and piecemeal testing of components dominate. Seldom are students asked to engage in sustained reasoning on significant tasks, nor are they asked to explain their reasoning. A major exception was the National Assessment exam, which asked students to interpret or evaluate literary selections in essay form. The appallingly few students who could offer even rudimentary support for their opinions highlighted the need for renewed emphasis on higher-order skills and constructed response test formats (NAEP, 1981).

"The essay format is especially useful for assessing how students reach and explain their conclusions."

Figure 4. Recommendations for Designing Instructionally Relevant Assessments of Higher-Order Reasoning Skills

Specification of Skills

- Identify skills common to several conceptualizations
- Identify skills that can be applied in academic, life, and novel tasks.
- Select a manageable number of skills.
- Place skills in a coherent framework.

Design of Tasks

- Identify significant, recurring problem types.
- Assess integrated skills, *not* just components.
- Include tasks that permit alternative interpretations or solutions.
- Design open formats that ask for explanations of reasoning.
- Build sets of tasks that represent a desired range of generalization and transfer.

In addition to identifying a manageable set of skills fundamental to reasoning in academic subjects and in life and novel situations, we need to design tests that include a broader array of items and tasks. Clearly, assessments of higher-order thinking skills call for more task formats and contents than those that exist on available tests. We need to:

1. *Design problems that reflect important recurring issues in the domain.* Cognitive research has found that experts and skilled individuals store knowledge in problem-type structures. Examples of appropriate problems include classes of algebra problems or sets of reading or writing tasks with distinct narrative or persuasive discourse aims and structures. The reasoning skills required, of course, also determine the content students must find and connect. The California Assessment Program's 1984 pilot of the direct assessment of writing gave students passages from science, social science, or literature and asked them to use information at various levels of interpretation, including cause-effect relationships; evaluating strengths and weaknesses of a position; and inferring character traits, atmosphere, and themes (Quellmalz, 1985).

2. *Redress the imbalance between tests of component skills and their integration.* A set of comprehension questions, for instance, could contribute to inducing a theme or generalization rather than asking for unrelated details. Students would be required to sustain a line of reasoning to link multiple steps or sets of information in order to draw a conclusion.

3. *Design tasks that permit multiple interpretations or solutions.* For example, "Was Jack greedy or curious when he kept going back up the beanstalk?" Even in closed formats, students can check off sets of information and explanations that relate to one conclusion rather than another.

4. *Design open formats in which students explain their reasoning.* Single-sentence answers as well as complete essays can reveal plausible explanations or conclusions that a test writer might not have considered. The essay format is especially useful for assessing how students reach and explain their conclusions. Large-scale assessments have tended to avoid essays both because of their expense and questionable scoring reliability. Advances in writing assessment methodology, however, illustrate that high rater agreement levels can be achieved (Quellmalz, 1984). The California Assessment Program's pilot of direct writing assessment found that readers required only 1.5 minutes to score one- to two-page essays in which they interpreted subject-matter passages (Quellmalz, 1985). Further, the rater training discussions of such topics as quality of support serve as invaluable staff development experiences. Partially constructed response formats are also promising. Students tested on the Connecticut Assessment of Educational Progress were given the opening and closing statements of arguments and asked to provide strong support (Baron, 1984).

5. *Build tasks that represent a range of generalization and transfer.* In a subject-matter test, problems calling for comparisons and analyses of salient components could ask students

John McNamara

"Assessments of higher-order skills must be clear, valid, and coordinated if teachers and students are to trust, understand, and use the information they yield."

to compare two stories, characters, or settings. Tasks on a generalized thinking test could include comparisons of two accounts of an accident (life task), the Battle of Lexington (subject matter), and an incident on a space station (novel task). The rationale is both to test for generalization and transfer and to illustrate to teachers and students the relevant contexts for transfer.

6. *Develop tasks to assess metacognitive skills.* For example, given a problem, what are some ways to plan? Given a description of a problem-solving process, how might it have proceeded more efficiently? Again, such tasks both assess the process and serve as model activities for teachers and students.

The typical test development process involves producing items, field-testing them on large numbers of students, and then selecting and revising items with desirable psychometric properties. Given the complex nature of reasoning skills, development of assessment tasks is likely to require significantly more time and pilot testing. Collecting protocols, interviewing, and observing will be essential to understanding the background knowledge and strategies students use on tasks. Constructed responses can also supply some of this information.

Development efforts would also benefit from companion staff development and instructional interventions. Many educators and citizens will question whether reasoning items are testing what *is* being taught, what *can* be taught, or what *should* be taught. Instructional interventions can address these concerns and provide additional input to the design process.

Conclusion

For assessments of higher-order thinking skills to be relevant to instruction, the rationales must be clear for the skills framework, the assessment task designs, and the plan for task coordination. The cognitive processes of analysis, comparison, inference, and evaluation seem to appear in the major conceptualizations of problem solving, critical thinking, and intellectual performance. These processes are, therefore, potential fundamental skills to include in an assessment framework that merges psychological and philosophical views of essential reasoning strategies.

Subject-matter achievement tests present few items that measure higher-order skills, and those items tend to require fractionated reasoning skills. Even specialized thinking tests limit reasoning to tasks that seek one right answer and allow no opportunity or requirement for explanation. We need to develop a broader range of tasks that present significant problems, require sustained reasoning, and require component information to be integrated to form a conclusion. Further, we should permit multiple interpretations, develop constructed response formats, and design tasks that illustrate desired ranges of generalization and transfer. Finally, test development should include protocol analyses of students' reasoning and be accompanied by staff development and instructional studies.

Assessments of higher-order skills must be clear, valid, and coordinated if teachers and students are to trust, understand, and use the information they yield. Clearly, the teaching and testing of reasoning must be reasonable.□

References

Baron, J. B. "A Holistic Eye Toward Analytic Assessment." *Educational Measurement: Issues and Practices* 3 (1984): 15–18.

Bloom, B. S., ed. *Taxonomy of Educational Objectives Handbook: Cognitive Domain.* New York: McGraw-Hill, 1971.

Bloom, B. S. "Learning for Mastery." *Evaluation Comment* 1, 2. Los Angeles: University of California, Center for the Study of Evaluation, 1968.

Bruner, J. S. *Toward a Theory of Instruction.* New York: W. W. Norton & Company, 1966.

Covington, M. V. "Promoting Creative Thinking in the Classroom." *Journal of Experimental Education* 37, 1 (1968).

Dewey, J. *How We Think.* Boston: D. C. Heath & Company, 1933.

Ennis, R. H. "An Alternative to Piaget's Conceptualization of Logical Competence." *Child Development* 47 (1976): 903–919.

Ennis, R. H. "A Concept of Critical Thinking." *Harvard Educational Review* 32 (1962): 81–111.

Ennis, R. H. "Large-Scale Assessment of Critical Thinking in the Fourth Grade." Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1985.

Gagne, R. M. *The Conditions of Learning*. 3rd ed. New York: Holt, Rinehart and Winston, 1977.

Gagne, R. M., and Briggs, L. J. *Principles of Instructional Design*. 2nd ed. New York: Holt, Rinehart and Winston, 1979.

Guilford, J. P. "The Structure of Intellect." *Psychological Bulletin* 53 (1956): 267–293.

Linn, M. C. "Theoretical and Practical Significance of Formal Reasoning." *Journal of Research in Science Teaching* 21 (1982): 235–254.

National Assessment of Educational Progress. "Reading, Thinking, and Writing: Results from the 1970–80 National Assessment of Reading and Literature." Denver, Colo., 1981.

Newell, A., and Simon, H. A. "Computer Science as Empirical Inquiry." *Communications of the ACM* 19 (1976): 113–126.

Phillips, D. C., and Kelley, M. E. "Hierarchical Theories of Development in Education and Psychology." *Harvard Educational Review* 45 (1975): 351–375.

Quellmalz, E. S. "An Investigation of Scoring and Prompt Variations." Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1985.

Quellmalz, E. S. "Successful Large-Scale Writing Assessment Programs: Where Are We Now and Where Do We Go from Here?" *Educational Measurement: Issues and Practices* 3 (1984): 29–32.

Smith, B. O. "The Improvement of Critical Thinking." *Progressive Education* 30 (1953): 129–134.

Sternberg, R. J. "What Should Intelligence Tests Test? Implications of a Triarchic Theory of Intelligence for Intelligence Testing." *Educational Researcher* 13 (1983): 5–15.

Suchman, J. R. "Inquiry and Education." In *Teaching Gifted Students: A Book of Readings*. Edited by J. J. Gallagher. Boston: Allyn and Bacon, 1965.

Taba, H. "Teaching Strategy and Learning." *California Journal for Instructional Improvement* (December 1963): 3–11.

Edys S. Quellmalz is Acting Associate Professor, School of Education, Stanford University, Stanford, California 94305.