GRE

GRADUATE RECORD EXAMINATIONS

ANALYTICAL REASONING SKILLS INVOLVED IN GRADUATE STUDY: PERCEPTIONS OF FACULTY IN SIX FIELDS

Donald E. Powers
Mary K. Enright

GRE Board Professional Report GREB No. 83-23P
ETS Research Report 86-43

December 1986

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

Analytical Reasoning Skills Involved in Graduate Study:
Perceptions of Faculty in Six Fields

Donald E. Powers
Mary K. Enright

GRE Board Professional Report GREB No. 83-23P

December 1986

Acknowledgments

The authors wish to thank the members of the GRE Research Committee for helpful suggestions; Laurie Barnett for programming the analyses; Neal Kingston, William Ward, and Cheryl Wild for helpful reviews of the report.

Abstract

This study was intended to provide information on the role of analytical abilities in graduate study. Graduate faculty in six fields of study (chemistry, computer science, education, English, engineering, and psychology) were asked to judge:

(a) the importance for academic success of a wide variety of analytical skills

(b) the seriousness of various reasoning errors

(c) the degree to which a variety of "critical incidents" had affected their estimations of students' analytical abilities.

Faculty members were generally able to discriminate among the various skills, errors, and incidents they were asked to consider, although the vast majority of skills, for example, were rated to be at least moderately important on average. Some skills were viewed as extremely important in all disciplines. More typically, however, disciplines varied quite markedly with respect to faculty perceptions of the importance, seriousness, and impact of these skills, errors, and incidents.

Several relatively interpretable dimensions were found to underlie particular clusters of reasoning skills. These dimensions or factors involved (a) the analysis and evaluation of arguments, (b) the drawing of inferences and the development of conclusions, (c) the definition and analysis of problems, (d) the ability to reason inductively, and (e) the generating of alternative explanations or hypotheses. These dimensions were judged to be differentially important for success in the six disciplines included in the study.

The study results would appear to have definite implications for developing future editions of the GRE analytical ability measure. For instance, some reasoning abilities that were perceived as very important are well represented in the current version of the analytical measure but others are not. That some abilities appear to be much more important in some fields than in others would imply the need to carefully balance the measurement of various reasoning skills to ensure fairness to test takers from different fields of study. Fairness should be aided by including for measurement those skills rated as consistently important across all disciplines. Finally, the several dimensions underlying various clusters of reasoning skills should be applicable to extending the test specifications for the current version of the analytical measure.

Analytical Reasoning Skills Involved in Graduate Study:
Perceptions of Faculty in Six Fields

Despite the complexity of human cognitive abilities, standardized admissions tests have tended to focus almost exclusively on the measurement of broadly applicable verbal and quantitative aptitudes. One criticism of such omnibus verbal and quantitative ability measures is that they provide only limited descriptions of students' academic strengths and weaknesses, and that they do not therefore adequately reflect test takers' differential development in other important cognitive areas.

In 1974 the GRE Board approved a plan to restructure the GRE Aptitude Test in order to allow examinees to demonstrate a broader range of academic skills (Altman, Carlson, & Donlon, 1975). A survey of constituents revealed that, of several possible new areas of measurement (e.g., abstract reasoning, scientific thinking, and study skills), graduate faculty, administrators, and students were most receptive to assessing analytical or abstract reasoning skills (Miller & Wild, 1979). Developmental activities then followed and, after careful psychometric study of several alternative analytical item types, four distinct kinds of items were selected for the new analytical section of the GRE Aptitude Test, which was introduced operationally in the 1977-78 testing year. Graduate institutions were cautioned against using the scores from the new analytical section until further evidence could be generated on the validity of the new measure. Subsequently, the administration of the new measure to large numbers of examinees under operational conditions enabled the further collection of information about the new measure.

Some research strongly suggested the promise of the analytical section: it appeared to measure an ability that was distinguishable from the verbal and quantitative abilities measured by the test (Powers & Swinton, 1981), and the score derived from it was related to successful performance in graduate school (Wilson, 1982). Unfortunately, however, further research suggested serious problems with the two item types (analysis of explanations and logical diagrams) that comprised the bulk of the analytical section. Performance on these item types was shown to be extremely susceptible to special test preparation (Swinton & Powers, 1983; Powers & Swinton, 1984) and to within-test practice (Swinton, Wild, & Wallmark, 1983). Consequently, in 1981 the two problematic item types were deleted from the test, and additional numbers of analytical reasoning and logical reasoning items, which constituted a very small part of the original analytical measure, were inserted.

The most recent research on the General Test (Stricker & Rock, 1985; Wilson, 1984) has given us some reason to question both the convergent and the discriminant validity of the two remaining item types. Specifically, the two currently used analytical item types correlate more highly with other verbal items or with other quantitative items than they do with each other. Moreover, after reviewing the psychological and educational research literature on reasoning, Duran, Powers, and Swinton (in press) concluded that the two currently used GRE analytical item types reflect only a limited portion of the reasoning skills that are required of graduate students. The most notable omission is the assessment of inductive reasoning skills, i.e., reasoning from incomplete knowledge, where the purpose is to learn new subject matter, to develop hypotheses, or to integrate previously learned materials into a more useful and comprehensive body of information. Thus, it seemed, the analytical ability measure of the GRE General Test might be improved through further effort.

The objective of the study reported here was to generate information that might guide the development of future versions of the GRE analytical measure. More specifically, the intention was to gain a better understanding of what reasoning (or analytical) skills are involved in successful academic performance at the graduate level, and to determine the relative importance of these skills or abilities both within and across academic disciplines. It was thought that this information might be especially useful for developing additional analytical item types.

As mentioned earlier, the initial version of the GRE analytical ability measure was developed after a survey had suggested the importance of abstract reasoning to success in graduate education. This survey, however, was not designed to provide any detailed information on the importance of specific analytical skills, as was the intention here.

Method

Questionnaire Development

Initially, 30 department chairs (in English, education, engineering, chemistry, computer science, or psychology) were contacted in 30 graduate institutions, and asked to identify three faculty members in their departments who would be willing to provide their insights into the analytical or reasoning skills that are most critical for successful performance in graduate school. These 30 institutions were chosen from the GRE Directory of Graduate Programs in such a way as to ensure some degree of geographical representation. All of these departments require or recommend that applicants submit GRE General Test scores; it was felt that these departments might be more interested than nonrequiring departments in efforts to improve the GRE General Test.

At this preliminary stage, faculty members were informed of the purpose of the project and asked to give, in an open-ended fashion, examples of:

(a) the analytical, reasoning, or thinking skills they perceived as most important for successful graduate study in their fields (e.g., identifying assumptions on which an argument is based), particularly as these skills differentiate successful from marginal students

(b) specific critical incidents related to thinking or reasoning that caused them to either raise or lower their estimation of a student's analytical ability (e.g., failing to qualify a conclusion as appropriate)

(c) particular reasoning or thinking "flaws" they have observed in their students (e.g., using the conclusion as the premise of an argument).

Useable responses were obtained from 33 faculty members, who suggested a total of 138 important reasoning or thinking skills, 86 critical incidents, and 75 reasoning "flaws." Some of these responses were duplicates. Several other respondents did not specify discrete skills or errors but chose rather to send helpful discursive replies to our inquiry. All responses were condensed and edited, and generally evaluated with respect to whether they should be included in the larger, more structured questionnaire that was planned. Some responses constituted usable questionnaire items essentially as stated by respondents (e.g., "the ability to break complex problems into simpler components"). Other responses were revised or eliminated because they were too general (e.g., "the ability to think independently"), and others because they were too specific or applied only to a particular field (e.g., "the ability to resolve into enthymemic form any argumentative work" or "the ability to take ecological validity into account").

The structured questionnaire was constructed on the basis of this preliminary survey, on a review of relevant literature (Duran, Powers, & Swinton, in press) and on a number of additional books or texts on reasoning (e.g., Campbell, 1974; Fischer, 1970; Johnson & Blair, 1983; Kahane, 1984, Nosich, 1982; Salmon, 1984; Scriven, 1976; Toulmin, Rieke, & Janik, 1984; Wason & Johnson-Laird (1972); and Weddle, 1978). Several other articles, e.g., a seminal work by Ennis (1962) and a list of skills by Arons (1979), proved especially useful. Various issues of <u>CT News</u>, published by the Critical Thinking Project at California State University at Sacramento, were also perused. Previous work on critical incidents in graduate student performance (Reilly, 1974a, 1974b) was also consulted, and several of the incidents related to critical facility were included in the present study. Finally, the list generated by Tucker (1985), who gathered the impressions of ETS test development staff, philosophers, and cognitive psychologists, also proved to be a valuable resource.

The final questionnaire (see Appendix A) was structured to include questions about the importance and frequency of various reasoning skills, of commonly observed errors in reasoning, and of specific incidents that may have led faculty to adjust their estimates of students' analytical abilities. Questions were grouped under several headings, mainly to give respondents some sense of their progress in responding to the rather lengthy questionnaire.

The Sample

Six academic fields (English, education, psychology, chemistry, computer science, and engineering) were included in the final survey. These fields were thought to represent the variety of fields of graduate study and the variation in the kinds of reasoning abilities involved in graduate education. Using the data tapes of the Higher Education General Information Survey (HEGIS), nonoverlapping samples of 64 graduate institutions with doctoral programs were drawn for each of the six graduate fields. A random sampling procedure was used such that eight institutions from each of the eight HEGIS geographic regions were selected for each field. This sampling was greatly facilitated by the work of Oltman (1982). The admission requirements of these institutions were determined from the Directory of Graduate Programs (GRE/CGS, 1983), and only those that either required or recommended GRE General Test scores were included in the sample. In this manner, 40 institutions were selected for the final sample for each field. In addition, one institution with a relatively large proportion of Black students and one with a relatively large percentage of Hispanic students were included in the samples for each field, thus raising the total number of institutions to 42 per field. Letters were then sent to departmental chairpersons, who were asked to nominate two faculty members who would be willing to complete the questionnaire. Respondents were paid $25 for their participation.

Data Analysis

Means and standard deviations were calculated for each question by academic field of study, and analyses of variance were run for each question to assess differences among the six fields. The various ratings were correlated within questionnaire sections. For example, within the section on reasoning skills, the ratings of frequency and importance were correlated; within the section on reasoning errors, the ratings of frequency and seriousness were correlated.

Finally, within each section (and for each kind of rating), the data were factor analyzed to effect some reduction in the large number of questions. A principal axis factoring, with squared multiple correlations as the initial estimates of communalities, was used to determine the number of factors to be retained for each section, according to both the magnitude of the eigenvalues and the breaks in their size. (Our inclination was to err on the side of retaining too many factors at this exploratory stage.) Various numbers of factors were then rotated according to the varimax criterion. Although other oblique rotations could have been used also, it was felt that the detection of uncorrelated factors would best serve the objectives of further test development.

Results

The Sample

A total of 165 chairpersons (65% of those contacted) nominated a total of 297 faculty members, of whom 255 (86%) returned usable questionnaires. The response rates across fields were generally comparable.

Full professors constituted a slight majority of the responding sample (51%); associate professors made up the next largest proportion (34%). About 13% were assistant professors, and the remaining small proportion were deans, associate deans, or lecturers.

Item-level Results

Tables 1-3 show the mean ratings by discipline for each question included in the survey instrument. The numbers in the total column are the grand means for all disciplines. Numbers under each discipline represent for each item the deviations from these means. The F tests in the right-most column indicate whether the means are significantly different among the six disciplines. Because the average ratings, over all respondents, for "frequency of use" and "importance for success" correlated .99, only the importance ratings are presented for reasoning skills. Likewise, only the "seriousness" ratings are presented for reasoning errors, since their correlation with frequency ratings was .98, and, for critical incidents, only the average "effect" ratings are presented, since their correlation with frequency ratings was .94.

Tables 1-3 show a substantial number of significant differences among disciplines with respect to the importance placed on various reasoning skills (Table 1), the seriousness with which they regard particular kinds of reasoning errors (Table 2), and the impact that various critical incidents have on the estimation of students' analytical abilities (Table 3). Table 4, showing only the very highest rated skills and most critical errors and incidents, gives a flavor of the differences among these six disciplines. For example, chemistry faculty placed a high premium on being able to generate hypotheses, questions, or experiments, to draw sound inferences from observations, and to analyze and evaluate previous research. English faculty, on the other hand, saw greater importance in skills involving argumentation--being able to understand, evaluate, analyze, elaborate, recognize, and support aspects of an argument.

Faculty in the six disciplines also appeared to have quite different views as to the numbers of skills that were important in their respective disciplines. The numbers of reasoning skills that received average ratings of 4.0 or higher varied markedly by discipline as follows: 23 for chemistry, 5 for computer science, 27 for education, 22 for engineering, 29 for English, and 26 for psychology. These differences may have arisen, for example, from our particular choice of questions, from differences in standards among disciplines, or from some other factor(s).

It can be seen, even from Table 4, however, that some skills were viewed as very important by several disciplines. For example, "breaking down complex problems into simpler ones" was rated as the single most important skill (of the 56 skills listed) in both computer science and engineering. "Determining whether conclusions are logically consistent with, and adequately supported by, the data" was rated as one of the three most important skills by both education and psychology faculty; "drawing sound inferences from observations" was the highest rated skill in chemistry and nearly the highest in education.

The extent to which faculty in different disciplines agreed on the importance of various skills, errors, or incidents can be examined in a slightly different manner. To get some idea of the skills, errors, and incidents that were viewed as relatively important, and for which average ratings did not differ significantly across disciplines, Table 5 was prepared. This table shows only those skills that received average ratings of importance of more than 3.5 over all six disciplines combined, and for which analyses of variance did not detect any significant differences among disciplines.

"Reasoning or problem solving in situations in which all the needed information is <u>not</u> known" was the skill rated as most important overall. Such skills as "detecting fallacies and logical contradictions in arguments," "deducing new information from a set of relationships," and "recognizing structural similarities between one type of problem or theory and another" were the next most highly rated skills. These were followed closely by "taking well-known principles and ideas from one area and applying them to a different specialty," "monitoring one's own progress in solving problems," and "deriving from the study of single cases structural features or functional principles that can be applied to other cases."

Table 6 lists the reasoning errors and critical incidents that were judged overall to be the most serious or to have the most effect on the estimation of students' abilities. Three errors/incidents were judged to be most serious or critical: "accepting the central assumptions in an argument without questioning them," "being unable to integrate and synthesize ideas from various sources," and "being unable to generate hypotheses independently."

It should be noted that there are many other decision rules, based on average ratings and differences among disciplines, that could have been used here to form a "common core" of skills or errors/incidents. Tables 1-3 could be consulted to apply alternative rules.

Factor Analytic Results

To condense the many questions into a more manageable form, factor analyses were computed for each section of the questionnaire.

For the section on reasoning skills, only the importance ratings were analyzed because they were so highly correlated with frequency ratings. Because frequency ratings were slightly less correlated with ratings of seriousness and criticality in the other two sections, they too were analyzed for the questionnaire sections on reasoning errors and critical incidents.

The reader should bear in mind that the factors resulting from this analysis should not be construed as representing dimensions of analytical ability, but rather only as reflecting the dimensions that underlie faculty perceptions of analytical abilities. These dimensions merely reflect the extent to which graduate faculty tended to rate certain skills as about equally important (or equally unimportant), not the degree to which these dimensions represent "factors of the mind." Thus, the results presented below are intended to provide a parsimonious representation of faculty perceptions rather than a basis for postulating distinct analytical abilities.

Reasoning skills. For the ratings of importance of reasoning skills, the largest eigenvalues were 16.4, 3.9, 2.4, 1.6, and 1.1, and the application of a scree test (Cattell, 1966) suggested the appropriateness of a five-factor solution, which was then rotated according to the varimax criterion (Kaiser, 1958). The five-factor varimax rotation accounted for 80% of the common variance. The factor loadings and communalities are given in Appendix B. Table 7 summarizes the variables that were most instrumental in defining each factor.

Factor I, which accounted for about a third of the common variance, was characterized by highest loadings, generally, from skills involving arguments. Thus, Factor I seems to involve a kind of critical thinking related to argumentation.

Factor II accounted for about 29% of the common variance, and was defined primarily by variables related to the drawing of conclusions, e.g., generating valid explanations, supporting conclusions with sufficient data, and drawing sound inferences from observations. The conclusion-oriented skills that define this second critical thinking factor would seem to be of a more active or productive nature, involving the construction of inferences or conclusions, rather than evaluating the soundness of arguments or inferences, as is the case for Factor I.

Factors III-V each accounted for a somewhat smaller proportion of common variance (10% - 15%) than did Factors I and II. Factor III is best defined by skills related to defining and setting up problems or analyzing their components as a prelude to solving them. Factor IV is best characterized by inductive reasoning skills, i.e., the drawing of conclusions that have some evidential support, but not enough to indicate logical necessity. Factor V is somewhat difficult to define, but, by virtue of its two highest loadings, it seems to reflect an ability to generate alternatives.

Reasoning errors. For the ratings of seriousness of reasoning errors, the largest eigenvalues were 6.5 and 1.1, and the two factors accounted for 96% of the common variance. (Frequency ratings were also factor analyzed and are presented in Appendix B. Because the results were so similar to the analysis of seriousness ratings, they are not discussed here.) As shown in Table 8, Factor I, which explained about 52% of the common variance, was characterized by loadings from errors involved in the evaluation of evidence, e.g., offering irrelevant evidence to support a point. Factor II, on the other hand, seemed to involve more formal logical errors, particularly as related to reasoning with more statistically oriented material--for example, failing to take account of a base rate, failing to recognize differences between populations and samples, and confusing correlation with causation.

Critical incidents. As for reasoning errors, only ratings of the effects of critical incidents, not their frequencies, were factor analyzed. This analysis yielded eigenvalues of 8.2, 1.6, and 0.9 for the largest factors, and this three-factor solution accounted for 89% of the common variance. Table 9 summarizes the results, and the complete set of loadings is given in Appendix B.

Factor I, explaining about 38% of the common variance, was best defined by highest loadings from such incidents as accepting/supporting arguments based more on emotional appeal than evidence, offering nonconstructive or unsound criticism of other students' presentations, and confusing anecdote and/or opinion with "hard data." This factor appears to involve critical facility.

Factor II, accounting for about 34% of the common variance, appears to involve the ability to consider or to generate alternatives, being defined primarily by high loadings from such incidents as accepting conclusions without critically evaluating them, being able to criticize but unable to suggest better alternatives, being unable to integrate ideas from various sources, and being unable to generate hypotheses.

Factor III, defined by such incidents as applying a formula or rule without sufficient justification, and searching for a complicated solution when a simpler one is obvious, is difficult to interpret. One possible characterization might be a kind of rationality or critical facility that is sometimes referred to as practical judgment or perhaps "common sense."

Scales based on reasoning skill factors. Table 10 gives the average scores for each discipline on scales composed of the questionnaire items that best defined the reasoning skill factors discussed earlier. As is clear, there are substantial differences among disciplines on these scales. Skills involved in analyzing/evaluating arguments (Scale 1) were rated as extremely important in English ($m = 4.53$), quite important in education ($m = 3.83$) and psychology ($m = 3.73$), and somewhat less important in the other three disciplines, particularly computer science ($m = 2.97$).

Critical thinking skills involved in developing or otherwise dealing with conclusions (Scale 2) were viewed as very important (means greater than 4.0) in all disciplines except computer science.

Abilities involved in analyzing and defining problems (Scale 3) were rated as extremely important in computer science (m = 4.05) and engineering (m = 4.00), but less important in other disciplines, especially English (m = 2.70).

The inductive reasoning skills reflected on Scale 4 were rated as moderately important on each of the six disciplines. The skills composing Scale 5, generating alternatives/hypotheses, were rated very high in psychology (m = 4.21) and in education (m = 3.93), and as somewhat less important in other disciplines, particularly computer science (m = 2.90).

Other Comments from Respondents

A number of general comments were made about the study--some positive and some negative. The study was described alternately as "very well done" and "interesting," but also, by one respondent, as a "complete waste of time." Most of the comments were positive, however, and many pertained more specifically to the kinds of questions that were asked. The consensus seemed to be that the questionnaire was not easy to complete. Moreover, faculty in the several disciplines sometimes had different ideas as to what kinds of questions would have been appropriate. For example, one English faculty member noted the lack of questions on the use of language in critical writing, and a computer science faculty member observed that questions on abilities involved in formulating proofs, which are vital to success in computer science, were only partially covered in the questionnaire. An education faculty member noted that the survey did a better job of assessing skills associated with hypothesis-testing than with other research skills.

Along these same lines, a number of other respondents also believed that the questions were more relevant to other disciplines than to theirs. Several computer science professors, for example, characterized the questions as oriented more toward argument than problem solving, in which they had greater interest. An engineering professor said that some of the questions were more pertinent to educational research than to scientific or technical research, and one English faculty found that questions seemed "geared to the hard sciences." Finally, some noted ambiguities or redundancies, or lamented that the questions were "too fine." Even with these difficulties, however, most of the comments about questions were positive: "Items seem especially well chosen," "questions are appropriate," "questions were quite thorough," "a good set of questions," "topics covered are critical," and "your lists are right on target." The majority of comments, therefore, suggested that the questionnaire was pitched at about the right level and included appropriate kinds of reasoning skills.

A number of comments were made about the relationship between subject matter and analytical skills, e.g., that successful problem solving is predicated on having specific knowledge in a field. One respondent believed that the questionnaire downplayed the importance of "context effects" in favor of "strict reasoning ability," and another noted that the measurement of analytical abilities is quite discipline specific. Another commented on the difficulty of measuring analytical ability without regard to the amount of knowledge available.

Several faculty commented on the development of analytical skills in graduate school and on the differential importance of these skills at various stages of graduate education. One respondent said, "I rated entering behavior or behavior across the entire program (courses, internships, dissertation). If I were to rate the dissertation experience alone, the ratings would have been much higher." Many noted that by the end of their programs, skills would be expected to increase and various reasoning errors could be expected to occur less frequently: "Entering students are more likely to make these errors and graduates to make far fewer." Another said, "In some sense, the essence of graduate training is analytical skills." "These are skills which students acquire. When they enter they make most of the mistakes you mentioned. If they can't learn, they leave the program." Another said, "I'm more concerned about the presence of these behaviors after my course than before it. One simply does not harshly judge a beginning student who makes an error, but one could be very critical of a student about to finish a Ph.D. thesis...."

Discussion

Summary

Some 255 graduate faculty in six fields of study (chemistry, computer science, education, engineering, English, and psychology) (a) rated the importance for success in their programs of a wide variety of reasoning skills, (b) judged the seriousness of various reasoning errors, and (c) indicated the degree to which a variety of "critical incidents" affected their estimations of students' analytical abilities.

Faculty members were able to discriminate among the various skills they were asked to consider, although the vast majority of skills were seen, on average, as at least moderately important. Faculty also made distinctions with respect to both the seriousness of different kinds of reasoning errors and the effects of various critical incidents. Individually, such general skills as "reasoning or problem solving in situations in which all the needed information is not known" were viewed as extremely important in all disciplines. Such reasoning errors and critical incidents as "accepting the central assumptions in an argument without questioning them," "being unable to integrate and synthesize ideas from various sources," and "being unable to generate hypotheses independently" were judged to be very serious in all disciplines.

More typically, however, disciplines varied quite markedly with respect to faculty perceptions of the importance, seriousness, or impact of various reasoning skills, errors, or incidents. As might be expected, for example, whereas "knowing the rules of formal logic" was rated as one of the most important skills in computer science, this skill was rated as quite unimportant in all other disciplines.

Several relatively interpretable dimensions were found to underlie faculty perceptions of the substantial number of reasoning skills, errors, and incidents that were rated in the study. Skills involved in (a) the analysis and evaluation of arguments, (b) the drawing of inferences and the development of conclusions,, (c) the definition and analysis of problems, (d) the ability to reason inductively, and (e) the generating of alternative explanations or hypotheses formed five distinct dimensions, which were perceived as differentially important for success in each of the six disciplines included in the study. Analysis and evaluation of arguments was judged to be most important in English, defining and analyzing problems most important in computer science and engineering, and generating alternatives most important in psychology and education. Inductive reasoning skills were judged to be about equally important in all disciplines, and drawing inferences/developing conclusions was rated as relatively important in all disciplines except computer science.

Implications

In providing some information on faculty perceptions of the involvement of various reasoning skills in their disciplines, the study has, we hope, implications for developing future versions of the GRE analytical ability measure. Converting this information to operational test items will represent a significant step, however, and it is not crystal clear at this stage exactly how helpful these results may be eventually. Nonetheless, the findings do seem to contain several useful bits of information:

1. Among the specific reasoning skills perceived as the most important were several, e.g., "deducing new information from a set of relationships" and "understanding, evaluating, and analyzing arguments," that seem well represented in the two item types (analytical reasoning and logical reasoning) currently included in the analytical section of the General Test. This suggests that these item types should continue to play a role in future editions of the GRE General Test.

2. Some skills that are not measured by the current version of the analytical measure were rated as very important. "Reasoning or problem solving in situations in which all the needed information is not known" was among the skills rated as most important in each discipline, but currently unmeasured, at least in any explicit manner, by the analytical measure. In this regard, however, the previous

GRE-sponsored work of Ward, Carlson, and Woisetschlager (1983) is noteworthy. These investigators studied "ill-structured" problems, i.e., problems that do not provide all the information necessary to solve the problem, and noted the resemblance of these problems to one variant of the logical reasoning item type used in the analytical measure. They concluded that there was no indication that "ill-structured" problems measure different aspects of analytical ability than do "well-structured" problems, and therefore that "ill-structured" problems could not be expected to extend the range of cognitive skills already measured by the GRE General Test. They did note, however, that the "ill-structured" item type could be used to increase the variety of items types in the test. The findings of the current study suggest that the inclusion of this item type would probably meet with faculty approval in most fields of study.

3. With respect to their perceived importance, skills involving the generation of hypotheses/alternatives/explanations tended to cluster together, and the inability to generate hypotheses independently was one of the incidents rated consistently as having a substantial effect on faculty perceptions of students' analytical abilities.

   A number of years ago the GRE Board sponsored a series of studies (Frederiksen & Ward, 1978; Ward, Frederiksen, & Carlson, 1978; Ward & Frederiksen, 1977; Frederiksen & Ward, 1975) that explored the development and validation of tests of scientific thinking, including one especially promising item type called "Formulating Hypotheses," which required examinees to generate hypotheses. Although the research suggested that this item type complemented the GRE verbal and quantitative measures in predicting success in graduate school, the work was discontinued, largely because of problems in scoring items that require examinees to construct, not merely choose, a correct response. Carlson and Ward (1986) have proposed to renew work on the "Formulating Hypotheses" item type in light of recent advances in evaluating questions that involve constructed responses. The results of the faculty survey reported here would appear to support this renewal.

4. Some of the highly important skills that are currently well represented in the analytical measure are viewed as more important for success in some disciplines than in others. For example, "understanding, analyzing, and evaluating arguments" was seen as more important in English than in computer science. However, some skills seen as highly important in some disciplines but not in others may not be as well represented currently. For example, "breaking down complex problems into simpler ones" was perceived as

extremely important in computer science and engineering but not at all important in English. This would suggest, perhaps, the need to balance the inclusion of items reflecting particular skills, so that skills thought to be important (or unimportant) in particular disciplines are neither over- nor underrepresented.

5. The several dimensions that appear to underlie clusters of reasoning skills may provide an appropriate way to extend the current test specifications for the analytical measure, especially if new item types are developed to represent some of these dimensions.

6. The reasoning skills that were rated as very important, and consistently so, across disciplines point to a potential common core of skills that could be appropriately included in an "all-purpose" measure like the GRE General Test. Other skills judged to be very important in only a few disciplines might best be considered for extending the measurement of reasoning skills in the GRE Subject Tests. Faculty comments about the difficulty in separating reasoning from subject matter knowledge would seem to support this strategy.

Limitations

Any study of this nature is necessarily limited in several respects. First of all, the survey approach used here is but one of several that can be used to inform decisions about extending the measurement of analytical abilities. Tucker's (1985) results provide useful information from different perspectives--those of cognitive psychologists and philosophers. Other approaches that might also be informative include the methods of cognitive psychology, which could be used not only to supplement but also to extend the survey results reported here. These methods would seem especially appropriate because they relate more directly to actual skills and abilities than to perceptions.

Second, the diversity that characterizes graduate education renders the results of this study incomplete. Some clues have been gained as to similarities and differences among a limited sample of graduate fields. However, the substantial differences found among fields are a source of concern, since we cannot be certain whether or not some other sample of fields might exhibit even greater variation.

Finally, as several survey respondents pointed out, many of the reasoning skills about which we asked are expected to, and do, improve as the result of graduate study. In some sense these skills may represent competencies that differ from, say, the verbal skills measured by the GRE General Test in the respect that these analytical skills may develop much more rapidly. A question of interest, then, is how to accommodate the measurement of these skills in the context of graduate admissions testing, which currently focuses on the predictive effectiveness of abilities that are presumed to develop slowly over a significant period of time.

Future Directions

The study suggested several possible future directions. Because of the substantial variation among fields, one possibility would involve extending the survey to include additional fields of graduate study. Some refinements could now be made on the basis of past experience. For example, ratings of the frequency with which skills are used, as well as the frequencies of errors and critical incidents, could probably be omitted without much loss of information. On the other hand, it would seem desirable to add categories allowing ratings of the differential importance of various reasoning skills at different stages of graduate education, ranging from entry level to dissertation writing.

Finally, based on the reasoning skills identified as most important, criterion tasks might be developed against which the validity of the current GRE analytical measure could be gauged. This strategy would make especially good sense for those important skills that may not be measurable in an operational test like the GRE General Test, but which might correlate highly with the abilities now measured by the test. One specific possibility would be the development of rating forms, which could be used by faculty to rate the analytical abilities of their students. These ratings could then be used as a criterion against which GRE analytical scores could be judged.

References

Altman, R. A., Carlson, A. B., & Donlon, T. F. (1975). Developing a plan for the renewal of the GRE norming test: A report on investigations into the possible shortening of the GRE-V and GRE-Q and the creation of a modular aptitude test (GRE Report GREB No. 74-3). Princeton, NJ: Educational Testing Service.

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

Campbell, S. K. (1974). Flaws and fallacies in statistical thinking. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Carlson, S. B., & Ward, W. C. (1986). A new look at formulating hypotheses items (proposal to the GRE Research Committee). Princeton, NJ: Educational Testing Service.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.

Duran, R. P., Powers, D. E., & Swinton, S. S. (in press). Construct validity of the GRE analytical test (Draft final report to the GRE Research Committee, January 1984). Princeton, NJ: Educational Testing Service.

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

Fischer, D. H. (1970). Historians' fallacies: Toward a logic of historical thought. New York: Harper & Row, Publishers.

Frederiksen, N., & Ward, W. C. (1975). Development of measures for the study of creativity (GRE Board Professional Report GREB No. 72-2P; also ETS RB No. 75-18). Princeton, NJ: Educational Testing Service.

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

Graduate Record Examinations Board and Council of Graduate Schools in the United States. (1983). Directory of graduate programs: 1984 and 1985 (4 volumes). Princeton, NJ: Educational Testing Service.

Johnson, R. H., & Blair, J. A. (1983). Logical self defense (2nd ed.). Canada: McGraw-Hill Ryerson Limited.

Kahane, H. (1984). Logic and contemporary rhetoric: The use of reason in everyday life (4th ed.). Belmont, CA: Wadsworth Publishing Company, Inc.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187-200.

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

Nosich, G. M. (1982). Reasons and arguments. Belmont, CA: Wadsworth Publishing Company, Inc.

Oltman, P. K. (1982). Content representativeness of the Graduate Record Examinations advanced tests in chemistry, computer science, and education (GRE Board Professional Report GREB No. 81-12P). Princeton, NJ: Educational Testing Service.

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

Powers, D. E., & Swinton, S. S. (1984). Effects of self-study for coachable item types. Journal of Educational Psychology, 76, 266-278.

Reilly, R. R. (1974a). Factors in graduate student performance (GRE Board Professional Report GREB No. 71-2P; also RB 74-2). Princeton, NJ: Educational Testing Service.

Reilly, R. R. (1974b). Critical incidents of graduate student performance (GREB Board Research Report GREB No. 70-5). Princeton, NJ: Educational Testing Service.

Salmon, M. H. (1984). Introduction to logic and critical thinking. New York: Harcourt Brace Jovanovich, Publishers.

Scriven, M. (1976). Reasoning. New York: McGraw-Hill Book Company.

Stricker, L. J., & Rock, D. A. (1985). The factor structure of the GRE General Test for older examinees: Implications for construct validity (GRE Board Research Report GREB No. 83-10R). Princeton, NJ: Educational Testing Service.

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

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

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

Tucker, C. (1985). Delineation of reasoning processes important to the construct validity of the analytical test (GRE Report submitted to the GRE Research Committee). Princeton, NJ: Educational Testing Service.

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

Ward. W. C., & Frederiksen, N. (1977). A study of the predictive validity of the tests of scientific thinking (GRE Board Professional Report GREB No. 74-6P; also ETS RB No. 77-6). Princeton, NJ: Educational Testing Service.

Ward, W. C., Frederiksen, N., & Carlson, S. (1978). Construct validity of free-response and machine-scorable tests of scientific thinking (GRE Board Professional Report GREB No. 74-8P; also ETS RB No. 78-15). Princeton, NJ: Educational Testing Service.

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

Weddle, P. (1978). Argument: A guide to critical thinking. New York: McGraw-Hill Book Company.

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

Wilson, K. M. (1984). The relationship of GRE General Test item-type part scores to undergraduate grades (GRE Board Professional Report GREB No. 81-22P). Princeton, NJ: Educational Testing Service.

Table 1

Mean Ratings of Importance of Reasoning Skills by Disciplines

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	F
		Discipline						
Reasoning Skills--General								
Deriving principles from facts or cases	4.02	-.21	-.04	-.09	-.23	.23	.35	3.80**
Reasoning when facts are known explicitly	3.63	.31	.55	-.06	.44	-.59	-.65	10.81***
Reasoning when all information is not known	4.24	-.11	-.15	-.10	.29	-.06	.13	1.60
Reasoning when inconsistencies are present	4.03	.07	-.59	-.03	-.01	.22	.34	5.92***
Knowing rules of formal logic	2.76	-.46	1.03	-.21	.10	-.31	-.15	9.62***
Applying principles to a different specialty	3.75	-.05	-.08	-.13	.23	-.11	.14	1.32
Supporting or refuting a given position	3.52	-.23	-.62	.47	-.34	.47	.23	7.74***
Analyzing and evaluating previous research	4.33	.13	-.66	.05	.02	.21	.25	7.63***
Incorporating isolated data into a framework	3.37	.38	-.40	-.16	-.09	.04	.23	3.19**
Monitoring progress in solving problems	3.71	-.20	.15	.29	.12	-.24	-.12	1.71
Deriving principles that can be applied to other cases	3.68	-.11	.16	-.06	.09	.11	-.18	0.79
Recognizing the probabilistic nature of events	3.24	-.40	-.63	.31	.44	-.40	.70	9.49***
Understanding and evaluating arguments	4.05	.00	-.56	.09	-.38	.70	.14	10.51***
Deducing information from relationships	3.86	.03	-.16	.02	.16	-.07	.01	0.60
Knowing what evidence will support a hypotheses or thesis	4.19	.32	-.75	.14	-.26	.31	.22	10.35***

Table 1 (Continued)

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	F
		Discipline						
Specific Skills--Problem Definition								
Identifying central issues or hypotheses	4.34	-.13	-.34	.16	.01	.09	.22	3.20**
Recognizing sides of an issue	3.58	-.15	-.75	.39	-.38	.62	.26	10.59***
Generating alternative hypotheses	3.87	.05	-.50	.16	-.24	.13	.39	4.80***
Breaking down complex problems	4.06	.05	.57	-.13	.54	-.85	-.17	14.34***
Identifying variables in a problem	3.88	.36	.02	.16	.51	-.59	-.47	7.65***
Identifying approaches to solving problems	3.98	-.25	-.05	.43	.07	-.20	.00	3.03*
Recognizing similarities between problems	3.83	-.25	.03	-.02	.17	-.06	.13	1.16
Developing operational definitions	3.66	-.66	.36	-.07	-.11	.02	.46	5.49***
Setting up formal models	3.36	-.12	.59	-.22	.73	-.93	-.04	13.49***
Recognizing the historical context	2.97	-.70	-.74	.41	-.55	1.36	.22	26.65***
Specific Skills--Constructing Theses or Arguments								
Supporting assertions with details	3.92	-.08	-.60	.13	-.22	.74	.04	11.22***
Making explicit components in a chain of reasoning	3.57	-.17	.06	-.10	.01	.29	-.09	1.10
Distinguishing relevant and irrelevant information	4.20	.05	-.45	.09	-.10	.30	.11	3.78**

Table 1 (Continued)

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	F
		Discipline						
Perceiving relationships among observations	4.10	.15	-.70	.03	.04	.15	.33	7.74***
Drawing distinctions between similar ideas	3.67	-.48	-.30	.19	-.13	.56	.16	6.67***
Using analogies appropriately	3.36	-.38	-.26	.12	.18	.53	-.18	3.93**
Elaborating an argument	3.80	-.37	-.57	.30	-.43	1.02	.04	18.59***
Drawing sound inferences	4.25	.32	-.74	.25	.03	-.04	.18	8.84***
Producing a consistent argument	4.12	.15	-.54	.10	-.05	.35	-.01	4.67***
Synthesizing different positions	3.10	-.34	-.26	.33	-.10	.22	.16	2.90*
Translating graphs or symbolic statements	3.55	.40	.45	.02	.50	-1.44	.07	18.17***
Specific Skills--Analyzing Arguments								
Recognizing the central argument	4.04	.12	-.57	-.02	-.11	.57	.00	7.13***
Identifying assumptions	3.98	-.17	-.12	-.03	.06	.47	-.20	3.14***
Testing the validity of an argument	3.57	-.09	.27	-.23	.02	-.09	.13	1.19
Finding alternative explanations	3.82	.16	-.56	.28	-.16	-.27	.55	7.89***
Recognizing supporting points	3.64	.04	-.52	.01	-.20	.61	.06	6.84***
Detecting fallacies and contradictions	3.92	.08	-.36	.06	-.22	.22	.21	2.18
Distinguishing major points in an argument	3.28	-.12	-.52	-.14	.02	.63	.11	5.84***

Table 1 (Continued)

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	F
		Discipline						
Being sensitive to different evidence	3.31	-.34	-.96	.71	-.42	.05	.95	18.86***
Recognizing shifts in meaning of an argument	2.90	-.90	-.69	.52	-.55	1.48	.14	33.50***
Judging whether a thesis has been supported	3.86	-.08	-.81	.24	-.32	.73	.25	12.76***
Determining whether conclusions are supported	4.25	.10	-.79	.27	.05	.11	.27	8.65***
Distinguishing among factors	3.45	-.20	-.35	.10	.31	-.24	.38	3.31**
Specific Skills--Drawing Conclusions								
Generating valid explanations	4.16	.18	-.76	.27	.07	.10	.14	8.26***
Supporting conclusions	4.24	.15	-.77	.31	.02	.22	.07	8.92***
Recognizing implications of an interpretation	4.07	-.07	-.72	.26	-.28	.59	.23	14.82***
Qualifying conclusions	3.99	.06	-.73	.13	-.15	.40	.29	8.90***
Generating new questions or experiments	3.88	.53	-.58	.07	.21	-.61	.38	10.08***
Considering alternative conclusions	3.85	.24	-.59	.25	-.16	-.14	.41	6.33***
Comparing conclusions	3.91	.28	-.47	.04	.18	-.09	.05	3.25**
Revising views for new observations	3.84	.30	-.77	.14	.07	.07	.18	6.38***

Note. Entries under individual disciplines are deviations from the grand means given under "Total." Standard deviations for totals range from .78 to 1.31, and 33 of 56 are between .9 and 1.1.

*$p < .01$

**$p < .05$

***$p < .001$

Table 2

Mean Ratings of Seriousness of Reasoning Errors by Disciplines

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	F
		Discipline						
Error								
Failing to recognize differences between sample and population	2.92	-.59	-.55	.64	.41	-.82	.91	16.90***
Failing to account for base rate	2.64	-.18	-.62	.36	.14	-.03	1.34	24.59***
Offering irrelevant statements	3.05	-.46	-.66	.58	-.53	.41	.66	8.27***
Failing to recognize a concealed conclusion	3.34	-.42	-.45	.34	-.24	.45	.32	6.31***
Confusing coincidence with causation	3.60	-.47	-.44	.49	-.22	-.20	.83	9.51***
Using an ambiguous term	3.02	-.66	-.62	.18	-.38	.93	.54	15.13***
Arguing to accept an unproven conclusion	3.28	-.20	-.07	.10	-.21	-.26	.65	3.96**
Accepting assumptions with questioning	3.96	-.23	-.22	.09	-.12	.24	.24	2.14
Failing to evaluate the credibility of a source	3.74	-.23	-.63	.35	-.09	.37	.23	5.55***
Offering irrelevant evidence	3.48	-.40	-.53	.27	-.18	.61	.23	7.49***
Making generalizations from insufficient evidence	3.86	-.24	-.42	.31	-.14	.30	.20	3.97**
Inappropriately interpreting text	3.49	-.52	-.61	.39	-.28	.62	.40	11.22***

Table 2 (Continued)

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	_F_
		Discipline						
Allowing anecdotal information	3.23	-.64	-.33	.72	.01	-.51	.74	10.32***
Basing conclusions on partial data	3.57	-.08	-.36	.09	.00	.16	.19	1.50
Failing to recognize similarities in analogies	3.06	-.33	-.42	.28	.03	.17	.28	3.46**

Note. Entries under individual disciplines are deviations from the grand means given under "Total." Standard deviations for totals range from .98 to 1.48, and 8 of 15 are between 1.0 and 1.2.

*p < .01
**p < .05
***p < .001

Table 3

Mean Ratings of Effects of Critical Incidents by Disciplines

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	F
		Discipline						
Incidents								
Making irrelevant remarks	3.23	-.23	-.46	.26	-.33	.38	.39	4.18**
Offering nonconstructive criticism	2.79	-.22	-.40	.35	-.23	.41	.10	3.20**
Submitting a nonresponsive paper	3.67	.17	-.32	-.05	-.23	.38	.05	2.25
Accepting an argument on emotional appeal	3.34	-.23	-.77	.49	-.46	.52	.46	9.71***
Confusing anecdote with "hard data"	3.24	-.24	-.45	.28	-.14	.12	.43	3.57**
Accepting uncritically conclusions of authorities	3.46	-.11	-.42	.15	-.23	.40	.21	4.63**
Being unable to integrate ideas	3.96	-.15	-.19	-.05	-.10	.32	.17	2.04
Failing to recognize that evidence can support conclusions	3.39	-.12	-.39	.11	-.18	.24	.34	3.76**
Using inconsistent statements	3.33	.12	-.38	-.17	-.00	.25	.18	1.96
Arguing for the obvious	2.76	-.09	-.16	-.22	-.14	.79	-.19	5.58**
Unable to perceive analogy breakdown	2.98	-.03	-.35	.05	.12	.24	-.02	1.58
Failing to draw conclusions about a pattern	3.41	.18	-.18	-.29	.14	.33	-.19	2.44*
Making implausible assumptions	3.44	.23	-.42	.01	.09	.10	-.03	1.55

Table 3 (Continued)

Variable	Total (N=255)	Chemistry (N=37)	Computer Science (N=43)	Education (N=42)	Engineering (N=43)	English (N=44)	Psychology (N=46)	F
		Discipline						
Applying rules without justification	3.38	.70	.11	-.33	.72	-1.06	-.14	16.24***
Relying solely on narrative	3.68	-.12	-.11	.13	-.25	.50	-.16	2.99*
Preferring complex to simple explanations	3.23	.07	-.11	-.06	-.23	.16	.18	.81
Unable to see a pattern	3.76	.13	-.32	.02	-.05	.23	-.00	1.61
Failing to qualify conclusions	3.44	.07	-.56	.39	-.17	-.01	.28	4.99**
Writing a one-sided essay	3.01	-.30	-.60	.47	-.28	.47	.25	7.46***
Searching for a complicated solution	3.28	.02	.16	-.18	.23	-.19	-.04	1.11
Ignoring details	3.74	.20	-.37	-.12	-.04	.26	.082	.23
Unwilling to respond in an unknown situation	3.28	.56	.06	-.28	.20	-.35	-.19	3.76**
Unable to generate hypotheses	3.94	.14	-.27	.08	-.06	-.12	.23	1.54
Unable to suggest alternatives	3.43	-.11	-.15	.30	-.04	-.00	.00	.90
Unable to suggest tests of hypotheses	3.52	.31	-.28	.17	-.19	-.64	.63	8.15***

Note. Entries under individual disciplines are deviations from the grand means given under "Total." Standard deviations for totals range from .93 to 1.27, and 15 of 25 are between 1.0 and 1.2.

*p < .01

**p < .05

***p < .001

Table 4

Reasoning Skills, Errors, and Incidents Rated as Most Important or Critical by Disciplines

Discipline	Skills	Errors/Incidents
Chemistry	Drawing sound inferences from observations (4.57)	Applying a formula, algorithm, or other rule without sufficient justification (4.08)
	Critically analyzing and evaluating previous research or reports in a field (4.46)	Being unable to generate hypotheses independently (4.08)
	Generating new questions or experiments to extend or support the interpretation of data (4.42)	
Computer Science	Breaking down complex problems or situations into simpler ones (4.62)	Being unable to integrate and synthesize ideas from various sources (3.77)
	Reasoning or solving problems in situations in which all facts underlying a problem situation are known (4.19)	
Education	Supporting conclusions with sufficient data or information (4.55)	Making generalizations on the basis of insufficient evidence (4.17)
	Determining whether the conclusions drawn are logically consistent with, and adequately supported by, the data (4.52)	Confusing coincidence and/or correlation with causation (4.10)
		Failing to evaluate the credibility or reliability of a source or text (4.10)

Table 4 (Continued)

Discipline	Skills	Errors/Incidents
	Clearly identifying central issues and problems to be investigated or hypotheses to be tested (4.50)	
	Drawing sound inferences from observations (4.50)	
Engineering	Breaking down complex problems or situations into simpler ones (4.60)	Applying a formula, algorithm, or other rule without sufficient justification (4.09)
	Reasoning or solving problems in situations in which all the needed information is <u>not</u> known (4.53)	
	Identifying all the variables involved in a problem (4.40)	
English	Elaborating an argument and developing its implications (4.82)	Being unable to integrate and synthesize ideas from various sources (4.27)
	Understanding, analyzing, and evaluating arguments (4.75)	Accepting the central assumptions in an argument without questioning them (4.20)
	Supporting general assertions with details (4.66)	Relying solely on narrative or description in papers and reports when analysis is appropriate (4.18)
	Recognizing the central argument or thesis in a work (4.61)	

Table 4 (Continued)

Discipline	Skills	Errors/Incidents
Psychology	Critically analyzing and evaluating previous research or reports in a field (4.58)	Confusing coincidence and/or correlation with causation (4.43)
	Clearly identifying central issues and problems to be investigated or hypotheses to be tested (4.57)	Accepting the central assumptions in an argument without questioning them (4.20)
	Determining whether the conclusions drawn are logically consistent with, and adequately supported by, the data (4.52)	

Note. Numbers in parentheses are the average ratings for each skill, error, or incident for each discipline.

Table 5

Reasoning Skills Rated Consistently As At Least Moderately Important

Skill	Mean Rating
Reasoning or solving problems in situations in which all the needed information is <u>not</u> known	4.24
Detecting fallacies and logical contradictions in arguments	3.92
Deducing new information from a set of relationships	3.86
Recognizing structural similarities between one type of problem or theory and another	3.83
Taking well-known principles and ideas from one area and applying them to a different specialty	3.75
Monitoring one's own progress in solving problems	3.71
Deriving from the study of single cases structural features or functional principles that can be applied to other cases	3.68
Making explicit all relevant components in a chain of logical reasoning	3.57
Testing the validity of an argument by searching for counterexamples	3.57

<u>Note</u>. Moderately important is defined as having an average rating over all disciplines of 3.5 or greater. There were no significant differences among disciplines with respect to the average importance of these skills.

Table 6

Errors or Incidents Rated Consistently as at Least Moderately Serious or Critical

Error/Incident	Mean Rating
Accepting the central assumptions in an argument without questioning them	3.96
Being unable to integrate and synthesize ideas from various sources	3.96
Being unable to generate hypotheses independently	3.94
Being unable to see a pattern in results or to generalize when appropriate	3.76
Ignoring details that contradict an expected or desired result	3.74
Submitting a paper that failed to address the assigned issues	3.67
Basing conclusions on analysis of only part of a text or data set	3.57

Note. Moderately critical is defined as having an average rating over all disciplines of 3.5 or greater. There were no significant differences among disciplines with respect to the average ratings of seriousness or criticality.

Table 7

Summary of Variables Defining Factors Underlying
Ratings of Importance of Reasoning Skills

Factor	Variables Loading Highest on Factors	
I	Recognizing shifts in the meaning of a word in the course of an argument	(73)
	Elaborating an argument and developing its implications	(72)
	Recognizing supporting points in an argument	(70)
	Recognizing the historical context of a problem	(67)
	Understanding, analyzing, and evaluating arguments	(64)
	Judging whether a thesis has been adequately supported	(64)
	Recognizing the central argument or thesis in a work	(63)
II	Generating valid explanations to account for observations	(72)
	Supporting conclusions with sufficient data or information	(69)
	Drawing sound inferences from observations	(69)
	Determining whether the conclusions drawn are logically consistent with, and adequately supported by, the data	(68)
	Comparing conclusions with what is already known	(67)
	Considering alternative conclusions	(59)
III	Setting up formal models for problems under consideration	(69)
	Breaking down complex problems or situations into simpler ones	(59)
	Translating graphs or symbolic statements into words and vice versa	(57)
	Identifying all the variables involved in a problem	(56)
	Knowing the rules of formal logic	(53)

Table 7 (Continued)

Factor	Variables Loading Highest on Factors	
IV	Recognizing structural similarities between one type of problem or theory and another	(61)
	Drawing distinctions between similar but not identical ideas	(53)
	Synthesizing two different positions into a third position	(44)
	Deriving from the study of single cases structural features or functional principles that can be applied to other cases	(41)
V	Finding alternative explanations for observations	(53)
	Generating alternative hypotheses	(46)

Note. Only the variables that best characterize the factors are given. Loadings are given in parentheses for the factor on which the variable's loading was most predominant.

Table 8

Summary of Variables Defining Factors Underlying Ratings of Seriousness of Reasoning Errors

Factor	Variables Loading Highest on Factors	
I	Offering irrelevant evidence to support a point	(75)
	Making generalizations on the basis of insuffiecent evidence	(72)
	Failing to evaluate the credibility or reliability of a source or text	(71)
	Accepting the central assumptions in an argument without questioning them	(70)
II	Failing to take into account the base rate for a phenomenon in a population	(76)
	Failing to recognize differences between a sample and a population of interest	(68)
	Offering irrelevant statements about a person's character or circumstances to oppose his/her conclusion	(63)
	Confusing coincidence and/or correlation with causation	(61)

Note. Only the variables that best characterize the factors are given. Loadings are given in parentheses for the factor on which the variable's loading was most predominant.

Table 9

Summary of Variables Defining Factors Underlying Ratings of the Effects of Critical Incidents

Factor	Variables Loading Highest on Factors	
I	Accepting or supporting an argument based more on emotional appeal than on evidence	(72)
	Offering criticism of other students' presentations that was not constructive or well founded	(65)
	Confusing anecdote and/or opinion with "hard data"	(65)
	Submitting a paper that failed to address the assigned issues	(62)
II	Accepting the conclusions of recognized authorities without critically evaluating them	(71)
	Being able to criticize but unable to suggest better alternatives	(59)
	Being unable to integrate and synthesize ideas from various sources	(58)
	Being unable to generate hypotheses independently	(56)
	Failing to recognize that evidence can support more than one conclusion	(55)
III	Applying a formula, algorithm, or other rule without sufficient justification	(67)
	Searching for a complicated solution when an obvious simple one exists	(58)
	Being unwilling to respond in a new or unknown situation when the limits of a student's knowledge has been reached	(54)

Note. Only the variables that best characterize the factors are given. Loadings are given in parentheses for the factor on which the variable's loading was most predominant.

Table 10

Means of Scales Based on Items that Best Defined Factors by Disciplines

Discipline	Scale 1	2	3	4	5
Chemistry	3.34	4.29	3.57	3.38	3.43
Computer Science	2.97	3.37	4.05	3.58	2.90
Education	3.83	4.29	3.44	3.73	3.93
Engineering	3.25	4.11	4.00	3.61	3.44
English	4.53	4.10	2.70	3.87	3.46
Psychology	3.73	4.26	3.37	3.78	4.21
Total	3.61	4.07	3.52	3.66	3.56

Key. Scales are unweighted composites of the following variables:

Scale 1 = Recognizing shifts in the meaning of a word in the course of an argument
Elaborating an argument and developing its implications
Recognizing supporting points in an argument
Recognizing the historical context of a problem
Judging whether a thesis has been adequately supported
Recognizing the central argument or thesis in a work
Understanding, analyzing, and evaluating arguments

Scale 2 = Generating valid explanations to account for observations
Drawing sound inferences from observations
Supporting conclusions with sufficient data or information
Determining whether the conclusions drawn are logically consistent with, and adequately supported by, the data
Comparing conclusions with what is already known
Considering alternative conclusions
Revising a previously held view to account for new observations

Scale 3 = Setting up formal models for problems under consideration
Breaking down complex problems or situations into simpler ones
Translating graphs or symbolic statements into words and vice versa
Identifying all the variables involved in a problem
Knowing the rules of formal logic

Scale 4 = Recognizing structural similarities between one type of problem or theory and another
Drawing distinctions between similar but not identical ideas
Deriving from the study of single cases structural features or functional principles that can be applied to other cases
Synthesizing two different positions into a third position
Deriving general or abstract principles from disparate facts or cases

Scale 5 = Finding alternative explanations for observations
Being sensitive to the strength of different types of evidence (correlational, causal, testimony)
Generating alternative hypotheses
Recognizing the probabilistic nature of most events

Note. Means have been divided by the number of items on each scale.

Appendix A

Questionnaire

I. REASONING SKILLS

The following are checklists of some general and specific reasoning skills that may be important to success in your graduate program. Please rate these skills with respect to:

(1) frequency--how often do students in your graduate program need to use this skill?

(2) importance for success--to what extent does this skill differentiate between marginal and successful students in your program?

For "importance for success," please refer to these descriptions:

1. This skill is not relevant to my field of teaching
2. There is little or no difference between marginal and successful students with respect to this skill
3. There is a moderately important difference between marginal and successful students with respect to this skill
4. There is a very important difference between marginal and successful students with respect to this skill
5. There is a critically important difference between marginal and successful students with respect to this skill

	Frequency					Importance for Success/difference				
A. Reasoning Skills--General	Never/ Hardly Ever				Very Frequently	Not Relevant	Little	Moderate	Important	Critical
Deriving general or abstract principles from disparate facts or cases.	1	2	3	4	5	1	2	3	4	5
Reasoning or problem solving in situations in which all facts underlying a problem solution are known explicitly.	1	2	3	4	5	1	2	3	4	5
Reasoning or problem solving in situations in which all the needed information is not known . . .	1	2	3	4	5	1	2	3	4	5
Reasoning when inconsistencies are present in the information	1	2	3	4	5	1	2	3	4	5
Knowing the rules of formal logic	1	2	3	4	5	1	2	3	4	5
Taking well known principles and ideas from one area and applying them to a different specialty . .	1	2	3	4	5	1	2	3	4	5
Being able to both support and refute a given position.	1	2	3	4	5	1	2	3	4	5
Critically analyzing and evaluating previous research or reports in a field.	1	2	3	4	5	1	2	3	4	5
Incorporating isolated instances or data into a preexisting framework	1	2	3	4	5	1	2	3	4	5
Monitoring one's own progress in solving problems .	1	2	3	4	5	1	2	3	4	5

	Frequency					Importance for Success/difference				
	Never/ Hardly Ever				Very Frequently	Not Relevant	Little	Moderate	Important	Critical
Deriving from the study of single cases structural features or functional principles that can be applied to other cases.	1	2	3	4	5	1	2	3	4	5
Recognizing the probabilistic nature of most events.	1	2	3	4	5	1	2	3	4	5
Understanding, analyzing, and evaluating arguments.	1	2	3	4	5	1	2	3	4	5
Deducing new information from a set of relationships	1	2	3	4	5	1	2	3	4	5
Knowing what kind of evidence will support or refute a hypothesis or thesis	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5
B. Specific Skills--Problem Definition										
Clearly identifying central issues and problems to be investigated or hypotheses to be tested.	1	2	3	4	5	1	2	3	4	5
Recognizing two or more sides of an issue	1	2	3	4	5	1	2	3	4	5
Generating alternative hypotheses	1	2	3	4	5	1	2	3	4	5
Breaking down complex problems or situations into simpler ones.	1	2	3	4	5	1	2	3	4	5
Identifying all the variables involved in a problem	1	2	3	4	5	1	2	3	4	5
Identifying more than one approach to solving a problem	1	2	3	4	5	1	2	3	4	5
Recognizing structural similarities between one type of problem or theory and another	1	2	3	4	5	1	2	3	4	5
Developing operational (or very precise) definitions of concepts	1	2	3	4	5	1	2	3	4	5
Setting up formal models for problems under consideration	1	2	3	4	5	1	2	3	4	5
Recognizing the historical context of a problem . .	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5

	Frequency					Importance for Success/difference				
	Never/ Hardly Ever				Very Frequently	Not Relevant	Little	Moderate	Important	Critical
C. Specific Skills--Constructing Theses or Arguments										
Supporting general assertions with details.	1	2	3	4	5	1	2	3	4	5
Making explicit all relevant components in a chain of logical reasoning.	1	2	3	4	5	1	2	3	4	5
Distinguishing between relevant and irrelevant information	1	2	3	4	5	1	2	3	4	5
Perceiving relationships among observations	1	2	3	4	5	1	2	3	4	5
Drawing distinctions betweem similar but not identical ideas	1	2	3	4	5	1	2	3	4	5
Using analogies appropriately	1	2	3	4	5	1	2	3	4	5
Elaborating an argument and developing its implications.	1	2	3	4	5	1	2	3	4	5
Drawing sound inferences from observations.	1	2	3	4	5	1	2	3	4	5
Producing an argument that is internally consistent.	1	2	3	4	5	1	2	3	4	5
Synthesizing two different positions into a third position.	1	2	3	4	5	1	2	3	4	5
Translating graphs or symbolic statements into words and vice versa.	1	2	3	4	5	1	2	3	4	5
Other (specify)________________	1	2	3	4	5	1	2	3	4	5
D. Specific Skills--Analyzing Arguments										
Recognizing the central argument or thesis in a work.	1	2	3	4	5	1	2	3	4	5
Identifying both stated and unstated assumptions. .	1	2	3	4	5	1	2	3	4	5
Testing the validity of an argument by searching for counterexamples	1	2	3	4	5	1	2	3	4	5
Finding alternative explanations for observations .	1	2	3	4	5	1	2	3	4	5
Recognizing supporting points in an argument. . . .	1	2	3	4	5	1	2	3	4	5
Detecting fallacies and logical contradictions in arguments.	1	2	3	4	5	1	2	3	4	5

	Frequency					Importance for Success/difference				
	Never/ Hardly Ever				Very Frequently	Not Relevant	Little	Moderate	Important	Critical
Distinguishing major and minor points in an argument.	1	2	3	4	5	1	2	3	4	5
Being sensitive to the strength of different types of evidence (correlational, causal, testimony). . .	1	2	3	4	5	1	2	3	4	5
Recognizing shifts in the meaning of a word in the course of an argument	1	2	3	4	5	1	2	3	4	5
Judging whether a thesis has been adequately supported	1	2	3	4	5	1	2	3	4	5
Determining whether the conclusions drawn are logically consistent with, and adequately supported by, the data.	1	2	3	4	5	1	2	3	4	5
Distinguishing among contributing, necessary and sufficient factors.	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5
E. Specific Skills--Drawing Conclusions										
Generating valid explanations to account for observations.	1	2	3	4	5	1	2	3	4	5
Supporting conclusions with sufficient data or information	1	2	3	4	5	1	2	3	4	5
Recognizing the implications or consequences of an interpretation.	1	2	3	4	5	1	2	3	4	5
Qualifying conclusions appropriately and recognizing ways in which conclusions could be challenged . . .	1	2	3	4	5	1	2	3	4	5
Generating new questions or experiments to extend or support the interpretation of data	1	2	3	4	5	1	2	3	4	5
Considering alternative conclusions	1	2	3	4	5	1	2	3	4	5
Comparing conclusions with what is already known. .	1	2	3	4	5	1	2	3	4	5
Revising a previously held view to account for new observations.	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5

II. REASONING ERRORS

Please indicate which of the following kinds of reasoning errors you have observed in your graduate students and the frequency with which you have noted these errors.

Error	Frequency: Never/ Hardly Ever				Very Frequent	Seriousness: Not Serious				Very Serious
Failing to recognize differences between a sample and a population of interest.	1	2	3	4	5	1	2	3	4	5
Failing to take into account the base rate for a phenomenon in a population	1	2	3	4	5	1	2	3	4	5
Offering irrelevant statements about a person's character or circumstances to oppose his/her conclusion.	1	2	3	4	5	1	2	3	4	5
Failing to recognize when a conclusion is concealed within a premise	1	2	3	4	5	1	2	3	4	5
Confusing coincidence and/or correlation with causation	1	2	3	4	5	1	2	3	4	5
Using an ambiguous term to shift senses in the same argument.	1	2	3	4	5	1	2	3	4	5
Arguing that a conclusion should be accepted (rejected) because it has not been disproved (proved).	1	2	3	4	5	1	2	3	4	5
Accepting the central assumptions in an argument without questioning them.	1	2	3	4	5	1	2	3	4	5
Failing to evaluate the credibility or reliability of a source or text	1	2	3	4	5	1	2	3	4	5
Offering irrelevant evidence to support a point	1	2	3	4	5	1	2	3	4	5
Making generalizations on the basis of insufficient evidence.	1	2	3	4	5	1	2	3	4	5
Reading into a text one's own views or inappropriately interpreting text or data on the basis of one's own experience.	1	2	3	4	5	1	2	3	4	5
Allowing anecdotal information to override more extensive statistical data.	1	2	3	4	5	1	2	3	4	5
Basing conclusions on analysis of only part of a text or data set	1	2	3	4	5	1	2	3	4	5
Failing to recognize relevant similarities and dissimilarities when using analogies to argue	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5

III. CRITICAL INCIDENTS

How frequently have you observed the following kinds of incidents or tendencies in your students? What effect have these had on your estimation of the analytical ability of your students?

Incidents	Never/ Hardly Ever				Very Frequent	No Effect				Great Effect
Repeatedly making irrelevant remarks during class discussions . .	1	2	3	4	5	1	2	3	4	5
Offering criticism of other students' presentations that was not constructive or well founded.	1	2	3	4	5	1	2	3	4	5
Submitting a paper that failed to address the assigned issues . .	1	2	3	4	5	1	2	3	4	5
Accepting or supporting an argument based more on emotional appeal than on evidence	1	2	3	4	5	1	2	3	4	5
Confusing anecdote and/or opinion with "hard data".	1	2	3	4	5	1	2	3	4	5
Accepting the conclusions of recognized authorities without critically evaluating them.	1	2	3	4	5	1	2	3	4	5
Being unable to integrate and synthesize ideas from various sources	1	2	3	4	5	1	2	3	4	5
Failing to recognize that evidence can support more than one conclusion.	1	2	3	4	5	1	2	3	4	5
Using inconsistent statements to support the same position. . . .	1	2	3	4	5	1	2	3	4	5
Arguing in support of the obvious	1	2	3	4	5	1	2	3	4	5
Being unable to perceive when an analogy breaks down.	1	2	3	4	5	1	2	3	4	5
Recognizing a pattern but failing to draw any conclusions about it or attempting to explain it.	1	2	3	4	5	1	2	3	4	5
Making implausible assumptions.	1	2	3	4	5	1	2	3	4	5
Applying a formula, algorithm, or other rule without sufficient justification	1	2	3	4	5	1	2	3	4	5
Relying solely on narrative or description in papers and reports when analysis is appropriate.	1	2	3	4	5	1	2	3	4	5
Preferring complex or far-fetched explanations over simpler ones.	1	2	3	4	5	1	2	3	4	5
Being unable to see a pattern in results or to generalize when appropriate	1	2	3	4	5	1	2	3	4	5
Failing to recognize that a definite conclusion cannot be reached (or failing to qualify a conclusion appropriately). . . .	1	2	3	4	5	1	2	3	4	5

	Never/ Hardly Ever				Very Frequent	No Effect				Great Effect
Writng a one-sided essay when a more balanced treatment is appropriate	1	2	3	4	5	1	2	3	4	5
Searching for a complicated solution when an obvious simple one exists.	1	2	3	4	5	1	2	3	4	5
Ignoring details that contradict an expected or desired result. .	1	2	3	4	5	1	2	3	4	5
Being unwilling to respond in a new or unknown situation when the limits of a student's knowledge has been reached.	1	2	3	4	5	1	2	3	4	5
Being unable to generate hypotheses independently	1	2	3	4	5	1	2	3	4	5
Being able to criticize but unable to suggest better alternatives.	1	2	3	4	5	1	2	3	4	5
Being unable to suggest appropriate tests of hypotheses	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5
Other (specify)____________________	1	2	3	4	5	1	2	3	4	5

IV. BACKGROUND INFORMATION

1. Name ______________________________

2. Academic rank (e.g., Associate Professor)____________________

3. Institution______________________________

4. Department______________________________

5. Years of teaching experience at the undergraduate/graduate levels____________________

6. Your current teaching load (number of contact hours per week) undergraduate____________________
   graduate ____________________

7. Altogether, about how many graduate students at each of the following levels have you advised over the past three years?

   Masters' ________
   Doctoral ________

Do you have any comments that you'd like to make about measuring students' analytical abilities?

____

____

____

Thanks for your contribution to our study.

Appendix B

Factor Analyses

Appendix B.1

Varimax Factor Loadings for Ratings of Importance of Reasoning Skills

Appendix B.1

Varimax Factor Loadings for Ratings of Importance of Reasoning Skills

Variable	Factor I	II	III	IV	V	Communality
Reasoning Skills--General						
Deriving general or abstract principles from disparate facts or cases				36		28
Reasoning or solving problems in situations in which all facts underlying a problem solution are known explicitly			48			27
Reasoning or solving problems in situations in which all the needed information is <u>not</u> known				38	35	31
Reasoning when inconsistencies are present in the information		34		32	38	41
Knowing the rules of formal logic			53			38
Taking well-known principles and ideas from one area and applying them to a different specialty						16
Being able to both support and refute a given position	51				31	40
Critically analyzing and evaluating previous research or reports in a field	40	45			30	46
Incorporating isolated instances or data into a preexisting framework						23
Monitoring one's own progress in solving problems			36			20
Deriving from the study of single cases structural features or functional principles that can be applied to other cases				41		21
Recognizing the probabilistic nature of most events					32	25
Understanding, analyzing and evaluating arguments	64					49
Deducing new information from a set of relationships		33	33			32
Knowing what kind of evidence will support or refute a hypothesis or thesis	53	43				52

Appendix B.1 (Continued)

Variable	Factor I	II	III	IV	V	Communality
Specific Skills--Problem Definition						
Clearly identifying central issues and problems to be investigated or hypotheses to be tested	32	30				24
Recognizing two or more sides of an issue	58				36	54
Generating alternative hypotheses	32				46	39
Breaking down complex problems or situations into simpler ones			59			44
Identifying all the variables involved in a problem		32	56			43
Identifying more than one approach to solving a problem		31	36			39
Recognizing structural similarities between one type of problem or theory and another				61		47
Developing operational (or very precise) definitions of concepts	30		48			33
Setting up formal models for problems under consideration			69			53
Recognizing the historical context of a problem	67					55
Specific Skills--Constructing Theses or Arguments						
Supporting general assertions with details	57	40				52
Making explicit all relevant components in a chain of logical reasoning	40		40			42
Distinguishing between relevant and irrelevant information	36	51				46
Perceiving relationships among observations		58		37		52
Drawing distinctions between similar but not identical ideas	46			53		54
Using analogies appropriately	35			39		29

Appendix B.1 (Continued)

Variable	Factor I	II	III	IV	V	Communality
Elaborating an argument and developing its implications	72					64
Drawing sound inferences from observations		69				56
Producing an argument that is internally consistent	51	43				49
Synthesizing two different positions into a third position	32			44		34
Translating graphs or symbolic statements into words and vice versa			57			36
Specific Skills--Analyzing Arguments						
Recognizing the central argument or thesis in a work	63	34				53
Identifying both stated and unstated assumptions	52					47
Testing the validity of an argument by searching for counterexamples	36		44		33	46
Finding alternative explanations for observations		38			53	53
Recognizing supporting points in an argument	70	33				64
Detecting fallacies and logical contradictions in arguments	58	32				54
Distinguishing major and minor points in an argument	53					43
Being sensitive to the strength of different types of evidence (correlational, causal, testimony)	41	35			42	52
Recognizing shifts in the meaning of a word in the course of an argument	73					65
Judging whether a thesis has been adequately supported	64	41				61

Appendix B.1 (Continued)

Variable	Factor I	II	III	IV	V	Communality
Determining whether the conclusions drawn are logically consistent with, and adequately supported by, the data	36	68				64
Distinguishing among contributing, necessary, and sufficient factors		43	35	34		49
Specific Skills--Drawing Conclusions						
Generating valid explanations to account for observations		72				59
Supporting conclusions with sufficient data or information	31	69				58
Recognizing the implications or consequences of an interpretation	47	45				54
Qualifying conclusions appropriately and recognizing ways in which conclusions could be challenged	48	56				57
Generating new questions or experiments to extend or support the interpretation of data		57			37	52
Considering alternative conclusions		59			37	60
Comparing conclusions with what is already known		67				56
Revising a previously held view to account for new observations		58				49
Common Variance	8.41	7.37	3.84	3.21	2.51	25.34

Note. Loadings less than .30 have been omitted, as have all decimal points.

Appendix B.2

Varimax Factor Loadings for Ratings of Frequency and Seriousness of Reasoning Errors

Appendix B.2

Varimax Factor Loadings for Ratings of
Frequency of Reasoning Errors

Error	Factor I	Factor II	Communality
Failing to recognize differences between a sample and a population of interest		74	56
Failing to take into account the base rate for a phenomenon in a population		75	58
Offering irrelevant statements about a person's character or circumstances to oppose his/her conclusion	48	40	39
Failing to recognize when a conclusion is concealed within a premise	51	40	43
Confusing coincidence and/or correlation with causation	30	53	38
Using an ambiguous term to shift senses in the same argument	65		47
Arguing that a conclusion should be accepted (rejected) because it has not been disproved (proved)	37	34	25
Accepting the central assumptions in an argument without questioning them	63		40
Failing to evaluate the credibility or reliability of a source or text	63		44
Offering irrelevant evidence to support a point	67		47
Making generalizations on the basis of insufficient evidence	63		46
Reading into a text one's own views or inappropriately interpreting text or data on the basis of one's own experience	70		55
Allowing anecdotal information to override more extensive statistical data	42	53	46
Basing conclusions on analysis of only part of a text or data set	58	30	42
Failing to recognize relevant similarities and dissimilarities when using analogies to argue	50	40	42
Common Variance	4.06	2.60	6.66

Note. Loadings less than .30 have been omitted, as have all decimal points.

Appendix B.2

Varimax Factor Loadings for Ratings of Seriousness of Reasoning Errors

Error	Factor I	Factor II	Communality
Failing to recognize differences between a sample and a population of interest		68	48
Failing to take into account the base rate for a phenomenon in a population		76	59
Offering irrelevant statements about a person's character or circumstances to oppose his/her conclusion	30	63	48
Failing to recognize when a conclusion is concealed within a premise	40	56	47
Confusing coincidence and/or correlation with causation	38	61	51
Using an ambiguous term to shift senses in the same argument	50	48	48
Arguing that a conclusion should be accepted (rejected) because it has not been disproved (proved)	35	60	48
Accepting the central assumptions in an argument without questioning them	70		53
Failing to evaluate the credibility or reliability of a source or text	71		56
Offering irrelevant evidence to support a point	75		61
Making generalizations on the basis of insufficient evidence	72		56
Reading into a text one's own views or inappropriately interpreting text or data on the basis of one's own experience	65	33	53
Allowing anecdotal information to override more extensive statistical data	43	56	50
Basing conclusions on analysis of only part of a text or data set	55		39
Failing to recognize relevant similarities and dissimilarities when using analogies to argue	46	44	40
Common Variance	3.97	3.58	7.56

Note. Loadings less than .30 have been omitted, as have all decimal points.

Appendix B.3

Varimax Factor Loadings of Frequency and Effects of Critical Incidents

Appendix B.3

Varimax Factor Loadings of Effects of Critical Incidents

Incident	Factor I	Factor II	Factor III	Communality
Repeatedly making irrelevant remarks during class discussions	60			36
Offering criticism of other students' presentations that was not constructive or well founded	65			44
Submitting a paper that failed to address the assigned issues	62			42
Accepting or supporting an argument based more on emotional appeal than on evidence	72			60
Confusing anecdote and/or opinion with "hard data"	65			52
Accepting the conclusions of recognized authorities without critically evaluating them		71		56
Being unable to integrate and synthesize ideas from various sources		58		40
Failing to recognize that evidence can support more than one conclusion		55		40
Using inconsistent statements to support the same position	56		40	51
Arguing in support of the obvious	38	39		37
Being unable to perceive when an analogy breaks down	49		38	47
Recognizing a pattern but failing to draw any conclusions about it or attempting to explain it		34	48	40
Making implausible assumptions	48		53	53
Applying a formula, algorithm, or other rule without sufficient justification			67	45
Relying solely on narrative or description in papers and reports when analysis is appropriate	39	41		37
Preferring complex or far-fetched explanations over simpler ones	32		40	33

Appendix B.3 (Continued)

Incident	Factor I	Factor II	Factor III	Communality
Being unable to see a pattern in results or to generalize when appropriate		40	42	40
Failing to recognize that a definite conclusion cannot be reached (or failing to qualify a conclusion appropriately)	42	48	32	51
Writing a one-sided essay when a more balanced treatment is appropriate	48	41		44
Searching for a complicated solution when an obvious simple one exists			58	40
Ignoring details that contradict an expected or desired result	34	40	39	42
Being unwilling to respond in a new or unknown situation when the limits of a student's knowledge have been reached			54	39
Being unable to generate hypotheses independently		56	35	44
Being able to criticize but unable to suggest better alternatives		59		40
Being unable to suggest appropriate tests of hypotheses		41	37	31
Common Variance	4.15	3.66	3.05	10.86

Note. Loadings less than .30 have been omitted, as have all decimal points.

Appendix B.3

Varimax Factor Loadings of Frequency of Critical Incidents

Incident	Factor I	Factor II	Factor III	Communality
Repeatedly making irrelevant remarks during class discussions	52			31
Offering criticism of other students' presentations that was not constructive or well founded	46			22
Submitting a paper that failed to address the assigned issues	37		36	27
Accepting or supporting an argument based more on emotional appeal than on evidence	71			53
Confusing anecdote and/or opinion with "hard data"	59			39
Accepting the conclusions of recognized authorities without critically evaluating them	41	47		39
Being unable to integrate and synthesize ideas from various sources		44		29
Failing to recognize that evidence can support more than one conclusion	40	36		35
Using inconsistent statements to support the same position	53		53	58
Arguing in support of the obvious	47		32	35
Being unable to perceive when an analogy breaks down	47	33	30	42
Recognizing a pattern but failing to draw any conclusions about it or attempting to explain it		42		28
Making implausible assumptions	32		50	42
Applying a formula, algorithm, or other rule without sufficient justification			54	34
Relying solely on narrative or description in papers and reports when analysis is appropriate		39		28
Preferring complex or far-fetched explanations over simpler ones			49	30

Appendix B.3 (Continued)

Incident	Factor I	Factor II	Factor III	Communality
Being unable to see a pattern in results or to generalize when appropriate		47	45	45
Failing to recognize that a definite conclusion cannot be reached (or failing to qualify a conclusion appropriately)	41	42		43
Writing a one-sided essay when a more balanced treatment is appropriate	65			48
Searching for a complicated solution when an obvious simple one exists			64	43
Ignoring details that contradict an expected or desired result	39	32	41	42
Being unwilling to respond in a new or unknown situation when the limits of a student's knowledge have been reached		48	36	36
Being unable to generate hypotheses independently		72		54
Being able to criticize but unable to suggest better alternatives alternatives	35	52		39
Being unable to suggest appropriate tests of hypotheses		68		52
Common Variance	3.75	3.27	2.72	9.75

Note. Loadings less than .30 have been omitted, as have all decimal points.