Assessing Critical Thinking in a Student-Active Science Curriculum

Neil Stillings
Mary Anne Ramirez

School of Cognitive Science
Hampshire College

Laura Wenk
School of Natural Science
Hampshire College

Presented at the
1999 NARST Annual Meeting
Boston, MA
March 31, 1999

Abstract

The desired student outcomes of a mature inquiry-oriented college science curriculum were identified by conducting and analyzing a series of faculty interviews. Both motivational and cognitive-intellectual learning goals were identified. The present study concerns two classes of target cognitive skills: skills involved in the cycle of scientific inquiry and skills involved in constructing quantitative models and interpreting quantitative data. A paper-and-pencil inventory consisting of open-ended questions about simple scientific scenarios was constructed to assess these skills. The effects of one semester of inquiry-oriented instruction on the skills was assessed by administering the inventory pre- and post-semester to a group of students that took inquiry-oriented science courses, a group that took no science courses, and a group that took more traditional biology courses. The scores of students in the inquiry courses improved significantly, while the scores of the other three groups showed no change. The results suggest that inquiry-oriented science instruction can improve students' ability to reason scientifically about issues that are outside the domain of instruction.

Assessing Critical Thinking in a Student-Active Science Curriculum

Neil Stillings, School of Cognitive Science, Hampshire College
Mary Anne Ramirez, School of Cognitive Science, Hampshire College
Laura Wenk, School of Natural Science, Hampshire College

Characterizing the Desired Student Outcomes of an Inquiry-Oriented Curriculum

Institutional Context

This study is part of a larger research project on science teaching and learning at Hampshire College in Amherst, Massachusetts. The College opened in 1970 following several years of planning that was influenced by the higher education reform ideas of the 1960s. Since its inception the college's educational program has been characterized by narrative and portfolio evaluation, individualized plans of study, highly interdisciplinary curricular organization, and an emphasis on putting knowledge to work in the world.

The founders of the college also articulated a commitment to teaching and learning skills of inquiry through hands-on practice. The emphasis on hands-on inquiry learning was intended to apply to all fields and to all levels of instruction. Every student, for example, is required to complete an undergraduate thesis project in the final year. Also, for many years, the entire curriculum for the first three semesters was organized around a concept of modes of inquiry, under which students were required to acquire and demonstrate skills in four broad areas of inquiry: natural science, social science, humanities & arts, and communication & cognitive science.

Although the first-year program at Hampshire has changed in some ways, the inquiry requirement in the sciences has remained a part of the curriculum. The requirement is overseen by the School of Natural Science (known as NS), which houses the entire faculty in mathematics and the physical and biological sciences, comprising about twenty

---

This research was supported by a grant from the Learning and Intelligent Systems program of the National Science Foundation (Grant No. REC-9720363). It was also supported by grants from the NSF's Institution-Wide Reform Program and from the Howard Hughes Medical Institute. We are grateful for the research assistance of Barbara Reyes. Portions of this paper will be presented at the 1999 annual conference of the National Association of Research in Science Teaching, Boston, MA. Correspondence should be addressed to Neil Stillings at nstillings@hampshire.edu, or School of Cognitive Science, Amherst, MA 01002.

people. Students can fulfill the requirement by taking inquiry-oriented introductory courses or by doing independent projects under the supervision of individual faculty members.

First-year courses in NS typically involve students in realistic, often original, laboratory or field investigations or in reading the primary literature in a research area. For example, in a course titled *How People Move* students learn to design and conduct original studies in an electromyography lab. In *Aquatic Ecology* they conduct field investigations of a Cape Cod salt marsh and experimental studies in large laboratory fish tanks. These examples suggest a specific meaning in the Hampshire context of the notions of inquiry and student-active or hands-on science learning. This meaning is roughly captured by a common insistence of faculty members that students should work on problems to which the instructor does not know the answer. The Hampshire approach thus potentially contrasts with carefully-crafted constructivist curricula in which students rediscover established theories or knowledge.

The faculty has been challenged by its commitment to bringing an inquiry-oriented curriculum to the entire population of first-year students, who vary considerably in academic skills, scientific background, and initial motivation. The curriculum and approaches to classroom teaching in NS developed largely locally. The rationales offered for the inquiry approach and the faculty's classroom practices bore interesting resemblances, however, to constructivist K-12 reform curricula, as well as to inquiry-oriented experiments at other colleges.

Over the past 10 years the NS faculty has increased its professional involvement with educational change and research. These activities have included efforts to export Hampshire's approach to science education, outreach to local schools, collaboration with other reform and research groups, efforts to reflect on and incorporate teaching practices that have been adopted in other reform curricula, and attempts to assess classroom practices and student outcomes at Hampshire.

Research Context

The present study is part of a collaborative project to begin to characterize the philosophy, classroom practices, and student outcomes of introductory science instruction at Hampshire. Because it is possibly the most radical and mature reform curriculum in American higher education, the Hampshire experience should hold many lessons for those seeking to change college science teaching and learning. Hampshire's mix of local culture and interaction with national reform movements makes it possible to address many issues in a single context.

Our research is driven to a significant extent by the learning goals that have been articulated by the science faculty. We seek to understand what these goals are, how they are related in theory and in actuality to instructional practices, how they can be

conceptualized in terms of current psychological theory, and, finally, whether the goals are being met in the form of appropriate student outcomes.

Our work draws on a survey of publications by the NS faculty and a series of extended, structured interviews with a sample of faculty members. It is also tied to five goals and student-assessment criteria for beginning instruction, which were formally adopted by the science faculty: (1) Active engagement in, and ownership of, the work; (2) Understanding of the scientific process, including critical awareness of the limitations and strengths of scientific methods; (3) Seeing the work in a larger social and political context; (4) Use of quantitative methods; (5) Oral and written expression and use of the scientific literature.

The Motivational Goals of an Inquiry Curriculum

Although the present study concerns the cognitive-intellectual goals of introductory instruction, we note here that faculty members also stress motivational-attitudinal goals that they believe are best achieved through inquiry-oriented instruction. Early, active experience with scientific inquiry is said to increase interest in science, to increase the interest and participation in science of women and members of under-represented minority groups, and to increase students' confidence that they can do science or participate as citizens in science-related policy debates.

These motivational goals are not fully expressed in the formal list of five goals, which are oriented toward student assessment, but the first formal goal of *active engagement* does capture some of the motivational thrust of the curriculum.

Some statistical data show that the percentage of Hampshire students who graduate in the sciences is indeed higher than the percentage of entering students who express an interest in science, and the percentage of Hampshire students who go on to scientific careers compares favorably with other institutions in national samples. It is not clear at this point to what degree the inquiry aspect of the beginning curriculum is responsible for these results.

The Cognitive Goals of an Inquiry Curriculum

This study focused on the cognitive -intellectual goals of the Hampshire inquiry approach, which we broke into several broad categories. The categories arose from a content analysis of faculty interviews and probably do not refer to mutually-exclusive cognitive outcomes. They map quite well onto the formally-adopted list of five goals, which proved to have broad assent among individual faculty members. The relationships between the goals and the instructional strategies pursued by the faculty are complex and are not addressed in this study.

Cognitive Skills

Many of the goals mentioned by faculty members constitute cognitive skills in the sense that students are expected to acquire some of the active skills of a working scientist at a rudimentary level. Several interrelated clusters of cognitive skills emerged in our analysis.

The Inquiry Cycle. Several of the skills are related to the cycle of inquiry that is typical of scientific research and that is often cited as a foundation of reform-oriented curricula: Question-theory-hypothesis formation; research design; data gathering; data analysis & interpretation; theory-hypothesis critique and reformulation. The expectation is that by experiencing the doing of science students will learn to some degree how to engage in the various phases of the inquiry cycle. For example, they will be able to generate or recognize researchable questions or models and have some ability to generate or critique research designs. They will acquire some ability to organize and analyze data and to see the implications of data for hypotheses or theories. The faculty's formal list of instructional goals and student-assessment criteria includes *understanding the scientific process*.

Quantitative skills. A cluster of quantitative skills, which are related to the inquiry-cycle skills, emerged somewhat separately. One kind of quantitative skill refers to the descriptive and statistical handling of results, including the graphical presentation of data and some ability to deal with variability and error. An overlapping set of skills involves setting up and reasoning with quantitative models, either to make predictions or to interpret data. *Use of quantitative methods* is listed as a separate goal on the faculty's adopted list.

Primary literature skills. Teaching students to locate and read primary literature is a distinct goal of the curriculum (part of formal criterion 5). To some extent this goal follows from the emphasis on the inquiry cycle. Conducting research often requires finding out what other researchers have done in an area. Reading primary literature can also help students understand and begin to master the active process of research, which is almost never featured in textbook presentations. Finally, the critical review of literature is an independent ability that is critical in making policy decisions.

Scientific communication skills. A final cluster of skills concerns scientific communication (part of formal criterion 5). Students work on presenting their ideas in writing and in stand-up presentations. They comment on each others' work. They work collaboratively in group research projects.

This list of desired inquiry skills seemed to us to be thoughtful and clearly related to goals stated by other reform-oriented science educators.

Epistemological Stance

In addition to building cognitive skills the inquiry approach is intended to instill philosophical or epistemological views about the nature of science and the status of

scientific knowledge. The philosophy of science is contested territory, and a variety of defensible philosophical contexts could be constructed for the collection of inquiry activities listed above. Appropriately, perhaps, we found that our group of faculty members was not urging a full-blown or fully-articulated philosophy of science on students. Rather, they were hoping that students' beliefs, or epistemological stances, would move in certain general directions.

We divided the epistemological concerns into two categories, which concern the philosophy and sociology of science. This division reflects the content of faculty interviews overall and does not represent a claim on our part or by individual faculty members that philosophical and sociological issues are easy or even possible to separate.

On the philosophical side faculty members expressed a hope that students would shift from relatively naive to more sophisticated views of science. In the naive view science is a fixed body of unassailable knowledge possessed by scientific authorities, and the scientific method is a fixed procedure that can be applied piecemeal to natural phenomena to yield reliable knowledge. More sophisticated views recognize the uncertain, open-ended, and theory-laden nature of scientific knowledge, and they show a more nuanced appreciation of the uncertainties, details, and importance of the research process and the evidence it produces. The second formal goal, particularly its second clause, *critical awareness of the limitations and strengths of scientific methods*, suggests the larger epistemological goals of the curriculum.

It is worth noting here the assertion by the faculty that they see many beginning undergraduates who hold static and authority-based views of science, since reform-based K-12 curricula are in part also aimed at changing such views. Several faculty members also noted that shifts in students' epistemological stances can be complicated, sometimes including, for example, a radically relativistic phase.

On the sociological side the hope is that students will move toward a deeper appreciation of the social nature of the scientific enterprise. Scientific research is carried out within a social context by groups of researchers and it is therefore influenced by a complex of historical, cultural, political, and social factors. Further, the pursuit of scientific research and development has political and moral dimensions. The third formal goal, *seeing the work in a larger social and political context*, captures this concern.

Field-specific content

The above goals are stated very generally, but they are meant to be achieved in courses in particular scientific fields and often on rather specific topics within those fields, as is shown by the example of the courses on muscle electrophysiology and aquatic ecology cited above. Considerable field-specific knowledge is taught and learned in these courses. Faculty members report consistently, however, that in these courses content mastery is a means in the service of the more general ends sketched above. It is of some interest that

the formal list of five instructional goals, which is used in student assessment, does not explicitly mention the mastery of field-specific knowledge.

The de-emphasis of content mastery raises two important questions. First, it might be asked when and how students are going to achieve thorough content mastery if this is not a goal of introductory courses. The faculty expects students who go on in science to achieve during their concentrations the content mastery necessary to support an undergraduate thesis project. This is sometimes described as the diamond-shaped model of an undergraduate career. It begins with a focused exploration of a specific topic area which allows the student to experience the process and values of the research enterprise. This exploration is followed during the concentration by a broad exposure to and mastery of a field. The hope is that the earlier research experience will deepen students' encounters with textbook knowledge, leading them to think critically about its origins in the research process and to seek out primary sources. Finally, in the thesis project the student returns to do a more sophisticated research project.

Some students who do not go beyond the introductory curriculum in science may never achieve the kind of content mastery that is typical in conventional introductory courses. For example, a student in How People Move will probably not encounter the material on the circulatory system that might occur in a traditional introductory biology or physiology course. Faculty members feel that an appreciation of the scientific research process is more valuable than this kind of textbook knowledge. For example, a faculty member might say that a student armed with this appreciation and the associated confidence can always look up textbook information on the circulatory system and, further, will be in a position to critically evaluate current policy-related controversies about the causes and treatments of disease.

The second, deeper, question is whether, or to what extent, general scientific skills can exist independently of field-specific knowledge. This is a central topic of cognitive psychology and is taken up briefly in the following section.

Issues of Generality and Transfer

Introductory science instruction at Hampshire implicitly expresses a hypothesis that the human mind is capable of acquiring general scientific thinking skills that are not tied to specific scientific domains or areas of content. In addition, it constitutes a hypothesis that from one to two semesters of experience with inquiry-oriented, student-active instruction, often in highly focused research domains, can lead to significant general gains in critical thinking skills.

In this study we adopted a straightforward empirical approach, attempting to develop a measure for scientific thinking skills and to use it to assess the outcomes of a single semester of instruction. Our focus in the remainder of the paper is the description of the approach and the results. However, we must mention at the outset that the theoretical background of the approach is complex and unsettled.

Within modern psychology debate about the existence and nature of general mental skills extends at least back to Thorndike's claim in the early 1900s that the transfer of learning across tasks is mediated by identical elements and his consequent attack on the doctrine of formal discipline, which had been the foundation of liberal learning theory since the 18th Century and, arguably, since the time of Aristotle. Formal discipline was defended by many psychologists and educators who were contemporaries of Thorndike's, including Angell and Meiklejohn (a summary can be found in Singley & Anderson, 1989).

Since then, the major theories and bodies of empirical research on cognition and learning have had significant implications for the idea that people can acquire highly general or abstract thinking skills. Contemporary cognitive psychology presents a particularly complex picture. On the one hand there is abundant and diverse evidence for lack of generality in cognitive skills. The sources include failures of empirical support for Piaget's theory of development (Siegler, 1986), people's poor performance on some deductive reasoning tasks (Wason & Johnson-Laird, 1972), and the discovery of the limitations of Newell & Simon's theory of problem solving (GPS) and of the domain-specific nature of much expert knowledge (Larkin et al., 1980). On the other hand there are theoretical positions and empirical findings that support generality. The relevant theoretical ideas include claims that analogical mappings between representations can support transfer (Vosniadou & Ortony, 1989), the argument that schemas or production rules can contain variables at high levels of abstraction (Singley & Anderson, 1989), and theories of metacognitive capacities (Flavell, 1985; Astington, Harris, & Olson, 1989). These theories enjoy some empirical support.

Particularly relevant to the current context are Nisbett and colleagues' success in teaching the general application of the law of large numbers to situations involving variability (Nisbett, 1993), and Schoenfeld's (1985) success in teaching the use of general heuristics in mathematical problem solving. The law of large numbers and mathematical heuristics (derived from Polya's work) are at the level of generality that science faculty members stressed in our interviews with them.

A Preliminary Study of the Acquisition of Scientific Thinking Skills

In this study we began to develop and test a paper-and-pencil instrument to assess general scientific reasoning skills. Although we are also using finer-grained methods, such as clinical interviews and thinking-aloud protocols, we felt that a paper-and-pencil instrument would be an efficient, and hence invaluable, tool in further research.

One of the main purposes of the study was to explore the properties of the instrument and the feasibility of using it in a small-college setting. Assuming that we were able to develop a reasonably effective instrument, however, our analysis of the faculty's instructional goals and implicit educational theory entailed two hypotheses. First, students enrolled in science courses should show an improvement in scientific thinking

skills from the beginning to the end of the course. Second, the gains of science students should be greater than those of comparable students who have not taken a science course or who have taken a more traditional science course that does not stress inquiry skills.

Method

Item Development

We wrote a set of questions based directly on the goals extracted from faculty interviews. Each question presented a scenario and then asked the subject a series of open-ended, short-answer questions about the scenario that were related to the instructional goals. The scenarios were non-technical, and the questions were designed to engage general scientific reasoning skills.

Since any scenario has specific content, a subject with relevant technical knowledge could use it to answer the questions in a way that obviated the need to exercise general cognitive skills. However, our scenarios were sparse enough and varied enough that we doubted that this would occur on all questions for any individual or frequently across a population of subjects.

Since it was difficult to engage the full range of target skills in a single question, items were developed for three rough categories, which engaged different areas of skill. The first kind of item tapped the subject's ability to generate questions about a situation and to begin to turn the questions into researchable hypotheses or experimental designs. The second kind of item presented some data or predicted data for a scientific study and asked the subject to critically analyze the relationship between the data and a stated hypothesis or range of hypotheses. Such questions could engage the logic of experimental design or the qualitative properties of quantitative data (for example, thinking about whether a hypothesis predicts a linear or curvilinear bivariate relationship). A third kind of item presented a scenario in which scientists disagree about, or have obtained conflicting results about, some phenomenon. The subject is asked talk about how such a disagreement might come about and what might happen next. Due to a procedural difficulty, results from this third type of item were not analyzed for this study. Sample items from the first two categories appear in Appendix 1.

Early versions of some items were critiqued by faculty members in a teaching workshop. Faculty members also gave suggestions for further items. A group of items were pilot-tested on a sample of students enrolled in summer school science classes at the University of Massachusetts. Items that failed to elicit clear attempts at scientific reasoning were reworded or eliminated. Five items were selected for this study.

Development of a Scoring Procedure

A scoring procedure was developed that rated the quantity and logical quality of subjects' reasoning. For example, on a question that asked subjects to generate testable hypotheses, answers were awarded points both for the number of hypotheses generated and for hypotheses that were stated in a testable form. The procedure was refined to a simplified rubric by two researchers working on a sample of pilot data. Working independently, they then achieved 90% agreement on a separate sample of pilot data. Samples of the scoring protocol appear in Appendix 1.

Research Design

Performance on the critical-thinking questionnaire was assessed in the Fall 1998 first-year class at Hampshire College. A large sample of the incoming class was assessed on orientation day (n=187), prior to the beginning of the term, yielding a *pre-* measure. A second sample of the class was assessed at the end of the term (n=80), yielding a *post-* measure. From these two samples, we formed a smaller sample of students who had taken both the pre- and the post-test. We then split this sample into two groups: students who had been enrolled in a science course during the term (n=43) and students who had not been enrolled in a science course (n=17). The resulting data can be seen as a 2 x 2 factorial design, with a within-subjects factor of time-of-test (pre vs. post), and a between-subjects factor of enrollment-status (science vs. non-science), with the non-science students serving as a control group. All subjects answered the same questions, which were presented in the same order, on both the pre- and the post-test.

As an additional control we also administered pre- and post-tests to a sample of students enrolled in two introductory classes at another college. One of these classes was a traditional introductory biology class (n=18) with little emphasis on the kinds of inquiry skills described above. The other class (n=33), an interdisciplinary general science course, blended traditional instruction with some emphasis on the nature of science and on inquiry-oriented reforms.

Procedure

The Hampshire pre-test was administered at an orientation session, which the entire fall 1998 incoming class was invited to attend. A questionnaire on attitudes toward science was administered at the same time. Students were informed that they were being asked to participate in an ongoing study of science education at the college, that they might be asked to participate again in the future, that they had a right not to participate, and that their data would be kept anonymous. They were given one-half hour to complete the questionnaires.

During the last two weeks of the term the Hampshire post-test was administered in meetings of a broad sample of science and non-science classes. After a brief statement of the purpose of the research, students were given one-half hour to complete the questionnaires.

At the comparison college the pre- and post-tests were administered at the beginning and end of the term in class.

Answers were scored by two raters, who achieved 85% agreement on a sample of questionnaires that each coded independently.

Results

Mean raw scores on the pre- and post-tests for the three groups of subjects are presented in Table 1. The logically possible range of raw scores on the four questions that were analyzed was 0 to 30, while the actual range across all of the data was 4 to 26. Fifty per cent of the scores fell between 13 and 19, and 90 per cent fell between 9 and 20. The results were not affected by answers that displayed field-specific technical knowledge (For example, the perfume question in Appendix 1 could be answered in terms of current knowledge of the physiology, genetics, and psychophysics of the olfactory system, but no one answered in this way).

The difference between the pre- and the post-test scores for each of the four groups in Table 1 was assessed with a t-test. The scores of students in inquiry science classes increased significantly ($t(42)=2.55$, $p<.015$). The other pre- vs. post- differences were nonsignificant.

The questionnaire contained five questions, which were presented in the same order on both the pre- and post-tests: two questions tapped hypothesis generation, two tapped data interpretation, and one tapped the interpretation of scientific controversies. A substantial number of subjects failed to complete the controversies question. That question was therefore dropped, and the analysis above is based only on the hypothesis generation and data interpretation questions.

Table 1:
Mean critical-thinking raw scores for four groups of students

Group	Pre-test	Post-test
Students in inquiry science classes	15.53	17.21
Control: students not enrolled in science classes	15.12	14.94

Comparison institution: Students enrolled in traditional biology class	14.73	15.45
Comparison institution: Students enrolled in reform science class	15.61	14.38

The increase for students enrolled in inquiry courses occurred mainly on two questions. The first was the *perfume* question, reproduced in Appendix 1. This question taps students' ability to generate researchable hypotheses. The second question was the *eggs* question, also reproduced in Appendix 1. This question taps students' ability to see relationships in simple graphical displays of data and to relate them to hypotheses, research procedures, and elementary statistical considerations.

At the end of both the pre- and post-tests, a small number of students turned in their questionnaires without completing four questions. This source of attrition was not distributed unequally across groups, however.

Discussion

The critical thinking scores of first-semester students enrolled in inquiry-oriented science courses improved more than the scores of students at the same institution who did not take science courses, and more than students at the comparison institution who were enrolled in science courses with a more traditional orientation. These results confirm the hypothesis that one semester of inquiry-oriented instruction can have differential effects on general scientific reasoning ability.

Although the positive change in the inquiry group was statistically significant with a substantial effect size, the increase in mean raw score was modest. We have not tested a sample of working scientists, but we would expect their scores to be over 20 and probably in the range of 25 to 30. The increase from 15-plus to 17-plus could reflect the modest effects of one semester of instruction or characteristics of our methodology. Studies with more advanced students, a larger population of items, and revised scoring procedures will help answer this question.

The results must also be weighed against the objective expectations of the faculty that designed the curriculum. Our hypothesis was that inquiry instruction would lead to some increase in critical thinking performance. The faculty may conclude that it seeks a larger gain, although it should be noted that while our measurements were carried out over the course of the first semester, many students work beyond the first semester to meet the five assessment criteria, either by completing research projects or by taking a second course.

The results are potentially compromised by our inability to randomly assign students to the treatment groups. Although the pre-test scores of the two Hampshire groups did not differ, it might be argued that the students who elected science courses in their first term were better prepared to benefit from inquiry-oriented instruction. Since all Hampshire students are required to meet the natural science requirement, this problem could be largely remedied by following a student cohort throughout their experience with the introductory curriculum.

When the results from the comparison college are taken into consideration, we must consider not only the fact that the students in the two biology courses were self-selected, but also the fact that subjects were not randomly assigned to colleges. The potential effects of these factors is mitigated by the statistical equivalence of the pre-test scores in all four groups. This question could also be addressed by partialing out the effects of certain concomitant variables such as high school GPA, science background, and SAT scores.

The use of identical questionnaires for the pre- and post-test was a potential source of problems. The fact that the scores of three of the four groups did not change statistically eliminates a number of the possible problems. It could be argued, however, that Hampshire science students, the group that improved, had some unique opportunity or motivation to think over their original answers during the term or to discuss them with others. We have no evidence of this and received no comments from subjects about repeating the questions. In the future we plan to develop a larger population of items, which will support the development of alternate forms.

A final question about the positive results for students in inquiry-oriented courses is whether the gains in critical thinking will persist over time.

Overall, the results of this study suggest that an inquiry-oriented approach to beginning college science instruction can produce increases in general scientific reasoning ability. They suggest that further theoretical and empirical work is warranted.

References

Astington, J. W., Harris, P. & Olson, D. (1989). *Developing Theories of Mind*. New York: Cambridge Univ. Press.

Flavell, J. H. (1985). *Cognitive Development, 2nd Ed*. Englewood Cliffs, NJ: Prentice-Hall.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Nisbett, R. E. (1993). *Rules for Reasoning*. Hillsdale, NJ: Erlbuam.

Schoenfeld, A. H. (1985). *Mathematical Problem Solving*. New York: Academic Press.

Siegler, R. S. (1986). *Children's Thinking*. Englewood Cliffs, NJ: Prentice-Hall.

Singley, M. K. & Anderson, J. R. (1989). *The Transfer of Cognitive Skill*. Cambridge, MA: Harvard.

Vosniadu, S. & Ortony, A. (1989). *Similarity and Analogical Reasoning*. New York: Cambridge Univ. Press.

Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard.

Appendix 1: Sample Items & Scoring Rubric

Category 1: Question and hypothesis generation

Sample question:

Two people are sitting in a room at equal distances from a bottle of perfume. After the bottle is opened, one person smells the perfume, and the other person does not.

A. Write a list of questions that occur to you about the statement (that is, the reasons one person smells the perfume and the other does not).

B. Put a star (*) to the left of the question above that you think would yield the most fruitful, testable hypothesis.

C. For the question you chose in part B above, write a well-formulated hypothesis that could actually be investigated.

Scoring rubric:

Subpart	**Score**	**Explanation**
A	3	Greater than 3 logical questions
	2	3 logical questions (possibly some illogical)
	1	1 or 2 logical questions (possibly some illogical)
	0	Illogical questions or no answer
B	1	Testable hypothesis
	0	Untestable or no hypothesis
C	3	Hypothesis stated in a testable form
	2	Hypothesis not specific to this situation. States a question or experiment rather than hypoth. Modal statement (may be, could be, perhaps)

Category 2: Analysis of hypothesis-data relationships

Sample question:

It has been observed that some birds living close to the equator lay fewer eggs than birds living farther from the equator. A team of scientists wanted to test the hypothesis that there would be a linear (or straight-line) relationship between the average number of offspring that ducks have at a given time and their latitude (or number of degrees north of the equator). They counted the number of eggs produced by individuals of the same

species of ducks at different latitudes and calculated the average number of eggs per individual duck. The data are listed below and are also given in a graph.

Latitude (degrees)	Average No. of eggs per duck
0	6.4
15	8.9
22	10.3
30	10.9
42	12.3
48	12.2
61	12.0
71	11.8

INSERT FIGURE HERE

A. What is the relationship between the number of eggs laid and the latitude? That is, is it linear (straight line)? How would you describe it? Be as specific as possible.

B. What could you do to further support your statement about the mathematical relationship between latitude and number of eggs laid per duck?

Scoring rubric:

Subpart	**Score**	**Explanation**
A	4	Linear to 40 degrees then plateaus or falls off
	3	Increases to 40 degrees then plateaus or declines
	2	Increases then goes down
	1	Increases farther away from equator
		Decreases after 40 degrees
	0	Illogical or no answer
B	2	Replicate to get more reliable data
		Look at statistics of differences
	1	Repeat in So. hemisphere
		Repeat experiment; no explanation why
	0	Illogical or no answer