DOCUMENT RESUME

ED 283 840 TM 870 352

AUTHOR	Duran, Richard P.; And Others
TITLE	Construct Validity of the GRE Analytical Test: A Resource Document.
INSTITUTION	Educational Testing Service, Princeton, NJ. Graduate Record Examination Board Program.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-RR-87-11; GREB-PR-81-6P
PUB DATE	Apr 87
NOTE	108p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Abstract Reasoning; Academic Achievement; *Cognitive Ability; *Cognitive Processes; *Cognitive Tests; College Entrance Examinations; *Construct Validity; Critical Thinking; Educational Research; Graduate Study; Higher Education; *Language Skills; Literature Reviews; Logical Thinking; Measurement Techniques; Problem Solving; Psychometrics
IDENTIFIERS	*Analytical Tests; *Graduate Record Examinations

ABSTRACT

This study planned further research aimed at understanding how the current version of the Graduate Record Examinations (GRE) analytical measure reflects cognitive skills that are important to success in graduate education. Several activities were completed, including selective literature reviews of: (1) cognitive and psychometric research on reasoning; and (2) more general educational research on the role of reasoning in higher education. The two item types included in the current test (analytical reasoning and logical reasoning) were analy ed to determine more precisely the kinds of basic skills that underlie their solution. Based on these activities a summary was made of the major thinking and language skills that seem to be required for successful performance on the test. Two major conclusions were reached: (1) although there is a considerable body of research on reasoning, thinking, and problem solving, there are few well-developed models or theories to guide the measurement of analytical ability; and (2) although reasoning, critical thinking, and other such processes are assumed to be requirements of successful academic performance, there is little documentation of the specific involvement of these skills in graduate education. Based on these general conclusions and other more specific findings, further work has been planned. A 10-page bibliography follows the report. A simulated solution process for one analytical reasoning item is appended. (Author/JAZ)

GRE®

GRADUATE RECORD EXAMINATIONS

CONSTRUCT VALIDITY OF THE GRE ANALYTICAL TEST: A RESOURCE DOCUMENT

Richard Duran
Donald Powers
Spencer Swinton

GRE Board Professional Report No. 81-6P
ETS Research Report 87-11

April 1987

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

Construct Validity of the GRE Analytical Test:

A Resource Document

Richard Duran
Donald Powers
Spencer Swinton

GRE Board Professional Report No. 81-6P

April 1987

Educational Testing Service Princeton, NJ 08541

Abstract

The purpose of the study reported here was to plan further research aimed at understanding how the current version of the GRE analytical measure reflects cognitive skills that are important to success in graduate education. Several activities were completed, including selective literature reviews of (a) cognitive and psychometric research on reasoning and (b) more general educational research on the role of reasoning in higher education. The two analytical item types included in the current test were analyzed to determine more precisely the kinds of basic skills that underlie their solution. Based on these activities a summary was made of the major thinking and language skills that seem to be required for successful performance on the test. Two major conclusions that were reached are that (1) although there is a considerable body of research on reasoning, thinking, and problem solving, there are few well-developed models or theories to guide the measurement of analytical ability and (2) although reasoning, critical thinking, and other such processes are assumed to be requirements of successful academic performance, there is little documentation of the specific involvement of these skills in graduate education. Based on these general conclusions and other more specific findings, further work has been planned and is proposed in a separate document.

Table of Contents

List of Figures

List of Tables

Construct Validity of the GRE Analytical Test: A Resource Document

The goal of the study reported here was to understand better how the current GRE analytical test reflects basic cognitive skills that are important to success in graduate education. This goal was to be achieved by (a) reviewing relevant cognitive and psychometric research in order to model the basic cognitive processes that underlie the solution of analytical ability test items and (b) reviewing research on the role of higher level thinking and reasoning skills in advanced academic performance. In addition to learning more about the current analytical measure from these activities, we also hoped to be better able to suggest fruitful directions for further research on the construct validity of the analytical measure.

The report begins (Section I) with a description of the analytical test, its purpose, and its current structure. Appendix A presents a simulated solution process for one analytical reasoning item. Section II contains a brief discussion of construct validity as it applies to the test and a summary of Board-sponsored research that relates to the construct validity of the measure. Section III lays the framework for two selective surveys of research on reasoning. A summary is provided of some psychological and cognitive science research that bears on the construct validity of the test and the identification of skills required by it. Section III also contains a discussion of some previous research on the role of reasoning in academic performance and also previous research on the analytical test. It also summarizes some research on the training and measurement of thinking skills. Section IV presents an overview of the major thinking and language skills likely to be required on the test, given the previous educational, psychological, and cognitive science research. The final section of the report (V) is a summary of findings, especially as they relate to planning further research on the analytical test. One of the objectives suggested for further research is to investigate how the reasoning skills measured by the current analytical test relate to the academic demands of graduate school and to the outcomes of graduate school training. Another goal is to identify academic demands and outcomes that are not represented in the current analytical test.

I. The GRE Analytical Test

The analytical test is intended to assess examinees' developed ability to reason with complex information as might be required in any area of graduate study. The test does not presume any specialized academic content knowledge on the part of examinees. The current version of the test is made up of two parallel 30-minute sections. Two item types are used: analytical reasoning and logical reasoning. Nineteen analytical reasoning and six logical reasoning items occur within each section. Within sections, items of each type are arranged roughly in their order of difficulty.

Analytical Reasoning Questions

Figure 1 displays an analytical reasoning item and associated questions taken from the GRE 1982-83 Information Bulletin. Questions occur in sets of three or more. Each set of questions is preceded by a description of a situation involving fictitious entities--persons, places, things, or events. Descriptions consist of three to seven statements. Descriptions mention relationships and constraints in relationships among the entities referred to in the description. The three or more questions which accompany a description are independent of each other. That is to say, each question poses a problem that has no relation to either information given in other questions or to the problems posed by other questions. Nonetheless, all questions in a set rely on the same situation description.

Figure 1

An Analytical Reasoning Item and Associated Questions

There are seven co-workers, three women (Fran, Gladys, and Helen) and four men (Juan, Karl, Leonard, and Mark) who eat lunch only at restaurants N, O, or P.

Fran goes to a restaurant only on Wednesdays.

Gladys and Karl are never able to go to a restaurant together.

Helen will not go to a restaurant unless Fran also goes.

Juan and Mark will not go to the same restaurant together unless Gladys also goes.

Karl will not go to restaurant O.

1. If six of the co-workers have lunch together at one of the three restaurants on a Wednesday, who of the following must be absent?

   (A) Gladys (B) Helen (C) Juan (D) Karl (E) Leonard

2. What is the greatest number of the co-workers who can go out to lunch together on Tuesday at one of the three restaurants?

   (A) 3 (B) 4 (C) 5 (D) 6 (E) 7

3. If Juan and Mark eat lunch with a group of the co-workers, which of the following must be true?

   I. The group can include no more than three men.
   II. The group can include only two women.
   III. The group cannot eat at restaurant O.

   (A) I only (B) II only (C) III only (D) I and II only
   (E) I and III only

The relationships mentioned in analytical reasoning descriptions include: ordering in terms of space, time, or other continuous variables; set membership determined by given criteria; or cause and effect. Clear and simple language is used to describe relationships. All the information required to answer questions is given in a description, though examinees must already be familiar with the meanings of common words occurring in the description and questions. An attempt is made to avoid terminology that is specific to logic as a discipline or to other specialized academic areas. Answering questions typically involves establishing fewer than six intermediate relationships based on what is already stated explicitly.

The difficulty of analytical reasoning questions is related to the number and complexity of different relationships that simultaneously enter a solution and, to a lesser degree, to the number of conditions or number or entities specified for a single relationship. All analytical reasoning questions have a unique answer based on a description. However, for some questions there may be alternative schemes for interpreting information so as to arrive at the single correct answer.

Two basic types of questions are developed for analytical reasoning problem sets. One type requires examinees to infer new information that must necessarily be true given an accompanying situation description. A second type requires inferences about the possibility of a state of affairs given a situation description.

The foregoing information was based largely on the GRE program description of items, but there are also some interesting independent observations that might be added. For instance, analytical reasoning items tend to have a strong combinatorial reasoning flavor. They are puzzle-like in nature and are deliberately constructed so as not to require knowledge of specific situations and events. Training in formal mathematics, computer science, and informal logic, or experience with other formalisms for representing information are not necessary for solving these items. It could be the case, however, that skills achieved in such areas might aid in the solution of this problem type; this is an issue that seems amenable to further research. A detailed analysis of a solution of the analytical reasoning item in Figure 1 is given in Appendix A.

Logical Reasoning Questions

These questions test the ability to comprehend and assess logical relationships among either statements in a prose passage or to answer options referring to a passage. Several questions may be associated with a prose passage, but for a given passage questions are independent of each other. Figure 2 displays a sample prose passage and two questions based on this passage. These materials are drawn from the GRE 1982-83 Information Bulletin (Educational Testing Service, 1982).

Questions are based on short prose passages (no more than 100 words in length) and are drawn from popular publications, political speeches, conversations, advertisements, and academic texts in the arts and sciences. Understanding passages and questions does not require specialized knowledge in any academic field, including logic. However, some of the materials and issues that these questions present would not be out of place in textbooks or course materials on rhetoric, critical thinking, or formal logic.

Logical reasoning questions test the general ability to understand and analyze arguments, and more specifically, skills in:

- recognizing the point of an argument
- recognizing assumptions on which an argument is based
- drawing conclusions from given premises
- evaluating arguments and counterarguments
- analyzing evidence

The skills addressed by logical reasoning questions have obvious relevance to the academic reasoning activities required of graduate students. The skills outlined feature the comprehension and manipulation of information of a complex conceptual nature. Language comprehension skills, particularly reading comprehension skills, are also critical to this item type. As will be discussed in more detail later, logical reasoning questions require more than basic comprehension skills; they require sophisticated mental representation of entailment relationships between premises of arguments and conclusions drawn from premises. When we consider advanced reading, it becomes difficult, if not impossible, to distinguish reading comprehension skills from the kinds of reasoning skills required to solve logical reasoning questions. Nonetheless, the detailed specifications underlying logical reasoning questions permit a well-rounded and workable method for probing examinees' reasoning skills. In contrast, GRE reading comprehension questions are not designed to systematically tap the same specific set of skills as logical reasoning questions.

Figure 2

Two Logical Reasoning Items

1. Mrs. Thomas boarded an airplane and saw another passenger sitting in her assigned seat. The passenger showed her his boarding pass, and, indeed, he was in the right seat. Mrs. Thomas turned to the flight attendant and exclaimed, "How do you explain the fact that my assigned seat is not available?" "Because, Madam," replied the flight attendant, "there has been a multiple assignment."

   The flight attendant's reply is an example of:

   (A) rephrasing rather than explaining a problem
   (B) confusing company policy with employee practice
   (C) arguing from inconsistent points of view
   (D) failing to provide specific support for a general assumption
   (E) presupposing what is to be proved

2. State Commissioner of Environmental Affairs: "There is no cause for alarm at recent disclosures that cancer-causing agents have been found in the air near seven chemical plants in the state. Such substances have been found in the air at almost all chemical, petrochemical, and pharmaceutical plants in the country."

   The Commissioner's conclusion depends on his assumption that:

   (A) experts in the Environmental Affairs department have determined that the substances are not harmful
   (B) most of the substances found in the air near the plants were organic compounds that are not linked to cancer
   (C) such substances are not causing significant harm at the other places where they have been found
   (D) substances used in chemical, petrochemical, and pharmaceutical manufacturing are harmless
   (E) similar findings would result if similar tests were made downwind of many industrial sites

II. Construct Validity and Analytical Ability

The evolution of the analytical measure has not been driven explicitly by a cognitive analysis of the thinking skills that ought to be included in the analytical section of the test, but rather by other more traditional psychometric concerns.

Development

The early development of analytical item types was concerned primarily with establishing the face validity, criterion validity, degree of speededness, difficulty, and reliability of various item types (Miller & Wild, 1979). Figure 3 displays descriptions of the various item types that were investigated.

Item types were sought which possessed desirable psychometric properties and which exhibited discriminant validity relative to performance on the verbal and quantitative sections of the GRE General Test. The development plan for the analytical ability section focussed on four questions (Miller & Wild, 1979, p. 49):

1. Will the item types be of appropriate difficulty, reliability, and speededness?
2. Will the item types measure skills that are relatively independent of the verbal and quantitative ability measures?
3. Will the item types have criterion validity?
4. What combination of item types appears to be best in terms of: a) efficiency (high reliability in the available time), b) face validity, c) criterion validity, d) independence from the verbal

Figure 3

Item Types Originally Considered for the Analytical Ability Section

1. X21 (Letter Sets)

The Letter Sets item type originated in the French Factor Kit and was intended to test inductive reasoning. The item type and the instructions were revised to eliminate possible ambiguities, and new questions were written. Each question consists of five groups of letters, only one group of which is unlike the others in alphabetic arrangement. The examinee's task is to identify that dissimilar group.

Letter Sets questions seemed potentially useful because they were shown to be relatively independent of the verbal and quantitative ability measures in earlier studies. Although testing a narrow aspect of reasoning, Letter Sets were assumed to be efficient, requiring a minimal investment of time and students at or above the college level.

2. X22 (Logical Reasoning and Letter Sets)

Since Logical Reasoning questions had previously been pretested in a single module on the GRE population, pretesting of Logical Reasoning alone would have been repetitive. However, because one possible combination of reasoning item types was foreseen to be Letter Sets and Logical Reasoning, the two item types were pretested together to determine their relationship.

Logical Reasoning questions are based on brief arguments or statements presenting evidence or opinions. The questions require that the examinee recognize unstated presuppositions, logical flaws, methods of persuasion, conclusions logically following from given arguments, and the like. From earlier pretests, it was known that the item type, which had high face validity, correlated highly with the verbal ability score. However, in combination with an item type such as Letter Sets, it was expected to be appropriate as a measure of reasoning yielding a separate score. From experiments in the Law School Admission Test program, it was known that the item type had high criterion validity, with the highest validity coefficient among several experimental item types.

3. X23 (Analytical Reasoning and Letter Sets)

Analytical Reasoning questions had previously been pretested in a single module on the GRE population, and thus all necessary information on reliability, speededness, and relationship to verbal and quantitative ability was available. However, since an appropriate reasoning module might include both Letter Sets and Analytical Reasoning, the two item types were pretested together to determine their relationship. The Letter Sets questions used in X23 were identical to those used in X22.

Analytical Reasoning questions are based on brief sets of statements expressing relationships among abstract symbols (letters) or sets of rules governing processes or procedures having few concrete referents. The examinee is asked to draw inferences from and sometimes critically assess those sets of statements.

4. X24 (Evaluation of Evidence)

Evaluation of Evidence questions are based on brief narratives establishing a situation and conclusion drawn from the facts presented. The items consist of bits of evidence that, in relation to the situation described, strengthen, weaken, confirm, disprove, or fail to affect the conclusion. Newly developed for the GRE program, Evaluation of Evidence provided high face validity, but its statistical characteristics were unknown. The present options, which are the same for all questions, made the item type likely to be efficient.

5. X25 (Analysis of Explanations)

Analysis of Explanations questions are based on brief narratives establishing a situation in which an action is taken in order to have a specific effect. A later result, which may or may not be directly related to the action, is described in a brief statement. Each question is a piece of information that must be evaluated in terms of the facts and the result. The five options are the same for all questions. The first option that cannot be eliminated is the correct answer. Newly developed for the GRE program, the Analysis of Explanations item type provided high face validity, but its statistical characteristics were unknown.

6. Y1 (Logical Diagrams)

Logical Diagrams, an item type derived from Venn diagrams, presents several circle diagrams of possible relationships. Each question consists of three nouns, and the examinee is asked to select the circle diagram that best characterizes the relationship of the three nouns. Logical Diagrams had been pretested in a similar form in the French Factor Kit. Having proved to be efficient and somewhat independent of verbal and quantitative ability, the item type had promise as a component of a reasoning measure on which a separate score could be reported. It was noted, however, that the item type presented problems of multiple interpretations and key disagreement. To forestall such problems, the development process was expanded to include more than the usual number of independent reviews.

7. Y10 (Deductive Reasoning)

The Deductive Reasoning item type consists of a relatively complex set of rules the examinee is asked to apply in solving problems based on diagrams. Deductive Reasoning items had high face validity but unknown statistical characteristics.

and quantitative measures, and e) appropriateness for both science students and humanities and social science students.

These issues were pursued by research involving pretesting of items, statistical analyses of pretests, and surveys of students attempting new item types. As a result of this work the Analytical Ability section was introduced as part of the GRE examination on an experimental basis in October 1977. The item types selected for inclusion included: analysis of explanations; analytical reasoning; logical reasoning; and logical diagrams. Subsequent research on the Analytical Ability section is discussed later in this report.

Subsequent Board-Sponsored Studies of the Analytical Measure

The GRE Board has sponsored a number of investigations that bear directly on the construct validity of the analytical measure. A factor analysis of the restructured GRE Aptitude Test (Powers & Swinton, 1981) revealed that, as a result of the introduction of three analytical item types the new test contained a distinct, identifiable, analytical dimension that was highly related to the verbal and quantitative factors underlying performance on the test. This analytical factor was defined predominantly by the three analytical item types, but not by any of the verbal or quantitative types. The results of the study thus provided considerable evidence of both the discriminant and convergent aspects of the construct validity of the new measure.

Unfortunately, additional research on the new analytical module revealed that two of the four item types (55 to 70 items in the analytical section) were susceptible to both within-test practice (Swinton, Wild, &

Wallmark, 1983) and to short-term coaching (Swinton & Powers, 1983) and test familiarization (Powers & Swinton, 1982; 1984) and have since been deleted from the test. This state of affairs has, unfortunately, rendered much of the early research on the analytical measure not entirely applicable to the revised measure, which contains only logical reasoning and analytical reasoning items. In addition, too few of these two item types were included in the original analytical measure to permit a comprehensive assessment of their individual validity. The situation was further complicated by the fact that these items appeared primarily in the final positions in a particular separately timed section of the test, thereby confounding item difficulty with rate of work.

Wilson (1982) summarized the results of a series of predictive validity studies in which scores on the original analytical measure were included as a predictor. He found that in three of four quantitatively oriented fields (all but mathematics), analytical scores were slightly more predictive of first-year graduate grades than were either quantitative scores or verbal scores. In verbal fields, no consistent pattern was noted. In education, the analytical measure was the best of the three measures. In history, the analytical and verbal measures performed about equally, and in sociology the verbal measure was the best predictor.

Wilson (1982) concluded that "On balance, these findings suggest that, in the fields designated as verbal, the predictive value of the analytical score may tend to be about like that of the verbal score, whereas, in the fields designated as quantitative, the predictive value of the analytical score may parallel that of the quantitative score" (p. iv). With respect to incremental validity, the results were inconclusive, but Wilson (1982) suggested that "...the analytical score may prove to be somewhat more useful as an additional predictor in the quantitative than in the verbal areas..." (p. vi). It should be noted, especially with respect to incremental validity, that Wilson's results were based on the original analytical test, which contained a large proportion of analysis of explanations items, a highly verbally-loaded item type.

In contrasting examinees' profiles of performance on the three sections of the Aptitude (General) Test, Swinton and Powers (1979) found strikingly different patterns of abilities among candidates in different intended major fields. Predictable patterns were evident with respect to verbal and quantitative abilities for students intending to major in the humanities, social sciences, and physical sciences, with the first two groups having higher verbal than quantitative abilities and the opposite occurring for physical science majors.

The analytical ability factor (again based on the original four item types) did not distinguish particularly across the arts/science continuum, but rather provided more discrimination within each general category, distinguishing the more theoretical from the more applied majors. The basic, more theoretical majors tended to be higher in analytical ability than the more applied majors. For example, within the humanities, philosophy, religion, and art history were higher than music and art. Within the physical sciences, mathematics and physics majors were higher than chemistry, earth science, and engineering.

In his examination of criterion-related validities for item type part scores, Wilson (1984) found consistent differences in the performance profiles, as well as the validity, of item types across fields. In the primarily verbal fields, the validity was generally higher for the logical reasoning item type than for the analytical reasoning type. In the quantitative fields, the opposite was true. Wilson, however, also found the somewhat discouraging evidence that the two currently used analytical item types may be more highly related to either other verbal item types or to other quantitative item types than to each other, thus casting serious doubt as to whether these two analytical item types primarily reflect the same analytical dimension.

Recent Thinking about Construct Validity

As described above, the development of the analytical test did follow a carefully established plan of item type identification and research on the face validity and psychometric characteristics of item types. However, the development plan that was followed differs in some major ways from more recent suggestions on how construct validity might inform test development. Messick (1981), for example, points to the central role of a behavioral theory and a measurement model in establishing the construct validity of a test. Understanding what a test measures rests on some explicit or assumed theory of behavior tapped by a test. An adequate theory should include not only an account of subskills critical to performance on assessed constructs, but also an analysis of how performance on a test is related to performance in "real world" tasks. In addition, the measurement model has implications for interpreting the construct validity of a test. The measurement model not only operationalizes a construct in terms of item performance and test scores, but also establishes a framework for interpreting the skills of examinees with respect to the construct being assessed.

Frederiksen (1982) outlines seven steps characterizing an ideal construct validity approach to test development. To quote Frederiksen (1982, p.2):

> ...Construct validity implies that the first step in test development should be a scientific investigation of the performance to be assessed or predicted. The results of the investigation should then guide the development of both criterion measures and the desired assessment test. The new test should be validated by demonstrating that it reflects the same cognitive processes as the criterion.

The seven test development steps suggested by Frederiksen include:

Step 1: Develop a theory about criterion performance

Step 2: Develop a criterion test

Step 3: Verify the theory about criterion performance

Step 4: Develop a test for use in assessment or prediction

Step 5: Test the construct validity of the assessment test

Step 6: Test the construct similarity of the assessment for trained (expert) and untrained (novice) groups in the skill area under assessment

Step 7: Repeat steps 1 to 6

Frederiksen's steps describe an interactive procedure, starting with an attempt to identify the criterion behavior which is the target of assessment, followed by (a) the development of an instrument to assess the skills of interest, and (b) an investigation of the connections between test performance and criterion behavior. This approach also considers whether test performances mean the same thing for groups of examinees that differ in their expertise in the domain under assessment.

A significant obstacle to implementation of a construct validity approach to test development lies in how well educational, psychological, and assessment theories concretely inform the test development and test use process. To date, growth in the sophistication of cognitive theory has had virtually no effect on the development of admissions tests. The conceptualization of aptitude tests for college admissions rests on models of cognitive skills that have changed very little in the past 40-50 years. While testing programs have shown considerable responsiveness to psychometric developments over the years, the response to advances in cognitive psychology has been far more limited. This is true partly because educational and psychological research has only recently evolved to the point where sophisticated analysis of thinking and learning could be thought to have a significant impact on testing.

III. Two Selective Reviews of Research on Reasoning

Selected Psychological and Cognitive Science Research on Reasoning Problem Solving and Critical Thinking

In many way it is difficult to conceptualize reasoning skills without drawing attention to the notion of problem solving more generally. As Sternberg (1982) notes, the concepts of reasoning, problem solving, and intelligence overlap considerably, and our common notions of intelligence include both reasoning and problem solving. In turn, problem solving constitutes a more general framework for describing reasoning, since to understand and view the operations of the latter we must specify problem solving circumstances. The importance of problem solving in the conceptualization of higher level reasoning skills is evident in the definitions of "critical thinking" which have arisen in the field of educational psychology. For example, Dressel and Mayhew (1954) suggested that the following abilities are related to critical thinking:

Defining a problem

Selecting pertinent information for a solution

Recognizing stated and unstated assumptions

Formulating and selecting relevant and promising hypotheses

Drawing valid conclusions and judging the validity of inferences

In their information-processing description of problem solving, Newell and Simon (1972) call attention to the mental representation of problems and to the reasoning used in problem solving. They use the term "problem space" to refer to an individual's mental representation of a problem.

"Well-defined" problems have certain characteristics which affect the reasoning applied during problem solving. A well-defined problem is characterized by an initial state of affairs constituting a problem, a final state of affairs indicating that a problem has been solved, and a set of "operators" or mental manipulations that can be used to create new problem states from previous problem states.

Problem solving is goal-driven. Individuals create mental plans for solving a problem, and they search their long-term memory for operators and knowledge that will aid in working a problem. Newell (1980, p. 700) describes the process of problem solving as follows:

> Given a problem in a problem space, the only way a subject can solve the problem is by searching in the space: working out from the current state by applying operators, adding new states to the stock to be used at new points of search, evaluating whether the results help, etc. To accomplish this search requires performing repeatedly a fixed set of functions:
>
> Search Control: The following functions determine the behavior in a problem space while working on a given problem.
>
> Decide to quit the problem.
> Decide if a goal state has been produced.
> Select a state from the stock to be the current state.
> Select an operator to be the current operator.
> Decide to save the new state just produced by an operator.
>
> These functions operate within a cycle consisting of repeating the following three steps:
>
> 1. Select a state; select an operator.
> 2. Apply operator to state, producing new state.
> 3. Decide if a goal state is met; decide to quit; decide to save the new state.

Control of search processes during problem solving can involve some specific high level reasoning processes which have great generality across problem domains. Newell (1980, p. 703) indicates the following processes, which have been implemented in artificial intelligence computer programs for problem solving:

- <u>Generate and Test</u>: Generate in any way possible (e.g., systematically or haphazardly) a sequence of candidate states, testing each for whether it is the desired state.

- <u>Heuristic Search</u>: Apply heuristics to reject possible operators from the current state and to reject newly produced states; remember the states with untried operators in some systematic way (different schemes yield search strategies, such as <u>depth first</u>, <u>breadth first</u>, <u>progressive deepening</u>, and <u>best first</u>).

- <u>Hill Climbing</u>: Generate and apply operators from the current state; select one that produces a state with an improved evaluation and move to it.

- <u>Means-Ends Analysis</u>: Compare the current state with the desired state to detect any difference; use that difference to select an operator that reduces or eliminates it; otherwise proceed as in heuristic search.

- <u>Operator Subgoaling</u>: If an operator cannot be applied to a current state, set up a subgoal to find a state in which the operator can be applied; otherwise proceed as in heuristic search or means-ends analysis.

- <u>Planning</u>: Abstract from the present state (by processing only selected information throughout) and proceed to solve the simplified problem; use what is remembered of the path as guide to solving the unabstracted problem.

These problem solving strategies are termed "weak methods" by Newell because their use and utility are not necessarily dependent upon a detailed knowledge of a particular problem domain. Indeed, as will be discussed later in this section, when a great deal of information is known about a problem domain there is often less need for high level reasoning in problem solving. An awareness of such "weak methods" alerts us to a broad class of reasoning skills that problem solvers might use to integrate and monitor the use of reasoning skills that are more automatic. The ability to use more specific reasoning skills is often the control of more general problem solving strategies such as those which Newell discusses.

Sternberg's (1982) components of intelligence theory provide a very helpful framework within which to view the purposes and integration of different sorts of skills in problem solving. According to Sternberg (1980):

> A component is an elementary information process that operates upon internal representation of objects or symbols. What is "elementary" is determined by theoretical context: What is called "elementary" in one theory at one level of analysis might be called "complex" in another theory at another level of analysis. A component may translate a sensory input into a conceptual representation, transform one conceptual representation into another, or translate a conceptual representation into a motor output (Sternberg, 1977b; see also Newell & Simon, 1972). The basic idea is that components represent latent abilities of some kind that give rise to individual differences in measured intelligence and in real-world performance, and to individual differences in factor scores as well.

Sternberg (1982) mentions five kinds of cognitive components: metacomponents; performance components; acquisition components; retention components; and transfer components. Sternberg's (1982, p. 227) description of the various components is particularly apt in terms of its discussion of reasoning and problem solving processes:

> Metacomponents are higher-order control processes that are used for planning a course of action, for making decisions regarding alternative courses of action during reasoning or problem solving, and for monitoring the course of action.

> Performance components are processes that are used in the execution of a reasoning or problem-solving strategy.
>
> Acquisition components are processes used in learning how to reason or to solve problems.
>
> Retention components are processes used in retrieving previously stored knowledge, whether it be knowledge needed during reasoning or problem-solving or knowledge regarding the reasoning or problem-solving algorithm itself.
>
> Transfer components are processes used in generalization, that is, in carrying over knowledge from one reasoning or problem-solving task to another.

The five kinds of components mentioned can further be classified as "general components...required for performance of all tasks within the task universe...class components...required for performance of a proper subset of tasks within the task universe and specific components... required for the performance of single tasks within the task universe." Sternberg's overview of components suggests the complexity involved in conceptualizing reasoning performance. Clearly Newell's "search control" shares important attributes with Sternberg's "metacomponents," and Newell's reasoning processes involve many of Sternberg's components. The degree of overlap in these two accounts may be more important than their differences. Points of overlap suggest skills which are important and which might be assessed in college admissions tests. For example, planning or "executive" functions might be worthy of assessment though they are not explicitly tested in current verbal, mathematics, or reasoning aptitude tests.

A critical factor affecting the relevance of components to problem solving is whether problems are "well-structured" or "ill-structured." Newell's (1980) description of problem solving is appropriate to "well-structured" problems, i.e., problems in which all of the information required to interpret and to solve a problem are available in the memory of the problem solver or in the initial specification of a problem. Simon (1973) has defined "ill-structured" problems as problems in which, initially, all of the information necessary to solve a problem is <u>not</u> available to the problem solver. An example of a "well-structured" problem is an algebraic word problem in which the problem solver understands both the statement of the problem and the appropriate mathematical formulas to represent it and, therefore, needs only to interpret the problem in terms of an equation. In contrast, an example of an "ill-structured" problem is the problem of "designing a home." In this case there are many options and parameters, which lead dynamically to the refinement and solution of the problem. Problems that are well-structured place more emphasis on reasoning processes that resemble deduction. Problems that are ill-structured require not only deductive reasoning, but also inductive reasoning, i.e., the gaining of new knowledge based on experience and experiment. In reality, well-structured and ill-structured problems are on a continuum.

A critical factor affecting solution of both types of problems is knowledge about a problem and a problem domain. For an expert, with considerable knowledge about a problem domain, a problem is more likely to be well-structured than for a novice. Also, the expert is usually more skilled than a novice problem solver in converting ill-structured problems into well-structured ones. An expert's organization of both factual knowledge about a domain and procedural knowledge about problem solving is typically greater than that of a novice problem solver.

Current cognitive research on expert-novice differences points to some important issues that could be considered in designing assessments of reasoning ability. The following synopsis of findings drawn from Duran, Stricker, and Ward (1982) summarizes some of the major conclusions to be drawn from research on expert-novice differences in problem solving:

> <u>Finding 1</u>: Experts' knowledge in a problem solving area is more than knowledge of facts and isolated principles. Experts maintain knowledge about how facts and principles are organized and related to each other. When asked to categorize problems according to their similarity, experts sort problems according to the principles which different problems exemplify, while novices classify problems as similar on the basis of crude similarities related to their statement and embodiment (Chi, Feltovich, & Glaser, 1981). A further implication of this finding supported by research is that experts are more able than novices to analyze new problems and novel aspects of problems in terms of the most relevant principles (Lesgold, Feltovich, Glaser, & Wang, 1981).

> <u>Finding 2</u>: Experts are more automated in their problem solving behavior than novices. Given their greater knowledge in an area and understanding of how to apply knowledge, experts are more likely to "work forward" in problem solving rather than "backward" (Larkin, 1980). That is to say, experts are less likely to start working a problem by searching extensively for alternative ways in which a final solution might be formulated unless this is necessary. On the other hand, novices are more likely to search extensively for ways in which a problem might be conceptualized. Experts recognize quickly and "automatically" what principles apply, and they concentrate from the start on ways in which data fit into known problem solving algorithms. As a result of this difference, experts are quicker and more efficient in solving problems than novices.

Finding 3: Experts' development of knowledge in an area is an outgrowth of extensive experience which leads to development of automatic, but flexible, problem recognition strategies and procedural knowledge about how to solve problems. Lesgold, Feltovich, Glaser, and Wang (1981), for example, concluded that it took a number of years of experience in the field for radiologists to acquire rapid recognition skill in their examination and diagnosis of patient x-rays. Novice radiologists were more likely to ignore data that were important to consider in x-ray diagnosis. Lesgold, et al. (1981) also concluded that experts may be more flexible in their problem diagnosis strategies to account for anomalies encountered in a problem.

Finding 4: Novices often maintain naive models of knowledge in a problem solving domain. Research by Green, McCloskey, and Carammaza (1980), DiSessa (1982), and by Steven and Collins (1980) has found that students often maintain naive models explaining or predicting physical phenomena that affect their learning of scientific principles. While naive models are incorrect, and become displaced by a "correct" scientific model through further learning, the naive models nonetheless represent a basic important starting point for acquiring knowledge and problem solving ability in a problem domain.

Finding 5: Transfer of problem solving knowledge across similar problems occurs most successfully only when the relationship between two problems is explicit. Simon and Hayes (1976) and Reed, Ernst, and Banerji (1974), for example, have found that human subjects are unable to recognize and represent formal similarities between puzzle problems that are isomorphic to each other. Instruction in the formal similarities between a new problem and an old problem, however, did lead to a more efficient and correct solution of a new problem.

Finding 6: Very little research in this area has attempted to work with large groups of subjects. Most of findings 1 through 5 have been established by detailed careful study of problem solving behavior by individuals or small groups of subjects.

Deductive and Inductive Reasoning Research

In contrast to the richer view of reasoning stemming from a problem solving approach, most psychological research on reasoning has focussed on a more narrow view of reasoning processes with attention given to specific classes of reasoning. For example, in their review of thinking and concept attainment, Neimark and Santa (1975) listed the following subheadings under the general category of "logic and thought":

1. transitive inference
2. syllogistic reasoning
3. comprehension of logical symbols and terms
4. conditional reasoning
5. formal operations
6. generalization and abstraction

The range of topics mentioned by Neimark and Santa exemplify research topics in the area of deductive reasoning. Descriptions of recent research are provided in Wason and Johnson-Laird (1972), Revlin and Mayer (1978), Johnson-Laird and Wason (1977), Falmagne (1975), and Anderson (1982). The key questions of research have focussed on whether a human's reasoning behavior conforms with the operation of a well-specified model of deductive reasoning. Sometimes this latter model has been based closely on a formal logical system, while in other cases models have also introduced a human decision making component with reasoning described in information processing terms. The most obvious stimulus to research has been evidence that humans do not typically apply formal rules of inference in a logical manner. Researchers have sought to model and understand how it is that humans reason with materials drawn from formal logic. Special attention, for example, has been given to the solution of categorical syllogisms (e.g., All A are B. All B are C. Therefore....); conditional syllogisms (e.g., If A implies B and B implies C, then....); other non-syllogistic forms of conditional reasoning; reasoning about quantifiers (e.g., What are different meanings of "All," "Some," or "None"); and solution of Venn diagram problems. It is difficult to summarize concisely the findings of this research because researchers employ different models and conceptions of the underlying processes they study. Anderson (1982, p. 325) nonetheless provides a useful capsule summary:

> We have seen ample evidence for the conclusion that humans are in some sense illogical in their deductive reasoning. By this, we mean a number of different things:
>
> 1. People do not always interpret statements the way logicians prescribe they should. For instance, they sometimes interpret *if* as *if and only if* and they interpret *All As are Bs* as *A equals B*.
>
> 2. They fail to recognize certain logical conclusions as valid, either because they have not learned the appropriate rule of inference (for instance $p \rightarrow q \Rightarrow q \rightarrow p$) or because they cannot figure out how to combine a number of steps in a deduction (for instance, in the hard categorical syllogisms).
>
> 3. They accept certain contingent conclusions because in part this acceptance may reflect a misinterpretation of the premises (point 1 above), but in part it reflects an inability to find counter-examples to the conclusions.
>
> 4. Faced with the difficulties identified in points 2 and 3, subjects often fall back on various heuristics for solving logical problems. These heuristics often succeed, but sometimes they do not.

To this list we might add that people's judgment of how conclusions should be worded is biased by the manner and order of wording used in the premises of syllogistic arguments (see for example the work of Johnson-Laird and Steedman, 1978).

In addition to the phenomena cited above, logic and rhetoric textbooks cite a number of other errors in deductive reasoning, some of which have been investigated sporadically in psychological research. Table 1 lists typical deductive reasoning fallacies as given by Blumberg (1976).

There are several important and interesting implications of the foregoing discussion with respect to our knowledge of deductive reasoning and deductive reasoning errors. First, there is a wide range of deductive reasoning phenomena relevant to graduate school academic work, and these phenomena are not captured by a single test of deductive reasoning. A second implication is that many deductive reasoning "errors" are systematic; in some sense they don't represent "errors," but rather the influence of alternative or irrelevant sources of knowledge and frames of reference on reasoning. As will be discussed in the next section, the use of a variety of sources of knowledge and frames of reference in reasoning is quite natural, and indeed critical to everyday reasoning and to inductive reasoning generally. Such forms of mitigated deductive reasoning are critical to some of the skills that are most important to learning in graduate school and to production of new knowledge in academic areas and science.

While a wide range of deductive reasoning skills may be needed in graduate training, given the limitations of assessment situations what can be learned about the construct validity of testable deductive reasoning

Table 1

Fallacies in Deductive Reasoning

1. Argument ad baculum ("appeal to the stick"). This fallacy is usually defined as the appeal to force or to the threat of force to obtain acceptance of a conclusion.

2. Argument as hominem ("to the man"). This fallacy consists in offering statements about a person's character or circumstances as premises to oppose his conclusion.

3. Argument ad ignorantiam ("from ignorance"). To commit this fallacy is to argue that a conclusion should be accepted because it has not been disproved or that it should be rejected because it has not been proved.

4. Argument ad misericordiam ("appeal to pity"). Here the fallacy consists in presenting an appeal to pity as if it were a proper premise from which to infer a conclusion.

5. Argument ad populum ("to the people"). An argument is said to be an instance of the ad populum fallacy if it relies on an appeal to the "emotions of the crowd" to gain assent to its conclusion.

6. Argument ad verecundiam ("appeal to authority"). This is the fallacy of appealing to an authority in an area that lies outside of his or her competence.

7. Accident. The fallacy of accident is committed when we argue from some general principle to some particular case whose "accidental" features make it an exception to the principle.

8. Many questions, or complex question. This is the familiar "have you stopped beating your wife" fallacy. It arises when two or more questions are asked as one, with the insistence that they be given a single yes-or-no answer.

9. Petitio principii ("begging the question"). An argument that has its conclusion as one of its premises is said to commit the fallacy of petitio principii. It begs the question by assuming what is to be proved.

10. Ignoratio elenchi ("arguing beside the point"). In the most general sense, any argument that commits a fallacy of irrelevance is guilty of arguing beside the point, and thus all fallacies of irrelevance are cases of ignoratio. In a more specific sense, the fallacy is committed when premises are addressed to the wrong conclusion.

11. Equivocation. A fallacy of equivocation is committed when an ambiguous term is used to shift senses in the same argument.

12. Composition and division. Composition and division are paired fallacies and represent special cases of equivocation. Here the equivocation is on general terms, expressions that denote wholes or classes.

13. Amphiboly. This is a fallacy in which an ambiguity attaches not to a word but to an entire sentence. Such a sentence is said to be amphibolous; amphiboly, like ambiguity is relative to context.

14. Accent. Some sentences vary in meaning depending on which of their words are stressed. If a source of the ambiguity in meaning lies in stress, then the fallacy of amphiboly is sometimes called a fallacy of accent.

15. Fallacy of inconsistent premises. This is the case when not only are the premises not all true, but when in addition they cannot possibly all be true--in short, when the premises are inconsistent or mutually contradictory.

Blumberg, A. E. Logic: A first course. New York: Alfred A. Knopf, 1976.

skills? Here we will discuss steps toward identifying cognitive component skills that might underlie general classes of deductive reasoning tasks. Our attention for the movement is focused on very elementary information processing skills that are required in the application of higher level reasoning skills. Carroll (1976) presents the following description of syllogistic reasoning and general deductive reasoning in the context of factor and information processing models of cognitive ability:

> Factor Rs (Syllogistic Reasoning) involves both retrieving of meanings [of words and sentences] and algorithms [i.e., rules for solving problems] from relevant portions of LTM [long-term memory], and performing in STM [short-term memory] serial operations on materials retrieved. Individual differences [in performance] could appear in content and temporal aspects [e.g., speed] of both of these types of operations. They could also occur in the probability that the subject will give adequate attention to details of the stimulus material.* (p. 50)
>
> Factor R (General Reasoning) is very similar to Factor Rs (Syllogistic Reasoning) in that it involves both retrieval and serial operations. It would be distinguished from Factor Rs only with respect to the precise types of contents in LTM that are required to be retrieved and utilized in the serial operations. In the case of Factor Rs, these contents have to do with logical characteristics of certain linguistic quantifiers (all, some, no, etc.), whereas in Factor R the contents are more general algorithms concerned with concrete relations (time, rate, cost, etc.), and in addition the same types of number associations that are involved in Factor N (Number Facility).

*Bracketed material added for clarity.

Carroll's description of Factor R (General Reasoning) seems particularly apt for GRE analytical reasoning items. The design of a construct validity study of the analytical reasoning and logical reasoning study could give attention to assessing elementary cognitive skills such as those mentioned in Carroll's description.

Carroll's analysis of the 8 basic information processing skills required in deductive reasoning does not reflect influences on reasoning that would be capable of accounting for fallacies in reasoning of the sort outlined by Blumberg (1976). A different approach is required to assess these latter influences. This approach would require looking at higher-level reasoning and strategic thinking skills that could be represented by logical reasoning items. The earlier discussion of errors in deductive reasoning and the ensuing discussion on inductive reasoning mention some of the issues which should in future research on reasoning assessment be faced.

Reasoning from Incomplete Knowledge

Deductive reasoning, while critical to all problem solving, must be supplemented by other forms of reasoning in most realistic problem solving and learning settings. Problem solving situations that are ill-structured do not typically provide enough information at the outset for a well-defined problem representation. The term "induction," which is often applied to reasoning that leads to new knowledge, will not be used in this section since more explicit notions of this concept will be introduced in the context of current cognitive theory and research.

Collins, Warnock, Aiello, and Miller (1975) distinguish whether problem solving involves reasoning about "closed" or "open" worlds. A closed world is a knowledge domain in which the facts underlying the solution of problems are known explicitly by the problem solver. A question such as "Are there volcanos on Hawaii?" requires reasoning about closed world knowledge for most of us, since we know for a fact that there are volcanos on Hawaii. A question such as "Are there volcanos in Guatemala?" would be a question requiring open world reasoning for persons without enough factual knowledge to answer this question with certainty. In this case, a "reasonable" answer, exemplifying open world reasoning, might be "There probably are volcanos in Guatemala, because I know that Mexico has volcanos and because I know that volcanos tend to occur in chains along the North American continent."

Collins, et al. (1975) describe two general, non-exhaustive classes of open world reasoning strategies. Negative inference strategies are used to provide "No" answers to questions about hypothetical facts when attempts to derive "Yes" answers fail. A "No" answer is given when supporting a "Yes" answer is not readily available. Functional inference strategies occur when "people...figure out what they do not know by reasoning from their knowledge about what it depends on" (Collins, et al., 1975, p. 400). An answer to a question such as "Given that it rained yesterday, are the lawns wet?" exemplifies this sort of reasoning. Collins, et al. (1975) go on to describe negative inferences and functional inferences in terms of different sorts of information processing strategies; the reader is referred to their paper for further details.

Collins, et al. (1975) state that the sorts of open world inference strategies they observed in students are at the heart of learning to reason inductively. In undertaking open world reasoning, students can, with feedback from expert tutors, learn to explore a knowledge domain and to integrate their developing knowledge of an area.

In other relevant research, Collins (1977) has investigated the reasoning strategies of expert teachers who use the Socratic method. The objective of this work was to describe how expert teachers can instruct students in learning facts about a new domain of knowledge. By asking leading questions, expert teachers instructed students to reason about a new domain of facts and to reason about how to integrate accumulating facts. Table 2 lists 24 Socratic rules explored by Collins. Rules 1-15 emphasize students' learning of general principles from known facts, while Rules 16-23 teach students' to apply general principles, learned abstractly to unknown cases.

Interestingly, exposing students to use of the rules may conceivably lead to students acquiring some of the rules as self-motivated reasoning strategies. To the extent that such learning is possible, students may develop higher-level reasoning strategies that help them to monitor and extend the range and depth of their knowledge in a problem solving domain. Reasoning skills of the sort outlined by Collins (1977) do not appear to be assessed directly at present by any known psychometric instruments.

There are a number of additional research areas in cognitive science that pertain to reasoning from incomplete knowledge. The whole topic area is often subsumed under research on schema theory models of cognition--see Bobrow and Norman (1975) and Schank and Abelson (1977) for example.

Table 2

Rules in Socratic Tutoring

Rule 1: Ask about a known case.

If

(1) it is the start of a dialogue, or

(2) there is no other strategy to invoke,

then

(3) pick a well-known case and ask what the value of the dependent variable is for that case, or

(4) ask the student if he knows a case with a particular value of the dependent variable.

Example: Ask the student "Do they grow rice in China?" or "Do you know any place where rice is grown?"

Reason for Use: It brings out any well-known facts the student knows about such as rice growing in China.

Rule 2: Ask for any factors.

If

(1) a student asserts that a case has a particular value of the dependent variable,

then

(2) ask the student why.

Example: If a student says they grow rice in China, ask why.

Reason for Use: This determines what causal factors or chains the student knows about.

Rule 3: Ask for intermediate factors.

If

(1) the student gives as an explanation a factor that is not an immediate cause in the causal chain,

then

(2) ask for the intermediate steps.

Example: If the student mentions monsoons in China as a reason for rice growing, ask "Why do monsoons make it possible to grow rice in China?"

Reason for Use: This insures that the student understands the steps in the causal chain, for example, that rice needs to be flooded.

Rule 4: Ask for prior factors.

If

(1) the student gives as an explanation a factor on a causal chain in which there are also prior factors,

then

(2) ask the student for the prior factors.

Example: If the student mentions water as a factor in growing rice, ask him "What do you need to have enough water?"

Reason for Use: Same as Rule 3.

Rule 5: Form a general rule for an insufficient factor.

If

(1) the student gives as an explanation one or more factors that are not sufficient,

then

(2) formulate a general rule asserting that the factor given is sufficient and ask the student if the rule is true.

Example: If the student gives water as the reason they grow rice in China, ask him "Do you think any place with enough water can grow rice?"

Reason for Use: This forces the student to pay attention to other causal factors.

Rule 6: Pick a counterexample for an insufficient factor.

If
(1) the student gives as an explanation one or more factors that are not sufficient, or
(2) agrees to the general rule in Rule 5,
then
(3) pick a counterexample that has the right value of the factor(s) given, but the wrong value of the dependent variable, and
(4) ask what the value of the dependent variable is for that case, or
(5) ask why the causal dependence does not hold for that case.

Example: If a student gives water as the reason they grow rice in China or agrees that any place with enough water can grow rice, pick a place like Ireland where there is enough water and ask "Do they grow rice in Ireland?" or "Why don't they grow rice in Ireland?"

Reason for Use: Same as rule 5.

Rule 7: Form a general rule for an unnecessary factor.

If
(1) the student gives as an explanation one or more factors that are not necessary,
then
(2) formulate a general rule asserting that the factor is necessary and ask the student if the rule is true.

Example: If a student says rainfall is a reason for growing rice, ask "Do you think it is necessary to have heavy rainfall in order to grow rice?"

Reason for Use: This forces the student to consider the necessity of a particular factor.

Rule 8: Pick a counterexample for an unnecessary factor.

If
(1) the student gives as an explanation one or more factors that are not necessary, or
(2) the student agrees to the general rule in Rule 7,
then
(3) pick a counterexample with the wrong value of the dependent variable, and
(4) ask the student what the value of the dependent variable is for that case, or
(5) ask why the causal dependence does not hold in that case.

Example: If the student gives rainfall as a reason for growing rice, ask "Do you think they can grow rice in Egypt?" or "Why do they grow rice in Egypt when they don't have much rainfall?"

Reason for Use: Same as Rule 7.

Rule 9: Pick a case with an extreme value.

If
(1) the student is missing a particular factor,
then
(2) pick a case with an extreme value of that factor and ask why the dependent variable has a particular value in that case.

Example: If the student has not mentioned temperature with respect to rice growing, ask "Why don't they grow rice in Alaska?"

Reason for Use: This forces the student to pay attention to any factor he is ignoring.

Rule 10: Pose a misleading question.

If
(1) there is a case in which a secondary factor overrides the primary factors,
then

(2) pose a misleading question to the student, based on the fact that the value of the dependent variable is different from what would be predicted from the primary factors above, or

(3) pose a misleading choice as to the dependent variable between two cases in which consideration of the primary factors alone leads to the wrong prediction.

Example: Because the tree cover in the Amazon jungle keeps the temperature down to a high of about 85 degrees, ask the student "Do you think the temperatures in the Amazon jungle reach 100 degrees?" or "Do you think it gets hotter in the Amazon jungle or Texas?"

Reason for Use: This forces the student to learn about common exceptions, about secondary factors, and about the limitations of general rules.

Rule 11: Specify how the variable depends on a given factor.

If

(1) the student mentions a factor, but does not specify how the dependent variable varies with that factor, or

(2) only partially specifies the relationship,

then

(3) ask him to specify the relationship more precisely, or

(4) suggest a possible relationship to him.

Example: Ask the student "Can you say how temperature depends on latitude?" or "Does average temperature increase linearly the further south you go?"

Reason for Use: This forces the student to specify more precisely the functional relation between the factor in question and the dependent variable.

Rule 12: Probe for a necessary factor.

If

(1) a student makes a wrong prediction of the dependent variable because he has not identified one or more necessary factors,

then

(2) tell him he is wrong, and ask him to formulate a hypothesis about another factor that is necessary.

Example: If a student thinks they can grow rice in Ireland because of the heavy rainfall, point out they cannot grow rice there and ask "Can you make a hypothesis about what other factor is necessary for rice growing?"

Reason for Use: This forces the student to use hypothesis formation as a systematic strategy for dealing with unexplained problems.

Rule 13: Probe for a sufficient factor.

If

(1) a student makes a wrong prediction of the dependent variable because he treats a factor as necessary when it is not,

then

(2) tell him he is wrong, and ask him to formulate a hypothesis about another factor that might be sufficient.

Example: If a student thinks they cannot grow rice in Egypt because there is little rain, point out they can grow rice there and ask "Can you think of what other factor makes it possible to grow rice there?"

Reason for Use: Same as Rule 12.

Rule 14: Probe for differences between two cases.

If

(1) a student cannot think of a factor that could account for different values of the dependent variable between two cases,

then

(2) ask him to consider what the differences are between the two cases that might account for the difference in the dependent variable.

Example: If a student cannot think of why they can grow rice in China but not in Alaska, ask what the differences are between China and Alaska that might account for the difference in rice growing.

Reason for Use: Same as Rule 12.

Rule 15: Request a test of the hypothesis about a factor.

If

(1) the student has formulated a hypothesis about how the dependent variable is related to a particular factor,

then

(2) ask him how it could be tested.

Example: Ask the student "If you want to test whether distance from the ocean affects temperature, would you compare the temperature in January for St. Louis with Washington, D.C. or Atlanta?"

Reason for Use: By getting the student to test hypotheses, it forces him to learn to control other factors that might affect the variable.

Rule 16: Ask for a prediction about an unknown case.

If

(1) a student has identified all the primary factors that affect the dependent variable,

then

(2) pick a case that is either hypothetical or unlikely to be known and ask the student to predict the likely value of the variable for that case.

Example: If the student has identified the factors that affect rice growing, then ask "Do you think they can grow rice in Florida?"

Reason for Use: This forces the student to use the factors he has accumulated in a predictive way.

Rule 17: Ask what are the relevant factors to consider.

If

(1) the student cannot make a prediction,

then

(2) ask the student what are the relevant factors to consider.

Example: Ask the student "If you cannot predict whether they grow rice in Florida, what factors do you need to consider?"

Reason for Use: This teaches the student to ask the right questions in trying to make reasonable predictions about new cases.

Rule 18: Question a prediction made without enough information.

If

(1) a student makes a prediction as to the value of the dependent variable on the basis of some set of factors, and

(2) there is another value consistent with that set of factors,

then

(3) ask the student why not the other value.

Example: If the student predicts they grow wheat in Nigeria because it is fertile and warm, ask him why not rice.

Reason for Use: This forces the student not to jump to conclusions without enough information.

Table 2 (contd.)

Rule 19: Point out irrelevant factors.

If

(1) the student asks about the value of an irrelevant factor in trying to make a prediction,

then

(2) point out the factor is irrelevant, or

(3) ask whether the irrelevant factor affects the dependent variable.

Example: If the student asks whether Denver or Salt Lake City is further west in trying to decide which has the colder temperature, then point out that longitude does not matter, or ask whether longitude affects temperature.

Reason for Use: This forces the student to learn what is irrelevant, as well as what is relevant, in making any decision.

Rule 20: Point out an inconsistent prediction.

If

(1) a student makes a prediction about the dependent variable which is inconsistent with any of the values of the factors discussed,

then

(2) point out the inconsistency, or

(3) ask whether the value of the factor discussed is consistent with his prediction about the dependent variable.

Example: If the student predicts they grow rice in Spain after dryness of the climate has been discussed, either point out that a dry climate is incompatible with rice growing unless there is irrigation, or ask how he thinks they can grow rice when the climate is so dry.

Reason for Use: This reminds the student to consider all the relevant factors in making a prediction, and insures he understands the relation between the factor and the dependent variable.

Rule 21: Ask for consideration of a possible value.

If

(1) there is a value of the dependent variable that has not been considered and which either is consistent with several factors or important to consider a priori,

then

(2) ask the student to consider that value.

Example: If the student has not considered rice as a possible grain in Nigeria, ask him to consider it.

Reason for Use: This forces the student to actively consider alternatives in making any prediction.

Rule 22: Test for consistency with a given hypothesis.

If

(1) a particular value of the dependent variable is being considered, and

(2) the values of one or more relevant factors have been discussed, but

(3) whether these values are consistent with the particular value of the dependent variable has not been discussed,

then

(4) pick one or more of the factors that are consistent with the dependent variable and ask if they are consistent, or

(5) pick one or more of the factors that are inconsistent with the dependent variable and ask if they are consistent.

Example: If the hot climate and rainfall in Java have been discussed, the student can be asked "Is the heavy rainfall in Java consistent with growing wheat?" or "Are the hot climate and heavy rainfall consistent with growing rice?"

Reason for Use: This tests whether the student understands the functional relations between the various factors and the dependent variable.

Rule 23: Ask for consideration of relevant factors.

If

(1) a student makes a wrong prediction in a particular case, or

(2) cannot make a prediction,

then

(3) pick the most relevant factor not discussed and

(4) ask the student what the value of that factor is for the particular case.

Example: If the student predicts that the average temperature is very hot in Buenos Aires, ask if he knows what the latitude of Buenos Aires is.

Reason for Use: This forces the student to consider relevant factors in making a prediction, and elicits whether a mistake is due to wrong information about a case, or a mistake about how the dependent variable varies with different factors.

Rule 24: Trace the consequences of a general rule.

If

(1) a student agrees to a general rule such as Rule 5 or Rule 7,

then

(2) ask if he agrees with the consequences of that rule in a particular case.

Example: If the student states that it is necessary to have flat terrain to grow rice, then point out that since they do not have flat terrain in Japan, they must not grow much rice and so must import most of the rice they eat.

Reason for Use: Same as Rule 5 or Rule 7.

Collins, A. Processing in acquiring knowledge. In R. Anderson, R. Spiro, & W. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Erlbaum, 1977, 342-350.

According to Johnson-Laird (1983), reasoning should be studied in terms of persons' mental models of the situations and topics to be reasoned about. The underlying mental models are adapted from relevant memory schemata stored in long-term memory. Recently, there has been attention to mental models of dynamic systems, i.e., entities and situations which systematically change over time (Gentner, & Stevens, 1983). Also, special attention has been given to the mental models that novices develop about domains of knowledge related to academic course material likely to be encountered in college. Gentner and Steven's (1983) volume, for example, presents several studies of beginning learners' mental models of subject matter in physics and mathematics. Special attention is given in this work to students' naive mental models of a subject area and to misconceptions arising in reasoning.

Other research investigating reasoning in terms of mental models has stressed description of reasoning heuristics which deviate from probability theory models or alternative models based on formal measurement theory (see, e.g., Tversky and Kahneman, 1983 for a recent overview). All of the work which has been described tends to suggest that so called "errors" in reasoning often are not "errors" per se; they result instead from the systematic application of reasoning strategies that are deemed appropriate to a mental model for the situation to be reasoned about. The underlying mental model of the situation is often richer than that conveyed in the statement of a reasoning problem. The model may be based on knowledge stored in long-term memory that is explicit to a problem solving domain or that is based on a naive interpretation of a problem drawn from common sense beliefs.

While not explored in depth here, there has been a fair amount of research to suggest that sociocultural background can have a strong impact on reasoning correctly. For example, Luria (1976) and Scribner (1979) describe the difficulties faced by persons with little formal schooling in the solution of formal reasoning problems. Scribner (1979) claims that ability to solve abstract reason problems is possible only if one has learned such form of reasoning as a "genre" of thought through the medium of formal schooling or some other social medium.

Psychometric and Individual Differences Research on Reasoning from Incomplete Knowledge

The material discussed in the previous section has not focussed on assessment, psychometric theory, or differential psychology. The examples of research overviewed in the present section, however, have focussed on assessment and individual differences. The work of Sternberg (1977a) and Pellegrino and Glaser (1982) exemplify approaches to these issues which utilize tightly specified information processing models of task behavior in the domain of analogies problem solving. According to Pellegrino and Glaser (1982) expertise in solving analogy problems (of the sorts often encountered in aptitude tests) is affected by skill in three underlying cognitive factors. These factors include: management of memory load; organization of an appropriate declarative (or conceptual) knowledge base; and procedural knowledge of task constraints.

Frederiksen and Ward's (1978) GRE-sponsored work on the assessment of scientific problem solving ability represents a psychometric approach to reasoning skills assessment that does not rely on a formal information processing model. A strength of the approach was its emphasis on problems with high face validity; it used simulations of problems that might be encountered by social research scientists. Figure 5 lists the problem types and reasoning tasks presented examinees in the study.

This work (in part) probed relationships between examinees' ability to reason in an open-ended fashion as might be required in real social science research situations, and examinees' verbal and quantitative GRE scores, and GRE Advanced Psychology test scores. The results of the research indicated that several indices of the quality of examinees' open-ended responses correlated slightly better with GRE Verbal scores than with GRE Quantitative scores. Scores on the GRE Advanced Psychology test tended to correlate lower than GRE verbal or quantitative scores with measures of quality of reasoning on the criterion battery of open-ended scientific reasoning tests. Three measures of productivity of ideas in reasoning on the open-ended scientific tests showed relatively lower correlations with GRE test scores, but some evidence of higher correlation among each other. The results of the study indicated that the reasoning performance of examinees on the open-ended test was tapping skills which could be measured reliably and which reflected abilities that were only partially captured by existing GRE measures.

In a follow-up study to this research Ward, Frederiksen, and Carlson (1980) compared performance on the original free-response version of the Formulating Hypothesis test with performance on a machine-scorable version of the test. In the machine-scorable version examinees were asked to mark appropriate responses to problems, where responses were presented in a list format. The same quality and frequency of response measures were

Figure 4

Tests of Scientific Thinking

1. Formulating Hypotheses (FH). Each problem consists of a brief description of a research study, a graph or table showing the principal results, and a statement of the major finding. The task is to write hypotheses that might explain, or help to explain, the finding. The subject is asked to write not only the hypothesis he/she thinks is most likely to be correct, but also other hypotheses that should be considered in interpreting the data or in planning another investigation.

2. Evaluating Proposals (EP). The subject is asked to suppose that he/she is teaching a senior course in design and methodology, and that as a class exercise has asked the students to write brief proposals of research studies. Several proposals presumed to have been written by the students are presented as the test items. The subject's task is to write critical comments for each student regarding the design, methodology, or theoretical position of the proposal.

3. Solving Methodological Problems (SMP). Each problem is a brief statement of a methodological problem encountered by a graduate student or psychologist in planning a research study. The subject's task is to write suggested solutions to the methodological problem.

4. Measuring Constructs (MCH). Each problem consists of a name and definition of a psychological construct (e.g., conservatism, bigotry, leadership). The task is to suggest methods for eliciting the behavior so that it can be observed and measured without resorting to ratings or self-report methods.

Frederiksen, N., & Ward, W. C. Measures for the study of creativity in scientific problem-solving. Applied Psychological Measurement, 1978, 2(1), 1-24.

utilized in evaluating performance on the two tests. In addition to the Formulating Hypothesis test, subjects were administered a battery of tests representing eight cognitive factors. These factors included: vocabulary, quantitative reasoning, induction, logical reasoning, cognitive flexibility, expressional fluency, ideational fluency, and knowledge of psychology.

The results of the study suggested that the free-response and machine-scorable versions of the Formulating Hypothesis test were not alternate forms of the same test. Quality of response measures on the machine-scorable version of the test appeared to constitute information about examinees' aptitudes that was already measured by the GRE verbal and quantitative measures and by the GRE Advanced Psychology test. Quality of response measures was also found to be related to factor measures of knowledge of psychology, induction, and logical reasoning. Quality of response measures on the free-response version of the Formulating Hypothesis test contributed little information not already available on the machine-scorable version of the same test.

The various frequency of response measures on the free-response version of the Formulating Hypothesis test, however, did not show a very high relationship to measures of the same kind in the machine-scorable version of the test. Frequency of response measures on the free-response test showed relatively higher correlations with factors representing expressional fluency and ideational fluency. The GRE scores did not associate as strongly with the expressional fluency and ideational fluency factors. The outcome of the research thus suggests that reasoning from incomplete knowledge in Formulating Hypothesis items draws on some reasoning skills that are not currently represented by GRE measures and that these skills involved the ability to produce numerous ideas about the nature of the problem situations represented by items. However, the utility of such ideational fluency skills for graduate school performance has not been demonstrated.

Frederiksen, et al. (1981) investigated performance of first year and fourth year medical students on a set of branching formulating hypothesis problems. Some problems were presented in a multiple choice structured response format. Some problems had a medical problem solving content, whereas others did not. Subjects were also administered a battery of cognitive ability tests, and performance on this battery led to identification of six factors: reasoning, verbal comprehension, medical knowledge, ideational fluency, cognitive flexibility, and science achievement. The results of the study suggested that fourth year medical students rely noticeably on knowledge of medicine in both the free-response and structured tests as would be expected. On the free-response version of the medical problem solving test, students rely more on reasoning and verbal comprehension than they do on cognitive flexibility and ideational fluency.

Both first year and fourth year medical students showed an association between their ideational fluency factor scores and scores of their ability to generate numerous ideas on the free-response version of the medical problem solving test. In addition, fourth year students showed an association between their reasoning factor scores and scores of their ability to generate numerous ideas on the free-response test. The results support the hypothesis that as medical students gain in their knowledge of medicine they are more likely to rely on reasoning strategies based on their knowledge of medicine and that reasoning skills are also important to non-medical problem solving that involves generation of a hypothesis. Regardless of number of years of medical study, subjects showed an association between their propensity to generate a hypothesis in non-medical problem solving and their ideational fluency or cognitive flexibility.

Ward, Carlson, and Woisetschlager (1983) investigated the feasibility of a new "ill-structured" logical reasoning item type for the GRE Analytical Section test. The new ill-structured items were developed from earlier research on machine-scorable versions of the tests of scientific thinking (Ward, Frederiksen, & Carlson, 1980). The new item type presented examinees with a several sentence description of an everyday situation accompanied by a graph or table showing numerical relationships among variables described in the situation description. In addition, a finding and a question concerning a conclusion about the finding were stated. Examinees were required to select the best multiple choice answer option for the question. The new items did not require any specialized knowledge of an academic domain for their solution; although the items were multiple choice in nature, examinees needed to generate a number of intermediate inferences in order to arrive at an optimal conclusion to answer a question. In this regard these items resembled Simon's (1978) notion of ill-structured problems.

Ward, et al. (1983) found that performance on ten of the new items correlated .73 with performance on ten GRE logical reasoning items that had been classified beforehand as "ill-structured." Performance on the new items correlated .68 with performance on ten other GRE logical reasoning items that were judged beforehand to be "well-structured." Performance on all three of the aforementioned items correlated slightly lower with scores on the letter set's inductive reasoning test.

Ward, et al. (1983) concluded that performance on their new item type did not appear to tap skills that were different from those already assessed by the existing logical reasoning item type which showed "well-structured" problem solving. They did note that their new item type might be used to broaden the pool of item types for the present analytical section, without altering the constructs under assessment. Ward, et al. (1983) also pointed out that their results did not imply that there are not other ill-structured item types which would indeed assess skills not represented by items on the analytical section. They also indicated that evaluation of connections among steps that examinees take to solve ill-structured problems could lead to other useful measures of reasoning performances not represented by the current GRE analytical measure.

The Role of Thinking in Higher Education

Thinking as Both an Outcome and a Prerequisite

Intuitively, many of the tasks that are required of both undergraduate and graduate students involve some component of reasoning that may be distinct from both general verbal and general quantitative skills. In both the social and physical sciences, knowledge of the logic underlying experimental designs and research paradigms and such central concepts as causality and correlation would seem important. The manipulation of independent variables, the arranging of experimental conditions, and the drawing of appropriate conclusions from the resulting circumstances are central to scientific research. In mathematics, deductive reasoning is a key ability in proving theorems, understanding set theory, etc., and in learning foreign languages, rule learning may be involved. Many disciplines entail the development and evaluation of theories, which are involved networks of relations between concepts, many of which are abstract in nature. Examples of such theories can be found in numerous fields--kinetic theory, Gestalt psychology, the theory of evolution, and Keynesian economics (Ennis, 1962). Ennis (1962) has laid much of the foundation in education research for thinking about critical thinking, which, defined as "the correct assessing of statements" (Ennis, 1962), is at the heart of much scholarly work. Ennis (1962), in a most thoughtful analysis, listed twelve aspects of critical thinking as follows:

1. Grasping the meaning of a statement.
2. Judging whether there is ambiguity in a line of reasoning.
3. Judging whether certain statements contradict each other.

4. Judging whether a conclusion follows necessarily.
5. Judging whether a statement is specific enough.
6. Judging whether a statement is actually the application of a certain principle.
7. Judging whether an observation statement is reliable.
8. Judging whether an inductive conclusion is warranted.
9. Judging whether the problem has been identified.
10. Judging whether something is an assumption.
11. Judging whether a definition is adequate.
12. Judging whether a statement made by an alleged authority is acceptable.

Reasoning skills are recognized both as prerequisites for and intended outcomes of post-secondary education. Bowen (1977) for one has suggested that among the ten most desirable outcomes of higher education is the broad goal of "rationality," which is characterized by:

- the capacity for logical thinking based on useful assumptions
- the ability to see facts and events in an objective manner, distinguishing the normative, ideological, and emotional from the positive and factual
- the disposition to weigh evidence, evaluate facts and ideas critically, and to think independently
- the ability to analyze and synthesize.

Hartnett and Willingham (1979) also point out educators' long-standing contention that one of the most important outcomes of higher education is a developed ability "to think critically and analytically."

"Beyond the acquisition of knowledge, it has often been claimed that what one should hope to garner from education...is an ability to reason carefully, to recognize valid and invalid inferences and assumptions, to approach problems with an attitude of inquiry that demands evidence in support of assertions, to identify logical contradictions in arguments, and so on."

Thinking skills are also recognized as important instructional outcomes long before students reach the advanced levels of higher education. The College Board (1983) has undertaken a ten-year effort to strengthen the academic quality of secondary education. With the assistance of high school and college faculty, the Board has defined several basic academic competencies--broad, generally applicable intellectual skills thought to be essential for effective participation in all fields of college study and without which "knowledge of history, science, language, and all other subjects is unattainable" (p. 7). These basic competencies are reading, writing, speaking and listening, mathematics, studying, and reasoning.

Under reasoning, the following abilities have been identified:

- The ability to identify and formulate problems, as well as the ability to propose and evaluate ways to solve them.

- The ability to recognize and use inductive and deductive reasoning, and to recognize fallacies in reasoning.

- The ability to draw reasonable conclusions from information found in various sources, whether written, spoken, or displayed in tables and graphs, and to defend one's conclusions rationally.

- The ability to comprehend, develop, and use concepts and generalizations.

- The ability to distinguish between fact and opinion.

Arons (1979) identifies 16 reasoning capacities expected of college students; these capacities are listed in Figure C.1. Aron's list is intended to be suggestive of the kinds of skills which entering college students are expected to utilize. The list is rich in its citation of examples of various kinds of reasoning required in dealing with academic subject matter.

Other examples of explicit objectives related to thinking and reasoning skills can be found in various collections of instructional objectives. As one example, the Instructional Objectives Exchange (1971) has presented objectives dealing with "one side of judgment: deductive logic and assumption recognition." They are:

- o Given a series of statements (either emotionally laden or not) which are expressed in various conditional logic formats, the student will discriminate which of the conclusions are logically valid or invalid
- o Given statements expressed in class logic formats (either emotionally laden or not) the student will discriminate valid and invalid statements
- o Given a set of statements, the student will recognize whether or not particular assumptions are necessary and will be able to identify assumptions underlying arguments
- o Given descriptions of observations, the student will choose the observation with the highest reliability

On yet another front, in a strongly worded statement the Commission on Higher Education Issues concluded that no student who has not demonstrated fundamental competencies in five areas, including reasoning, should be admitted to full participation in any baccalaureate degree.

Role of Thinking in Reading and Writing

An ability to comprehend verbal material and a facility for written expression are two very general skills commonly expected of virtually all students. Exactly how these two general verbal skills relate to reasoning

Figure 5

Reasoning Skills Expected of College Students

Recognition, Identification, and Control of Variables

It is generally expected that students can recognize and control, or weigh variables in setting up experiments or in solving verbal problems in the natural and social sciences. For example, they should be able to account for the possible influence of length, diameter, and composition in the well-known Piagetian task of the bending of rods, then control two of these variables in establishing the effect of the third. Similarly, the interpretation of historical phenomena requires recognition and sorting of political and social factors within the information available, following which the student must decide whether cause-and-effect relationships are discernible or whether the variables are so confounded as to preclude a reliable inference. Economics, political science, and experimental psychology all depend upon student facility in dealing with similar considerations of variables and their relations.

Arithmetical Reasoning

Ratio reasoning is required for predicting the alteration of gravitational or electrical forces between point objects with changes of mass, charge, or spacing. Similarly, the use of scale ratios is necessary for interpreting maps, for determining sizes of objects viewed under a microscope, or for comparing relative changes in volumes and surface (or cross-sectional) areas with changes of scale. Further, it is useful to relate such arithmetical changes to, say, the disproportionately large cross-sections of bones in larger animals, or to the rapidity of the dissolving of a material under finer subdivision. The use of similar kinds of ratio reasoning in connection with demographic data may be required in political science or sociology, and in connection with scaling of factors or rates in economic problems.

Interpretation of division is needed in dealing with concepts of density, velocity, or acceleration in physics; moles, gas behavior in chemistry; and in the study of population and other growth rates in biology. Combinatorial problems arise in elementary genetics, in probability considerations, in control of variables, and in design of experiments.

Forming and Comprehending Propositional Statements

Formation of intelligible propositional statements requires an intuitive grasp of the rules of logic, and of the grammar in which such statements are to be made. For example, forming or understanding verbal statements involves inclusion, exclusion, and serial ordering. In addition, one must grasp syntactical constructions such as double negatives, subjunctive mood, and the capacity to deal with elementary one- or two-step syllogistic reasoning. This is **not** meant to include involved propositional logic in which one is forced towards symbols or Boolean algebra for elucidation. The **basic** skill indicated, however, applies to all areas of study which require the use of language.

Ability to Paraphase Paragraphs of Text in One's Own Words

This expectation is applicable to all areas of study. A word of warning is needed here; students may be able to rephrase a paragraph using language similar to that in the text without understanding its content. Thus the insistence that they put it in their own words is of critical importance.

Awareness of Gaps in Knowledge or Information

This problem has two dimensions--gaps in the student's own knowledge, or incompleteness of known information in a given area of study. In the former case, it is expected that when a student fails to recognize the meaning of a word or symbol used in an oral presentation or a passage of a text, he or she will sense the need for establishing its meaning, and have the motivation to do whatever is necessary to establish it.

When the problem is incompleteness of information in a particular context, the student should realize that a definite conclusion cannot be reached, or should note that conclusions or decisions are being reached in the face of incomplete data and hence that such conclusions must be qualified accordingly. This overlaps with problems of psychological maturity, upon which depends the capacity to recognize and tolerate ambiguity in the material under study.

In a sequence of development of a given subject matter, the student is expected to gradually distinguish what has become known or clearly established at any particular point from what has not been so established. This implies learning to anticipate some of the questions still to be asked.

Understanding the Need for Operational Definitions

In general, students are expected to learn the criteria by which it is possible to determine whether or not a definition is operational. These criteria include: realizing when a concept in a passage of text has not been clearly defined; recognizing the necessity of using only words of prior definition in forming a new definition; becoming aware of the appeal to shared experience in forming operational definitions.

Translating Words into Written Symbols and Written Symbols into Words

The skills necessary for such operations are more rigorous than those needed for paraphrasing textual passages. Examples include: transforming a verbal statement into its equivalent arithmetical, algebraic, or graphical form in any of the natural or social sciences; interpreting a graphical presentation or the results of a symbolic problem solution in words, extracting its content while expressing the relevant qualifications and restrictions.

Discriminating between Observation and Inference

Students in all academic disciplines must learn to recognize the observational, empirical, or experimental facts that are available in a text presentation or in laboratory work. The next step is to separate these clearly from the concepts that may be formed or the inferences that may be drawn from them. An example of this would be identifying observed facts concerning the extent of illumination of the moon relative to its position with respect to the sun and then separating these from the inference that moonlight is reflected sunlight. Another example would be distinguishing the observed behavior of electrical circuits (that bulbs get dimmer as more are put in series at the same source) from the concept of "resistance" which is induced from the observed behavior. In this particular case, further distinctions would then need to be made between inferences concerning the nature of electrical resistance, and the predictions that can be made concerning phenomena in more complicated circuits.

Other examples from quite different areas of inquiry include: separating Mendel's observations of nearly integral ratios of population members having different color or size characteristics from the inference of discrete elements controlling inheritance; the distinction--common in the study of literature--between analysis of the **structure** of a novel or poem and an **interpretation** of the work; and the historian's task of recognizing the distinction between primary historical data and his own interpretation of such data.

Analyzing a Line of Reasoning in Terms of Underlying Assumptions

Every line of reasoning has an underlying set of assumptions separate from the factual data it may utilize. Students need to develop the capacity first to discover and second to distinguish among assumptions, assertions, and reasoned conclusions.

Drawing Inferences from Data and Evidence, Including Correlational Reasoning

Separate from the analysis of another's line of reasoning comes the formulation of one's own. For example, given the observation that the spot formed on the screen by the cathode beam (Thompson's experiment) remains coherent (i.e., does not smear out) under both electric and magnetic deflection, what inferences can be drawn concerning the origin and character of the beam? Or: given the code of laws of Hammurabi, what--if anything--can be inferred about how the people subject to it lived, and what they held to be of value? Yet another example is the problem of recognizing possible functional or cause-and-effect relationships in either positive or negative correlations in the face of statistical scatter or uncertainty; for example, discerning relative or competing effects of light, heat, or moisture on a biological population (with a simultaneous awareness of whether or not the variables have been adequately controlled).

Ability to Discriminate between Inductive and Deductive Reasoning

Students should be able to follow inductive reasoning used in generating a concept, model, or theory, and use deductive reasoning in testing the validity of a construct. They should perceive the analogous patterns of scientific thought arising in such broadly diverse areas as the Newtonian Synthesis, wave versus particle models of light, atomic-molecular theory, gene theory, theory of evolution, economic or sociological models of society or its parts, and so on.

Performing Hypothetico-Deductive Reasoning

Students should be able to visualize, in the abstract, outcomes that might stem from changes imposed on a given system or situation, whether it be in scientific, literary, historical, economic, or political contexts, and effect such visualizations through reasoning within the basic principles or other rational constraints applicable to the system.

Performing Qualitative, Phenomenological Reasoning or Thinking

In science and mathematics, the ability to recall formulas and manipulate them algebraically does not by itself indicate complete understanding of a subject area. Students should also be able to give **qualitative** explanations of principles and make direct inferences from them without referring to the results of numerical calculations. They should be able to apply phenomenological reasoning without relying on mathematical formalism.

Checking Inferences, Conclusions, or Results

Skills in this category include: testing for internal consistency; using alternative paths of reasoning; examining extreme, limiting, or special cases.

In some instances, only initial or preliminary levels of the skills listed in the preceding section are actually presupposed in college work at introductory levels, while enhancement and further development of such skills are often implicit objectives of the courses of instruction. In addition to these objectives, others are at least implied, when not explicitly articulated, in most statements of the cognitive goals of higher education. Two of these more general goals which subsume many of the preceding objectives can be articulated as follows.

Developing Self-Consciousness Concerning One's Own Thinking and Reasoning Processes

It is generally desired that students learn to become explicitly aware of the mode of reasoning being used in particular situations. This provides the basis for consciously seeking to transfer familiar modes of reasoning to new and unfamiliar situations. In general, students should learn to attack a new problem or unfamiliar situation by first forming very simple related problems, by asking themselves questions derived from the simplest and most concrete aspects that seem to underlie the given situation.

Developing the Skills of One's Discipline

Finally, students are expected to combine the preceding modes and processes into the general skills of problem solving as practiced by the discipline(s) of choice.

Arons, A. B. Some thoughts on reasoning capacities implicitly expected of college students. In J. Lochhead & J. Clement (Eds.), ...e process instruction. Philadelphia, PA: The Franklin Institute Press, 1979, 210-213.

and thinking is complex, but it suffices to say that reasoning processes are involved in each. Thorndike (1973) concluded from three sets of evidence that, after basic decoding is accomplished, reading is primarily an indicator of the general level of an individual's thinking and reasoning processes rather than a separate set of specialized skills. The role of reasoning in reading was shown in a slightly different way by Baker (1979), who, after intentionally introducing confusions and inconsistent information into reading passages, found the college students often failed to recognize logical inconsistencies in reading. Because the ability to detect written fallacies in reasoning has been shown to be related to reading achievement (Holihan, 1980), the failure to recognize such inconsistencies would undoubtedly hinder comprehension.

Flower and Hayes (1980) have characterized writing as the act of balancing a number of simultaneous constraints (e.g., the demand for integrated knowledge, linguistic conventions, and the rhetorical problem itself) and suggest that a writer is very much like a busy switchboard operator trying to juggle a number of demands on attention and constraints on possible actions. (We note in passing that the current analytical reasoning items in the GRE General Test appear in many ways to demand the capacity to deal with various constraints.)

Research in Graduate Education Constituency Surveys. To assess the impact of a planned restructuring of the GRE Aptitude, surveys of both examinees and institutions were undertaken in 1975. Of the possible new measures listed (abstract reasoning, scientific thinking, and study style) abstract reasoning was the one most favored by faculty, administrators, and students in all fields (Miller & Wild, 1979).

Examinees were also subsequently asked for their perceptions of alternative test item types developed for possible inclusion in a new analytical measure. Questions asked of examinees were "Do the questions measure abstract reasoning?" "Do the abilities tested seem important?" "Are the skills measured different from verbal and quantitative skills?" The following results, based on samples of 300 to 1,000 examinees, were obtained. As can be seen (Table 3), a majority of examinees thought each item type (even those not used in the analytical measure) reflected abstract reasoning.[3] Except for the letter sets format, all item types were thought to reflect important abilities by a majority of examinees, and a majority of examinees rated each item type as measuring something other than verbal and quantitative skills. Test takers were least convinced that logical reasoning items reflected a different skill. Thus, those specifically involved in graduate education appear to recognize the importance of abstract reasoning and believe themselves able to distinguish among test items with respect to their relevance in measuring this attribute.

<u>Critical Incidents of Graduate Performance</u>. A series of studies by Reilly provides another perspective on the skills required of graduate students. Reilly (1976) asked graduate faculty in English, chemistry, and

---

[3] In any future surveys to determine the face validity of analytical item types, it would be useful to include for baseline purposes item types that are not intended to measure abstract reasoning.

Table 3

Examinees' Reactions to Item Types

Item Type	Measure abstract reasoning?	Abilities seem important?	Different from verbal and quantitative skills?
Letter sets	73%	45%	80%
Logical reasoning[1,2]	67%	66%	59%
Analytical reasoning[1,2]	79%	50%	80%
Evaluation of evidence	76%	57%	84%
Analysis of explanations[1]	71%	57%	81%
Logical diagrams[1]	83%	52%	81%

Note. From 6% to 10% of students were uncertain whether the item types measured abstract reasoning. From 10% to 18% were uncertain whether the abilities tested by each item type seemed important. From 4% to 9% were uncertain about whether the skills measured were different from verbal and quantitative skills.

[1]Included in the original analytical test.

[2]Included in the revised analytical test.

psychology to provide examples of specific incidents which, when they occurred, caused them either to lower or raise their estimates of the competence of graduate students. A total of 52 distinct incidents was gathered, a number of which seem related to critical thinking. From these incidents, Reilly (1976) identified the following eight factors:

1. Independence and initiative
2. Conscientiousness
3. Enthusiasm
4. Critical facility
5. Teaching skills
6. Research and experimentation
7. Communication
8. Persistence

These factors exhibited a relatively high degree of consistency across the three fields studied. The critical facility factor was defined by seven incidents:

1. Repeatedly made irrelevant remarks during class or seminar discussion.
2. Talks at great length in class but exhibits little understanding of material on papers and tests.
3. Was often unable to consider new ideas objectively because of strongly held prejudices.
4. Submitted a paper or report which failed to address the assigned issues.
5. Presented ideas in a seminar, paper, or test in a poorly organized and disjointed fashion.
6. Was unwilling or unable to accept criticism.

7. Although able to criticize studies with facility was unable to suggest better alternatives.

Although they did not specifically define the critical facility factor, several other relevant incidents were also mentioned. These include:

1. Consistently offered well-founded and constructive criticisms of other students' presentations.
2. When making a judgment or reaching conclusions, this student supported his or her position with carefully documented research.
3. Performed an experiment without making proper checks.
4. Was unable to formulate a testable hypothesis from a theoretical analysis.
5. When this student asked a question, it was always relevant and usually perceptive.
6. Showed an ability to examine carefully an author's premises and frame of reference before accepting conclusions.

The last incident (#6) was rated as especially important in all three subject fields. In chemistry, it was rated as the eighth most critical incident (of 52), and in English and psychology, it was the second most important. Of the eight factors listed above, "critical facility" was ranked as the second most important in English and the fifth most important in both chemistry and psychology. Several of these empirically determined incidents seem related to the various aspects of critical thinking, etc., that were summarized previously.

Training and Measurement of Reasoning

Training. Frederiksen (1983) reviewed research in cognitive psychology on the training of thinking skills. He found some evidence that some thinking skills can be taught, but that this possibility is less when dealing with ill-structured problems than with well-structured ones. He emphasized the importance of further research on ill-structured problem solving in order to arrive at better models of the reasoning skills involved in problem solving. He also contrasted approaches that attempt to teach skills explicitly with those that emphasize the teaching of thinking skills by discovery. Finally, he also mentioned the dilemma of deciding the level of generality at which skills should be taught. Much research in cognitive psychology indicates that learning of problem solving skills proceeds more effectively with hands-on problems in specific problem domains than with formal instruction in the abstract principles of problem solving (Sternberg, 1980). Snow (1982) and Sternberg (1980) have also emphasized the critical role of motivation, attitudes, and cognitive style factors in the effective training of thinking skills.

Neimark and Santa (1975) concluded one part of their review on thinking and concept attainment by stating that "...abstract thinking may well be a product of formal education" (p. 192). And there is a good deal of information that formal education does have some impact. Pace (1974), as reported by Bowen (1977), reported that a majority of alumni and upperclassmen, when surveyed, thought that their college experiences had benefitted them "very much" or "quite a bit" with respect to their ability to think critically. "Harder" data collected by Keeley, Browne, and Kreutzer (1982) and Lehmann and Dressel (1962; 1963) found changes in critical thinking skills from the freshman to the senior years in college in students' ability to define problems, recognize stated and unstated assumptions, select pertinent information, formulate and select relevant hypotheses, and draw valid conclusions.

More specific, and much briefer, efforts aimed at undergraduate students have also reported some success. Feibel (1978) found significant changes in reasoning after training based on two different theories of cognitive development, and, with only three to six hours of training, Wood (1980) had some success in improving the logical reasoning skills (as measured by the Watson-Glaser Critical Thinking Appraisal) of undergraduate psychology students by using a computer program based on the popular game "Mastermind."

Successful attempts have also been made to improve the reasoning process of medical students in medical problem solving. More experienced physicians seem to be more adept than first year medical students in making efficient use of information (Mandernach, 1979), and medical students can improve in a relatively short time with appropriate training in a diagnostic problem formulation. The critical thinking skills of law students have also been improved with relevant intervention, e.g., with a language arts program emphasizing critical thinking and reading (Pachtman, 1976). That different kinds of training may differentially affect critical thinking skills has also been shown--(e.g., Agne & Blick, 1972; Sorenson, 1966). The more effective treatments are typically nontraditional ones using laboratory-centered, or experimental, data-centered approaches instead of traditional classroom lectures.

Ennis (1962) suggests that each of three dimensions of critical thinking "appear to be learnable somewhat as academic subject matter is learnable..." (p. 108). These dimensions involve the ability to use logical operators, knowledge of criteria for judging statements, and the ability to recognize when ther[illegible] enough evidence. The other side of the coin is that some types of reasoning may be remarkably resistant to change, and particular "logical" biases (e.g., tendency to select irrelevant confirmatory instead of possibly relevant disconfirmatory evidence) are difficult to eradicate (Mynatt, Doherty, & Tweney, 1977; Wason, & Johnson-Laird, 1972).

Measurement of Thinking. A 1977 ETS Test Collection bibliography of "Measures of Reasoning, Logical Thinking, and Problem Solving Ability" contains more than 100 references to measures of these abilities. A brief inspection of the entries reveals such subtests as logical reasoning, inductive reasoning, conditional reasoning, critical thinking, abstract reasoning, analytical reasoning, and syllogistic reasoning. It is clear that there has been no shortage of attempts to measure these skills.

Two of the more interesting measures appropriate for adults are the Watson-Glaser Critical Thinking Appraisal (Watson, & Glaser, 1980) and the Ennis-Weir Argumentation Test: An Essay Test of Rational Ability (Ennis, 1982). The Critical Thinking Appraisal includes five subtests of:

1. Inference. Discriminating among degrees of truth or falsity of inferences drawn from data.

2. Recognition of assumptions. Recognizing unstated assumptions or presuppositions in given statements or assertions.

3. Deduction. Determining whether certain conclusions necessarily follow from information in given statements or premises.

4. Interpretation. Weighing evidence and deciding if generalizations or conclusions based on the given data are warranted.

5. Evaluation of arguments. Distinguishing between arguments that are strong and relevant and those that are weak or irrelevant.

This test has probably been researched about as much as any test of critical thinking (or related trait) and has been used in as many research studies as any other test. Various studies have shown that the measure reflects a dimension that is independent of those tapped by various other ability and achievement measures (Follman, Miller, & Hernandez, 1969) and by other measures, such as those designed to measure the "structure of intellect" (Landis, 1976). Others point out, however, that the various measures of critical thinking, including the Critical Thinking Appraisal, contain a substantial component of verbal ability (Hartnett, & Willingham, 1979).

The Ennis-Weir Test is intended to measure the ability to evaluate an argument and to formulate in writing an argument in response. The test takes the form of a fallacy-laden letter to the editor of a newspaper. Examinees read the letter and write an essay evaluating the arguments given in each paragraph. A theoretical analysis has been used to validate the construct, but to our knowledge no empirical studies have been conducted of the construct validity of the measure.

The point to be made here is that a wide variety of standardized measures are available (some of which have been researched more extensively than others) and that some of these measures may provide reference tests against which to gauge the construct validity of the GRE analytical measure.

Summary

One objective of this study was to provide a selective review of "higher level reasoning" skills. On the face of it, there would seem to be little problem in providing such a review, especially in light of the immense amount of educational and psychological research conducted on the topic. Difficulties arise, however, in developing a framework for defining different kinds of reasoning skills and for showing the relationships among different skills. These difficulties have more than one origin, but essentially they all center on ambiguities involving reasoning as an object of research and assessment. In addition, the situation is made more complex when we consider attempts to define and investigate the kinds of informal reasoning that people exercise in their everyday activities.

In our judgment, the most important factors affecting the interpretation of research on reasoning are the particular description (or theory) of mental functioning and the model of reasoning that guide the research. This notion is consistent with the views of Messick (1981) and Frederiksen (1982) concerning the centrality of both a cognitive theory and a performance model in the investigation of construct validity. In interpreting research on reasoning we need to appreciate how reasoning is defined in terms of cognitive skills and also how task structures in criterion settings invoke the use of reasoning skills. Few of the previous studies of reasoning offer highly developed descriptions or theoretical models of reasoning skills. Without both a cognitive model and a performance model of reasoning skills it is difficult to formulate an adequate measurement model that might be used to assess reasoning skills. In contrast, although many of the research studies cited do not embody a formal description or theory of reasoning, many of the tasks studied do show considerable face validity with respect to the reasoning demands faced by students in graduate work, especially when compared with that of some of the theoretically based work. Often there is a tendency for theoretically motivated research to focus on reasoning skills that are so tightly constrained in their purpose and application that they exhibit little similarity to the kinds of reasoning tasks faced by students in graduate school; on the other hand, focus on such "micro" skills might be usefully considered, to the extent that it may help predict performance on more general reasoning tasks. With these caveats in mind we have attempted to review research on reasoning in a way that might contribute to studies of the construct validity of the GRE analytical measure.

One area considered was "critical thinking," which might be deemed as the most general of reasoning skills important to successful graduate education. Some of the various terms that are used in reference to the skills we have in mind range from "reasoning," "judgment," and the very general "thinking" (which may be modified by such adjectives as critical, abstract, rational, logical, analytical, and creative), to highly specific labels such as "transitive inference." A variety of terms are often paired in the general area: for example, thinking and concept attainment (Neimark, & Santa, 1975), thinking and reasoning (Wason, & Johnson-Laird, 1970), deductive logic and assumption recognition (Instructional Objectives Exchange, 1971), and critical thinking/critical reading

(Pachtman, 1976). But, critical thinking is also distinguished from a variety of other kinds of thinking such as creative thinking, associative thinking, and problem solving (Ennis, 1962).

There is considerable evidence that reasoning skills are viewed as extremely important in higher education generally, and in graduate education in particular, and further that there are some definite expectations with respect to the kinds of reasoning skills that students should have. Both of the reviews touch on many of the categories or terms used in discussions of reasoning or thinking.

IV. An Overview of Skills Measured by the Analytical Section

Integrating the approaches and findings of previous research in terms of their implications for the construct validity of the GRE analytical section is not easy, nor can it be done comprehensively. In this section we will highlight what appears to be some of the most important cognitive and linguistic skills that should be required in solving the two item types on the analytical section. Based on our review of the literature, Figure 7 summarizes the skills that we believe are the most important for solving analytical reasoning and logical reasoning item types. The figure is organized so that the reader can readily compare similarities and differences in the skills required on the two item types.

The three classes of skills might be thought of in terms of a hierarchy. The top level of this hierarchy concerns examinees' general knowledge and expertise in multiple choice test taking. This knowledge guides the application of general reading comprehension skills represented by the second level of the skill hierarchy. At the bottom tier of the hierarchy are represented some of the specialized cognitive and linguistic skills which are required by an item type. These bottom-level skills are utilized in the service of the more general problem solving and reading strategies guiding multiple choice test problem solving.

It is important to note that the skills represented at the top two tiers of the hierarchy are important for all aptitude test taking. In the GRE General Test these skills are shared in common across not only the two analytical section item types, but also across all three sections of the

Figure 6

Skills Hypothetically Required for Solution of Analytical Section Items

	Analytical Reasoning Items	Logical Reasoning Items
Multiple Choice Test Problem Solving and Reasoning Strategies	o Understand test instructions o Manage time in problem solving o Establish the goal of identifying only a single correct answer option for each item o Follow strategies to eliminate distractor options when necessary	o Understand test instructions o Manage time in problem solving o Establish the goal of identifying only a single correct answer option for each item o Follow strategies to eliminate distractor options when necessary
General Reading Comprehension Skills	o Vocabulary recognition o Sentence recognition o Paragraph structure recognition (limited)	o Vocabulary recognition o Sentence recognition o Paragraph structure recognition (important)
Specialized Cognitive and Linguistic Skills	o Recognizing algebraic-like relationships and properties of these relationships - Conservation of quantity - Reflexivity - Symmetry - Transitivity - Class inclusion/non-inclusion and set membership/non-membership o Ability to reason deductively/inductively about relationships, including ability to enact serial operations o Ability to keep track of problem information and alternative models of a problem (short-term memory span) o Ability to use diagrams to encode problem information (optional) o Ability to generate combinations and to search problem space for counter-examples	o Recognizing the structure of an argument - Isolate premises - Isolate conclusions - Isolate noncritical information in an argument o Evaluating the validity of conclusions - Understanding how logical entailment is qualified in an argument - Reasoning deductively - Reasoning inductively

General Test. There is probably an important amount of overlap in the common, general skills required to work different sections and item types. This overlap contributes to the intercorrelation of performance across all item types. In addition, systematic overlap may occur across sections and item types because of commonality in specific cognitive and linguistic skills. This matter will be discussed in the ensuing overview of the specific cognitive and linguistic demands of each analytical section item type. Before starting this discussion it is useful to remember that at present we lack empirical evidence verifying the contribution of the cognitive and linguistic skills mentioned to performance on each analytical item type. A goal of the discussion is thus to identify skills that could become part of an empirical research study on the construct validity of analytical section item types.

The analytical reasoning item type and quantitative General Test items share a common emphasis on reasoning about relationships among variables. While analytical reasoning items do not require manipulation of numbers for their solution, the relationships among variables in these problems emphasize algebraic properties important to quantitative reasoning. For example, relationships of transitivity/intransitivity, symmetry/asymmetry, and conservation of quantity are often critical to solving analytical reasoning items. As with quantitative items, the solution of analytical reasoning items requires noticing key relationships among variables and interpretation of questions in terms of these relationships. The difficulty of problems has been found to increase as the number of relationships required to work items increases.

There are a number of information processing skills which seem critical to working analytical reasoning items. Ability to reason with serial operations seems important, as does short-term memory capacity, and the degree of proficiency in these skills would seem to become more important as problems become more complex and difficult. Ability to represent information spatially in terms of diagrams may be important, and this importance is likely to be a function of the difficulty of problems and individual examinee's preference or need for diagramming strategies. Representation with other symbol systems is an alternative strategy for some examinees.

The test development specifications for the logical reasoning item type are explicit as to the reasoning skills which are targeted for assessment by each item. The reasoning abilities tested emphasize skills in analyzing verbal arguments and in evaluating the degree to which arguments are supported by statements. Vocabulary and sentence recognition skills seem critical to the item types, since examinees must (a) derive a precise understanding of the propositions represented by sentences and (b) understand the way in which vocabulary qualifies the interpretation of statements. Examinees must be able to discern how the information in a passage is organized and connected together to form a logical argument. Thus, it is necessary for examinees to understand both the discourse structure of item passages and the semantic and logical relationships among sentences. They must also be able to distinguish information that is part of a set of premises from information that is implicitly or explicitly alleged to be a conclusion based on premises. Examinees' ability to detect grammatical structures and vocabulary terms signifying logical entailment would seem to be of critical importance. Examinees also have to distinguish information that is not relevant to conclusions that might be drawn from a passage. That these items correlate more strongly with verbal ability than do analytical reasoning items is thus not surprising. They represent a crucial subset of reading comprehension skills that may warrant separate testing, if not separate reporting.

V. Summary

The objective of this paper was to work towards a better understanding of the current version of the GRE analytical measure as an indicator of cognitive skills that are important for success in graduate education. This better understanding was to be accomplished within the context of the considerable developmental research that preceded the selection of test item types for the original version of the GRE analytical measure, which was introduced in 1977.

However, the choice and mix of item types for the original measure was dictated largely by traditional psychometric considerations, as was the measure's construct validity. Since the analytical measure was introduced, Messick, Frederiksen, and others have advanced the conceptions of construct validity, stressing the role of theory in understanding both the subskills that are critical to performance on a test and also the relationships of test performance to performance or "real world" criteria. Recent advances in cognitive psychology have also contributed to educational and psychological testing by focusing on the specific processes that underlie the solution of test items.

This paper has described the current version of the GRE analytical measure, its developmental history, and the research that led to revisions of the original measure. Selective reviews of literature have been presented for two relevant areas. One review focused on the psychological and cognitive science research on reasoning and problem solving. The other involved the somewhat more general educational research on the role of thinking or reasoning in higher education.

The results of these reviews suggested a number of things. First, there is a substantial amount of literature in these areas. However, there do not seem to be any comprehensive educational or psychological theories for reasoning, thinking, or problem solving. The absence of any well-developed theories renders the measurement of reasoning skills significantly more difficult. On the other hand, these reviews strongly suggest, at least indirectly, the importance of higher-level reasoning skills in higher education.

This paper also has analyzed each of the two item types that are used in the current version of the analytical measure. This analysis provides an overview of the major reasoning skills that seem to be required for the successful solution of these item types.

A major conclusion of this paper is that the current version of the GRE analytical measure appears to reflect skills that are important for success in graduate education, but that these skills constitute only a subset of many other important analytical skills. Further empirical research is recommended in order to document more precisely the demands of graduate education on students' analytical abilities.

References

Agne, R., & Blick, D. A comparison of earth science classes taught by using original data in a research approach technique versus classes taught by conventional approaches not using such data. Journal of Research in Science Teaching, 1972, 9(1), 83-89.

Anderson, J. R. Acquisition of cognitive skill. Psychological Review, 1982, 89, 369-406.

Arons, A. B. Some thoughts on reasoning capacities implicitly expected of college students. In J. Lochhead & J. Clement (Eds.), Cognitive process instruction. Philadelphia: The Franklin Institute Press, 1979, 209-215.

Baker, L. Comprehension monitoring: Identifying and coping with text confusions. Journal of Reading Behavior, 1979, 11(4), 365-374.

Blumberg, A. E. Logic: A first course. New York: Alfred A. Knopf, 1976.

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975, 131-149.

Bowen, H. R. Investment in learning: The individual and social value of American higher education. San Francisco: Jossey-Bass, 1977.

Carroll, J. Psychometric tests as cognitive tasks: A new "Structure of Intellect." In L. B. Resnick (Ed.), The nature of intelligence, Hillsdale, NJ: Lawrence Erlbaum Associates, 1976, 27-56.

Chi M., Feltovich, P. J., & Glaser, R. Categorization and representation of physics problems by experts and novices. Cognitive Science, 1981, 5, 121-152.

The College Board. Academic preparation for college: What students need to know and be able to do. New York: College Entrance Examination Board, 1983.

Collins, A. Processes in acquiring knowledge. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977, 339-363.

Collins, A., Warnock, E. H., Aiello, N., & Miller, M. L. Reasoning from incomplete knowledge. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975, 383-415.

DiSessa, A. A. Unlearning Aristotelian physics: A study of knowledge-based learning. Cognitive Science, 1982, 6, 37-75.

Dressel, P., & Mayhew, L. (Directors). General education: Exploration in evaluation. Final Report of the Cooperative Study of Evaluation in General Education. Washington, DC: American Council on Education, 1954.

Duran, R. P., Stricker, L. J., & Ward, W. C. Implications of cognitive psychology for assessment. Preliminary Report. Princeton, NJ: Educational Testing Service, 1982, unpublished internal memorandum.

Educational Testing Service. GRE Information Bulletin, 1982-83. Princeton, NJ: Educational Testing Service, 1982.

Ennis, R. H. A concept of critical thinking: A proposed basis for research in the teaching and evaluation of critical thinking ability. Harvard Educational Review, Winter, 1962, 32, 81-111.

Falmagne, R. J. Reasoning: Representation and process in children and adults. Hillsdale, NJ: Lawrence Erlbaum Associates, 1975.

Feibel, W. M. Effects of training on the transition from concrete formal reasoning in college students. Dissertation Abstracts, 1978, 3547-B.

Flower, L. S., & Hayes, J. R. The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), Cognitive processes in writing. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1980.

Follman, J., Miller, W., & Hernandez, D. Factor analysis of achievement, scholastic aptitude, and critical thinking subtests. The Journal of Experimental Education, 1969, 38(1), 48-53.

Frederiksen, N. Implications of theory for instruction in problem solving (RR-83-19). Princeton, NJ: Educational Testing Service, April 1983.

Frederiksen, N. Construct validity and construct similarity: Methods for use in test development and test validation (RR-82-45). Princeton, NJ: Educational Testing Service, November 1982.

Frederiksen, N., & Ward, W. C. Measures for the study of creativity in scientific problem solving. Applied Psychological Measurement, 1978, 2, 1-24.

Frederiksen, N., Ward, W., Case, S. M., Carlson, S., & Samph, T. Development of methods for selection and evaluation in undergraduate medical education (RR-81-4). Princeton, NJ: Educational Testing Service, March 1981.

Gentner, D., & Stevens, A. L. (Eds.). Mental Models. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

Green, B., McCloskey, M., & Carramaza, A. The relation of knowledge to problem solving, with examples from Kinematics. Paper presented at the NIE-LRDC Conference on Thinking and Learning Skills. University of Pittsburgh, October 1980.

Hartnett, R. T., & Willingham, W. W. The criterion problem: What measure of success in graduate education? (GREB Research Report No. 77-4R). Princeton, NJ: Educational Testing Service, 1979.

Holihan, D. D. The relationships between the ability to detect written fallacies in reasoning and realistic levels of career aspiration of students in grade twelve. Dissertation Abstracts, 2033-A, 1980.

Instructional Objective Exchange. Judgment: Deductive logic and assumption recognition: Grades 7-12. Los Angeles: Instructional Objective Exchange, 1971.

Johnson-Laird, P. N. Mental models. Cambridge, MA: Harvard University Press, 1983.

Johnson-Laird, P. N., & Steedman, M. The psychology of syllogism. Cognitive Psychology, 1978, 10, 64-99.

Johnson-Laird, P. N., & Wason, P. C. (Eds.). Thinking: Readings in cognitive science. New York: Cambridge University Press, 1977.

Keeley, S. M., Browne, M. N., & Kreutzer, J. S. A comparison of freshmen and seniors on general and specific essay tests of critical thinking. Research in Higher Education, 1982, 17, 139-154.

Landis, R. The psychological dimensions of three measures of critical thinking and twenty-four structures-of-intellect tests for a sample of ninth-grade students. Dissertation Abstracts, 1976, 5705-A.

Larkin, J. Differences in problem solving processes used by experts and novices. Paper delivered at the NIE-LRDC Conference on Thinking and Learning Skills. Pittsburgh, October 1980.

Lehmann, I. J., & Dressel, P. L. Changes in critical thinking ability, attitudes, and values associated with college attendance: Final report (U.S. Department of Health, Education, and Welfare Cooperative Research Project No. 1646). East Lansing, MI: Michigan State University, 1963.

Lehmann, I. J., & Dressel, P. L. Critical thinking, attitudes, and values in higher education. U.S. Department of Health, Education, and Welfare. East Lansing, MI: Michigan State University, 1962.

Lesgold, A. M., Feltovich, P. J., Glaser, R., & Wang, Y. The acquisition of perceptual diagnostic skill in radiology. Technical Report No. PDS-1, Learning and Research Development Center. Pittsburgh, PA: University of Pittsburgh, September 1981.

Luria, A. R. Cognitive development: Its cultural and social foundations. Cambridge: Harvard University Press, 1976.

Mandernach, J. B. The use of reasoning procedures in medical problem solving. Dissertation Abstracts, 1979, 761-A.

Messick, S. Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin, 1981, 89, 575-588.

Miller, R., & Wild, C. L. (Eds.). Restructuring the Graduate Record Examinations Aptitude Test (GREB Technical Report). Princeton, NJ: Educational Testing Service, 1979.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. Confirmation bias in a simulated research environment: An experimental study of scientific inference. Quarterly Journal of Experimental Psychology, 1977, 29, 85-95.

Neimark, E. D., & Santa, J. L. Thinking and concept attainment. Annual Review of Psychology, 1975, 26, 173-205.

Newell, A. Reasoning, problem-solving, and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), Attentional performance VIII. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

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

Pace, C. R. The demise of diversity. Berkeley, CA: Carnegie Commission on Higher Education, 1974.

Pachtman, A. B. The effects of a reading and language arts program on the critical thinking and critical reading of the first-year law student. Dissertation Abstracts, 1976, 2431-A.

Pellegrino, J. W., & Glaser, R. Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.), Advances in instructional psychology, Volume 2. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982, 269-345.

Powers, D. E., & Swinton, S. S. Effects of self-study for coachable item types. Journal of Educational Psychology, 1984.

Powers, D. E., & Swinton, S. S. Effects of self-study for coachable test item types. Journal of Educational Psychology, 1983.

Powers, D. E., & Swinton, S. S. Effects of self-study of test familiarization materials for the analytical section of the GRE Aptitude Test (GREB Research Report No. 79-9R; also ETS RR-82-17). Princeton, NJ: Educational Testing Service, 1982.

Powers, D. E., & Swinton, S. S. Extending the measurement of graduate admission abilities beyond the verbal and quantitative domains. Applied Psychological Measurement, 1981, 5(2), 141-158.

Reed, S. K., Ernst, G. W., & Banerji, R. The role of analogy in transfer between similar problem states. Cognitive Psychology, 1974, 6, 436-450.

Reilly, R. R. Factors in graduate student performance. American Educational Research Journal, 1976, 13(2), 125-138.

Revlin, R., & Mayer, R. E. (Eds.). Human reasoning. Washington, DC: V. H. Winston & Sons, 1978.

Schank, R., & Abelson, R. Scripts, plans, goals, and understanding. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Scribner, S. Modes of thinking and ways of speaking: Culture and logic reconsidered. In R. Freedle (Ed.), New directions in discourse processing. Norwood, NJ: Ablex Publishing Corporation, 1979, 223-243.

Simon, H. A. Information processing theory of human problem solving. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 5). Human information processing. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978.

Simon, H. A. The structure of ill-structured problems. Artificial Intelligence, 1973, 4, 181-201.

Simon, H. A., & Hayes, J. R. The understanding process: Problem isomorphs. Cognitive Psychology, 1976, 8, 165-190.

Snow, R. E. The training of intellectual aptitude. In D. K. Detterman & R. J. Sternberg (Eds.), How and how much can intelligence be increased? Norwood, NJ: Ablex Publishing Corporation, 1982, 1-37.

Sorenson, L. Watson-Glaser critical thinking appraisal: Changes in critical thinking associated with two methods of teaching high school biology. Test Data Report No. 51. New York: Harcourt, Brace and World, 1966.

Sternberg, R. J. Reasoning, problem solving, and intelligence. In R. J. Sternberg (Eds.), Handbook of human intelligence. New York: Cambridge University Press, 1982, 225-307.

Sternberg, R. J. Instrumental and componential approaches to the nature and training of intelligence. Paper delivered at the NIE-LRDC Conference on Thinking and Learning Skills, LRDC University of Pittsburgh, October 1980.

Sternberg, R. J. Component processes in analogical reasoning. Psychological Review, 1977a, 84, 353-378.

Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977b.

Steven, A. G., & Collins, A. Multiple conceptual models of a complex system. In R. Snow, P. Federico, & W. Montague (Eds.), Aptitude, learning and instruction (Vol. 2): Cognitive process analyses of learning and problem solving. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980, 177-197.

Swinton, S. S., & Powers, D. E. A study of the effects of special preparation on GRE analytical scores and item types. Journal of Educational Psychology, 1983, 75, 104-115.

Swinton, S. S., & Powers, D. E. Patterns of abilities among groups of intended graduate majors. Paper presented at the American Educational Research Association Meeting, San Francisco, April, 1979.

Swinton, S. S., Wild, C. L., & Wallmark, M. Investigation of practice effects on item types in the Graduate Record Examinations Aptitude Test (GRE Board Professional Report No. 80-1cP). Princeton, NJ: Educational Testing Service, 1983.

Thorndike, R. L. Reading as reasoning. Reading Research Quarterly, 1973, 9, 135-147.

Tversky, A., & Kahneman, D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 1983, 90(4), 293-315.

Ward, W. C., Carlson, S. B., & Woisetschlager, E. Ill-structured problems as multiple-choice items (GREB No. 81-18P). Princeton, NJ: Educational Testing Service, March 1983.

Ward, W., Frederiksen, N., & Carlson, S. Construct validity of free-response and multiple-choice versions of a test. Journal of Educational Measurement, 1980, 17, 11-29 .

Wason, P. C., & Johnson-Laird, P. C. Psychology of reasoning: Structure and content. Cambridge, MA: Harvard University Press, 1972.

Wason, P. C., & Johnson-Laird, P. C. (Eds.). Thinking and reasoning. Baltimore, MD: Penguin Books, 1970.

Watson, G., & Glaser, E. M. Manual for Watson-Glaser critical thinking appraisal. New York: Harcourt, Brace, Jovanovich, Inc., 1980.

Wilson, K. M. An exploratory study of the criterion-related validity of verbal and other item-type part scores on the GRE General Test (GRE draft report). Princeton, NJ: Educational Testing Service, 1984.

Wilson, K. M. A study of the validity of the restructured GRE Aptitude Test for predicting first-year performance in graduate study (GREB Research Report No. 78-6R; also ETS RR-82-34). Princeton, NJ: Educational Testing Service, 1982.

Wood, L. E. An "intelligent" program to teach logical thinking skills. Behavior Research Methods & Instrumentation, 1980, 12(2), 256-258.

Appendix A

Solution Protocol for an Analytical Reasoning Item

Let us consider a simulated analysis of one way in which the analytical reasoning situation in Figure 1 might be interpreted and used to answer questions. This analysis characterizes some of the cognitive operations required for the solution of questions.

I. With 7 individuals, there are 128 different possible combinations to consider, so any attempt to list or diagram all possible combinations would be impractical. An initial planning or search control decision is vital if any efficient strategy is to be adopted. Thus, ruling out impossibilities rather than systematically constructing all possibilities must be seen as a valid solution strategy. A cognitive style that requires more closure than the problem solution demands could be a great disadvantage at this initial stage.

II. Correct encoding of the information does not require knowledge of technical terms, but does require precise understanding of the logical meaning of ordinary terms of relationship. Such errors as interpreting "Fran goes to a restaurant only on Wednesdays," to mean "Fran always goes on Wednesdays," or "Helen will not go to a restaurant unless Fran also goes," to mean, "If Fran goes, then Helen must go," are clearly very serious, but research suggests that even advanced undergraduates are prone to making such errors with surprising frequency.

III. In analytical reasoning problems in which a complex structure must be modeled, finding a representation and convenient notation can be important. In this problem, only a few exclusion rules are needed, and it is not necessary to depart from verbal representation, if the rules have been correctly encoded. It may be helpful for some to abbreviate this verbal representation, as in:

1. No F if not Wed.
2. No G if K

   No K if G
3. No H if No F
4. No J and M if No G (But J <u>or</u> M ok if No G)
5. No K if Rest. O

IV. By applying transitive inference, several additional rules may be derived from these, e.g.:

No H if not Wed. (1 & 3)

No J and M if K (2 & 4)

V. Specializing to question 1: If six individuals have lunch in one of the restaurants on a Wednesday, we see that it is not necessary to work out all the exclusions to answer the question. We may merely trace through a branch until it results in fewer than six workers being available and then go to the next possibility. Here a kind of test-wiseness involving the expectation that one and only one answer choice is correct is important. If a particular chain excludes only one individual, and that individual is among the response choices, the search may be terminated without checking all other possibilities. It may be prudent to check other possibilities as a check on one's reasoning, but if another solution is found, it is evidence of a mistake in representation or inference (or conceivably, of an error in the question) rather than of two intended answer keys.

Thus we may work through the conditions until we find an individual whose presence implies the absence of at least two others, and we will have the solution. In this case, we are given that it is Wednesday, so by 1) Fran is not excluded.

By 2), either Gladys or Karl is excluded, so Fran must be included if there are to be six individuals at lunch. Assume Gladys is excluded.

By 3), Helen is not excluded, since Fran is included. Indeed, Helen must be included to make the total six.

By 4), Juan or Mark is also excluded if Gladys is excluded. Therefore, the assumption that Gladys is excluded makes it impossible to include six individuals, and Gladys must be included. Therefore, by 2), Karl must be excluded, and we have a solution. Checking this against 5) we see that excluding Karl does not lead to any inconsistency, so we select response A. This argument has employed an indirect argument, assuming one alternative and showing that it leads to an impossible outcome, so that the other alternative must be the case. This form of reasoning is extremely important in analytical reasoning items, and does not appear to be well-studied in the research literature, other than in Wason and Johnson-Laird's documentation of subjects' tendency to seek confirming, rather than disconfirming evidence. A search control strategy that recognizes that the search is over once it has been shown that one of two exhaustive alternatives cannot be the case is more efficient than one that continues to check possibilities after the problem is solved. A construct validation of this item type should explicitly examine the importance of this reasoning paradigm in the tasks confronting graduate students.

The second problem in this sample set asks for the greatest number of individuals who can go out to lunch together on a Tuesday at one of the three restaurants. Although it deals with maximizing the number of individuals present, rather than with finding conditions for exactly one individual to be absent, the solution is very strongly related to problem 1.

By 1), Fran is excluded, and so, by 3) Helen is also, so the number is less than six.

By the same reasoning used in the first problem, excluding Gladys also excludes Juan or Mark, but excluding Karl excludes no one else. To maximize the number, we again exclude Karl, so there can be no more than 4. There are no other conditions, so the greatest number is 4.

Problem 3 involves additional reasoning steps. If Juan and Mark eat lunch with a group of co-workers, we know by 4) that Gladys must be there, and again by 2) that poor Karl is shut out again. Therefore, Statement I, "The group can include no more than three men," must be true. Unless it is Wednesday, neither Fran nor Helen can be included. On the other hand, if it is Wednesday, Fran may or may not attend, and if Fran attends, Helen may or may not join her. Therefore, depending on circumstances which may vary within the situation described, there may be one, two, or three women present, and Statement II, "The group can include only two women," is false, because we have found a counterexample in a permissible case with all three able to be there. Since we already know that Karl is excluded, his adversion to restaurant O is not relevant, and Statement III, "The group cannot eat at restaurant O," is false. Note the need to deal with multiple negation in this solution. Karl will not go to restaurant O. Karl is not in the group. Therefore, it is not true that the group cannot eat at restaurant O. Dealing with negation is particularly troublesome for many students, and this question gets at this important reasoning skill without resorting to awkward wording. All three questions require the examinee to have a clear understanding of the difference among statements that <u>must</u> necessarily be true, statements that <u>may</u> be true, and statements that <u>cannot</u> be true. In questions 2 and 3 certain statements that <u>may</u> be true must be used to demonstrate the falsity of generalizations. However, in no case is it necessary to generate and search the entire list of combinations that may be true. The ability to go to just those relevant cases that could disconfirm a proposition reflects an understanding of how deductive logic proceeds. Additional confirming instances do not prove a statement. One must show that there are no disconfirming cases. Apparently analytical reasoning items tap this understanding without resort to mathematical or logical terminology or symbolism. This ability to "get to the heart of the matter" may be more related to induction or even to an intuition developed by practice in problem solving than are such more commonly emphasized components such as transitive inference or rules for negation. It can be most effectively tested with problems such as this example in which there are numerous possibilities not completely determined by the conditions. Because they are under-determined (but not ill-structured), such problems usually are not easily diagrammed. It may be that some of the disagreements which recur about the value of diagrams in solving logic problems could be resolved by clearer specification of the degree to which the problems are under-determined in this sense. Certainly any construct validity study for analytical reasoning skills should make this variable explicit.

A much more detailed account of the steps involved in solving these sample problems is of course well within the reach of, and indeed preferred by cognitive psychologists. As one focusses on more microscopic procedural levels such as encoding and representing each phrase in the stimulus, the links to research on thinking in educational contexts become longer and thinner. Although more detailed modelling might have considerable value for test development, it is not deemed appropriate for construct validation, in which links to more global variables emphasized in previous research and thinking about thinking are the major foci. On the other hand, very global accounts, based on taxonomical cognitive levels such as "comprehension, application, or evaluation," or on Baconian models of scientific method, with similar categories (e.g., observe, diagnose, plan, implement, evaluate) are so general as not even to admit of agreement among judges as to the categorization of items. Such very global specifications have failed to stand up to empirical study of the actual structure of problem-solving activity. It is hoped that the level of description attempted here is near that target.