DOCUMENT RESUME

ED 391 815 TM 024 154

AUTHOR Carlson, Sybil B.; Ward, William C.
TITLE A New Look at Formulating Hypotheses Items. GRE Board Professional Report No. 85-14P.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Graduate Record Examinations Board, Princeton, N.J.
REPORT NO ETS-RR-88-12
PUB DATE Apr 88
NOTE 113p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Adaptive Testing; *Computer Assisted Testing; *Costs; Pilot Projects; *Scoring; Test Construction; Test Format; *Testing Problems; *Test Items; Test Use
IDENTIFIERS *Graduate Record Examinations; *Hypothesis Formulation; Prototypes; Subjective Tests

ABSTRACT

Issues concerning the cost and feasibility of using Formulating Hypotheses (FH) test item types for the Graduate Record Examinations have slowed research into their use. This project focused on two major issues that need to be addressed in considering FH items for operational use: the costs of scoring and the assignment of scores along a range of values rather than the conventional number-right scoring. The first issue was addressed directly by seeking ways to increase the efficiency of scoring through computerized delivery and scoring. The second issue was addressed both directly and indirectly by recommending specific procedures for the computer recognition of responses and problem delivery that will be sufficiently reliable and well-rationalized to be acceptable to reasonable evaluators. After a series of analyses to explore the design and scoring of FH-type items, specific recommendations were suggested for developing a system to deliver computerized versions. When developed, the prototype will also serve as a computerized research tool to conduct further studies. Appendix A gives an example of an FH item and scoring system. Appendix B presents six tables of text analysis with the Writer's Workbench software. Appendix C gives the instructions and items for a small pilot test and eight graphs from test examples. (Contains 35 references.) (Author/SLD)

GRE®

GRADUATE RECORD EXAMINATIONS

A NEW LOOK AT FORMULATING HYPOTHESES ITEMS

Sybil B. Carlson
and
William C. Ward

GRE Board Professional Report No. 85-14P
ETS Research Report 88-12

April 1988

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

A New Look at Formulating Hypotheses Items

Sybil B. Carlson and William C. Ward
with Michael Canale, Lawrence Frase, and Glynda Hull

GRE Board Report No. 85-14P

April 1988

Educational Testing Service, Princeton N.J. 08541

ABSTRACT

In the past, the GRE Board supported research on an item type that measures higher-level cognitive abilities and that uses a free-response format--the Formulating Hypotheses (FH) item type. Further research was not recommended because of issues associated with the cost and feasibility of the operational use of a test composed of FH items. This project focused on the two major issues that need to be addressed in considering FH items for operational use: (1) the costs of scoring, and (2) rather than the conventional number-right scoring, the assignment of scores along a range of values. The first issue was addressed directly by seeking ways to increase the efficiency of scoring through computerized delivery and scoring. The second issue was addressed both directly and indirectly by recommending specific procedures for the computer recognition of responses and problem delivery that will be sufficiently reliable and well-rationalized to be acceptable to reasonable evaluators.

This project involved collaboration with experts who are closely involved in confronting the issues involved in the computer recognition and evaluation of open-ended responses. After a series of analyses to explore the design and scoring of FH-type items for computer delivery, we arrived at specific recommendations for developing a system to deliver computerized problems of the FH type. When developed, the prototype also will serve as a computerized research tool to conduct further investigations of potential variations in these types of items.

EXECUTIVE SUMMARY

In the past, the GRE Board supported research on an item type that measures higher-level cognitive abilities and utilizes a free-response format--the Formulating Hypotheses (FH) item type. Further research was not recommended because of issues associated with the cost and feasibility of operational use of a test composed of FH items. Two major issues need to be addressed in considering FH items for operational use: (1) the costs of scoring and (2) rather than the conventional number-right scoring, the assignment of scores along a range of values. The proposed research addressed these limiting factors. The first set of factors can be addressed directly by seeking ways to increase the efficiency of scoring through computerized delivery and scoring. The second issue can be addressed both directly and indirectly by developing procedures for the computer recognition of responses that are sufficiently reliable and well rationalized to be acceptable to reasonable evaluators.

This project involved collaboration with experts who are closely involved in confronting the challenges presented by the computer recognition and evaluation of open-ended responses. At the outset, we recognized that it would not be the goal of the project to design a computer program that would genuinely "understand" the natural language responses to FH items. We did, however, make some progress in designing the specifications for a prototype that could carry out an analysis of those responses, given the fact that we already knew a great deal about the kinds of responses that people were likely to make. When developed, the prototype also will serve as a computerized research tool to conduct further investigations of potential variations in these types of items. Our conclusions and recommendations are briefly reviewed in this summary.

Summary of Previous Research

For several years we have conducted research involving open-ended and problem-solving item formats. In research supported by the GRE Board (Frederiksen & Ward, 1975), four kinds of Scientific Thinking items were developed: Formulating Hypotheses, Evaluating Proposals, Solving Methodological Problems, and Measuring Constructs. These items were designed to elicit the types of reasoning behaviors that are applied to research problems in the graduate-level psychology curriculum and in the field of psychology. The FH item, for example, presented the results of a psychological investigation, such as a study showing that a disproportionately large number of children charged with juvenile delinquency come from disrupted families. The examinee was asked to list the many possible hypotheses that could explain the finding (see example in Appendix A). The responses were categorized, and several different scores were obtained. Results of the 1975 study indicated that scores based on number of responses, though highly reliable, were relatively uncorrelated with scores from conventional tests. Scores reflecting the quality of ideas produced on the FH test overlapped in variance with the GRE General Test verbal score, but the percent of true variance accounted for by the verbal score was less than 20%. Hence the fluency and quality scores on the FH test represent skills and abilities that appear to be largely unmeasured by conventional test items.

In a follow-up to the original study, students who had completed the FH test at the time they took the GRE General Test reported on their experiences in their first year of graduate work in psychology. FH scores were more effective than GRE General Test scores in predicting self-reports in two areas: self-appraisals of knowledge and skills in psychology, and professional accomplishments such as research, publication, and teaching.

In a subsequent construct validity study (Ward, Frederiksen, & Carlson, 1978), we examined the relationships of scores on "machine-scorable" and "free-response" forms of the FH tests with GRE General Test scores, a personality inventory, and a battery of cognitive process variables. The data indicated that the free-response and machine-scorable versions of FH clearly could not be considered alternate forms of the same test. The correlations between corresponding forms were low. Moreover, reasoning, the ability to think divergently, and cognitive flexibility in the context of relevant knowledge are brought to bear in the generated-response format but not in the conventional recognition-response format.

More recently, Carlson (1985) completed an exploratory investigation of the FH item format for the Law School Admission Test battery of the future, in which new FH problems designed to have face validity for law school candidates and FH problems previously developed for the GRE research were combined to create a test. The responses of a small sample of students to the test items appear to reflect performance dimensions that would serve as meaningful indicators of potential success in law school. The FH item type is still being considered by the Law School Admission Council, particularly as computerized delivery and scoring become practical and feasible. Another related study (Carlson, 1988), currently supported by the GRE Board, is exploring the identification of thinking skills exhibited by candidates in samples of their writing. The research may indicate that other, additional variables observed in verbal production tasks may contribute richer information about the reasoning skills of GRE candidates. If this result is obtained, FH items could be adapted to incorporate these reasoning skills also. These studies, as well as other ETS research and development activities, provide a solid basis of experience with open-ended response tests to guide the refinement and investigation of the measurement properties of an FH test for the GRE.

Powers and Enright (1986) recently conducted a GRE project to obtain information about the role of analytical abilities in graduate work. Graduate faculty in six fields of study were asked to make judgments about "(a) the importance for academic success of a wide variety of analytical skills, (b) the seriousness of various reasoning errors, and (c) the degree to which a variety of 'critical incidents' had affected their estimates of students' analytical abilities" (Powers & Enright, 1986). Data analyses yielded seven dimensions to represent clusters of reasoning skills that, on the basis of faculty responses, were differentially important for success in the different disciplines. One of the four dimensions consisted of skills involving the generation of hypotheses/alternatives/ explanations. "The ability to generate hypotheses independently was one of the incidents rated consistently as having a substantial effect on faculty perceptions of students' analytical abilities" (p. 12). Thus the results of this research further support our exploration of the potential of computer-delivered FH items as components of some form of a GRE instrument in the future.

Suggestions for Natural Language Analysis of FH Responses

Drawing on natural language processing research, we have experimented, both paper and pencil and online, with several techniques for analyzing FH responses.

One approach we explored was <u>pattern matching</u>. We created a computer program to search for single key words or combinations of them. An iterative procedure was followed to refine the program in a series of analyses of FH responses. These explorations produced better results than had been anticipated. Looking at the last set of 50 responses studied, the program was correct 35 times in identifying a response as either high quality or not high quality, wrong 4 times, and possibly wrong 3 times. That results in a correct assignment, on the basis of good versus poor responses, of 70% of all responses, and of 83-90% of the responses that were categorized--not far below useful.

Thus, the simplest form of pattern matching was not sufficiently accurate, given the complexity of the responses to FH items. The information gained suggested that it might be possible to combine a parts-of-speech analysis with keyword matching, since programs are available for syntactic parsing. With a view toward a scoring system that might use several levels of analysis, keyword analysis might be supplemented by additional forms of analysis, until reaching a cutoff point with a high enough confidence level.

Several analyses were conducted to explore the lexical, syntactic, and stylistic properties of a set of FH responses. The results show that these analyses, although not providing a complete picture of the quality of written responses, can be used to improve response measurement and to simplify scoring. For instance, analyses suggest where response categories might be combined. In addition, good responses contain complex syntactic structures, at least at the sentence level. Perhaps this feature of good writing could be used to select automatically certain texts for further analysis.

Measures of content similarity could also be used to detect individual differences in the ability to match what one writes to a problem statement. Our data did not allow us to study this possibility in depth; however, they show that lexical matching can be used to measure similarities among written samples. Hence, content similarity should be useful for studying individual differences in ability to respond to content domains.

The notion of case frames has been used in natural language understanding. According to Hayes and Carbonell (1984), the key advantage of this approach is that it combines a bottom-up recognition of structuring constituents with top-down instantiation of less structured, more complex constituents. Case frames, as used in parsing, actually consist of more than a predicate and a collection of cases. Each case also consists of a positional or lexical marker. A positional marker indicates that the case filler is preceded by a marker word, usually a preposition in the surface string. In case frame grammar, verbs are classified according to the cases that can occur with them. Case frame parsing proceeds by first looking for the verb in a sentence, then retrieving the case frame associated with that verb, and then attempting to recognize each expected case by relying on lexical and positional markings.

A further development in case frame parsing is the conceptual dependency theory (Schank and Abelson, 1977), which provided the rationale for grouping together the actions of several surface representations for verbs into primitive actions. Thus, the sentences "John gave Mary a ball" and "Mary took the ball from John," while differing syntactically in terms of case frame instantiation and verb choice, nonetheless are similar in terms of the action each sentence expresses--what Schank calls ATRANS, or the transfer of possession, control, or ownership. Thus, there exists a means of representing the semantic information derived from a case parse in a canonical form.

Certain aspects of the case frame approach seemed useful in our computer analysis task. First is the idea of relying upon verbs to provide a set of expectations about what the rest of a proposition will look like (keyword matching relies mostly upon nouns and adjectives). By letting verbs drive the analysis for this approach, and nouns show the way in the keyword matching, we could take advantage of as many lexical cues as possible. Second, the dependency relationships that are set up by the verb in case frame analysis might provide the necessary information to avoid false positives in some of the categories that proved nettlesome for keyword matching.

Finally, we experimented with a deeper kind of analysis, one that depends not solely on keyword strings or lexical and syntactic analyses, but also on features or semantic relationships. The problem with surface analyses of style and string matching is that they both have strong inherent limitations such that, beyond a certain point, they cannot be improved. While we recognize that surface-level analyses are efficient computationally and cost-effective, too, and that such approaches can certainly take us part of the distance, we were aware from previous research (Hull, Fox, Levin, & McCutchen, in press), and learned from our own experiments with actual responses to FH items, that such surface approaches will need to be supplemented.

We believe it is necessary to consider another sort of system, one that does not have such strong inherent limitations and can be upgraded and improved upon, and one that performs semantic analyses as well as syntactic analyses. In order to begin thinking about the design of such a system, we surveyed the computational techniques that are available for natural language analysis, and juxtaposed them to sample responses on the FH task. We did not expect to find a particular technique that we could import wholesale to solve our computational problem. Rather, we hoped to combine the strengths of whatever parsing strategies seemed useful into a single system.

The techniques we propose, which we call conceptual frames, begin with a linguistic analysis of the concepts and relationships that make up the semantic heart of the FH categories. The results of such an analysis then serve as predictors and constraints on our computational techniques, which combine features of case frame parsing and conceptual dependency theory.

Item Design for Computer Delivery

We also addressed the issue of creating --as an alternative to the conventional number-scoring system--a defensible scoring system in which scores on each item are assigned along a range of values. To create such a system, we need to demonstrate that the process by which scoring decisions are made is reasonable and rational by articulating specific, objective criteria for making these judgments. This can be accomplished by developing an accurate computerized scoring system such as that described in the preceding section. However, because the scoring system is dependent on the responses elicited by an FH task, we also need to demonstrate the validity of the values assigned to these responses. Thus, we will need to design FH-type items to obtain samples of performance that elicit rich and productive responses and are representative of what we intend to measure.

We briefly described the nature of hypothesis formation, a preliminary outline that might eventually contribute to test specifications for a computer-delivered FH test, and some approaches to refining the design of tasks of the FH type.

A small pilot test suggested that placing constraints on students' responses to FH items would not inhibit the production of high quality responses. Four experimental test booklets incorporating variations in FH problems and instructions were administered to a total of 60 students in different sections of an English composition course. The major results were as follows:

- Students conformed to the instruction to assume that the investigation described in a problem was methodologically sound. This constraint eliminated many hypotheses proposing flaws in the design or execution of a study as the basis for its findings. Such responses are highly variable, often trite, and frequently difficult to classify.

- In several problems students were instructed to respond with phrases rather than sentences; it was thought that this format would facilitate keyword matching. They ignored this instruction entirely, responding with complete sentences as they did in other problems.

- In several problems students were instructed to begin each hypothesis with one of two specified phrases. This limit did not seem to impose any constraints whatsoever--the responses were as varied as the responses to unconstrained problems. In fact, the students appeared to have sufficient freedom to provide the inverse cases of the different potential hypotheses, which may confound computer recognition considerably because these ideas can be expressed with so many variations of vocabulary and syntax.

This exploration suggests that the optimal format for responding may be one that requires the use of one specified introductory phrase. Because students did not appear to be constrained by using only one introductory phrase, and because it appears that syntax and vocabulary may be made more systematic without reducing the range and number of ideas, we may be able to achieve optimal conditions for computer recognition while maintaining the integrity of measurement we have previously experienced in FH problems.

Conclusions and Recommendations

Following a series of analyses to explore the design and scoring of FH-type items for computer delivery, we arrived at specific recommendations for developing a prototype system. The in-depth analyses of one set of FH responses demonstrated, in our judgment, that a computerized delivery and scoring system can be achieved with presently available tools and expertise. A number of computer-based linguistic analysis tools already have been developed, providing the basic components necessary for building a system using the conceptual analysis approach. In addition to scoring tools, the computerized adaptive testing system developed at ETS can be readily tailored to deliver items of the FH type. Because FH responses represent a high level of complexity and less well-structured verbal material, they serve as a good model for designing a system that is likely to deal more readily with the scoring of other forms of open-ended responding as well.

The report concludes with recommended stages for designing a test delivery and scoring system for open-ended, sentence-level responses, and for research during system development and after a prototype is functioning. Much of the research on constructing and refining the scoring system would take place while the prototype is being developed. More specifically, the following steps would constitute the next stages for the initial development of a delivery and scoring system for open-ended responses of the FH type:

1. Obtain a pool of responses to two or three FH-type items.

2. Create response categories by sorting and evaluating the responses.

3. Use the various computerized analytic tools to analyze pools of responses and assign them to categories.

4. Through many iterations, create a bank of common responses to develop a small domain for each FH-type item. A bank of common responses also will be created to develop a small domain across several FH-type items.

5. Combine the system for analyzing and identifying responses with the computerized adaptive testing delivery system to present the FH items, providing a prototype for further research.

Given the resources, this point in the development process could be accomplished in a year. Once a prototype is available, considerable research will be required to determine the optimal design of the FH-type items to support the measurement characteristics of the resulting instrument and to investigate human factor variables that influence responding on the computer.

The system we have recommended should be sufficiently flexible and powerful to accommodate a wide variety of sentence-level open-ended responses now and in the future.

Implications for the GRE

Successful completion of the research and development necessary to automate scoring of FH-type items would have both specific and general implications for GRE testing. Specifically, it would make it feasible for the GRE program to consider incorporating into its examinations an item type that requires a kind of reasoning that is important for success in graduate education and that is not well represented in the present General Test. Given the interest in increasing the breadth of the analytical section of the examination and in increasing its distinguishability from the abilities measured by the verbal and quantitative sections, this could be an important contribution to the redesign of the examination.

More generally, the FH work would serve as a model for the analysis of natural language responses that might be elicited by a variety of other item types. In reading comprehension, for example, questions posed in free-response form could be expected to result in responses whose analysis would involve issues almost identical to those posed by FH. The analytic techniques and computer programs developed for FH could thus serve to make free-response versions of a number of item types feasible, decreasing the test developer's dependence on the multiple-choice format and increasing the variety of tasks that could be considered for inclusion in the examination.

ACKNOWLEDGMENTS

This project was a collaborative endeavor in which the team of two ETS research psychologists and three external collaborators each performed independent analyses that were subsequently shared, evaluated, and synthesized by the entire group. The primary responsibilities of each collaborator (listed in alphabetical order) were as follows: Michael Canale and Sybil B. Carlson, item design for computer delivery; Lawrence T. Frase, analyses of the semantic, syntactic, and stylistic properties; Glynda Hull, semantic and case frame analyses; and William C. Ward, pattern matching analyses. Lillian Bridwell-Bowles, of the University of Minnesota, collected the data for the small pilot study. Stellan Ohlsson, of the Learning Research and Development Center at the University of Pittsburgh, reviewed and confirmed the feasibility of the case frame approach. Walter Emmerich, Mary Enright, and Juan Moran-Soto, ETS professionals, provided valuable insights.

Table of Contents

I. ISSUES CONFRONTED

In "Predicting Success in Graduate Education," Willingham (1985), poses as an important issue in improving prediction the identification of measures of higher-level cognitive abilities that are not well measured by present admission tests. He notes, however, that most machine-scorable tests measure cognitive abilities that are not greatly different from those measured by the GRE General Test, and suggests that other forms of assessment, such as free-response or work-sample formats, may show greater promise in measuring abilities that are both clearly different and useful. He concludes that the challenge in this area is to devise assessment procedures that are cost-effective with respect to admininistration and scoring.

The GRE Board has supported research on an item type that measures higher-level cognitive abilities and that utilizes a free-response format--the Formulating Hypotheses (FH) item type. A series of studies (Frederiksen & Ward, 1975; Ward & Frederiksen, 1977; Ward, Frederiksen, & Carlson, 1978; Ward, Frederiksen & Carlson, 1980) suggested the potential usefulness of this item type in predicting success in graduate education, particularly when "success" is defined by criteria other than the traditional one of first-year grade point average.

Further research on Formulating Hypotheses was not recommended to the GRE Board because of the cost and feasibility of operational use of a test composed of FH items. However, given the desirability of finding measurement approaches that may improve the prediction of success in graduate education, this line of research should not be abandoned prematurely. Instead of devoting research and development efforts toward refining the paper-and-pencil measures, these issues would be better addressed if placed in the context of computerized delivery and scoring. It is not too early to anticipate and prepare for testing conducted on the computer within the near future. A considerable number of organizations, including the U. S. armed forces and the National Board of Medical Examiners, are developing tests and work station centers that will deliver computerized standardized tests. At ETS, the College Board Computerized Placement Testing Program and the development of computerized tests for professional licensing are well underway. Testing also is moving in directions away from conventional item formats such as multiple-choice and toward the measurement of abilities and skills other than the conventional verbal and quantitative, particularly since the computer affords the capability to expand in these directions.

The following sections of the report present a review of previous research on FH items, the major issues we faced, and the rationale for the project. Members of the project staff worked with collaborators who have expertise in assessment, computer-assisted instruction, and practical applications of natural language processing. Through our investigations, we explored numerous approaches to the delivery and scoring of brief open-ended responses of the FH item type. These investigations are summarized briefly in the subsequent sections dealing with the critical issues of item design and scoring for computer delivery. Finally, we conclude the report with recommendations for the development of a prototype system based on our analyses.

Summary of Previous Research

For several years we have conducted research involving open-ended and problem-solving item formats. In research supported by the GRE Board (Frederiksen & Ward, 1975), four kinds of Scientific Thinking items were developed: Formulating Hypotheses, Evaluating Proposals, Solving Methodological Problems, and Measuring Constructs. These items were designed to elicit the types of reasoning behaviors that are applied to research problems in the graduate-level psychology curriculum and in the field of psychology. The FH item, for example, presented the results of a psychological experiment, and asked the student to list the many possible hypotheses that could explain the finding (see example in Appendix A). The responses were categorized, and several different scores were obtained. Results of the 1975 study indicated that scores based on number of responses (fluency), though highly reliable, were relatively uncorrelated with scores from the GRE General Test. Scores reflecting the quality of ideas produced on the FH test overlapped in variance with the GRE General Test verbal score, but the percent of true variance accounted for by the verbal score was less than 20%. Hence, the fluency and quality scores on the FH test represent skills and abilities that appear to be largely unmeasured by conventional test items.

In a follow-up to the original study, students who had completed the FH test at the time they took the GRE General Test reported on their experiences in their first year of graduate work in psychology. FH scores were more effective than GRE General Test scores in predicting self-reports in two areas: self-appraisals of knowledge and skills in psychology, and professional accomplishments such as research, publication, and teaching.

A subsequent construct validity study (Ward, Frederiksen, & Carlson, 1978) examined the relationships of scores on "machine-scorable" and "free-response" forms of the FH tests with GRE General Test scores, a personality inventory, and a battery of cognitive process variables. The data indicated that the free-response and machine-scorable versions of FH clearly could not be considered alternate forms of the same test, since the correlations between corresponding forms were low. To summarize the complex sets of relationships observed, the performance elicited by the free-response form of the FH test consists of more than the mere generation of random ideas that come to mind. Reasoning, the ability to think divergently, and cognitive flexibility in the context of relevant knowledge are brought to bear in the generated-response format but not in the conventional recognition-response format.

More recently, Carlson (1985) completed an exploratory investigation of the FH item format for the Law School Admission Test battery of the future, in which new FH problems designed to have face validity for law school candidates and FH problems with a psychological basis previously developed for the GRE research were combined to create a test. The data were analyzed descriptively; the responses of a small sample of students to the test items appear to reflect performance dimensions that would serve as meaningful indicators of potential success in law school. The FH item type is still being considered by the Law School Admission Council, particularly as computerized delivery and scoring become practical and feasible. Another related study, (Carlson, 1988) currently supported by the GRE Board, is exploring the identification of thinking skills exhibited by candidates in samples of their writing. The research may indicate that other variables observed in verbal production tasks may contribute richer information about the reasoning skills of GRE candidates. If this result is obtained, FH items could be adapted to incorporate these reasoning skills as well. These studies, as well as other ETS research and development activities, provide a solid basis of experience with open-ended response tests to guide the refinement and investigation of the measurement properties of an FH test for the GRE.

Powers and Enright (1986) recently conducted a GRE project to obtain information on the role of analytical abilities in graduate work. Graduate faculty in six fields of study were asked to make judgments about "(a) the importance for academic success of a wide variety of analytical skills, (b) the seriousness of various reasoning errors, and (c) the degree to which a variety of 'critical incidents' had affected their estimates of students' analytical abilities" (Powers & Enright, 1986). Data analyses yielded seven dimensions to represent clusters of reasoning skills that, on the basis of faculty responses, were differentially important for success in the different disciplines. One of the four dimensions consisted of skills involving the generation of hypotheses/alternatives/ explanations. "The ability to generate hypotheses independently was one of the incidents rated consistently as having a substantial effect on faculty perceptions of students' analytical abilities" (p. 12). Thus, the results of this research further support our exploration of the potential of computer-delivered FH items as components of some form of a GRE instrument in the future.

Factors Limiting the Use of FH Items

Two major issues were addressed in considering FH items for operational use: (1) the costs of scoring, and (2) the assignment of scores along a range of values rather than the conventional number-right scoring.

A major deterrent to the operational use of FH items is the effort required in scoring. In previous research, each protocol was scored independently by two readers, individuals with undergraduate or advanced training in fields related to the subject matter of the problems. Including time needed for quality control, the scorers spent about one hour in scoring for each hour an examinee spent in problem solving. This time expenditure has implications not only for cost but also for the feasibility of producing FH scores within a time period acceptable for preparation of score reports. With computerized delivery and scoring of responses to FH problems, these drawbacks can be dramatically minimized.

The second issue arises from the fact that problems of this sort do not have a single correct answer. Rather, there are multiple answers of various degrees of acceptability, each expressible in a variety of ways. As a consequence, judgment is involved in assigning a given response to the appropriate scoring category, and again in assigning quality values to each category. With good category lists and well-trained scorers, it is possible to achieve high reliability in these assignments, but not to secure perfect agreement. The defense of a scoring decision, therefore, cannot be made in the "absolute" terms in which it is made with items that have a single correct answer; instead, it must rest on a demonstration that the process by which scoring decisions are made is both reasonable and rational. In this respect, FH scoring is similar to that employed in the holistic evaluation of writing samples, rather than that used with multiple-choice items. Because responses to the problems can vary along several dimensions, we have not been able to articulate specific, objective criteria for making these judgments.

Throughout our explorations of applications of FH items to different contexts, however, we have obtained consistently high agreement among expert judges who assigned values to FH categories. Despite the complexity of this form of human judgment, individuals appear to perceive variations in the quality of responses along a relatively similar range of values, and thus appear to share a common scale of degrees of "goodness." The development of computer-delivered prototypes of FH-type problems will afford us the opportunity to confront the issue of multiple acceptable responses directly by (1) developing systems that recognize sentence-length open-ended responses, thus enabling us to explore the most valid and effective approaches to evaluating responses, and (2) providing us with an efficient tool for collecting data and analyzing responses, thus facilitating the design of valid psychometric approaches to test items of this kind.

Approach to the Task

The investigations we conducted were intended to address the limiting factors discussed above. The first set of factors, those associated with the time and effort required in scoring, can be addressed directly by seeking ways to increase the efficiency of scoring through computerized delivery and scoring of FH items as well as potential variations of FH items. The second, those related to the judgment involved in evaluating FH responses, can be addressed both directly and indirectly by developing procedures for the computer recognition of responses that are sufficiently reliable and well rationalized to be acceptable to reasonable evaluators.

The work focused on obtaining collaborative input from experts who are closely involved in confronting the issues involved in the computer recognition and evaluation of open-ended responses. These individuals have collaborated with us on previous and current research projects, and have been working for some time on projects that are related to our concerns, though primarily from the perspective of instruction.

Considerable research on computerized text analysis has been conducted by linguists and experts in artificial intelligence (Harris, 1985; King, 1983; Schank & Abelson, 1977). The kinds of systems they are working to develop are extremely complex, in that they are attempting to faithfully simulate human language in all the intricacies of extended prose (e.g., Hayes & Carbonell, 1984). They have designed elaborate dictionaries and parsers but do not anticipate that these efforts will have direct practical application in the near future. At the other extreme are those working with computer systems capable of matching single words or phrases to a relatively finite dictionary of terms--an overly simplistic approach to the recognition of responses that are generated in FH problems. Somewhere between the two extremes are the individuals who need to use computerized text recognition for instructional applications now, and thus are exploring accommodations to text recognition that, with relatively good reliability, can identify sentence-length responses (Hull, Ball, Fox, Levin, & McCutcheon, in press; Ross, 1986; Ross & Bridwell, 1985; Sager, 1981).

We were further encouraged to explore the feasibility of the computer delivery of FH item types by the success of computer games that employ limited forms of natural language recognition. One example of this approach is in the area of commercial microcomputer software termed "interactive fiction" (Addams, 1985). These consist of text adventures, "computer games for the literate" (e.g., Infocom's Cutthroats and The Hitchhiker's Guide to the Galaxy), with sophisticated parsers and language systems that enable the player to enter natural language commands. "As is typical with Infocom games, the vocabulary understood by the program is quite good and enhances the interaction with the story. The Infocom parser, that part of the program responsible for accepting and interpreting commands typed by the player, allows for normal sentences and ideas to be communicated to the game..." (Schulz, 1985, p. 160). Computer systems such as these appear to parallel the kinds of systems that would be appropriate for developing FH problems.

Thus, we selected collaborators for this project who are experts in their fields, particularly in areas in which they have developed computerized systems using modified versions of natural language systems. Our collaborators were Michael Canale, of the Ontario Institute for Studies in Education (OISE); Lawrence Frase, of AT&T Bell Laboratories, and Glynda Hull, of the University of California, Berkeley, and formerly of the Learning Research and Development Center (LRDC) of the University of Pittsburgh. Lillian Bridwell-Bowles, of the University of Minnesota, intended to work closely with us, but new and additional professional responsibilities prevented her from doing so. She has assisted us, however, by collecting student responses to modified FH-type items.

Our collaborative work involved intensive discussions in periodic meetings. Between meetings, the collaborators conducted data analyses to pursue explorations that grew out of the discussions as we progressed through the phases of the project. As we learned more about the complexities of the task, the final meetings involved interactions of individual collaborators with project staff that focused on specific aspects of the problem.

At the beginning of the project, all collaborators received copies of actual responses to several FH items (categorized and not categorized), in hard copy and on disks, as well as background materials and data collected in previous FH research. In the early stages of the study, discussions were based on the responses to several FH problems in order to understand the kinds of responses that students might generate. Toward the middle of the project, we agreed to conduct more extensive analyses of one relatively representative FH item, "Family Situation of Juvenile Delinquents" (Appendix A). The project staff and collaborators investigated different approaches to analyzing the FH responses with prototype computer systems (e.g., parsing and pattern matching) and with paper-and-pencil analyses that represented hypothetical but feasible computer systems (e.g., item design and delivery). Finally, we all worked together to prepare the final report with recommendations.

This exploratory project resulted in specific recommendations for developing a system to deliver computerized problems of the FH type. This exploration of potential systems capabilities also suggested further research to study potential variations in FH items using the computer prototype as a research tool. With a computer delivery system for experimental items, we will be able to investigate more efficiently the potential to vary the problems, category scoring lists, and scoring procedures. The experimental problems will be designed to extend our paper-and-pencil investigations in order to determine the parameters of the problem task that most effectively and efficiently elicit the performance we wish to assess. The eventual outcome of this subsequent work would be a presentation and rationale for the most attractive testing and scoring scheme uncovered in the investigation, along with order of magnitude estimates of the time and cost involved in its application and of the resulting reliability and generalizability of scores. Precise evaluations of these factors would be deferred to be completed in the context of the further, more formal studies that would be required prior to the introduction of a computerized form of the FH item type into the operational examination.

The following sections describe preliminary analyses of FH responses that led to the refinement of a design for a computerized scoring system for a prototype computerized scoring system (Section II), logical analyses of FH item design for computer delivery and for making scoring criteria explicit and defensible (Section III), and our recommendations (Section IV).

II. SUGGESTIONS FOR NATURAL LANGUAGE ANALYSIS OF FH RESPONSES

We investigated not only the ways in which FH items could be administered by computer, but also the feasibility of the computerized scoring of responses. This section summarizes our preliminary investigations in approaches to scoring. Our explorations drew on our experience and on suggestions from previous work in natural language analysis. Analyses using FH responses obtained in previous research enabled us to determine which approaches might be applied to scoring these kinds of responses.

One of the limitations of the FH paper-and-pencil item is that it takes human readers a considerable amount of time to score responses. For every hour of test taking, an hour of scoring is required--a ratio that is unacceptable both monetarily and in terms of turnaround time for reporting scores. Thus it would be desirable to perform the scoring operation by computer.

But here is where the difficulty begins. The characteristic of FH items that sets them apart from other measures--that students respond in natural language rather than by selecting a response from a multiple-choice format--is also the characteristic that makes machine scoring so complex. Research on natural language processing, or creating computational mechanisms for communicating through English and other human languages, engages the interest of many scientists in the artificial intelligence (AI) community. Yet the difficulties inherent in such an attempt have long been a thorn in the side of AI research. The problem, as it is currently conceived, is the relationship between real-world knowledge and natural language input. To understand unrestricted natural language input, a computer program must possess a vast amount of knowledge about the world--so great an amount, in fact, that such a representation is considered infeasible (see, for example, Winograd & Flores, 1986). Thus, instead of creating general language understanding systems, systems that could operate on unconstrained input, AI researchers have built "toy" systems to demonstrate that particular approaches to knowledge representation might be feasible. These systems work in the laboratory on carefully selected examples.

Clearly, it could not be the goal of our project to design a computer program that would genuinely "understand" the natural language responses to FH items. We did, however, make some progress in designing the specifications for a prototype that could carry out an analysis of those responses, given the fact that we already knew a great deal about the kinds of responses people were likely to make. We were encouraged in such an attempt by the progress made in what might be called "applied natural language processing," attempts to allow people to communicate with computers in natural language in restricted domains. Here the emphasis is not on making the machine understand natural language, but on making it respond helpfully to users who are engaged in a particular task, e.g., categorizing moon rocks or making airline reservations. Examples include, then, interfaces for data bases and expert systems. In applied natural language processing, the knowledge a computer program must own about language and the world is constrained.

Likewise, it is crucial to remember that computer scoring of FH responses in a way that is very useful for our task need not mean analyzing them in any complete manner, either syntactically or semantically. Although test takers will be allowed to respond to FH items in natural language, we will not have to concern ourselves with building a program that can process unrestricted input. It will not even be necessary to build a program that can completely process input in restricted knowledge domains. Our computerized analysis needs to proceed only so far as to determine which of several predetermined categories each response comes closest to fitting. Indeed, the task is simpler still, since we are ultimately interested in which of each student's responses fit into "high-quality" categories; this again reduces the distinctions the program will have to make. Thus, we are really concerned only with separating out responses that fit into "high," "middle," and "low" categories in terms of quality. Although this is not a trivial task, it appears much easier than complete natural language understanding. As we demonstrate below, certain other constraints on the task help to make it practically achievable.

Our suggestions for computer-aided analysis of FH responses take two forms. First, we have experimented on paper or online with several techniques that show promise as analytic tools for our task. These techniques vary in terms of how sophisticated an analysis they attempt and how easily they can be implemented. They also can be viewed as a continuum, with one level of analysis building upon and extending the previous. Our modus operandi would be to use as big a hammer as we needed to drive each nail. That is, if we could carry out an analysis for certain items with minimal effort, we certainly would do so, but we would also be prepared to deepen the analysis as necessary. Most of the following section is a description of these analytic techniques.

Second, we also have considered how we might constrain natural language input. Thus, we have explored alternative approaches to presenting FH type items such that we reduce the great syntactic diversity of responses. We have hoped to constrain syntactic structure without suppressing the quality and quantity of divergent responses. These possibilities are discussed in Section III, along with some initial concerns that such contraints may alter the nature of the task. Because of the complex interactions of computer scoring and FH item design, many substantive issues will need to be addressed in research employing a computerized prototype that integrates our recommendations.

Natural Language Processing

The analyses undertaken for this project can be placed in the perspective of recent research and development efforts in AI, particularly applied natural language processing. Hayes and Carbonell (1984) provide the following categories of natural language analysis systems:

1. Pattern matching (e.g., ELIZA)
2. Syntactically driven (e.g., ATNs)
3. Semantic grammars (e.g., SOPHIE)
4. Case frame instantiation (e.g., ELI)
5. Wait and see (e.g., Marcus)
6. Word expert (e.g., Small)
7. Connectionist (e.g., Small)
8. Skimming (e.g., Frump)

The first four categories represent most of the work on natural language systems. Our explorations fall in categories 1, 2, and 4. In the following sections, we first give brief overviews of these approaches to natural language analysis, then report our applications of these techniques to the FH task. The latter are relatively detailed in order to communicate the kinds of thinking that were required to study the situation in depth. The conclusion of this section summarizes our findings.

Each of the applications described focused on one FH problem, "Family Situation of Juvenile Delinquents." The problem and the categories employed in analyzing responses to it are presented in Appendix A; these materials should be reviewed prior to reading the analyses that follow.

Pattern Matching

In the kind of parsing or language analysis technique known as pattern matching, input utterances are recognized as a whole by matching them against patterns of words. The most famous system that uses this technique is ELIZA, the simulation of a Rogerian psychologist. ELIZA is actually a pattern matcher that can key on certain patterns like "X you Y me" and provide realistic responses on that basis. For example, this pattern would match on a sentence like "You don't like me" and would provide the response, "Why do you think I don't like you?" The program worked well enough to make many people believe it actually understood and responded to them.

An Attempt at Pattern Matching

Even a cursory examination of responses to FH items reveals much repetition of key concepts--words and phrases that represent the semantic heart of the responses. In category 15 of the Juvenile Delinquency item, for example, phrases that have to do with love, affection, and so forth, are key. An obvious first attempt at categorizing the responses was to develop a simple pattern matcher that could flag such phrases. Although we understood that programs that simply match single words or phrases against a dictionary of terms and phrases would not be sufficiently powerful to serve as a sole analytic technique, it would be instructive to explore how well the technique would work and where it would fail. We created a Pascal program to search for single key words or for combinations of them. The program occasionally makes use of the order in which two sets of characters appear. For example, one rule includes the specification that if the string "angry" appears in a sentence

--somewhere before "self" or "selves," the response is categorized as #31, "Child feels responsible" [a paraphrase of the category];

--if the string precedes "parent" or "family," the response is categorized as #32, "Delinquency to punish the parent";

--if it appears without either of these conditions being met, the response is categorized as #20, "Emotional problems."

Most rules are not as complex as this example, involving only the presence of a single string of characters.

No attempt was made to deal with the general categories, 1-14; these might prove very difficult because they often are vague and poorly formed, and can be very diverse in content. These categories represented criticism of the design of a study, and frequently were presented by students when they were unable to deal with the specifics of the problem in the FH item.

Several changes and elaborations were made in the category lists in order to classify the responses more accurately: a few categories in which the vocabulary and quality of ideas were similar were combined; responses were not assigned to vague and unelaborated categories until the other categories were first considered. For most of the categories, there appeared to be a few key words or combinations of words that appropriately detected many of the responses. For some categories, there were many ways in which to express a response. For other categories, the rules required only a search for a few specific words or phrases.

First approximations. The first version of the program was written using a set of 184 examples that had been chosen, independent of this effort, to represent responses to the Juvenile Delinquency problem. The program was then tested for its ability to classify those examples correctly, excluding responses that fell into one of the general categories. Results were as follows:

Responses correctly classified: 55%

Responses classified in part correctly: 27%

Responses incorrectly classified: 13%

Responses not classified: 4%

Thus, the program assigned slightly more than half of the responses to exactly the same category or categories as did human judges. Another quarter received partially correct assignments; that is, the program assigned the response to a correct category but also assigned it to one that was incorrect, or failed to assign it to a second that was also appropriate. Seventeen percent were classified wholly incorrectly or were not classified at all. Consistent with judgmental scoring, many responses were given multiple category assignments; of all assignments made, 67% were correct.

Revisions. An iterative procedure was followed in revising the program. A new set of 20 responses were classified judgmentally by a human judge, excluding responses that were fragmentary or that belonged to one of the general categories, 1-14. These responses were then classified by the program and errors were examined. Revisions were made whenever possible to deal with any errors that appeared likely to recur; no changes were made to accommodate responses judged to be idiosyncratic. Another set of 20 responses were then examined in the same way. Including the responses used in the initial development of the program, about 400 responses were employed in testing and revising it.

Cross validation. A new set of 263 responses were used for cross validation; no further changes in the program were made as these were examined. Results were as follows:

Responses correctly classified: 46%

Responses classified in part correctly: 8%

Responses incorrectly classified: 13%

Responses not classified: 19%

Responses not attempted: 15%

Fifteen percent of these responses, those that were idiosyncratic or that belonged to one of the general categories, were excluded from the analysis. Most such responses are vague or fragmentary, or represent general criticisms of the design or execution of the investigation; no automated scoring procedure is likely to be able to deal with these. Proposed changes in FH instructions, described in Section III of this report, should reduce the number of such responses that are obtained.

Overall, the program made completely correct assignments of 46% of the cross validation responses, and partly or completely wrong assignments of 21%; it failed to classify 19% in addition to the 15% that were not attempted.

It should be noted that the procedure employed has several flaws. First, the human judgments that serve as a standard against which to compare automated classification are themselves imperfect. No systematic data are available on the degree to which two judges would agree in classifying these responses, but it is unlikely to be greater than 90%. Moreover, the same individual classified the responses and wrote the computer program, which might create a bias toward resolving doubtful cases in the same way the program would have operated. (Remembering the key words employed for each category, however, is not easy; the program is relatively complex, about 600 lines in length.)

Considerations regarding categorizations of responses. Beginning with a well-developed category list for scoring the Juvenile Delinquency problem, this exercise required approximately 40 hours of effort. In an ongoing testing program, perhaps half that time could be saved by creating software utilities to facilitate rule creation and testing, and some of the work could be carried out by clerical or key entry staff.

The classification program has not reached the limit of accuracy possible, but it has probably reached a point of diminishing returns. Some improvement might be made by introducing contingencies among rules. For example, a rule such as the following might be added: "If category 23 applies, do not score also as category 22 unless the following words are present...." The gain in accuracy resulting from such changes, however, would be fairly slight.

In setting up an automated classification scheme, there is a question of trade-offs: Are misclassifications more serious than failures to classify? An example of such a choice is that of whether to assign a response to category 29, "low socioeconomic status," on the basis of the word "poor." The rule in that instance produces many correct classifications, but it also introduces errors, such as assigning "poor discipline" to category 29.

Note also that one source of difficulty in classifying these responses is that an examinee can give equivalent responses in two more or less opposite ways--"Children from broken homes lack..." or "Children from intact homes have...." If the response format were restricted to one of these, a number of ambiguities might disappear.

The general categories also pose difficulty. Perhaps the problem content of items could be constructed in such a way as to make quarrels with design and sample size less reasonable, or the instructions could indicate that the design and interpretation of the study should be assumed to be correct.

Based on these explorations, Ward estimates that the keyword approach, applied without any restriction or restructuring of the FH item type, will be limited to approximately 75-80% accuracy in categorizing approximately 75% of all responses encountered. That is not sufficient, but a further point should be considered. The objective of classifying is not to classify each response correctly, but to judge the quality of the examinee's responses. The categories were designed to facilitate human judgment; the kinds of matches that a computer recognition system might use would not necessarily parallel these classifications. Suppose, for example, the score to be derived for an individual is the number of high-quality responses given, where high quality means the response has a quality value in the upper one third of the quality values assigned to the 35 categories for the item. Looking at the last set of 50 responses studied, the program was correct 35 times in identifying a response as either high quality or not high quality, wrong 4 times, and possibly wrong 3 times. That results in a correct assignment, on the basis of good versus poor responses, of 70% of all responses, and of 83-90% of the responses that were categorized--not far below useful.

Conclusions. These explorations produced better results than had been anticipated, since we were aware that the simplest form of pattern matching was unlikely to be accurate, given the complexity of the responses to FH items. The information gained suggested that it might be possible to combine a parts-of-speech analysis with keyword matching, since programs are available for syntactic parsing. We thus needed to consider whether pattern matching would be strengthened by knowing a general word category rather than simply a literal word string. With a view toward a scoring system that might use several levels of analysis, keyword analysis might be supplemented by additional forms of analysis, until reaching a cut-off point with a high enough confidence level.

A major difficulty with pattern matching is the enormous number of patterns that must be specified, and also the impossibility of imagining every possible pattern that one might need to specify. The first difficulty can be reduced through hierarchical pattern matching, where input is gradually canonicalized through pattern matching against subphrases (see discussion in Hayes & Carbonell, 1984). Some patterns match only part of the input and replace that part with some canonical result. Then, other, higher-level patterns match on the canonical elements in a similar way. Finally, a top-level pattern matches the canonicalized input as a whole. In this way, similar parts of different utterances can be matched by the same patterns, and the total number of patterns is greatly reduced. For instance, "children from disrupted families" would be replaced by canonical notation for "children from broken homes," which also would replace surface strings like "children with one parent," "children of divorced parents," and so on.

Another higher level of analysis that we had begun to consider was that of case frame instantiation. This is described in the last summary in this section. The present analyses provided information to suggest that a case frame approach would have improved the matching of responses to categories.

Syntactic Parsing

Syntactic parsing works very differently from pattern matching, by constructing intepretations of larger groups of words contingent on the relationships between individual words and phrases. An interpretation is derived, then, by applying a grammar or set of specifications of what constitutes an acceptable sentence in a language to natural language input. Thus, for "children from broken homes lack affection," a syntactic parse would determine that the sentence consists of a noun phrase ("children from broken homes," in which "children" is the plural of "child," and in which "broken homes" is a noun phrase, "homes" is plural of "home," and "broken" is a past participle of "break") and a verb phrase ("lack affection," in which the verb is "lack" and "affection" is a noun phrase).

Educational Testing Service is exploring the use of several computer tools that can contribute to automated analysis of Formulating Hypotheses responses. A staff member, Juan Moran-Soto, has analyzed a sample of responses using Fidditch, a syntactic parser on loan from AT&T Bell Laboratories. Its failures arose most often through encountering words not stored in its dictionary, a deficiency that can be remedied by linking the parser to a large machine-readable dictionary. Moran-Soto has also, with the assistance of Carl Frederiksen of McGill University, begun working with Frederiksen's Coda Program. Coda assists a human judge in the propositional analysis of a text, resulting in a representation closely related to the case frame analysis discussed earlier. Eventually, these two tools will be tied together, so that the syntactic parse will provide the information needed to automate fully the propositional analysis. The result of that analysis will be representations in a form suitable for more complex analyses, including canonical pattern matching and matching to conceptual frames.

Lexical, Syntactic, and Stylistic Explorations

The computer is, in fact, used more and more, and with increasing sophistication, for applied linguistic analysis, especially the analysis of written text (Frase, in press; Frase & Dieli, 1986). Another of our explorations involved carrying out more than 100 computer analyses to determine what variables might be useful for automated analysis of FH free-response items, and to get a sense of the role that automated scoring might best play in the analysis of written responses. Primary software tools for our studies were the UNIX WRITER'S WORKBENCH software,* programs that exist as part of the shell language of the UNIX* operating system, and other programs that we created. The Writer's Workbench performs a syntactic analysis of written texts, assigning parts-of-speech to understand words, and uses that analysis and others to assess stylistic, lexical, and syntactic features.

Resources and limitations. We analyzed FH responses (all categories of response) in one problem domain: Family Situation of Juvenile Delinquents. The contents of different problem domains were clearly different (shown by human inspection and by a computer measure of content similarity); hence, to limit the complexity of what we studied, we concentrated our work within one content domain. We would have liked a larger sample of words in the various response categories; however, the available sample, with certain adjustments (for instance, taking samples of equal size, and sampling from within a response category as well as across response categories), could be used to detect major factors for further study.

*Trademark of AT&T

Appendix B includes detailed data from our computer analyses. The first page lists major variables and briefly describes them. The second page shows measures for six response categories, using samples of roughly equal numbers of words from the categories, with two categories broken into two subsections to provide two samples from identical categories. (Sampling within the two categories gave us a feel for variability within a category, in contrast to variability across categories.) The third page of the appendix shows measures on equal size samples from the seven best and seven poorest response categories. Good responses contained many more words than poor responses; hence, the good responses were divided into four subsections to equate roughly the total words in good and poor samples. (Total number of words can influence linguistic variables.) The remaining pages show the intercorrelations of response categories, intercorrelations of the 84 text measures, and finally the 84 measures obtained for each response category.

Analysis 1: lexical properties. One question was whether the various response categories overlap in content. This overlap seemed clear from human inspection of the responses and the computer analysis of lexical similarity (see Frase, in press, for a detailed description of the similarity measure), which involved calculating the lexical overlap for 325 pairs of categories and showed that 28% of the response categories were highly related to each other (.22 or higher on a scale extending from 0 to 1.0). Extremely high relations among categories are a sign that they might be combined. Analysis of the similarity of words among the categories shown on page two of the appendix indicated that the categories were highly related (mean=.31), and samples drawn from within categories were very highly related (.48 and .50). In other words, the measure of lexical overlap was sensitive enough to show strong relations where one might expect to find them and to suggest where categories might be combined.

What about the relations among good and poor quality response categories? The poor response categories were related to the good response categories to the same extent that the good were related to each other (mean=.37, in both cases). We also looked at the lexical overlap of response categories after the words used in the problem statement had been removed. (This adjustment reduced the relatedness among categories by 35%.) After eliminating problem statement words, the average relation between poor quality responses and good responses was .24; among good responses the relation averaged .23. Good and poor responses were thus equally and substantially related among themselves. Therefore, we concluded that the individual content words of a response are not an adequate basis for distinguishing its quality. One must measure more. Our conclusion is not surprising, since research has shown that students use vocabulary they have encountered in a course well before they understand the content.

The vocabulary size (number of unique words exclusive of function words and words used in the problem statement) averaged 150 words for the good quality responses and 152 words for the poor responses--a close correspondence. (If the problem statement words are included in the measure of vocabulary size, those totals increase by 15 words.)

There were only 53 content words in the problem statement; hence, it is surprising that the overlap between words in the problem statement and student responses was high (mean=.24). However, the good and poor responses did not differ in overlap with the problem statement. The means were .23 and .25, respectively. We further explored the relation between occurrence of words in a sample and response quality. Two different measures were used. The "template" approach was one in which a set of problem-statement words was matched against the response sample, yielding a similarity score based on number of template words found in the sample (repetitions were not counted). The "distribution" approach was also one in which a set of problem-statement words were matched against the response sample; however, the template contained repeated words in the frequency in which they occurred in a "good sample" of responses. Repetitions of words were allowed in the distribution approach. With the template approach, the correlation between the similarity measure and quality was .51. With the distribution approach, the correlation was .45. The similarity measure correlated .95 with size of the word sample in the distribution approach, whereas the similarity measure correlated .26 with sample size in the template approach. In other words, it was possible to obtain higher relations between the similarity measure and criterion scores (quality ratings) by making similarity less dependent on sample size.

Analysis 2: syntactic properties. We looked at the diversity of grammatical sequences (parts of speech bigrams--e.g., adjective-noun, adjective-adjective) in the good and poor responses, but found no differences, although there was high overlap among all response categories (mean overlap on a scale of 0 to 1.0 was .58). In short, the grammatical sequences in good and poor responses were very similar. Thus, we found no evidence that higher-level (sequential) grammatical structures play a role in response quality.

Another analysis concentrated on the verbs used in good and poor response categories, because the verb is the occasion for different sentence frames. The response categories varied widely in verb similarity, but the good and poor response categories did not differ systematically.

Most poor responses consisted of simple sentences (variable 25 in the appendix). No compound-complex sentences were found among the poor responses, while all good response samples contained them (one good response sample had 25% compound-complex sentences). Grammatical complexity, at the sentence level, was clearly one component of a good response.

Analysis 3: stylistic properties. A major difference between good and poor responses was in the amount written. The total sample of poor responses contained 351 content words, while that of good responses contained of 1,540 content words.

Generally, the good responses were more complex. This is suggested by the higher average readability for the good responses (13th grade level versus 11th grade level for the poor responses). As the appendix shows, the sentences of poor responses were shorter than those of good responses (variable 8). And, as has been mentioned, most poor responses consisted of simple sentences (variable 25). We compared the length of each sentence (in words) with the response quality sample (good or bad) from which the sentence was drawn. The biserial correlation between sentence length and quality varied from .09 to .30 in the different samples. These data suggest that sentence length has a slight relation to response quality.

In addition, word length did not differ for good and poor response categories; thus, "simplicity" of response was a property of sentences and not of the vocabulary.

Several variables showed less correlation with quality of response then we had expected. These included the ratio of verbs to adjectives (variable 79) and the average length (in characters) of meaningful word groups (variable 82). (See Frase, Macdonald, & Keenan, 1985, for a description of the program that determines meaningful word groups.)

The data collected in our project show that single predictors of response quality contribute only part of the picture of hypothesis formulation. The data are complex, but they come into focus when we consider the criterion--quality of response. The correlations between response quality and the computer measures that we collected for all response categories help answer the question, "What combination of measures predicts the criterion?" Confirming our previous statements, we see that the number of simple sentences predicted quality ($r= -.58$). Complex explanations leave little room for simple sentences. Percentages of conjunctions and adverbs were positively related to response quality ($r= .55$ and .52, respectively), while the percentage of faulty phrases and the diversity of content words were negatively related to response quality ($r= -.50$ for both). Some of these relationships were influenced by the size of the word sample; for instance, a partial correlation between faulty phrases and quality reduced the relation from -.50 to -.32. The type-token ratio, however, increased from -.50 to -.52 with effects of length removed.

What picture emerges from all this? If we were to tell someone how to write a high-quality response, we could say the following based on the data described above: express your ideas in complex sentences (sentence complexity), stick to the point (type-token ratio), and avoid trite and awkward phrases (faulty diction). This may not completely describe what we want in a high quality response, but those elements of style certainly go along with good thinking.

Conclusions. The results show that lexical, syntactic, and stylistic analyses of responses, although not enough to provide a complete picture of the quality of written responses, can be used to improve response measurement and to simplify scoring. For instance, analyses suggest where response categories might be combined. In addition, good responses contain complex syntactic structures, at least at the sentence level. Perhaps this feature of good writing could be used to select automatically certain texts for further analysis.

Measures of content similarity could also be used to detect individual differences in the ability to match what one writes to a problem statement. Our data did not allow us to study this possibility in depth; however, they show that lexical matching can be used to measure similarities among written samples. Hence, content similarity should be useful for studying individual differences in ability to respond to content domains.

Semantic Analysis

It is clear from our work on keyword matching and surface features of style, lexicon, and syntax that a deeper level of analysis that captures semantic relationships is necessary to categorize FH responses. As one of the major recent developments in natural language processing, case frame parsing offers a starting point for thinking about how to accomplish this kind of analysis.

The modern notion of "case" in AI and linguistics is similar to the notion of "case" in traditional grammar. That is, the case of a noun in Latin (and Old English) was indicated by an inflectional ending, and this ending indicated how the noun functioned in the sentence--for instance, a subject or object. In modern English, case is indicated not primarily by word endings, but by word order and by prepositions preceding a noun.

Charles Fillmore (1968) introduced case frame grammar. His notion was that a proposition in a simple sentence has a deep structure that consists of a verb and one or more noun phrases, each associated with the verb in a particular semantic-syntactic relationship (a case). For example, in the sentence, "John opened the door with the key," "John" is the AGENT of the verb, "opened," "the door" is the OBJECT, and "the key" is the INSTRUMENT. For the sentence, "The door was opened by John with the key," the case assignments would be the same even though the surface structure is different. Verbs are classified according to the cases that can occur with them. For example, "open" must have an OBJECT, and it may also take an INSTRUMENT and an AGENT. The cases for any particular verb comprise what is called a case frame. Fillmore (1968) proposed the following cases:

Agent	the instigator of the event
Counter-Agent	the force or resistance against which the action is carried out
Object	the entity that moves or changes or whose position or existence is in consideration
Result	the entity that comes into existence as a result of the action
Instrument	the stimulus or immediate physical cause of an event
Source	the place from which something moves
Goal	the place to which something moves
Experiencer	the entity that receives or accepts or experiences or undergoes the effect of an action

It is important to note that the relations between the case frame head (the verb) and the individual cases are defined semantically, not syntactically, and that each case frame requires some cases, allows others optionally, and forbids others.

The notion of case frames has been used in natural language understanding. According to Hayes and Carbonell (1984), the key advantage of this approach is that it combines a bottom-up identification of structuring constituents with a top-down instantiation of less structured, more complex constituents. Case frames, as used in parsing, actually consist of more than a predicate and a collection of cases. Each case also consists of a positional or lexical marker. A positional marker indicates that the case filler is preceded by a marker word, usually a preposition in the surface string. In case frame grammar, verbs are classified according to the cases that can occur with them. Case frame parsing proceeds by first looking for the verb in a sentence, then retrieving the case frame associated with that verb, and then by attempting to recognize each expected case by relying on lexical and positional markings.

A further development in case frame parsing is conceptual dependency theory (Schank and Abelson, 1977), which provided the rationale for grouping together the actions of several surface representations for verbs into primitive actions. Thus, the sentences "John gave Mary a ball" and "Mary took the ball from John," while differing syntactically in terms of case frame instantiation and verb choice, nonetheless are similar in terms of the action each sentence expresses--what Schank calls ATRANS, or the transfer of possession, control, or ownership. Thus, there exists a means of representing the semantic information derived from a case parse in a canonical form.

Certain parts of the case frame approach seemed useful in our computer analysis task. First is the idea of relying upon verbs to provide a set of expectations about what the rest of a proposition will look like. Keyword matching, you will recall, relies mostly upon nouns and adjectives. By letting verbs drive the analysis for this approach, and nouns show the way in the keyword matching, we could take advantage of as many lexical and semantic cues as possible. Second, the dependency relationships that are set up by the verb in case frame analysis might provide the necessary information to avoid false positives in some of the categories that proved nettlesome for keyword matching. And, finally, the concept of semantic primitives could provide a means of canonicalizing our pattern matching.

Conceptual Frames

Drawing on computational techniques like pattern matching, syntactic parsing, and case frame instantiation, and combining that information with an analysis of actual responses to the Juvenile Delinquency FH item, we have converged on some general specifications for an approach to analyzing FH responses. We believe that this approach, which we call "conceptual frames," would increase the accuracy of what can be accomplished with keyword matching and lexical/style analysis. As we illustrate below, we have worked through examples of the analysis by hand and are sufficiently encouraged by this piloting to propose it as a promising technique. However, the only real test of the approach would be a computer implementation of it.

It is important to make clear that the feasibility of our attempt to analyze natural language depends upon our knowing a great deal about the task and possible responses to it. We began, then, by looking closely at the categories that had been derived for a given FH item, Juvenile Delinquency, and a set of natural language responses to that item. Specifically, our procedure was to identify the concepts that made up each category for that item and the relationships among the concepts in actual responses assigned to the categories, drawing on the analytic techniques described above. The following examples give the flavor of this kind of analysis.

When we analyzed the categories and responses for Juvenile Delinquency, we identified three broad kinds of concepts or lexical items. Most obvious were the nouns and noun phrases that we used in keyword matching--the various surface forms of such concepts as affection, attention, cruelty, boredom, stability, supervision, socialization, role models, peer influence, socioeconomic status. Such concepts represent the semantic core of various categories and can be used to distinguish one category from another, as illustrated above, with a fair degree of success. Second, the categories and responses also share certain generic concepts: children, parents, traditional home, broken home, causality, delinquency, more/less. These generic concepts are in large part supplied by the test item and do not by themselves distinguish one response from another.

Third, the categories and responses contain certain concepts that are actions or states and often take the form of verbs expressing concepts, such as: provide, need, receive, possess, experience, examine. Unlike the generic concepts or the category-specific keywords, these concepts, because they are verbs, determine certain semantic relationships, such as the necessity for an AGENT or an OBJECT in a sentence. For our own FH analysis task, we also can specify, in addition to the cases that are allowed given a particular verb, the particular filler that is expected for those cases, relying on our specification of generic concepts and category-specific keywords. This will be made clearer with the following specific example.

> The concept "receive" appears in several categories, but it plays a particularly important role in category 15: "Children from broken homes lack affection (care, warmth)." In fact, we can derive from this concept all the other concepts and relationships among them that we need in order to analyze all the responses to category 15. Beginning with the cases that belong to "receive," we see that it requires cases called an EXPERIENCER and an OBJECT. More for our purposes, the EXPERIENCER slot for category 15 must be some variant of the phrase "children from intact families," and the OBJECT slot must be filled with "love, affection, warmth," etc.

(Notice the same would be true for synonyms of "receive," like "get".) In its negative form, "don't receive," "don't get," or "lack," the filler for the EXPERIENCER slot changes to some variant of "children from broken homes." (Note that the negation could occur in the OBJECT slot, as in "less love and affection.") The antonym of "receive," which is "provide," would require as its AGENT some variant of "parents of intact families" and, as its EXPERIENCER, "children," and, as its OBJECT, "love," "affection," "warmth," etc. The negation of "provide," which could be "don't provide" or "fails to provide," would require as its AGENT the phrase "parents of broken homes."

With such information, we can categorize sentences like these as being instances of category 15:

> Children from disrupted families receive much less love and attention than those with two parents, and therefore resort to delinquency for attention.
>
> The husband-wife family provides an environment where there is care, love, supervision, and guidance, resulting in less delinquency.
>
> The traditional family situation provides love and stability, and the child grows up to be responsible.

Part of the same knowledge base could be used to analyze these responses from other categories:

> Category 19: Disrupted families provide a less stable environment in which to grow up, and thus a child is more prone to delinquent behavior.
>
> Category 21: The husband-wife family provides a better atmosphere for a child to grow and develop a positive self-image.
>
> Category 23: A traditional family--one containing a mother and a father--provides the setting most reflective of our societal values, and, therefore, children raised in such families are less apt to violate social norms.

That is, the verb "provide" or its synonyms (or antonyms in negations) appears in various responses for different categories, but for different categories the verb would activate different fillers in the SUBJECT and OBJECT slots.

A computerized analysis of FH items would begin, then, with a linguistic analysis of the kind we have illustrated above. This analysis would identify the primary concepts that make up each category, the semantic relationships among them, and syntactic signals for the concepts and the relationships. More specifically, as we have initially imagined the analysis, the basic tasks to be accomplished in the scoring of FH responses are these:

1. We need first to sort through the natural language input and identify the main verb of a given response.

2. That verb would then be mapped canonically onto a verb category or frame.

3. The verb frame would activate concept clusters attached to various categories.

4. The clusters, which would have tests associated with them, would be run to predict what the fillers of the verb should be, given a particular category.

5. The predicted fillers would be matched against the natural language input.

6. If a match occurred for a given category or categories, fine; if not, attempted matches would be made for predicted fillers for other possible categories.

To give these steps some practical force, the next example illustrates how the steps might be followed with the paraphrase of category 15, "children from broken homes lack affection." (This rough hypothetical example is but one illustration of how the above steps could be played out computationally. We do not offer it as an actual procedure to be followed, but as a mock up of the sorts of procedures involved in this kind of analysis.)

Let us imagine that we first perform a syntactic parse, looking up each word in our lexicon and applying rules of syntax that allow us to build noun phrases and verb phrases. Our lexicon will have entries like the following:

```
(GenericConcept(NAME "child")(CLASS NOUN)(NUMBER
SINGULAR)(TYPE ANIMATE))
```

that allow us to identify particular words. Our grammar will have definitions of noun phrases like

```
NP=(PREP)(DET)ADJ*N*N (S|NP)*
```

that allow us to group words into phrases. We learn from this parse that the main verb of the sentence is "lack."

Lack will have several sets of information and actions attached to it.

One set of information will associate lack with the verb type "DON'T HAVE," containing all the surface forms of the concept of not possessing yet needing, including lack, need, don't have, want, and others.

Another set of information will associate lack with the verb type DON'T HAVE with FH categories in which this verb type is expected to appear--in this instance, categories 15, 19, 23, 21.

A third set of information associated with type would supply the cases that the verb type allows or requires. In this instance, the DON'T HAVE verb type requires an EXPERIENCER and an OBJECT.

A fourth set of information associated with the verb type, as it is instantiated in a particular category, is expected surface fillers for the slots EXPERIENCER and OBJECT. In this case, we would expect some form of the Generic Concept Childword (kids, child, young people, teenagers, etc.) to fill the EXPERIENCER slot, and some form of a Specific Concept Loveword (like love, fondness, affection, care) to fill the OBJECT slot.

We can represent the information and actions this way:

```
(DON'T HAVE (CAT 15 (EXPERIENCER (GenericConcept (CLASS
   CHILDWORD))
   (OBJECT (SpecificConcept (NAME LOVEWORD)))
```

Action: If there is an animate noun that is a CHILDWORD in front of the DON'T HAVE structure, put it in the EXPERIENCER slot.

Action: If there is an inanimate noun following the DON'T HAVE structure that is a LOVEWORD, put it in the OBJECT slot.

If this frame can be constructed, it is a match for category 15; if not, we can go next to another category where the concept DON'T HAVE is expected.

In addition, this approach would need to be flexible, in order to handle other features of FH responses:

- o It must be able to ignore parts of a sentence.
- o It must be able to work with whatever information is available. For example, if a verb cannot be

identified, the search must begin with noun concept clusters.

- Possibly the most challenging analysis problem is that of dealing with the paraphrase, although this challenge may be partially met by the conceptual frame approach. Occasionally a student will repeat a concept, using similar but not identical wording and syntax.

  Moreover, the paraphrase problem involves negations and oppositions like, "Children from disrupted families receive less attention," and "Children from traditional families receive more attention." It might help to constrain syntactic choices somewhat. In the Juvenile Delinquency item, for example, we might limit sentence beginnings to four phrases: "children from broken homes," "children from traditional homes," "parents from broken homes," "parents from traditional homes." These constraints might produce some distorted syntax, but might simplify the analysis task. We would not, then, have to deal with sentences like, "Two is better than one in terms of sharing responsibility--more time to spend with the child." We also might consider limiting responses to one sentence. The responses that extend beyond one sentence appear to be unnecessarily verbose and sometimes include a restatement of the findings. (Our experimentation with some of these options is presented in Section III).

- Each category will draw on core concepts. For the Juvenile Delinquency item, those concepts will obviously include "children from broken homes," "children from traditional homes," and the corresponding parent categories. The same will be true for every category. This can be facilitated by constraining sentence beginnings to core concepts, as mentioned above.

- Another element of the problem of doing a semantic analysis of the categories is that the separate categories include too many concepts. For example, category 15 includes "lack of affection" and also "lack of parental time." For purposes of analysis, these might be treated as separate categories that can be combined, if appropriate, when computing quality scores. A formal semantic analysis might, in fact, make the categories less nebulous.

We have outlined here in broad strokes an approach to analyzing FH responses that emphasizes semantic relationships rather than syntactic structures or lexical items. We believe that the FH responses require a semantic analysis; more superficial treatements of surface features of style or syntax will not suffice. This is not to suggest, however, that stylistic or syntactic or lexical analyses are not helpful or will not be of some use. Ideally, it would be possible to draw on several sources of information in determining category assignments for FH responses.

It may be obvious from the above examples that the approach we are suggesting here can be used to categorize, in a more complex computational fashion, many of the same responses that could be categorized more simply using a keyword/string matching approach. The interesting and important question is how successfully the approach can extend what can be accomplished through keyword matching. Can it, that is, eliminate or decrease false positive rates, and can it increase the number of category assignments? It seems to us, based on our paper-and-pencil tests, that the answer is yes. But again, the only way to confirm this conclusion is to implement the approach with computer tools. We should make clear, however, that there will be some percentage of responses that this or any other analysis tool will fail to categorize or will categorize improperly; the richness and variety of language and the interpretive skills of human readers will assure that.

Summary and Conclusions

Drawing on natural language processing research, we have experimented, with paper and pencil and online, with techniques for analyzing FH responses.

> Pattern matching. The simplest form of pattern matching produced better results than were anticipated but was not sufficiently accurate, given the complexity of the responses to FH items. The information gained suggested that it might be possible to combine a parts-of-speech analysis with keyword matching, since programs are available for syntactic parsing. With a view toward a scoring system that might use several levels of analysis, keyword analysis might be supplemented by additional forms of analysis, until a cut-off point with a high enough confidence level was reached.
>
> A major difficulty with pattern matching is the enormous number of patterns that must be specified, and also the impossibility of imagining every possible pattern that

one might need to specify. The first difficulty can be reduced through hierarchical pattern matching, where input is gradually canonicalized through pattern matching against subphrases. Some patterns match only part of the input and replace that part with some canonical result. Then other, higher-level patterns match on the canonical elements in a similar way. Finally, a top-level pattern matches the canonicalized input as a whole. In this way, similar parts of different utterances can be matched by the same patterns, and the total number of patterns is greatly reduced.

Lexical, syntactic, and stylistic properties. The results show that lexical, syntactic, and stylistic analyses of responses, although they cannot provide a complete picture of the quality of written responses, can be used to improve response measurement and to simplify scoring. For instance, analyses suggest where response categories might be combined. In addition, good responses contain complex syntactic structures, at least at the sentence level. Perhaps this feature of good writing could be used to select automatically certain texts for further analysis.

Measures of content similarity could also be used to detect individual differences in the ability to match what one writes to a problem statement. Our data did not allow us to study this possibility in depth; however, they show that lexical matching can be used to measure similarities among written samples. Hence, content similarity should be useful for studying individual differences in ability to respond to content domains.

Case frame analysis. Certain parts of the case frame approach seemed useful in our computer analysis task. First is the idea of relying upon verbs to provide a set of expectations about what the rest of a proposition will look like (keyword matching relies mostly upon nouns and adjectives). By letting verbs drive the analysis for this approach, and nouns show the way in the keyword matching, we could take advantage of as many lexical cues as possible. Second, the dependency relationships that are set up by the verb in case frame analysis might provide the necessary information to avoid false positives in some of the categories that proved nettlesome for keyword matching. And, finally, the concept of semantic primitives could provide a means of canonicalizing our pattern matching.

Finally, we have experimented with a deeper kind of analysis, one that depends not solely on keyword strings or lexical and syntactic analyses, but also on features or semantic relationships. The problem with surface analyses of style and string matching is that they both have strong inherent limitations such that, beyond a certain point, they cannot be improved. While we recognize that surface-level analyses are efficient computationally and cost-effective, too, and that such approaches can certainly take us part of the distance, we were aware from previous research (Hull et al., in press), and learned from our own experiments with actual responses to FH items, that such surface approaches will need to be supplemented.

We believe it is necessary to consider another sort of system, one that does not have such strong inherent limitations, that can be upgraded and improved upon, and that performs semantic analyses as well as syntactic analyses. To begin thinking about the design of such a system, we surveyed the computational techniques that are available for natural language analysis and juxtaposed them to sample responses on the FH task. We did not expect to find a particular technique that we could import wholesale to solve our computational problem. Rather, we hoped to combine the strengths of whatever parsing strategies seemed useful into a single system.

The technique we propose, which we call conceptual frames, begins with a linguistic analysis of the concepts and relationships that make up the semantic heart of the FH categories. The results of such an analysis then serve as predictors and constraints on our computational techniques, which combine features of case frame parsing and conceptual dependency theory.

III. ITEM DESIGN FOR COMPUTER DELIVERY

This section addresses the issue of creating a defensible scoring system in which scores are assigned along a range of values as an alternative to the conventional number-right scoring system. To create such a system, we will need to demonstrate that the process by which scoring decisions are made is reasonable and rational by articulating specific, objective criteria for making these judgments. This can be accomplished by developing an accurate computerized scoring system such as that described in Section II. However, because the scoring system is dependent on the responses elicited by an FH task, we also need to demonstrate the validity of the values assigned to these responses. Thus, we will need to design FH-type items to obtain samples of performance that elicit rich and productive responses, and that are representative of what we intend to measure.

In planning for the delivery of Formulating Hypotheses item types on the computer, we need to seek optimal solutions that will balance the need to (1) provide realistic conditions under which examinees can generate hypotheses and (2) obtain more efficient computerized scoring of the responses. In order to design efficient scoring systems, the problems should be constrained so that they can be presented effectively on the computer screen, and will elicit a range of hypotheses that are neither too broad nor too narrow. Because this system creates conditions that differ considerably from paper-and-pencil testing, we need to determine to what extent these delivery conditions may influence hypothesis-generating performance.

The Formulating Hypotheses problems, when delivered as paper-and-pencil instruments, have provided information about an aspect of the problem-solving process that is not obtained when the test material is presented in a multiple-choice format. The earlier research demonstrated that these problems assessed an examinee's ability to generate ideas (fluency) as well as to generate ideas of high quality. To elicit these performances in a computer format, we agreed that the problems could be constrained generally in the following ways:

- By choosing topics that generate 10-12 good, more narrowly defined hypothesis categories
- By the computer recognition and scoring of only the good ideas, ignoring the poor ideas
- By instructing the examinee to eliminate hypotheses regarding flawed methodology or design, unless the methodology or design are the focus of the problem (The original FH categories included a list of general categories related to these criticisms).

- By trying out parsing procedures that identify conceptually rich responses to the problems
- By trying out different variations of problem situations to extend the potential of this item type
- By focusing on the products; we could then understand more about the process of hypothesis generation as a measurement construct within the context of GRE candidate performance

It follows that the next step in the subsequent project would be to experiment with these different approaches to constraining the problems in order to optimize scoring efficiency and elicit the kinds of performance that are valuable skills to be demonstrated by a candidate when generating hypotheses during problem solving.

At the outset, it was necessary to describe the nature of the process of hypothesis formation, then to translate this process into practical testing conditions that would permit us to observe and evaluate the products of hypothesis generation. The following sections discuss (1) the process of hypothesis formation, (2) the characteristics of the original FH problems, (3) possible characteristics of item formats for computer delivery, (4) the tools of computer delivery necessary to hypothesis generation, and (5) possible sources of item content. These considerations will be essential for structuring studies to support inferences regarding the validity of FH testing.

The Nature of Hypothesis Formation

Hypothesis formation is a cognitive process that involves the generation and manipulation of representations of a problem. This process contributes to both the early stages and later stages of problem solving when it is necessary to refine preliminary hypotheses and to generate new hypotheses in the light of new or additional information. Hypothesis formation requires divergent thinking skills. Fluency in divergent thinking influences the generation of different ideas without evaluation. When ideas are evaluated, then refined, by the problem solver, the quality of the ideas emerges from the process.

During hypothesis generation, the individual formulates and manipulates representations of the data that are presented in the problem. The fluency and quality of these ideas are influenced by many factors: domain knowledge (in the form of existing representations), interest preferences, psychosocial content, response style, learning style, and, in computer delivery, interfacing with the computer.

Furthermore, fluency involves the processes of "playing around" with, interrelating, and classifying ideas (e.g., as relevant vs. irrelevant, general vs. specific). Ideational quality involves several processes:

- Clarifying the language in which the ideas are expressed
- Expanding on ideas
- Eliminating ideas
- Combining ideas
- Placing ideas into hierarchical and other relational structures
- Moving between analysis and synthesis

According to Hildyard and Olson (1978), as the individual generates ideas about a problem, the inferences that might be drawn can be derived (1) from logical propositions inherent in the information presented by the problem; (2) by imposing general knowledge-based organization or schemata on the information; or (3) by bringing personal knowledge and experience to bear on the problem. The free-form hypotheses that could be generated are likely to be increasingly more difficult, in that order, for the computer scoring algorithm to match for evaluation.

The following discussions summarize our preliminary efforts to begin to articulate test specifications that would serve to make explicit our expectations for performance on FH tasks.

Characteristics of the Original FH Problems

The FH problems can be described as having the following characteristics:

- The problem situation presents data that are sufficiently ambiguous (ill-structured) to elicit divergent responses.

- The data in the problem situations can represent different kinds of logical relationships (e.g., cause-effect, associative, hierarchical) that should be sampled in any set of problems in a testing situation.

- The subject matter of the problems can be discipline specific, and are naturally problem specific. For the GRE candidate population, problem content should provide information that does not require specialized knowledge or experience.

- Data or findings in the problems should sometimes fit and sometimes not fit common beliefs and expectations.

- The FH problems need to be sufficiently engaging to encourage examinees to demonstrate their optimal skills of ideational fluency and quality.

- Length and quantity of data in a problem should be minimized to labels reflecting authentic yet constrained presentations of problems.

Possible Characteristics of Computer Delivered FH Items

Several concerns need to be addressed/incorporated:

- Provide examinees with an opportunity to "warm up"--to become accustomed to the computer environment, to the types of tasks, to the language of the tasks, and to expectations for performance on the tasks.

  Consider allowing examinees to select the first task from a thematic list of the set of tasks, in order to begin with a more familiar topic and have an overview of the topics being covered, and possibly conclude with a more familiar topic.

- The tasks should be problem situations that can elicit a candidate's best performance.

  Consider using probes, at least at the item development stage, to confirm what the candidate cannot do, to confirm that performance is not being underestimated.

- Take into account the influences of local interdependence of a sequence of problem situations or of sequential presentation of information within a single problem, and assumptions of unidimensionality. As the field is beginning to explore new psychometric approaches to test items that are not appropriate to classical test theory, we also will need to investigate the consequences of local interdependence and multidimensionality in this context of performance.

- The question posed by the problem could provide a prompt that suggests the structure of the responses (a phrase with which each hypothesis should begin), to facilitate computer recognition.

- Investigate the potential impact of computer delivery and restructured FH problems on minority populations.

Computer Tools Necessary to Enable Hypothesis Generation

The delivery of FH problems by computer will require attention to computer capabilities that will enable the examinee to generate hypotheses naturally and easily:

- The capability to represent ideas in language and/or graphic form by entering them as they freely emerge (mental vs. visual representations), especially at the stage of "playing around" with ideas, where fluency is of concern.
- Editing capabilities that permit the manipulation of blocks of ideas in order to transform an initial, less well-formed list of ideas
- Prompts or help that can be interposed, primarily during experimentation with item designs, that attempt to elicit optimal (and machine readable) performance (e.g., explanation of a key term, or suggestion to rule out overly general hypotheses)
- Computer manipulation devices such as PageUp, PageDown, Undo, windows or split screens, highlighting (providing ability to review the problem, ideas previously generated)
- Instructions preceding the specific problems that assist the examinee to become familiar with computer capabilities and the nature of the task of typing in responses

Formulating Hypotheses Item Designs to be Explored

With the previously developed perspectives, several approaches to the design of FH items for computer delivery could be investigated. Some of these designs can be explored in paper and pencil format, others can be presented by computer. The investigations would involve small-scale studies intended to provide additional information about the nature of hypothesis generation and the kinds of tasks that would elicit optimal candidate performance.

The FH items should be constrained for three reasons: (1) to focus the examinee so he or she can provide the opportunity to produce more quality (and more specific) responses, (2) to test taker from a content domain in order to eliminate the effects of prior knowledge and experience by encouraging a more generic approach to a problem, and (3) to simplify computer recognition. The FH items can be constrained in two ways: (1) refining the design of the original items and instructions and (2) using an modified item format.

Refining the Design of FH Items

In general, FH items can be constrained by directing attention to the design elements of the item format and by implementing specific strategies for controlling and manipulating the task demands of the problem presented in the item. These item elements and strategies follow (example in Appendix C):

Item Elements

Test title. The test would more appropriately be named "Generating Hypotheses" than "Formulating Hypotheses."

General instructions. The overall instructions would briefly describe the purpose and nature of ability being measured. A sample problem and sample responses would provide an example of the task and expectations for the responses (e.g., the number of responses, responses of high quality, responses with varying length and syntax). Additional comments following the sample would provide further focus: the list of responses is not exhaustive, the methodology and data are not flawed, the sample responses reflect certain major themes (metacategories).

Framing of the problem. Each problem should have a title that focuses the context. The content of the problem should present as realistic a situation as possible, but should avoid undue specificity (e.g., proper names and dates that would suggest that specific knowledge and experience might be advantageous to idea generation). On the other hand, the content should not be overly general or fictitious. Data would be presented in numerical, tabular, or verbal form. The data should be realistic, and should not be any more complex than the finding. The statement of the finding and the prompt (indicating the form of the response within the question the examinee should address) also should provide focus.

Strategies for Constraining the Task Demands

Language. The vocabulary of the problem and the sentence structure of the findings should convey expectations about the form and breadth of the responses. An alternative strategy to be explored would involve presenting the structure for one sample hypothesis (e.g., "One factor could be ______. What other factors?" [fill in the blank]).

Metacategories. Focus the responses by suggesting metacategories that still allow for a sufficient number of quality responses. For example, for the original Aggressive TV FH problem:

> "Think about the effects of the content of TV programs on the behavior of young children as you formulate your hypotheses."
>
> or "Some of the factors that influenced the finding might involve the content of the TV programs. What other factors might have contributed to this outcome?"
>
> (Metacategories of responses to the Aggressive TV problem include TV programming, the group studied, TV vs. other activities, child development and behavior.)

Identifying the metacategories of possible responses for the different FH-type problems also would be useful for content validity--an approach to representing the domain of hypotheses/ideas for a set of problems that would make up one test form.

Logical relationships. Different logical relationships connect the findings with the hypotheses. For the original FH problems, for example, prompts could be used to focus on the structure of these logical relationships and thus influence the structure of the responses:

> "What are the most direct causes for the observed effect?"
>
> "What psychological and social effects might be associated with TV viewing in general?"
>
> "What other factors (other than TV viewing) might have influenced the results?"
>
> "What other factors (other than TV viewing) might have caused a preschool child to behave aggressively?"

Types of responses. Different probing strategies should be explored to determine how the structure of the response could be constrained without unnaturally limiting hypothesis generation. These strategies are described in the following section.

Exploring Modified Item Formats

Before we can determine which item format would serve as an optimal design for hypothesis-generation items, some small-scale exploratory studies that address specific research questions should be conducted. The following were the possible formats that we considered.

Basic FH formats

- An FH problem that would prompt the examinee to generate hypotheses beginning with one introductory phrase (e.g., "Children from intact families....")

  Question: To what extent does this format facilitate computer scoring but suppress fluency and originality?

- An FH problem that would prompt the examinee to generate hypotheses beginning with any or all of a set of possible introductory statements (e.g., "Children from intact families..., "Children from broken homes....")

  Question: To what extent does this format facilitate computer scoring but suppress fluency and originality?

- An FH problem that offers one hypothesis. This hypothesis would be presented as a complete statement, then as a phrase that captures the theme of the idea. Examinees would be asked to list other ideas as phrases.

  Question: To what extent does this presentation facilitate computer scoring but possibly induce an artificial approach to the expression of ideas?

FH Item in Steps

- A sequential problem that provides additional information in stages. We rejected this approach because it is likely to yield less information per unit of testing time than would multiple "one shot" problems.

- A sequential problem that provides additional data from different perspectives (e.g., the data are plotted differently). This was also rejected because examinees might perceive the problem as a test-developer "trick."

- A sequential problem that provides all information in the first step. The next step, however, suggests that the examinee focus on a specific aspect of the <u>data</u> that might be overlooked.

Successive Probes

Problems with probes would be used in the item design process to determine whether, given another opportunity, the examinee could increase ideational fluency or quality.

With an additional prompt. After generating responses, the examinee is given a second opportunity to generate more or higher-quality ideas after receiving information that provides focus. The examinee then is asked to add to or revise (or make more specific or general) the original list of responses. (This prompt is more directed than in a sequential problem design.)

Question: Given the opportunity, can the examinee generate more and better ideas?

With cueing. After generating responses, the examinee can select further ideas from a keylist of phrases that represent a broader range of different ideas (good to poor), then state each idea as a hypothesis.

Question: To what extent does cueing influence production in generating hypotheses? If the examinee can recognize and state another idea, is it possible that he or she needed assistance in stating an idea that was difficult to put into words? (Consider issues of test wiseness, learning effect, coaching to learn the strategy.)

With confirmation. After the examinee generates responses, the computer matches each response to one or more categories, then presents the possible restatements. The examinee is asked to confirm which idea is the closest in meaning to the idea that was generated (or "none of the above"). (Asking the examinee to restate was ruled out, since examinees should not be penalized for language problems or perceive that responses are "right" or "wrong".)

Question: What can we learn about hypothesis generation with and without specific prompts by observing the discrepancy between an original (unprompted) hypothesis and a hypothesis the examinee selects?

Exploring Different Forms of Responses

In addition to the variations in item formats, examinees might be asked to generate ideas in one of several different forms:

- Hypotheses
- Recommendations
- Nonverbal manipulation of problem materials (deductive reasoning?)
- Predictions of outcomes, conclusions (A basic FH problem might eliminate the statement of the finding, where several findings are possible.)
- Reasons/criteria for making a decision/selection
- Views/premises held by the author/investigator of the problem
- Application of the features of a construct presented (e.g., "leadership") to a concrete situation (e.g., what concrete observations could be made for evidence of leadership?)

Possible Sources of Problem Content

Previously developed FH items used content appropriate to students in psychology, with slight adaptations for law school students. To address the full GRE candidate population, we will need to develop problems that are compelling and do not require knowledge and experience within a specific domain. The original FH problems included numerical data, usually in the form of charts. Some new versions of the problems need not present information numerically. We also will need to investigate whether it is possible to elicit optimal hypothesis generation that is generic across fields, or whether the problems need to be field specific to some extent (e.g., social science/ hard science sets of problems).

We have explored, and need to explore further, some additional sources of content for the problems:

- "Current" issues that involve strategies for dealing with problem situations, such as the the restructuring of American education; examinees could generate ideas that suggest possible outcomes of a recommended strategy, or other strategies that might be feasible.

- "Tips" offered by "experts" (e.g., about well-being, fitness, nutrition); examinees could generate hypotheses about how these tips might or might not be applied to a situation.

- A sequential problem in which the same data are presented in different ways; this approach was ruled out, since candidates might perceive the problems as "tricks," or would not bother to generate and refine ideas until the last page of data was presented.

- Given a short text, the examinee would be asked to predict what possible outcomes might occur, or what hypotheses were being tested in relation to the findings of a specific investigation, or what further hypotheses the author might support (but not a reading comprehension task).

A Small Pilot Test of FH Item Constraints

Four experimental test booklets were assembled to enable us to conduct a preliminary evaluation of our hypothesis that placing constraints on FH items would not inhibit the production of a reasonable number of quality responses. The FH instructions and problems were redesigned to eliminate responses corresponding to the "general" categories and to constrain responding using some of the approaches discussed previously in this section. These are presented in Appendix C; the problems are summarized rather than presented in test booklet format.

The problems in each booklet were placed in different sequences to avoid possible order effects. In all booklets, one FH problem required no constraints and one FH problem constrained responding by requesting that one introductory phrase be used in forming the hypotheses. In three booklets, responses to one FH problem could be introduced by either of two phrases. In one problem in each of two booklets, students were asked either to state their responses as phrases or to rule out one generic category of responses. In one booklet, responses could begin with any of four phrases.

Two forms of the test booklets were administered to a total of 30 students in two different sections of an English composition course, "Writing in the Quantitative Social Sciences." The other two forms were administered to a total of 30 students in another composition class.

Overall, students conformed to the constraints to introduce their responses in specific ways, with very few exceptions. However, in the two instances in which they were asked to answer with a phrase rather than a fully developed hypothesis, this instruction was totally ignored and responses were written as sentences. With only one or two exceptions, students also conformed to the instructions to assume that the situation in the problem was methodologically sound. Thus, they did not propose hypotheses that might have been categorized as "general" in our earlier studies. This instruction eliminated a great many responses that vary considerably, often are trite, and were difficult to classify.

The most striking result of this preliminary exploration was the lack of evidence that any of the constraints imposed by the various formats inhibited or affected the students' fluency and quality of responding. Generally students who produced a considerable number and variety of responses to one problem did so consistently throughout the test booklet, regardless of the content or order of the problems. Conversely, students who had little to say had little to say on any one problem. This result will need to be tested with larger samples and with other FH problems, but it suggests that FH problems can be constructed with constraints on responses in order to facilitate computer recognition without restricting performance. In fact, constraining responses to begin with only one introductory phrase appeared to elicit the number and range of responses that were obtained in previous studies with problems of the same subject matter (e.g., juvenile delinquency, violence).

Limiting responding to two or more introductory phrases did not seem to impose any constraints whatsoever--the responses were as varied as the responses to unconstrained problems. Students appeared to have sufficient freedom to provide the inverse cases of the different potential hypotheses, which are likely to confound computer recognition considerably because these ideas can be expressed with so many variations of vocabulary and syntax.

This exploration suggests that the optimal format for responding may be one that requires the use of one introductory phrase. Because students did not appear to be constrained by using only one introductory phrase, and because it appears that syntax and vocabulary may be more systematic without reducing the range and number of ideas, we may be able to achieve optimal conditions for computer recognition while maintaining the integrity of measurement we have previously demonstrated in FH problems.

Summary

In this section, we have briefly described the nature of hypothesis formation, a preliminary outline that might eventually contribute to test specifications for a computer-delivered FH test, and some approaches to refining the design of tasks of the FH type. Subsequent work in these areas would contribute, in conjunction with a computerized scoring system, toward articulating specific, objective criteria that would be applied to explicit performance expectations for a defensible scoring system.

IV. CONCLUSIONS AND RECOMMENDATIONS

Following a series of analyses to explore the design and scoring of FH-type items for computer delivery, we arrived at specific recommendations for developing a prototype system. The in-depth analyses of one set of FH responses demonstrated, in our judgment, that a computerized delivery and scoring system can be achieved with presently available tools and expertise. A number of computer-based linguistic analysis tools already have been developed, providing the basic components necessary for building a system using a conceptual frame approach. In addition to scoring tools, the computerized adaptive testing system developed at ETS can be readily tailored to deliver items of the FH type. Because FH responses represent a high level of complexity and less-well-structured verbal material, they serve as a good model for designing a system that is likely to more readily deal with the scoring of other forms of open-ended responding as well.

This section describes our recommendations for the design of a test delivery and scoring system for open-ended, sentence-level responses, and for research during system development and after a prototype is functioning. Much of the research on constructing and refining the scoring system would take place while the prototype is being developed. Once a prototype is available, considerable research will be required to determine the optimal design of FH-type items to support the measurement characteristics of the resulting instrument and to investigate human factor variables that influence responding on the computer.

Development of the Scoring and Delivery Systems

From our work on keyword matching and surface features of style, lexicon, and syntax, we concluded that a deeper level of analysis that captures semantic relationships is necessary to categorize FH responses. The techniques we propose, which we termed conceptual frame analysis, begin with a linguistic analysis of the concepts and relationships that make up the semantic heart of the FH categories. The results of case frame analyses would then serve as predictors of and constraints on our computational techniques, which draw on features of case frame parsing and conceptual dependency theory. This system has the potential to be upgraded and enhanced, in contrast to other systems that would have strong inherent limitations. It would take advantage of whatever information is available in the response input, affording considerable flexibility and ensuring higher accuracy in recognition of responses.

Since a variety of tools that would make up the components of the system are currently available, the development test is feasible. Further work will be required, using response databases, to determine how and when the tools should be linked to optimize their capabilities as an integrated system. It may be that different heuristics will be required for different forms of verbal input, but the combination of techniques we recommend would be adaptable to such potential variations. We anticipate that a bank of common responses will be constructed to serve as a small domain for each FH-type item, and that a bank of common responses also may be constructed across a somewhat larger domain of many FH items. As development progresses with different, expanded databases, it is likely that a "bank" of common responses will be created that will facilitate more immediate recognition of a considerable number of responses.

More specifically, the following steps would constitute the next stages for the initial development of a delivery and scoring system for open-ended responses of the FH type:

1. A pool of responses to two or three FH-type items will be obtained.

2. Human judges will create response categories by sorting and evaluating the responses.

   Response quality will be evaluated in order to develop categories that represent levels of quality.

   Category development will be facilitated by the design requirements of the scoring analysis system.

      A tool may be available (e.g., Kintch's programs that analyze propositions into meaningful chunks) for category development.

      The categories are likely to differ somewhat from the kinds of categories that were used for purposes of human judgment; in fact, a greater number of categories may be developed in order to make the necessary discriminations among ideas that vary in quality.

3. The various computerized analytic tools will be used to analyze pools of responses and assign them to categories.

   Experimentation with the tools will be required to determine in what ways and at what points they could be linked.

   This step will require numerous iterations using more and additional pools of data to revise and refine the system.

Above all, the system should be designed to be dynamic and evolutionary, so that it will accommodate to a variety of open-ended assessment instruments.

4. Through many iterations, a bank of common responses will be created to develop a small domain for each FH-type item.

   A bank of common responses also will be created to develop a small domain across several FH-type items.

5. After a relatively successful paradigm for this process has been developed, the creation of subsequent categories and linguistic domains for additional FH items should proceed more efficiently.

6. By combining the system for analyzing and identifying responses with the computerized adaptive testing delivery system to present the FH items, a prototype will be available for further research. Given the resources, this point in the development process could be accomplished in a year.

Recommendations for Research

Once a prototype has been developed, additional research can be conducted more efficiently by collecting data using the computer system. Several major areas will require investigation:

- A large pool of responses to more FH-type problems will be obtained. The responses can be collected in paper and pencil format for entry by clerical staff, or input directly by students who represent the GRE candidate population. These data would be used to create and refine small linguistic domains and to further refine the scoring system.

- Additional computer programs may be developed to facilitate data collection (e.g., time on task, reactions to the presentation mode).

- Variations of FH item designs will be investigated to determine which approaches to constraining the task promote optimal responding (possibly making comparisons with paper and pencil versions of the same tasks).

- The content of the problems will need to be investigated to determine the kinds of problem content that are most appropriate to the task and accessible to most examinees. Generalizability of scores across problems having different content, as well as generalizability across candidates in different graduate fields, also will need to be investigated. In addition, we will need to investigate the potential impact of the computer-delivered, restructured FH problems on minority populations.

- Validity studies will be required to demonstrate that the FH-type items, combined as a test instrument, possess the psychometric characteristics of the construct that is intended to be measured. Additional information can be collected to determine how the hypothesis-generation task relates to the domain of human abilities and reasoning skills.

- A specific approach to the construct validity of the task would involve investigations of the process of hypothesis formation. Studies could be conducted to examine how individuals form and refine hypotheses. A system designed to deliver instruction or tutoring would provide information pertaining to the coachability of the responses within the computerized delivery environment. These data would lend support to inferences regarding the construct of hypothesis formation, and would inform additional test development efforts.

- Human factor investigations should be undertaken to ensure that candidates taking a test on the computer are able to perform optimally. The need for a warm-up period prior to taking the test should be considered.

- In the future, research efforts might lead to the development of an adaptive form of the test in which candidates would be presented with problem situations at their respective levels of ability.

During the early stages of the project, the research team anticipated that our final recommendations might consist of more than one possible approach to the computerized analysis and recognition of FH responses. Instead, our investigations and the achievements of scientists in the field of applied linguistic analyses enabled us to provide a more convergent solution. The system we have recommended should be sufficiently flexible and powerful to accommodate a wide variety of sentence-level open-ended responses now and in the future.

Implications for the GRE

Successful completion of the research and development necessary to automate scoring of FH-type items would have both specific and general implications for GRE testing. Specifically, it would make it feasible for the GRE program to consider incorporating into its examinations an item type that requires a kind of reasoning that is important for success in graduate education and that is not well represented in the present General Test. Given the interest in increasing the breadth of the analytical section of the examination and in increasing its distinguishability (with respect to abilities measured) from the verbal and quantitative sections, this could be an important contribution to the redesign of the examination.

More generally, the FH work would serve as a model for the analysis of natural language responses that might be elicited by a variety of other item types. In reading comprehension, for example, questions posed in free-response form could be expected to result in responses whose analysis would involve issues almost identical to those posed by FH. The analytic techniques and computer programs developed for FH could thus serve to make free-response versions of a number of item types feasible, decreasing the test developer's dependence on the multiple-choice format and increasing the variety of tasks that could be considered for inclusion in the examination.

V. REFERENCES

Addams, S. (March 1985). Interactive fiction. Popular Computing, 4(5), 97-182.

Baker, F. B. (1984). Technology and testing: state of the art and trends for the future. Journal of Educational Measurement, 21(4), 399-406.

Bruce, B. (1975). Case systems for natural language. Artificial Intelligence, 6, 327-360.

Burton, R. R., & Brown, J. S. (1982) An investigation of computer coaching for informal learning activities. In D. H. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems. London: Academic Press.

Canale, M. (1984). Considerations in the testing of reading and listening proficiency. Foreign Language Annals. 17(4).

Canale, M. (1986). The promise and threat of computerized adaptive testing of reading comprehension. In C. Stansfield (Ed.), Techonology and language testing. Washington, DC: TESOL.

Carlson, S. (1988). Relationships of reasoning and writing skills to GRE analytical ability scores (GRE No. 84-23). Princeton, NJ: Educational Testing Service.

Carlson, S. B. (1985). Pilot study of an application of formulating hypotheses problems within the context of law school admissions. Final report to the LSAC/LSAS Test Battery Workgroup.

Carlson, S. B., Bridgeman, B., Camp, R., & Waanders, J. (1985). Relationship of admissions test scores to writing performance of native and nonnative speakers of English (GRE General Report, ETS RR, & TOEFL Report No. 19). Princeton, NJ: Educational Testing Service.

Case, R., & Bereiter, C. (1984). From behaviourism to cognitive behaviourism to cognitive development: Steps in the evolution of instructional design. Toronto: OISE, Centre for Applied Cognitive Science.

Fillmore, C. (1968). The case for case. In E. Bach & R. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart, & Winston, 1-88.

Frase, L. T. (in press). Computer analysis of written materials. In D. Reinking (Ed.), Computers and reading: Issues for theory and practice. New York: Teachers College Press.

Frase, L. T., & Dieli, M. (1986). Reading, writing, and the UNIX Writer's Workbench software. T.H.E. (Technological Horizons in Education), 14(3), 74-78.

Frase, L. T., Macdonald, N. H., & Keenan, S. A. (1985). Intuitions, algorithms, and a science of text design. In T. Duffy & R. Waller (Eds.), Designing usable texts. New York: Academic Press.

Frederiksen, N., & Ward, W. C. (1975). Development of measures for the study of creativity (ETS RB No. 75-18). Princeton, NJ: Educational Testing Service.

Harris, M. D. (1985). Introduction to natural language processing. Reston, VA: Reston Publishing.

Hayes, P., & Carbonell, J. (1984). Natural language tutorial-- IJCAI-83: Supplementary notes. Pittsburgh: PA: Carnegie-Mellon University.

Hildyard, A., & Olsen, D.R. (1978). Memory and inference in the comprehension of oral and written discourse. Discourse Processes, 1, 91-117.

Hull, G., Ball, C., Fox, J., Levin, L., & McCutchen, D. (in press). Detecting errors in natural language texts: Some research on pattern-matching. Computers and the humanities. Osprey, FL: Paradigm Press.

King, M. (Ed.). (1983). Parsing natural language. New York: Academic Press.

Powers, D. E., & Enright, M. K. (1986, December). Analytical reasoning skills involved in graduate study: Perceptions of faculty in six fields (GRE Board Professional Report GREB No. 83-23P, ETS RR 86-43). Princeton, NJ: Educational Testing Service.

Ross, D. (1986). What surface-structure parsing can tell us about style. In W. S. Greaves & J. D. Benson (Eds.), Systematic perspectives in discourse: Selected papers from the 9th international systemic workshop. Norwood, NJ: Ablex.

Ross, D., & Bridwell, L. (1985). Computer-aided composing: Gaps in the software. In S. Olson (Ed.), Computer-aided instruction in the humanities. New York: MLA.

Sager, N. (1981). Natural language information processing: A computer grammar of English. New York: Addison-Wesley,

Scardamalia, M., & Bereiter, C. (1986). Computer supported intentional learning environments. Mimeo. Toronto: Centre for
Applied Cognitive Science, OISE.

Schank, R., & Abelson, R. (1977). Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale, NJ: Lawrence Erlbaum.

Schulz, M. (March 1985). The hitchhiker's guide to the galaxy. Popular Computing, 4(5), 160-162.

Swets, J. A., Rubin, A., & Feurzeig, W. (1987). Cognition, computers, and statistics: Software tools for curriculum design. Cambridge: Bolt, Beranek and Newman Laboratories, Inc. (Report No. 6447.)

Tomita, M. (1986). Efficient parsing for natural language. Boston: Kluwer Academic Publishers.

Ward, W. C., Carlson, S. B., & Woisetschlaeger, E. (1983). Ill-structured problems as multiple-choice items (GRE Board Professional Report No. 81-18P and ETS RR 83-6). Princeton, NJ: Educational Testing Service.

Ward, W. C., & Frederiksen, N. (1977). A study of the predictive validity of the tests of scientific thinking (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 and ETS RB No. 78-15). Princeton, NJ: Educational Testing Service.

Ward, W. C., Frederiksen, N., & Carlson, S. (1980). Construct validity of free-response and machine-scorable versions of a test. Journal of Educational Measurement, 17, 11-29.

Willingham, W. (1985). Predicting success in graduate education--A framework of research issues. (Draft of internal document) Princeton, NJ: Educational Testing Service.

Winograd, T., & Flores, F. (1986). Understanding computers and Cognition. Norwood, NJ: Ablex.

Appendix A

Example of FH Item and Scoring System

Family Situation of Juvenile Delinquents

Instructions
Sample Problem and Responses
Sample Category Lists
Sample Score Sheet and Scored Responses
Quality Values for the Problem
Responses to One FH Problem Used in Analyses

FORMULATING HYPOTHESES

Directions

Each problem in this test consists of a brief description of a psychological investigation, a figure or table presenting the data from the study, and a short statement of an important finding. Your task is to think of hypotheses (possible explanations) to account for the finding.

For each problem think of the hypothesis you believe is most likely to provide the correct explanation or interpretation for the finding, and additional competing hypotheses that ought to be considered in interpreting the study or in planning further research. Write your hypotheses in the answer spaces. Mark the hypothesis you consider most likely to be correct by placing an X in the box at its right.

Now study the sample problem and sample answers. Then complete the six test problems, allowing yourself about eight minutes for each problem.

SAMPLE PROBLEM

Novelty of imaginative productions

The effects of two types of verbal discourse on novelty of imaginative productions were studied. One group of subjects (the "Monotony Group") listened to a dull, monotonous 12-minute recording of verbal discourse. Another group (the "Novelty Group") listened to an interesting, novel recording of verbal discourse of the same length.

Both groups were then shown a series of pictures of people and were asked to write a story suggested by each picture. The stories were scored for the degree of novelty of the imaginative productions. Results are given in the following table:

Groups	Number of subjects at three levels of novelty of imaginative productions		
	Low	Middle	High
Monotony Group	21	22	10
Novelty Group	7	18	24

Finding: The Monotony Group produced less novel imaginative productions than did the Novelty Group.

SAMPLE ANSWERS

Novelty of imaginative productions

Suggested Hypotheses

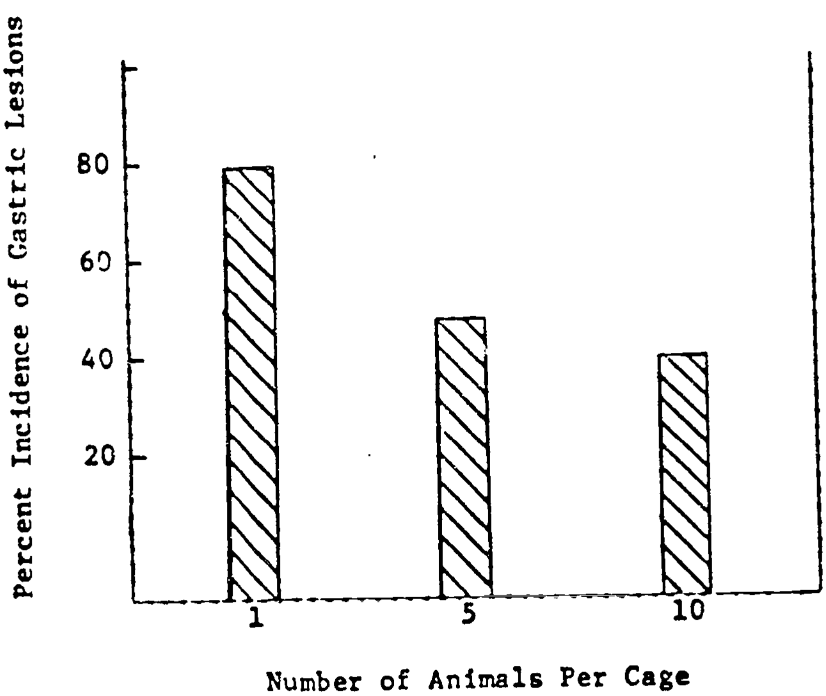
There was bias in assigning subjects to groups, so that the Novelty group was more imaginative to begin with.	☐
Subjects modeled their productions after the style to which they had been exposed.	☒
Novel discourse provided ideas which subjects could incorporate into their stories.	☐
The monotonous condition caused resentment toward the experimenter, which resulted in lack of cooperation by the subjects.	☐
Novel discourse heightened motivation, while monotonous discourse produced boredom which impaired performance.	☐
	☐
BEST COPY AVAILABLE	☐

Mark the hypothesis you think is best by putting an X in the box at its right.

Mice ulcers and housing

Sixty male mice were housed 10 to a cage (17x28x13 cm) from weaning to 45 days of age. Then they were randomly assigned to different housing conditions in identical cages of the same size. They were housed either 1 per cage (N=20), 5 per cage (N=20), or 10 per cage (N=20) for one month.

At the end of the month, the mice were examined for the presence of gastric lesions (ulcers). Results are shown in the table below:



Incidence of Ulcers in Relation to

Housing Conditions

Finding: The number of ulcers decreased as the number of animals per cage increased.

Ex. 1

The mice were accustomed to being with 9 other mice, so the change from pre-weaning to experimental treatment caused ulcers proportional to the degree of change. However, judging from the number of ulcers, I think the results are due to something wrong with the design of the experiment. ☒

Mice may be innately gregarious animals, living together in small bands. If this is the case, then "solitary confinement" (the one mouse per cage condition) may be a stressful situation (in this strain of mice, at least), leading to more ulcers. ☐

☐

Single mice suffer from the lack of physical warmth provided by the presence of other mice when living together. ☐

By the way, I'm getting a little bored with this task! ☐

Being in a large group reduces stress for mice, as they are basically "social" animals. ☐

Perhaps mice were housed in different-sized cages, which led to a difference in ulcer-formation. ☐

Mark the hypothesis you think is best by putting an X in the box at its right.

Perhaps single mice didn't get enough

PLEASE GO ON TO THE NEXT PAGE

Item 6

FH ANSWER CATEGORIES

General

1. There were too few cases to draw conclusions.

2. There was bias (unspecified) in assigning Ss to treatments.

3. The sample was not typical (representative) of the population (in ways unspecified).

4. Errors (unspecified) in the design or conduct of the study could account for the finding.

5. The experimenter, knowing the purpose of the experiment, was biased in his treatment of the groups.

6. The experimenter (observer, evaluator), knowing the purpose of the experiment, was biased in his assessment of the results.

7. The measurement procedure (instrument, test) was inadequate (not valid, unreliable).

8. The statistical method was inappropriate (inadequate).

9. The results are not statistically significant.

10. [The response is incomprehensible.]

11. [The response is essentially a restatement of the finding.]

12. [The response is an erroneous criticism of the experimental design or procedure.]

13. [The examinee apparently misread or misunderstood the problem.]

2/77

FH ANSWER CATEGORIES

Mice Ulcers as a Function of Housing Condition

15. The contrast (change) in crowding between initial housing and new housing produced stress.

16. Since mice are social animals, separation from the group produced stress (anxiety, fear) [due to disruption of important social structures, such as dominance hierarchies].

17. Separation from the group resulted in lack of stimulation (loneliness, boredom, loss of appetite, reduced activity).

18. Separation from the group caused sexual frustration.

19. Mice living in larger groups had other mice upon which to release tension and stress; [in single mice, this stress was inwardly directed].

20. The change in housing condition occurred at a critical period in the lives of the mice, [when they need to be with other mice or when major neurological development takes place].

21. As the number of mice decreased and the available food increased, eating habits changed.

22. Lack of social grooming (e.g., licking, lice removal) in the isolated mice produced stress.

23. Mice in larger groups had less room for movement, so they became less active (more relaxed).

24. Ulcers were caused by excessive space for single mice.

25. [Single mice have difficulty coping by themselves], whereas several mice work together (cooperate) to survive.

26. Mice emit (produce) [chemical] substances that inhibit formation of (reduce susceptibility to) ulcers in other mice; [therefore, the more mice, the fewer ulcers].

4
Ex. 1

Sample

SCORE SHEET

Scorer	Problem	Test	Registration No.
0 9	12	1	0 9 8

		B or bl	2 digits	N	1 digit
		Col. 1	Cols. 2,3	Col. 2	Col. 3
Space No.	Response	Best?	Category No.	New?	Rating
1	+	✓	15		
(2)	+		16		
3	—		—		
4	+		—	✓	6
5	—		—		
(6)	—		—		
7	+		12		
8	—		—		
1	+		4		

No "Best" ☐

Quality Values

Mice Ulcers & Housing

	Category Number	Quality Value	Rare	Creative
General	1	11.6	✓	
	2	6.8	✓	
	3	7.4	✓	
	4	10.4	✓	
	5	10.4	✓	
	6	9.0	✓	
	7	8.6	✓	
	8	9.4	✓	
	9	12.6	✓	
	10	omit		
	11	4.4		
	12	3.4		
	13	2.0		
	14	omit		
Specific	15	22.2		
	16	24.0		
	17	21.2		
	18	20.2	✓	✓
	19	23.0	✓	✓
	20	17.2	✓	✓
	21	18.2	✓	✓
	22	20.0	✓	✓
	23	15.8	✓	
	24	14.6	✓	
	25	17.0	✓	✓
	26	14.6	✓	

15. Children from broken homes lack affection (care, warmth).
Includes parents not having time to care for the child.

Children from disrupted families receive much less love and affection than those with 2 parents, and therefore resort to delinquency for attention.

The husband-wife family provides an environment where there is care, love, supervision, and guidance resulting in less delinquency.

Children from disrupted homes blame themselves for the family problems and become involved in delinquent activities because they think no one cares about them.

Juvenile delinquents come from homes which are broken in some way nearly 40% of the time. A single parent must support family alone. This parent has little remaining time and energy to spend on children.

When the head and only parent in the household is a man the children are frequently left alone an not nurtured and counseled through the hard years of adolescence causing a high rate of delinquency.

Parents of children from a broken home do not care what their children do, and hence, the children have no restraints on their behavior.

The traditional family situation provides love and stability and the child grows up to be responsible.

Failing to get sufficient love and attention at home, children may behave delinquently to attract attention to themselves.

Families that are strong tend to spend more time together than broken families. Children have less time to get into trouble.

A single parent does not have time to discipline a child as well as two parents can.

Children from disrupted families strick out against society, because they feel less loved.

Parents who are happily married are more likely to behave in caring and loving manner toward their children, thus children develop more respectful and caring.

[Same subject as above.] Children without a mother feel the most anxiety ridden because the father is unable to provide the child with the affection she needs and therefore rebels.

Two is better than one in terms of sharing responsibility--more time to spend with child.

Appendix B

Computerized Text Analysis with the Writer's Workbench

1. Variables measured
2. Contrasts among categories
3. Good and poor responses
4. Correlations among response categories
5. Correlations among variables
6. Data for all response categories

Readability
1 kin Kincaid readability grade level
2 aut Auto readability grade level
3 c-l Coleman-Liau readability grade level
4 fle1 Flesch readability grade level
5 fle2 Flesch readability index

Sentence information
6 nosnt number sentences
7 nowds number words
8 avs-len average sentence length
9 avw-len average word length
10 noqst number questions
11 noimp number imperatives
12 nocwds number content words
13 %cwds percentage content words
14 avlencw average length content words
15 shtsent short sentences < n words
16 %shts percentage short sentences
17 noshts number short sentences
18 lngs long sentences > n words
19 %lngs percentage long sentences
20 nolngs number long sentences
21 lgests longest sentence
22 where location
23 shtsts shortest sentence
24 where location

Sentence types
25 smpl% percentage simple sentences
26 simpno number simple sentences
27 cplex% percentage complex sentences
28 plexno number complex sentences
29 pound% percentage compound sentences
30 poundno number compound sentences
31 c-c% percentage compound-complex sentences
32 c-cno number compound sentences

Word usage
33 verbt total verbs
34 tobe% percentage to be verbs
35 tobeno number to be
36 aux% percentage auxiliary
37 auxno number auxiliary
38 inf% percentage infinitives
39 infno number infinitives
40 pass% percentage passives
41 passno number passives
42 prep% percentage prepositions
43 prepno number prepositions
44 conj% percentage conjunctions
45 conjno number conjunctions
46 adv% percentage adverbs
47 advno number adverbs
48 noun% percentage nouns
49 nounno number nouns
50 adj% percentage adjectives
51 adjno number adjectives
52 pron% percentage pronouns
53 pronno number pronouns
54 nom% percentage nominalizations
55 nomno number nominalizations

Sentence beginnings
56 nouno number noun subject openers
57 pronop number pronoun subject openers
58 posop number possessive subject openers
59 adjop number adjective subject openers
60 artop number article subject openers
61 totop total openers
62 prep% percentage prepositions
63 prepno number prepositions
64 adv% percentage adverbs
65 advno number adverbs
66 verb% percentage verbs
67 verbno number verbs
68 sub-c% percentage subordinate conjunctions
69 sub-cno number subordinate conjunctions
70 conj% percentage conjunctions
71 conjno number conjunctions
72 exp% percentage expletives
73 expno number expletives

Additional measures
74 abst percentage abstractness
75 dict number diction hits
76 dict%s percentage diction hits (hits/sentences)
77 tt.con content words diversity (unique words/total words)
78 tt.fun function words diversity (unique words/total words)
79 var verb/adjective ratio
80 sylrep syllabic repetitions of words (high score=monotone)
81 ch.n number of chunks
82 ch.m mean chunk length in characters
83 ch.sd standard deviation of chunk length
84 quality quality of response

BEST COPY AVAILABLE

CONTRASTS AMONG CATEGORIES
(SAMPLES OF ROUGHLY EQUAL LENGTH)

	categories-->	15	17	20	22a	22b	23	29a	29b
1	win	11.1	11.4	12.0	13.3	13.7	13.2	13.6	15.2
2	aut	11.6	11.9	12.3	14.6	13.8	13.9	13.6	14.2
3	c-l	12.5	13.0	14.3	13.7	12.7	14.7	13.3	14.3
4	fle1	13.3	13.8	14.8	14.4	14.9	15.2	15.2	17.0
5	fle2	48.2	44.9	38.0	40.4	37.6	35.3	35.3	22.9
6	nosnt	17	16	19	14	13	16	15	15
7	nowds	310	287	310	319	296	310	319	311
8	evslen	18.2	17.9	16.3	22.8	22.8	19.4	21.3	20.7
9	evwlen	5.08	5.17	5.43	5.24	5.05	5.45	5.18	5.36
10	noqst	0	0	0	0	0	0	0	0
11	noimp	0	0	0	0	0	0	0	0
12	nocwds	188	176	189	197	167	194	193	207
13	%cwds	60.6	61.3	61.0	61.8	56.4	62.6	60.5	66.6
14	avlencw	6.19	6.43	7.00	6.48	6.64	6.99	6.58	6.54
15	shtsent	13	13	11	18	18	14	16	16
16	%shts	18	13	26	50	31	19	20	20
17	noshts	3	2	5	7	4	3	3	3
18	ings	28	28	26	33	33	29	31	31
19	%ings	6	0	5	21	15	6	13	0
20	noings	1	0	1	3	2	1	2	0
21	lgests	35	28	32	63	60	33	48	28
22	where	7	5	8	10	13	10	1	8
23	shtsts	7	8	7	13	9	12	10	9
24	where	5	1	11	1	6	1	7	11
25	smpl%	29	63	53	50	31	31	27	27
26	simpno	5	10	10	7	4	5	4	4
27	cplex%	53	31	21	29	23	44	67	53
28	plexno	9	5	4	4	3	7	10	8
29	pound%	6	6	11	7	8	6	0	13
30	poundno	1	1	2	1	1	1	0	2
31	c-c%	12	0	16	14	38	19	7	7
32	c-cno	2	0	3	2	5	3	1	1
33	verbt	40	32	35	38	39	40	45	44
34	tobe%	23	25	31	24	54	43	38	41
35	tobeno	9	8	11	9	21	17	17	18
36	aux%	10	9	3	13	15	8	7	7
37	auxno	4	3	1	5	6	3	3	3
38	inf%	25	31	14	29	18	20	29	30
39	infno	10	10	5	11	7	8	13	13
40	pass%	10	9	10	11	16	9	3	6
41	passno	3	2	3	3	5	3	1	2
42	prep%	10.3	13.9	13.9	13.5	11.5	9.0	11.3	10.6
43	prepno	32	40	43	43	34	28	36	33
44	conj%	4.8	3.5	6.1	4.7	4.1	3.9	3.1	2.3
45	conjno	15	10	19	15	12	12	10	7
46	adv%	4.5	3.1	3.9	4.7	5.4	6.5	5.0	4.5
47	advno	14	9	12	15	16	20	16	14
48	noun%	28.7	31.7	31.3	27.6	27.7	28.1	26.3	24.8
49	nounno	89	91	97	88	82	87	84	77
50	adj%	16.1	16.4	17.1	19.1	14.5	19.4	18.8	28.3
51	adjno	50	47	53	61	43	60	60	88
52	pron%	4.5	5.6	2.3	5.0	5.1	2.9	4.1	2.3
53	pronno	14	16	[illegible]	16	15	9	13	7
54	nom%	3	7		4	3	1	1	1
55	nomno	9	19	0	12	9	4	3	2
56	nouno	10	10	[illegible]	8	[illegible]	5	4	5
57	pronop	0	2	[illegible]	0	0	0	0	1
58	posop	0	0	0	0	0	0	0	0
59	adjop	2	2	1	5	3	5	8	6
60	artop	4	1	6	1	2	5	1	3
61	totop	94	94	95	100	92	94	87	100
62	prep%	0	6	5	0	8	0	0	0
63	prepno	0	1	1	0	1	0	0	0
64	adv%	0	0	0	0	0	0	7	0
65	advno	0	0	0	0	0	0	1	0
66	verb%	0	0	0	0	0	0	0	0
67	verbno	0	0	0	0	0	0	0	0
68	sub-c%	6	0	0	0	0	6	7	0
69	sub-cno	1	0	0	0	0	1	1	0
70	conj%	0	0	0	0	0	0	0	0
71	conjno	0	0	0	0	0	0	0	0
72	exp%	0	0	0	0	0	0	0	0
73	expno	0	0	0	0	0	0	0	0
74	abst	1.3	0.7	1.6	0.6	0.6	3.7	1.5	0.6
75	dict	4	3	5	8	2	2	3	2
76	dict%s	23.52	18.75	26.31	57.14	15.38	12.50	20.00	13.33
77	tt.con	.64	.55	.59	.58	.60	.67	.54	.48
78	tt.fun	.41	.37	.35	.34	.31	.36	.29	.36
79	var	.90	.80	.60	.72	.77	.58	.77	.47
80	svlrep	.39	.39	.37	.39	.41	.35	.38	.33
81	ch.n	53	55	51	60	51	57	57	55
82	ch.m	35.08	35.13	37.94	32.03	34.57	34.12	34.28	35.58
83	ch.sd	13.60	18.66	15.00	12.10	17.94	13.36	10.43	13.27
84	quality	28.20	17.00	27.40	30.60	30.60	29.20	29.60	29.60

GOOD AND POOR QUALITY RESPONSES
(SAMPLES OF ROUGHLY EQUAL LENGTH)

quality-->		good	good	good	good	poor
1	kin	12.6	12.8	13.1	13.0	11.3
2	eut	12.9	13.4	13.1	12.7	11.2
3	c-l	12.9	13.6	13.0	14.3	13.1
4	flel	14.3	14.6	14.8	15.6	14.1
5	fle2	41.3	39.5	37.9	32.8	42.9
6	nosnt	31	28	31	40	35
7	ncwds	633	569	642	691	564
8	evslen	20.4	20.3	20.7	17.3	16.1
9	evwlen	5.12	5.24	5.14	5.40	5.22
10	noqst	0	0	0	0	0
11	noimp	0	0	0	0	0
12	nocwds	384	345	390	421	351
13	%cwds	60.7	60.6	60.7	60.9	62.2
14	evlencw	6.43	6.72	6.56	6.87	6.52
15	shtsent	15	15	16	12	11
16	%shts	25	25	23	23	26
17	noshts	8	7	7	9	9
18	lngs	30	30	31	27	26
19	%lngs	16	11	10	10	11
20	nolngs	5	3	3	4	4
21	lgests	63	50	48	33	41
22	where	18	4	8	38	29
23	shtsts	3	9	6	7	4
24	where	6	9	31	18	9
25	smpl%	48	21	29	50	60
26	simpno	15	6	9	20	21
27	cplex%	29	39	58	30	37
28	plexno	9	11	18	12	13
29	pound%	10	14	6	5	3
30	poundno	3	4	2	2	1
31	c-c%	13	25	6	15	0
32	c-cno	4	7	2	6	0
33	verbt	74	79	87	79	64
34	tobe%	32	46	36	30	27
35	tobeno	24	36	31	24	17
36	eux%	14	11	8	9	9
37	euxno	10	9	7	7	6
38	inf%	24	22	25	19	27
39	infno	18	17	22	15	17
40	pess%	14	13	8	6	6
41	passno	8	8	5	4	3
42	prep%	13.1	9.7	11.4	12.7	12.4
43	prepno	83	55	73	88	70
44	conj%	4.1	4.4	3.3	4.5	2.8
45	conjno	26	25	21	31	16
46	edv%	5.4	5.8	4.7	3.8	4.6
47	edvno	34	33	30	26	26
48	noun%	28.3	27.1	26.2	28.9	30.9
49	nounno	179	154	168	200	174
50	edj%	17.2	18.3	19.8	19.5	16.8
51	edjno	109	104	127	135	95
52	pron%	4.6	3.5	3.7	3.9	3.7
53	pronno	29	20	24	27	21
54	nom%	3	2	1	3	5
55	nomno	20	14	5	18	27
56	nouno	17	12	8	24	18
57	pronop	0	0	0	3	2
58	posop	0	0	0	0	0
59	edjop	9	7	11	3	6
60	ertop	3	6	7	9	5
61	totop	94	89	84	98	89
62	prep%	6	0	0	3	3
63	prepno	2	0	0	1	1
64	edv%	0	0	3	0	6
65	edvno	0	0	1	0	2
66	verb%	0	0	0	0	0
67	verbno	0	0	0	0	0
68	sub-c%	0	11	10	0	3
69	sub-cno	0	3	3	0	1
70	conj%	0	0	0	0	0
71	conjno	0	0	0	0	0
72	exp%	0	0	3	0	0
73	expno	0	0	1	0	0
74	ebs%	0.9	2.1	2.8	1.6	1.8
75	dict	11	6	4	7	8
76	dict%s	35.48	21.42	12.90	17.50	22.85
77	tt.con	.48	.56	.49	.49	.55
78	tt.fun	.21	.26	.22	.24	.25
79	var	.73	.68	.69	.56	.77
80	sylrep	.36	.40	.39	.35	.38
81	c .n	118	103	112	118	103
82	ch.m	32.30	34.02	34.54	36.98	35.32
83	ch.sd	14.78	12.90	12.76	13.22	17.40

CORRELATIONS AMONG RESPONSE CATEGORIES

category-->	[03]	[11]	[12]	[13]	[15]	[16]	[17]	[18]	[19]	[20]
[03]	1.00	0.62	0.46	0.51	0.45	0.53	0.43	0.74	0.57	0.35
[11]	0.62	1.00	0.73	0.71	0.75	0.84	0.79	0.73	0.77	0.71
[12]	0.46	0.73	1.00	0.79	0.29	0.49	0.36	0.68	0.36	0.25
[13]	0.51	0.71	0.79	1.00	0.28	0.42	0.33	0.53	0.35	0.25
[15]	0.45	0.75	0.29	0.28	1.00	0.97	0.99	0.57	0.98	0.99
[16]	0.53	0.84	0.49	0.42	0.97	1.00	0.98	0.69	0.97	0.95
[17]	0.43	0.79	0.36	0.33	0.99	0.98	1.00	0.58	0.96	0.99
[18]	0.74	0.73	0.68	0.53	0.57	0.69	0.58	1.00	0.66	0.50
[19]	0.57	0.77	0.36	0.35	0.98	0.97	0.96	0.66	1.00	0.95
[20]	0.35	0.71	0.25	0.25	0.99	0.95	0.99	0.50	0.95	1.00
[21]	0.66	0.82	0.60	0.50	0.83	0.90	0.84	0.84	0.89	0.79
[22]	0.27	0.62	0.12	0.14	0.97	0.90	0.96	0.40	0.92	0.99
[23]	0.36	0.69	0.20	0.21	0.99	0.94	0.98	0.49	0.95	0.99
[24]	0.80	0.63	0.60	0.49	0.52	0.62	0.50	0.79	0.65	0.42
[25]	0.73	0.89	0.67	0.66	0.84	0.91	0.84	0.61	0.89	0.78
[26]	0.69	0.88	0.55	0.61	0.86	0.92	0.88	0.75	0.91	0.84
[27]	0.78	0.88	0.62	0.56	0.82	0.90	0.84	0.62	0.86	0.77
[28]	0.66	0.88	0.54	0.48	0.92	0.96	0.93	0.78	0.94	0.66
[29]	0.31	0.63	0.14	0.15	0.98	0.91	0.96	0.43	0.93	0.99
[30]	0.79	0.67	0.70	0.52	0.40	0.53	0.42	0.87	0.50	0.32
[31]	0.55	0.71	0.67	0.77	0.43	0.53	0.46	0.53	0.50	0.39
[32]	0.75	0.81	0.83	0.86	0.51	0.67	0.55	0.81	0.59	0.44
[33]	0.77	0.84	0.79	0.62	0.65	0.77	0.66	0.85	0.73	0.57
[34]	0.54	0.71	0.82	0.94	0.33	0.47	0.38	0.83	0.41	0.30
[35]	0.76	0.77	0.66	0.62	0.51	0.61	0.51	0.59	0.60	0.43

category-->	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]
[03]	0.66	0.27	0.36	0.80	0.73	0.69	0.78	0.88	0.31	0.79
[11]	0.82	0.62	0.89	0.63	0.89	0.88	0.88	0.88	0.63	0.67
[12]	0.60	0.12	0.20	0.60	0.67	0.55	0.62	0.54	0.14	0.70
[13]	0.50	0.14	0.21	0.49	0.66	0.61	0.56	0.48	0.15	0.52
[15]	0.83	0.97	0.99	0.52	0.84	0.88	0.82	0.92	0.98	0.40
[16]	0.90	0.90	0.94	0.62	0.91	0.92	0.90	0.96	0.91	0.53
[17]	0.84	0.96	0.98	0.50	0.84	0.88	0.84	0.93	0.96	0.42
[18]	0.84	0.40	0.49	0.79	0.81	0.75	0.82	0.78	0.43	0.87
[19]	0.89	0.92	0.95	0.65	0.89	0.91	0.88	0.94	0.93	0.50
[20]	0.79	0.99	0.99	0.42	0.78	0.84	0.77	0.88	0.99	0.32
[21]	1.00	0.71	0.77	0.81	0.93	0.90	0.93	0.93	0.72	0.71
[22]	0.71	1.00	0.99	0.33	0.70	0.76	0.68	0.82	1.00	0.21
[23]	0.77	0.99	1.00	0.42	0.77	0.82	0.76	0.87	0.99	0.30
[24]	0.81	0.33	0.42	1.00	0.80	0.70	0.78	0.68	0.37	0.72
[25]	0.93	0.70	0.77	0.80	1.00	0.93	0.94	0.93	0.72	0.72
[26]	0.90	0.76	0.82	0.70	0.93	1.00	0.95	0.94	0.78	0.62
[27]	0.93	0.68	0.76	0.78	0.94	0.95	1.00	0.96	0.70	0.76
[28]	0.93	0.82	0.87	0.68	0.93	0.94	0.96	1.00	0.83	0.68
[29]	0.72	1.00	0.99	0.37	0.72	0.78	0.70	0.83	1.00	0.24
[30]	0.71	0.21	0.30	0.72	0.72	0.62	0.76	0.68	0.24	1.00
[31]	0.66	0.30	0.37	0.66	0.72	0.64	0.64	0.61	0.31	0.56
[32]	0.80	0.32	0.41	0.81	0.80	0.75	0.86	0.76	0.35	0.85
[33]	0.85	0.46	0.55	0.85	0.90	0.81	0.90	0.82	0.50	0.80
[34]	0.55	0.19	0.26	0.52	0.70	0.62	0.61	0.55	0.20	0.60
[35]	0.66	0.33	0.42	0.74	0.82	0.76	0.75	0.67	0.37	0.63

category-->	[31]	[32]	[33]	[34]	[35]
[03]	0.55	0.75	0.77	0.54	0.76
[11]	0.71	0.81	0.84	0.71	0.77
[12]	0.67	0.83	0.79	0.82	0.68
[13]	0.77	0.66	0.62	0.94	0.82
[15]	0.43	0.51	0.65	0.33	0.51
[16]	0.53	0.67	0.77	0.47	0.61
[17]	0.46	0.55	0.66	0.38	0.51
[18]	0.53	0.81	0.85	0.63	0.69
[19]	0.50	0.59	0.73	0.41	0.60
[20]	0.39	0.44	0.57	0.30	0.43
[21]	0.66	0.80	0.85	0.55	0.66
[22]	0.30	0.32	0.46	0.19	0.33
[23]	0.37	0.41	0.55	0.26	0.42
[24]	0.66	0.81	0.85	0.52	0.74
[25]	0.72	0.80	0.90	0.70	0.82
[26]	0.64	0.75	0.81	0.62	0.76
[27]	0.64	0.86	0.90	0.61	0.75
[28]	0.61	0.76	0.82	0.55	0.67
[29]	0.31	0.35	0.50	0.20	0.37
[30]	0.56	0.85	0.80	0.60	0.63
[31]	1.00	0.72	0.63	0.73	0.67
[32]	0.72	1.00	0.91	0.67	0.71
[33]	0.63	0.91	1.00	0.64	0.83
[34]	0.73	0.67	0.64	1.00	0.79
[35]	0.67	0.71	0.83	0.79	1.00

CORRELATIONS AMONG VARIABLES

		kin [, 1]	aut [, 2]	c-l [, 3]	flel [, 4]	fle2 [, 5]	nosnt [, 6]	nowds [, 7]	avsle [, 8]	avwle [, 9]	noqst [,10]
kin	[1,]	1.00	0.95	0.72	0.86	-0.93	0.02	0.08	0.55	0.57	NA
aut	[2,]	0.95	1.00	0.69	0.81	-0.84	0.11	0.18	0.64	0.53	NA
c-l	[3,]	0 72	0.69	1.00	0.81	-0.88	0.07	0.06	-0.11	0.97	NA
flel	[4,]	0.86	0.81	0.81	1.00	-0.93	0.17	0.19	0.24	0.73	NA
fle2	[5,]	-0.93	-0.84	-0.88	-0.93	1.00	0.01	-0.02	-0.22	-0.79	NA
nosnt	[6,]	0.02	0.11	0.07	0.17	0.01	1.00	0.99	0.11	0.02	NA
nowds	[7,]	0.08	0.18	0.06	0.19	-0.02	0.99	1.00	0.20	0.0	NA
avslen	[8,]	0.55	0.64	-0.11	0.24	-0.22	0.11	0.20	1.00	-0.31	NA
avwlen	[9,]	0.57	0.53	0.97	0.73	-0.79	0.02	0.0	-0.31	1.00	NA
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
no'mp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	0.08	0.18	0.07	0.20	-0.03	0.99	1.00	0.20	0.01	NA
%cwds	[13,]	0.16	0.16	0.57	0.47	-0.40	-0.02	-0.05	-0.45	0.68	NA
avlencw	[14,]	0.65	0.59	0.93	0.74	-0.82	0.07	0.06	-0.16	0.90	NA
shtsent	[15,]	0.56	0.65	-0.10	0.26	-0.23	0.11	0.21	1.00	-0.30	NA
%shts	[16,]	0.11	0.15	-0.10	-0.20	0.06	0.11	0.14	0.32	-0.17	NA
noshts	[17,]	0.05	0.15	0.04	0.09	0.02	0.93	0.94	0.18	-0.02	NA
lngs	[18,]	0.56	0.65	-0.10	0.26	-0.23	0.11	0.21	1.00	-0.30	NA
%lngs	[19,]	0.09	0.15	-0.21	-0.13	0.07	0.0	0.05	0.40	-0.26	NA
nolngs	[20,]	0.09	0.20	0.0	0.07	0.01	0.77	0.83	0.28	-0.07	NA
lgests	[21,]	0.28	0.39	-0.05	0.19	-0.07	0.77	0.83	0.61	-0.19	NA
where	[22,]	-0.09	0.0	0.02	0.03	0.09	0.57	0.54	-0.01	0.0	NA
shtsts	[23,]	0.45	0.45	0.01	0.21	-0.28	-0.37	-0.29	0.60	-0.10	NA
where	[24,]	0.06	0.13	0.09	0.19	-0.05	0.88	0.89	0.11	0.05	NA
smpl%	[25,]	0.05	0.01	0.37	0.12	-0.20	-0.23	-0.26	-0.39	0.45	NA
simpno	[26,]	-0.05	0.03	0.11	0.13	0.02	0.89	0.82	-0.06	0.09	NA
cplex%	[27,]	-0.11	-0.09	-0.26	-0.20	0.17	0.09	0.10	0.17	-0.30	NA
plexno	[28,]	0.04	0.12	0.04	0.18	-0.01	0.90	0.91	0.15	-0.02	NA
pound%	[29,]	-0.02	0.02	-0.05	0.08	0.05	0.45	0.45	0.10	-0.09	NA
poundno	[30,]	0.04	0.12	0.02	0.13	0.01	0.83	0.85	0.16	-0.03	NA
c-c%	[31,]	0.08	0.10	-0.22	0.06	0.06	0.07	0.11	0.37	-0.28	NA
c-cno	[32,]	0.09	0.19	0.02	0.13	0.0	0.80	0.83	0.25	-0.04	NA
verbt	[33,]	0.07	0.16	0.05	0.17	-0.01	0.97	0.99	0.19	-0.02	NA
tobe%	[34,]	0.36	0.22	0.12	0.17	-0.30	-0.05	-0.02	0.21	0.04	NA
tobeno	[35,]	0.13	0.20	0.07	0.20	-0.07	0.92	0.96	0.22	0.01	NA
aux%	[36,]	-0.14	-0.06	-0.32	-0.46	0.34	-0.21	-0.17	0.26	-0.37	NA
auxno	[37,]	-0.01	0.12	-0.07	-0.01	0.12	0.81	0.86	0.25	-0.13	NA
inf%	[38,]	-0.20	-0.11	-0.12	-0.17	0.22	-0.04	-0.02	-0.07	-0.06	NA
infno	[39,]	0.07	0.16	0.02	0.17	-0.01	0.90	0.94	0.22	-0.04	NA
pass%	[40,]	-0.27	-0.34	-0.24	-0.27	0.22	0.25	0.22	-0.19	-0.22	NA
passno	[41,]	-0.01	0.09	0.03	0.06	0.04	0.88	0.89	0.11	-0.02	NA
prep%	[42,]	0.33	0.25	0.56	0.45	-0.49	0.02	-0.02	-0.24	0.56	NA
prepno	[43,]	0.08	0.18	0.09	0.20	-0.04	0.98	0.99	0.17	0.03	NA
conj%	[44,]	-0.10	0.01	-0.39	-0.19	0.29	0.32	0.32	0.42	-0.46	NA
conjno	[45,]	0.05	0.17	0.04	0.13	0.02	0.95	0.95	0.21	-0.03	NA
adv%	[46,]	-0.23	-0.18	-0.38	-0.33	0.35	-0.01	0.0	0.15	-0.39	NA
advno	[47,]	0.06	0.17	0.02	0.14	0.01	0.94	0.97	0.22	-0.04	NA
noun%	[48,]	0.22	0.23	0.45	0.44	-0.38	0.09	0.03	-0.18	0.48	NA
nounno	[49,]	0.07	0 17	0.08	0.19	-0.02	0.99	0.99	0.18	0.01	NA
adj%	[50,]	0.54	0.42	0.58	0.56	-0.62	-0.11	-0.08	-0.03	0.56	NA
adjno	[51,]	0.13	0.21	0 10	0.24	-0.08	0.96	0.98	0.21	0.03	NA
pron%	[52,]	-0.34	-0.26	-0.42	-0.42	0.46	0.07	0.09	0.12	-0.45	NA
pronno	[53,]	0.0	0.14	-0.01	0.08	0.09	0.90	0.93	0.23	-0.08	NA
nom%	[54,]	-0.07	-0.07	0.23	0.11	-0.10	-0.03	-0.04	-0.38	0.33	NA
nomno	[55,]	0.01	0.11	0.05	0.11	0.02	0.75	0.75	0.11	0.01	NA
nouno	[56,]	-0.05	0.06	0.10	0.10	0.04	0.89	0.84	0.01	0.06	NA
pronop	[57,]	-0.26	-0.26	-0.18	-0.19	0.24	0.30	0.23	-0.14	-0.17	NA
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	0.14	0.19	0.06	0.22	-0.09	0.80	0.86	0.21	0.01	NA
artop	[60,]	0.08	0.17	0.08	0.20	-0.04	0.76	0.72	0.16	0.02	NA
totop	[61,]	0.23	0.34	0.39	0.20	-0.24	-0.08	-0.06	0.07	0.35	NA
prep%	[62,]	-0.18	-0.26	-0.45	-0.30	0.27	-0.08	-0.07	0.12	-0.45	NA
prepno	[63,]	-0.16	-0.14	-0.23	-0.18	0.21	0.43	0.40	0.06	-0.24	NA
adv%	[64,]	-0.15	-0.27	-0.08	0 02	0.04	0.0	-0.05	-0.29	-0.01	NA
advno	[65,]	-0.09	-0.19	-0.04	0.10	-0.01	0.16	0.13	-0.22	0.0	NA
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	-0.09	-0.09	-0.02	0.06	0.05	0.43	0.41	-0.11	-0.01	NA
sub-cno	[69,]	0.01	0.08	0.04	0.10	0.01	0.67	0.69	0.08	0.01	NA
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	0.02	0.09	0.06	0.04	0.0	0.27	0.27	0.06	0.04	NA
expno	[73,]	0.02	0.09	0.06	0.04	0.0	0.27	0.27	0.06	0.04	NA
abst	[74,]	-0.08	-0.05	0.29	0.11	-0.10	-0.16	-0.17	-0.45	0.43	NA
dict	[75,]	0.11	0.16	0.15	0.19	-0.10	0.86	0.87	0.07	0.11	NA
dict%s	[76,]	0.43	0.27	0.52	0.40	-0.54	-0.12	-0.11	-0.17	0.52	NA
tt.con	[77,]	0.02	-0.06	-0.07	-0.20	-0.02	-0.86	-0.82	-0.05	-0.02	NA
tt.fun	[78,]	-0.07	-0.18	0.01	-0.21	-0.01	-0.84	-0.83	-0.30	0.11	NA
var	[79,]	-0.59	-0.50	-0.45	-0.57	0.59	-0.08	-0.09	-0.24	-0.36	NA
sylrep	[80,]	-0.57	-0.52	-0.70	-0.78	0.72	-0.10	-0.08	0.06	-0.70	NA
ch.n	[81,]	0.07	0.18	0.06	0.18	-0.02	0.98	1.00	0.20	0.0	NA
ch.m	[82,]	0.61	0.57	0.68	0.76	-0.71	0.07	0.04	0.05	0.64	NA
ch.sd	[83,]	0.33	0.23	0.22	0.27	-0.37	0.05	0.03	0.11	0.15	NA
qualty	[84,]	0.01	0.12	-0.16	-0.08	0.16	0.48	0.51	0.37	-0.26	NA

CORRELATIONS AMONG VARIABLES

		noimp [,11]	nocwd [,12]	%cwds [,13]	evlen [,14]	shtse [,15]	%shts [,16]	nosht [,17]	lngs [,18]	%lngs [,19]	no ng [,20]
kin	[1,]	NA	0.08	0.16	0.65	0.56	0.11	0.05	0.56	0.09	0.09
eut	[2,]	NA	0.18	0.16	0.59	0.65	0.15	0.15	0.65	0.15	0.20
c-l	[3,]	NA	0.07	0.57	0.93	-0.10	-0.10	0.04	-0.10	-0.21	0.0
fiel	[4,]	NA	0.20	0.47	0.74	0.26	-0.20	0.09	0.26	-0.13	0.07
fle2	[5,]	NA	-0.03	-0.40	-0.82	-0.23	0.06	0.02	-0.23	0.07	0.01
nosnt	[6,]	NA	0.99	-0.02	0.07	0.11	0.11	0.93	0.11	0.0	0.77
nowds	[7,]	NA	1.00	-0.05	0.06	0.21	0.14	0.94	0.21	0.05	0.83
evslen	[8,]	NA	0.20	-0.45	-0.16	1.00	0.32	0.18	1.00	0.40	0.28
evwlen	[9,]	NA	0.01	0.68	0.90	-0.30	-0.17	-0.02	-0.30	-0.26	-0.07
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	NA	1.00	-0.03	0.06	0.20	0 13	0.94	0.20	0.04	0.82
%cwds	[13,]	NA	-0.03	1.00	0.41	-0.42	-0.40	-0.10	-0.42	-0.23	-0.12
evlencw	[14,]	NA	0.06	0.41	1.00	-0.15	-0.07	0.07	-0.15	-0.31	-0.02
shtsent	[15,]	NA	0.20	-0.42	-0.15	1.00	0.33	0.19	1.00	0.40	0.29
%shts	[16,]	NA	0.13	-0.40	-0.07	0.33	1.00	0.31	0.33	0.60	0.32
noshts	[17,]	NA	0.94	-0.10	0.07	0.19	0.31	1.00	0.19	0.15	0.91
lngs	[18,]	NA	0.20	-0.42	-0.15	1.00	0.33	0.19	1.00	0.40	0.29
%lngs	[19,]	NA	0.04	-0.23	-0.31	0.40	0.60	0.15	0.40	1.00	0.37
nolngs	[20,]	NA	0.82	-0.12	-0.02	0.29	0.32	0.91	0.29	0.37	1.00
lgests	[21,]	NA	0.82	-0.28	-0.08	0.61	0.44	0.83	0.61	0.42	0.87
where	[22,]	NA	0.53	-0.04	0.10	0.0	0.07	0.58	0.0	0.0	0.50
shtsts	[23,]	NA	-0.30	-0.18	-0.02	0.60	-0.26	-0.33	0.60	-0.09	-0.22
where	[24,]	NA	0.90	0.02	0.10	0.11	0.15	0.84	0.11	0.06	0.73
smpl%	[25,]	NA	-0.26	0.50	0.43	-0.38	-0.12	-0.22	-0.38	-0.09	-0.25
simpno	[26,]	NA	0.82	0.03	0.15	-0.05	0.07	0.79	-0.05	-0.07	0.55
cplex%	[27,]	NA	0.10	-0.39	-0.33	0.16	0.27	0.06	0.16	0.14	0.05
plexno	[28,]	NA	0.92	-0.02	0.0	0.14	0.10	0.78	0.14	0.0	0.65
pound%	[29,]	NA	0.45	-0.04	0.0	0.10	0.12	0.43	0.10	-0.05	0.35
poundno	[30,]	NA	0.85	-0.03	0.03	0.16	0.15	0.86	0.16	0.05	0.81
c-c%	[31,]	NA	0.10	-0.25	-0.24	0.37	-0.25	0.09	0.37	-0.03	0.18
c-cno	[32,]	NA	0.82	-0.13	0.04	0.25	0.12	0.89	0.25	0.12	0.91
verbt	[33,]	NA	0.99	-0.05	0.03	0.19	0.13	0.93	0.19	0.06	0.84
tobe%	[34,]	NA	-0.02	-0.27	0.34	[illegible]	0.04	0.02	0.21	-0.37	-0.06
tobeno	[35,]	NA	0.96	-0.06	0.09	0.22	0.16	0.92	0.22	0.02	0.84
aux%	[36,]	NA	-0.17	-0.37	-0.40	0.26	0.35	-0.09	0.26	0.18	0.01
auxno	[37,]	NA	0.87	-0.14	-0.10	0.25	0.22	0.87	0.25	0.14	0.89
inf%	[38,]	NA	[illegible]	0.31	-0.35	-0.05	0.15	-0.05	-0.05	0.42	0.07
infno	[39,]	NA	0.94	0.0	-0.05	0.22	0.11	0.83	0.22	0.09	0.78
pass%	[40,]	NA	[illegible]	-0.27	-0.20	-0.20	0.30	0.30	-0.20	0.39	0.29
passno	[41,]	NA	0.88	-0.08	0.04	0.12	0.21	0.96	0.12	0.12	0.91
prep%	[42,]	NA	-0.02	0.33	0.63	-0.23	-0.32	-0.04	-0.23	-0.44	-0.18
prepno	[43,]	NA	0.98	-0.04	0.10	0.18	0.12	0.94	0.18	0.01	0.80
conj%	[44,]	NA	0.31	-0.39	-0.44	0.42	0.30	0.33	0.42	0.66	0.39
conjno	[45,]	NA	0.94	-0.10	0.03	0.21	0.17	0.95	0.21	0.14	0.86
adv%	[46,]	NA	0.0	-0.10	-0.49	0.17	0.18	0.04	0.17	0.40	0.13
advno	[47,]	NA	0.97	-0.03	0.0	0.23	0.18	0.95	0.23	0.13	0.88
noun%	[48,]	NA	0.04	0.51	0.49	-0.18	-0.43	-0.03	-0.18	-0.39	-0.15
nounno	[49,]	NA	0.99	-0.04	0.08	0.19	0.12	0.94	0.19	0.03	0.80
adj%	[50,]	NA	-0.07	0.44	0.59	-0.03	-0.16	-0.10	-0.03	-0.40	-0.13
adjno	[51,]	NA	0.99	-0.01	0.08	0.21	0.12	0.90	0.21	0.02	0.78
pron%	[52,]	NA	0.07	-0.62	-0.40	0.11	0.38	0.11	0.11	0.43	0.15
pronno	[53,]	NA	0.92	-0.19	-0.02	0.23	0.18	0.89	0.23	0.10	0.82
nom%	[54,]	NA	-0.05	0.35	0.27	-0.39	-0.30	-0.11	-0.39	-0.34	-0.14
nomno	[55,]	NA	0.74	-0.08	0.07	0.11	0.02	0.71	0.11	-0.05	0.65
nouno	[56,]	NA	0.84	-0.03	0.07	0.01	0.06	0.82	0.01	-0.05	0.66
pronop	[57,]	NA	0.24	-0.17	-0.19	-0.15	0.12	0.15	-0.15	0.16	-0.03
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	NA	0.87	0.05	0.04	0.21	0.08	0.77	0.21	0.03	0.75
artop	[60,]	NA	0.72	-0.07	0.16	0.17	0.15	0.68	0.17	0.02	0.48
totop	[61,]	NA	-0.06	0.09	0.24	0.06	0.16	-0.07	0.06	0.13	-0.04
prep%	[62,]	NA	-0.08	-0.15	-0.38	0.11	-0.18	-0.06	0.11	-0.09	-0.05
prepno	[63,]	NA	0.39	-0.08	-0.20	0.07	0.04	0.47	0.07	-0.02	0.37
adv%	[64,]	NA	-0.04	0.07	0.05	-0.28	-0.07	-0.05	-0.28	-0.11	-0.09
advno	[65,]	NA	0.14	0.10	0.05	-0.21	-0.06	0.06	-0.21	-0.10	0.0
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	NA	0.41	0.0	0.10	-0.09	0.05	0.43	-0.09	-0.02	0.37
sub-cno	[69,]	NA	0.68	-0.04	0.10	0.10	0.13	0.71	0.10	0.06	0.68
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	NA	0.26	-0.05	0.15	0.07	0.08	0.30	0.07	0.02	0.21
expno	[73,]	NA	0.26	-0.05	0.15	0.07	0.08	0.30	0.07	0.02	0.21
abst	[74,]	NA	-0.17	0.57	0.26	-0.44	-0.08	-0.13	-0.44	0.12	-0.04
dict	[75,]	NA	0.87	0.05	0.15	0.08	0.09	0.89	0.08	0.0	0.86
dict%s	[76,]	NA	-0.11	0.36	0.58	-0.17	-0.20	-0.08	-0.17	-0.33	-0.07
tt.con	[77,]	NA	-0.83	-0.07	0.0	-0.05	0.09	[illegible]	-0.05	0.17	-0.51
tt.fun	[78,]	NA	-0.83	0.10	0.01	-0.29	-0.07	-0.7[illegible]	-0.29	-0.08	-0.59
var	[79,]	NA	-0.09	-0.10	-0.57	-0.24	0.20	-0.06	-0.24	0.60	0.09
sylrep	[80,]	NA	-0.10	-0.71	-0.63	0.03	0.13	-0.07	0.03	0.13	0.02
ch.n	[81,]	NA	1.00	-0.05	0.06	0.20	0.14	0.94	0.20	0.05	0.83
ch.m	[82,]	NA	0.05	0.38	0.58	0.07	-0.16	-0.02	0.07	-0.15	-0.08
ch.sd	[83,]	NA	0.03	-0.15	0.29	0.10	-0.17	0.01	0.10	-0.03	-0.03
qualty	[84,]	NA	0.50	-0.40	-0.21	0.37	0.24	0.49	0.37	0.18	0.44

CORRELATIONS AMONG VARIABLES

		lgest [,21]	where [,22]	shtst [,23]	where [,24]	smpl% [,25]	simpn [,26]	cplex [,27]	plexn [,28]	pound [,29]	pound [,30]
kin	[1,]	0.28	-0.09	0.45	0.06	0.05	-0.05	-0.11	0.04	-0.02	0.04
aut	[2,]	0.39	0.0	0.45	0.13	0.01	0.03	-0.09	0.12	0.02	0.12
c-i	[3,]	-0.05	0.02	0.01	0.09	0.37	0.11	-0.26	0.04	-0.05	0.02
fle1	[4,]	0.19	0.03	0.21	0.19	0.12	0.13	-0.20	0.18	0.08	0.13
fle2	[5,]	-0.07	0.09	-0.28	-0.05	-0.20	0.02	0.17	-0.01	0.05	0.01
nosnt	[6,]	0.77	0.57	-0.37	0.88	-0.23	0.89	0.09	0.90	0.45	0.83
nowds	[7,]	0.83	0.54	-0.29	0.89	-0.26	0.82	0.10	0.91	0.45	0.85
avslen	[8,]	0.61	-0.01	0.60	0.11	-0.39	-0.06	0.17	0.15	0.10	0.16
avwlen	[9,]	-0.19	0.C	-0.10	0.05	0.45	0.09	-0.30	-0.02	-0.09	-0.03
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	0.82	0.53	-0.30	0.90	-0.26	0.82	0.10	0.92	0.45	0.85
%cwds	[13,]	-0.28	-0.04	-0.18	0.02	0.50	0.03	-0.39	-0.02	-0.04	-0.03
avlencw	[14,]	-0.08	0.10	-0.02	0.10	0.43	0.15	-0.33	0.0	0.0	0.03
shtsent	[15,]	0.61	0.0	0.60	0.11	-0.38	-0.05	0.16	0.14	0.10	0.16
%shts	[16,]	0.44	0.07	-0.26	0.15	-0.12	0.07	0.27	0.10	0.12	0.15
noshts	[17,]	0.83	0.58	-0.33	0.84	-0.22	0.79	0.06	0.78	0.43	0.86
lngs	[18,]	0.61	0.0	0.60	0.11	-0.38	-0.05	0.16	0.14	0.10	0.16
%lngs	[19,]	0.42	0.0	-0.09	0.06	-0.09	-0.07	0.14	0.0	-0.05	0.05
nolngs	[20,]	0.87	0.50	-0.22	0.73	-0.25	0.55	0.05	0.65	0.35	0.81
lgests	[21,]	1.00	0.43	-0.10	0.71	-0.43	0.53	0.23	0.72	0.45	0.77
where	[22,]	0.43	1.00	-0.44	0.46	-0.22	0.52	0.07	0.46	0.34	0.47
shtsts	[23,]	-0.10	-0.44	1.00	-0.30	-0.05	-0.40	-0.12	-0.33	-0.30	-0.29
where	[24,]	0.71	0.46	-0.30	1.00	-0.19	0.70	0.13	0.90	0.26	0.65
smpl%	[25,]	-0.43	-0.22	-0.05	-0.19	1.00	0.06	-0.78	-0.36	-0.29	-0.29
simpno	[26,]	0.53	0.52	-0.40	0.70	0.06	1.00	-0.11	0.66	0.31	0.66
cplex%	[27,]	0.23	0.07	-0.12	0.13	-0.78	-0.11	1.00	0.30	0.06	0.03
plexno	[28,]	0.72	0.46	-0.33	0.90	-0.36	0.66	0.30	1.00	0.39	0.68
pound%	[29,]	0.45	0.34	-0.30	0.26	-0.29	0.31	0.06	0.39	1.00	0.75
poundno	[30,]	0.77	0.47	-0.29	0.65	-0.29	0.66	0.03	0.68	0.75	1.00
c-c%	[31,]	0.19	0.10	0.39	0.03	-0.41	-0.06	-0.19	0.0	-0.06	0.09
c-cno	[32,]	0.78	0.57	-0.13	0.67	-0.30	0.61	-0.04	0.62	0.34	0.82
verbt	[33,]	0.83	0.49	-0.29	0.90	-0.29	0.76	0.13	0.93	0.43	0.84
tobe%	[34,]	-0.02	-0.12	0.43	-0.02	0.16	-0.08	-0.15	-0.05	0.02	0.01
tobeno	[35,]	0.81	0.45	-0.21	0.90	-0.27	0.70	0.11	0.89	0.39	0.82
aux%	[36,]	0.04	-0.14	0.25	-0.16	-0.33	-0.30	0.20	-0.17	-0.02	-0.05
auxno	[37,]	0.83	0.51	-0.22	0.71	-0.38	0.55	0.13	0.73	0.54	0.90
inf%	[38,]	0.03	-0.16	-0.19	-0.04	0.05	-0.06	-0.01	-0.01	-0.22	-0.07
infno	[39,]	0.81	0.32	-0.24	0.85	-0.28	0.67	0.15	0.91	0.35	0.78
pass%	[40,]	0.21	0.28	-0.55	0.22	-0.09	0.27	0.23	0.20	0.08	0.19
passno	[41,]	0.77	0.62	-0.33	0.76	-0.19	0.77	0.04	0.70	0.36	0.83
prep%	[42,]	-0.25	-0.13	0.07	-0.02	0.50	0.22	-0.48	-0.11	0.15	0.02
prepno	[43,]	0.79	0.50	-0.28	0.86	-0.20	0.88	0.04	0.86	0.46	0.85
conj%	[44,]	0 50	0.20	-0.07	0.18	-0.26	0.30	-0.01	0.21	0.26	0.31
conjno	[45,]	0.82	0.55	-0.27	0.77	-0.23	0.85	0.01	0.77	0.44	0.87
adv%	[46,]	0.15	0.09	-0.04	0.07	-0.19	-0.05	0.18	0.04	-0.09	-0.05
advno	[47,]	0.84	0.57	-0.29	0.91	-0.28	0.74	0.12	0.90	0.38	0.81
noun%	[48,]	-0.16	0.12	0.0	0.0	0.48	0.30	-0.37	-0.04	0.10	0.02
nounno	[49,]	0.81	0.56	-0.30	0.87	-0.23	0.87	0.07	0.89	0.45	[illegible]
adj%	[50,]	-0.18	-0.21	0.29	-0.03	0.31	-0.18	-0.34	-0.05	-0.04	-0.03
adjno	[51,]	0.81	0.46	-0.26	0.91	-0.28	0.75	0.12	0.94	0.44	0.82
pron%	[52,]	0.23	0.04	-0.31	0.01	-0.31	0.05	0.29	0.07	0.14	0.07
pronno	[53,]	0.82	0.50	-0.30	0.75	-0.27	0.75	0.08	0.79	0.44	0.81
nom%	[54,]	-0.26	-0.06	-0.13	-0.14	0.26	[illegible]	-0.21	-0.13	-0.05	-0.02
nomno	[55,]	0.60	0.41	-0.20	0.48	-0.08	0.77	-0.06	0 52	0.38	0.74
nouno	[56,]	0.62	0.46	-0.35	0.63	-0.17	0.89	0.02	0.67	0.45	0.80
pronop	[57,]	0.05	-0.08	-0.28	0.19	0.09	0.50	0.02	0.21	0.13	0.15
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	0.75	0.30	-0.19	0.86	-0.24	0.51	0.12	0.87	0.36	0.72
artop	[60,]	0.56	0.76	-0.30	0.68	-0.25	0.69	0.15	0.72	0.33	0.49
totop	[61,]	0.02	-0.17	-0.01	-0.17	-0.08	-0.13	0.27	-0.04	0.12	0.03
prep%	[62,]	-0.09	-0.09	0.30	-0.09	0.04	-0.04	-0.35	-0.15	-0.06	-0.05
prepno	[63,]	0.29	0 23	-0.09	0.21	0.06	0.59	-0.28	0.16	0.22	0.44
adv%	[64,]	-0.09	0.01	-0.27	0.13	0.16	0.08	-0.06	-0.01	-0.14	-0.11
advno	[65,]	0.04	-0.05	-0.26	0.35	0.09	0.15	0.02	0.24	-0.11	-0.02
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	0.30	0.63	-0.34	0.53	-0.04	0.37	0.05	0.44	-0.01	0.20
sub-cno	[69,]	0.57	0.74	-0.27	0.74	-0.17	0.50	0.09	0.67	0.12	0.46
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	0.17	0.75	-0.14	0.37	-0.09	0.22	0.07	0.31	0.0	0.04
expno	[73,]	0.17	0.75	-0.14	0.37	-0.09	0.22	0.07	0.31	0.0	0.04
abst	[74,]	-0.29	0.08	-0.33	-0.08	0.38	-0.12	-0.34	-0.17	-0.16	-0.18
dict	[75,]	0.72	0.44	-0.24	0.71	-0.10	0.74	-0.06	0.68	0.31	0.83
dict%s	[76,]	-0.23	-0.13	0.21	-0.09	0.58	-0.07	-0.53	-0.17	-0.25	-0.09
tt.con	[77,]	-0.59	-0.48	0.34	-0.72	0.28	-0.78	-0.12	-0.79	-0.50	-0.73
tt.fun	[78,]	-0.75	-0.61	0.29	-0.69	0.27	-0.74	-0.15	-0.78	-0.58	-0.73
var	[79,]	-0.03	-0.05	-0.42	-0.07	-0.06	-0.06	0.18	-0.06	-0.17	-0.10
sylrep	[80,]	0.0	-0.09	0.10	-0.16	-0.42	-0.17	0.39	-0.07	-0.06	-0.02
ch.n	[81,]	0.83	0.55	-0.30	0.89	-0.26	0.82	0.10	0.91	0.45	0.85
ch.m	[82,]	0.0	-0.03	0.13	0.0	0.01	0.12	-0.04	0.03	-0.17	-0.04
ch.sd	[83,]	0 02	0.0	0.18	-0.02	0.11	0.17	0.01	-0.01	0.0	0.0
qualty	[84,]	0.56	0.34	-0.10	0.42	-0.58	0.32	0.29	0.48	0.40	0.44

CORRELATIONS AMONG VARIABLES

		c-c% [,31]	c-cno [,32]	verbt [,33]	tobe% [,34]	toben [,35]	aux% [,36]	auxno [,37]	inf% [,38]	infno [,39]	pass% [,40]
kin	[1,]	0.08	0.09	0.07	0.36	0.13	-0.14	-0.01	-0.20	0.07	-0.27
aut	[2,]	0.10	0.19	0.16	0.22	0.20	-0.06	0.12	-0.11	0.16	-0.34
c-l	[3,]	-0.22	0.02	0.05	0.12	0.07	-0.32	-0.07	-0.12	0.02	-0.24
fle1	[4,]	0.06	0.13	0.17	0.17	0.20	-0.46	-0.01	-0.17	0.17	-0.27
fle2	[5,]	0.06	0.0	-0.01	-0.30	-0.07	0.34	0.12	0.22	-0.01	0.22
nosnt	[6,]	0.07	0.80	0.97	-0.05	0.92	-0.21	0.81	-0.04	0.90	0.25
nowds	[7,]	0.11	0.83	0.99	-0.02	0.96	-0.17	0.86	-0.02	0.94	0.22
avslen	[8,]	0.37	0.25	0.19	0.21	0.22	0.26	0.25	-0.07	0.22	-0.19
avwlen	[9,]	-0.28	-0.04	-0.02	0.04	0.01	-0.37	-0.13	-0.06	-0.04	-0.22
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	0.10	0.82	0.99	-0.02	0.96	-0.17	0.85	-0.01	0.94	0.21
%cwds	[13,]	-0.25	-0.13	-0.05	-0.27	-0.06	-0.37	-0.14	0.31	0.0	-0.27
avlencw	[14,]	-0.24	0.04	0.03	0.34	0.09	-0.40	-0.10	-0.35	-0.05	-0.20
shtsent	[15,]	0.37	0.25	0.19	0.21	0.22	0.26	0.25	-0.05	0.22	-0.20
%shts	[16,]	-0.25	0.12	0.13	0.04	0.16	0.35	0.22	0.15	0.11	0.30
noshts	[17,]	0.09	0.89	0.93	0.02	0.92	-0.09	0.87	-0.05	0.83	0.30
lngs	[18,]	0.37	0.25	0.19	0.21	0.22	0.26	0.25	-0.05	0.22	-0.20
%lngs	[19,]	-0.03	0.12	0.06	-0.37	0.02	0.18	0.14	0.42	0.09	0.39
nolngs	[20,]	0.18	0.91	0.84	-0.06	0.84	0.01	0.89	0.07	0.78	0.29
lgests	[21,]	0.19	0.78	0.83	-0.02	0.81	0.04	0.83	0.03	0.81	0.21
where	[22,]	0.10	0.57	0.49	-0.12	0.45	-0.14	0.51	-0.16	0.32	0.28
shtsts	[23,]	0.39	-0.13	-0.29	0.43	-0.21	0.25	-0.22	-0.19	-0.24	-0.55
whern	[24,]	0.03	0.67	0.90	-0.02	0.90	-0.16	0.71	-0.04	0.85	0.22
smpl%	[25,]	-0.41	-0.30	-0.29	0.16	-0.27	-0.33	-0.38	0.05	-0.28	-0.09
simpno	[26,]	-0.06	0.61	0.76	-0.08	0.70	-0.30	0.55	-0.06	0.67	0.27
cplex%	[27,]	-0.19	-0.04	0.13	-0.15	0.11	0.20	0.13	-0.01	0.15	0.23
plexno	[28,]	0.0	0.62	0.93	-0.05	0.89	-0.17	0.73	-0.01	0.91	0.20
pound%	[29,]	-0.06	0.34	0.43	0.02	0.39	-0.02	0.54	-0.22	0.35	0.08
poundno	[30,]	0.09	0.82	0.84	0.01	0.82	-0.05	0.90	-0.07	0.78	0.19
c-c%	[31,]	1.00	0.42	0.12	-0.06	0.11	0.27	0.21	0.03	0.10	-0.21
c-cno	[32,]	0.42	1.00	0.83	0.0	0.82	-0.03	0.86	-0.02	0.74	0.16
verbt	[33,]	0.12	0.83	1.00	-0.02	0.97	-0.14	0.87	0.02	0.97	0.21
tobe%	[34,]	-0.06	0.0	-0.02	1.00	0.14	-0.09	-0.06	-0.60	-0.07	-0.10
tobeno	[35,]	0.11	0.82	0.97	0.14	1.00	-0.14	0.84	-0.04	0.93	0.18
aux%	[36,]	0.27	-0.03	-0.14	-0.09	-0.14	1.00	0.20	0.23	-0.10	-0.20
auxno	[37,]	0.21	0.86	0.87	-0.06	0.84	0.20	1.00	0.05	0.84	0.17
inf%	[38,]	0.03	-0.02	0.02	-0.60	-0.04	0.23	0.05	1.00	0.18	0.05
infno	[39,]	0.10	0.74	0.97	-0.07	0.93	-0.10	0.84	0.18	1.00	0.18
pass%	[40,]	-0.21	0.16	0.21	-0.10	0.18	-0.20	0.17	0.05	0.18	1.00
passno	[41,]	0.10	0.90	0.87	0.0	0.86	-0.13	0.86	-0.04	0.78	0.40
prep%	[42,]	-0.21	-0.14	-0.07	0.30	-0.03	-0.37	-0.18	-0.43	-0.13	-0.04
prepno	[43,]	0.07	0.81	0.96	0.0	0.93	-0.20	0.82	-0.06	0.90	0.23
conj%	[44,]	0.34	0.36	0.29	-0.45	0.18	0.15	0.31	0.24	0.26	0.23
conjno	[45,]	0.18	0.90	0.92	-0.08	0.87	-0.13	0.85	-0.01	0.84	0.23
adv%	[46,]	0.10	0.03	0.01	-0.47	-0.03	0.31	0.06	0.33	0.04	-0.03
advno	[47,]	0.12	0.85	0.97	-0.06	0.96	-0.12	0.87	0.02	0.92	0.21
noun%	[48,]	-0.30	-0.11	-0.04	0.05	-0.05	-0.53	-0.17	-0.35	-0.10	-0.17
nounno	[49,]	0.08	0.82	0.97	-0.02	0.93	-0.19	0.83	-0.04	0.91	0.23
adj%	[50,]	-0.01	-0.07	-0.05	0.60	0.06	-0.14	-0.10	-0.13	-0.02	-0.35
adjno	[51,]	0.10	0.77	0.99	0.03	0.97	-0.17	0.82	0.0	0.96	0.17
pron%	[52,]	0.04	0.07	0.09	-0.31	0.01	0.09	0.13	0.15	0.10	0.50
pronno	[53,]	0.14	0.81	0.92	-0.10	0.86	-0.13	0.86	0.01	0.88	0.27
nom%	[54,]	-0.10	-0.07	-0.07	-0.05	-0.08	-0.32	-0.10	0.11	-0.05	-0.10
nomno	[55,]	0.06	0.68	0.71	-0.05	0.64	-0.17	0.71	0.01	0.68	0.20
nouno	[56,]	0.06	0.74	0.81	-0.17	0.70	-0.15	0.73	-0.05	0.73	0.24
pronop	[57,]	-0.23	-0.06	0.20	-0.21	0.10	-0.09	0.05	0.23	0.23	0.42
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	0.08	0.67	0.90	0.09	0.93	-0.14	0.77	0.04	0.92	0.08
artop	[60,]	0.05	0.57	0.67	0.02	0.65	-0.24	0.46	-0.15	0.52	0.14
totop	[61,]	-0.32	-0.09	-0.06	-0.14	-0.10	0.07	-0.01	0.20	-0.01	-0.08
prep%	[62,]	0.49	0.02	-0.08	0.17	-0.05	0.15	-0.04	-0.08	-0.09	-0.08
prepno	[63,]	0.23	0.40	0.35	0.03	0.32	0.03	0.36	-0.08	0.29	0.20
adv%	[64,]	[illegible]	-0.10	-0.04	-0.01	-0.03	-0.25	-0.14	-0.18	-0.06	0.15
advno	[65,]	[illegible]	-0.05	0.16	0.01	0.18	-0.27	-0.04	-0.13	0.18	0.12
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	0.0	0.37	0.40	0.03	0.42	-0.23	0.32	-0.18	0.29	0.26
sub-cno	[69,]	0.09	0.66	0.68	0.05	0.71	-0.15	0.62	-0.10	0.57	0.24
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	0.06	0.23	0.23	0.06	0.28	-0.05	0.19	-0.11	0.09	0.13
expno	[73,]	0.06	0.23	0.23	0.06	0.28	-0.05	0.19	-0.11	0.09	0.13
abst	[74,]	-0.07	-0.08	-0.17	-0.36	-0.16	-0.27	-0.18	0.34	-0.19	-0.09
dict	[75,]	0.11	0.88	0.87	0.08	0.86	-0.20	0.82	-0.01	0.83	0.26
dict%s	[76,]	-0.08	-0.01	-0.10	0.60	-0.02	-0.32	-0.14	-0.24	-0.08	-0.08
tt.con	[77,]	-0.08	-0.58	-0.81	0.23	-0.71	0.13	-0.68	0.02	-0.78	-0.13
tt.fun	[78,]	0.02	-0.60	-0.79	0.13	-0.71	0.26	-0.65	0.11	-0.73	-0.16
ver	[79,]	-0.10	-0.06	-0.06	-0.72	-0.16	0.21	0.02	0.67	0.02	0.44
sylrep	[80,]	0.16	0.04	-0.05	-0.07	-0.06	0.52	0.13	0.08	-0.02	0.06
ch.n	[81,]	0.09	0.82	0.99	-0.02	0.96	-0.16	0.87	-0.01	0.94	0.23
ch.m	[82,]	0.13	0.08	0.03	-0.07	0.0	-0.25	-0.13	-0.06	0.01	-0.39
ch.sd	[83,]	-0.19	-0.06	-0.01	0.31	0.01	-0.33	-0.09	-0.59	-0.06	0.28
quality	[84,]	0.38	0.47	0.49	-0.21	0.44	0.31	0.50	0.06	0.44	0.21

CORRELATIONS AMONG VARIABLES

		passno [,41]	prep% [,42]	prepno [,43]	conj% [,44]	conjno [,45]	adv% [,46]	advno [,47]	noun% [,48]	nounno [,49]	adj% [,50]
k n	[1,]	-0.01	0.33	0.08	-0.10	0.05	-0.23	0.06	0.22	0.07	0.54
aut	[2,]	0.09	0.25	0.18	0.01	0.17	-0.18	0.17	0.23	0.17	0.42
c-i	[3,]	0.03	0.56	0.09	-0.39	0.04	-0.38	0.02	0.45	0.08	0.58
fle1	[4,]	0.06	0.45	0.20	-0.19	0.13	-0.33	0.14	0.44	0.19	0.56
fle2	[5,]	0.04	-0.49	-0.04	0.29	0.02	0.35	0.01	-0.38	-0.02	-0.62
nosnt	[6,]	0.88	0.02	0.98	0.32	0.95	-0.01	0.94	0.09	0.99	-0.11
nowds	[7,]	0.89	-0.02	0.99	0.32	0.95	0.0	0.97	0.03	0.99	-0.08
avslen	[8,]	0.11	-0.24	0.17	0.42	0.21	0.15	0.22	-0.18	0.18	-0.03
avwlen	[9,]	-0.02	0.56	0.03	-0.46	-0.03	-0.39	-0.04	0.48	0.01	0.56
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	0.88	-0.02	0.98	0.31	0.94	0.0	0.97	0.04	0.99	-0.07
%cwds	[13,]	-0.08	0.33	-0.04	-0.39	-0.10	-0.10	-0.03	0.51	-0.04	0.44
avlencw	[14,]	0.04	0.63	0.10	-0.44	0.03	-0.49	0.0	0.49	0.08	0.59
shtsent	[15,]	0.12	-0.23	0.18	0.42	0.21	0.17	0.23	-0.18	0.19	-0.03
%shts	[16,]	0.21	-0.32	0.12	0.30	0.17	0.18	0.18	-0.43	0.12	-0.16
noshts	[17,]	0.96	-0.04	0.94	0.33	0.95	0.04	0.95	-0.03	0.94	-0.10
lngs	[18,]	0.12	-0.23	0.18	0.42	0.21	0.17	0.23	-0.18	0.19	-0.03
%lngs	[19,]	0.12	-0.44	0.01	0.66	0.14	0.40	0.13	-0.39	0.03	-0.40
nolngs	[20,]	0.91	-0.18	0.80	0.39	0.86	0.13	0.88	-0.15	0.80	-0.13
lgests	[21,]	0.77	-0.25	0.79	0.50	0.82	0.15	0.84	-0.16	0.81	-0.18
where	[22,]	0.62	-0.13	0.50	0.20	0.55	0.09	0.57	0.12	0.56	-0.21
shtsts	[23,]	-0.33	0.07	-0.28	-0.07	-0.27	-0.04	-0.29	0.0	-0.30	0.29
where	[24,]	0.76	-0.02	0.86	0.18	0.77	0.07	0.91	0.0	0.87	-0.03
smpl%	[25,]	-0.19	0.30	-0.20	-0.26	-0.23	-0.19	-0.28	0.48	-0.23	0.31
simpno	[26,]	0.77	0.22	0.88	0.30	0.85	-0.05	0.74	0.30	0.87	-0.18
cplex%	[27,]	0.04	-0.48	0.04	-0.01	0.01	0.18	0.12	-0.37	0.07	-0.34
plexno	[28,]	0.70	-0.11	0.86	0.21	0.77	0.04	0.90	-0.04	0.89	-0.05
pound%	[29,]	0.36	0.15	0.46	0.26	0.44	-0.09	0.38	0.10	0.45	-0.04
poundno	[30,]	0.83	0.02	0.85	0.31	0.87	-0.05	0.81	0.02	0.85	-0.03
c-c%	[31,]	0.10	-0.21	0.07	0.34	0.18	0.10	0.12	-0.30	0.08	-0.01
c-cno	[32,]	0.90	-0.14	0.81	0.36	0.90	0.03	0.85	-0.11	0.82	-0.07
verbt	[33,]	0.87	-0.07	0.96	0.29	0.92	0.01	0.97	-0.04	0.97	-0.05
tobe%	[34,]	0.0	0.30	0.0	-0.45	-0.08	-0.47	-0.06	0.05	-0.02	0.60
tobeno	[35,]	0.86	-0.03	0.93	0.18	0.87	-0.03	0.96	-0.05	0.93	0.06
aux%	[36,]	-0.13	-0.37	-0.20	0.15	-0.13	0.31	-0.12	-0.53	-0.19	-0.14
auxno	[37,]	0.86	-0.18	0.82	0.21	0.85	0.06	0.87	-0.17	0.83	-0.10
inf%	[38,]	-0.04	-0.43	-0.06	0.24	-0.01	0.33	0.02	-0.35	-0.04	-0.13
infno	[39,]	0.78	-0.13	0.90	0.26	0.84	0.04	0.92	-0.10	0.91	-0.02
pass%	[40,]	0.40	-0.04	0.23	0.23	0.23	-0.03	0.21	-0.17	0.23	-0.35
passno	[41,]	1.00	0.0	0.90	0.28	0.92	0.01	0.90	0.02	0.90	-0.13
prep%	[42,]	0.0	1.00	0.10	-0.27	-0.03	-0.38	-0.10	0.71	0.03	0.28
prepno	[43,]	0.90	0.10	1.00	0.31	0.95	-0.04	0.94	0.12	0.99	-0.09
conj%	[44,]	0.28	-0.27	0.31	1.00	0.46	0.46	0.32	-0.24	0.33	-0.64
conjno	[45,]	0.92	-0.03	0.95	0.46	1.00	0.02	0.91	0.03	0.96	-0.16
adv%	[46,]	0.01	-0.38	-0.04	0.46	0.02	1.00	0.16	-0.29	-0.02	-0.57
advno	[47,]	0.90	-0.10	0.94	0.32	0.91	0.16	1.00	-0.04	0.95	-0.12
noun%	[48,]	0.02	0.71	0.12	-0.24	0.03	-0.29	-0.04	1.00	0.10	0.09
nounno	[49	0.90	0.03	0.99	0.33	0.96	-0.02	0.95	0.10	1.00	-0.10
adj%	50	-0.13	0.28	-0.09	-0.64	-0.16	-0.57	-0.12	0.09	-0.10	1.00
adjno	[51,]	0.82	-0.03	0.96	0.24	0.89	-0.02	0.95	0.0	0.96	0.02
pron%	[52,]	0.10	-0.32	0.08	0.52	0.13	0.05	0.05	-0.48	0.08	-0.57
pronno	[53,]	0.88	-0.03	0.93	0.40	0.91	-0.04	0.89	-0.02	0.92	-0.22
nom%	[54,]	-0.03	0.30	0.01	-0.29	-0.03	-0.57	-0.16	0.45	-0.01	0.16
nomno	[55,]	0.78	0.16	0.81	0.28	0.81	-0.18	0.66	0.23	0.79	-0.16
nou no	[56,]	0.82	0.12	0.88	0.37	0.91	-0.10	0.75	0.18	0.88	-0.20
pronop	[57,]	0.12	0.05	0.30	0.37	0.25	-0.10	0.09	-0.08	0.27	-0.26
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	0.69	-0.09	0.81	0.08	0.70	0.08	0.89	-0.10	0.81	0.12
artop	[60,]	0.60	-0.07	0.69	0.31	0.70	0.04	0.71	0.11	0.74	-0.10
totop	[61,]	-0.11	-0.09	-0.06	-0.08	-0.04	-0.19	-0.11	-0.19	-0.07	0.17
prep%	[62,]	-0.04	0.02	-0.06	0.20	-0.04	0.20	-0.05	-0.04	-0.07	-0.04
prepno	[63,]	0.49	0.13	0.47	0.30	0.49	0.08	0.37	0.14	0.44	-0.15
adv%	[64,]	-0.03	0.17	-0.03	-0.18	-0.11	0.02	-0.04	0.35	-0.03	-0.19
advno	[65,]	0.02	0.13	0.12	-0.18	-0.01	0.02	0.14	0.27	0.12	-0.10
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	0.48	0.02	0.38	0.0	0.35	0.07	0.48	0.22	0.42	-0.18
sub-cno	[69,]	0.74	-0.07	0.64	0.11	0.62	0.08	0.77	0.04	0.68	-0.09
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	0.30	-0.09	0.22	0.05	0.20	0.08	0.34	-0.01	0.27	-0.03
expno	[73,]	0.30	-0.09	0.22	0.05	0.20	0.08	0.34	-0.01	0.27	-0.03
abst	[74,]	-0.12	0.02	-0.19	-0.08	-0.15	-0.17	-0.14	0.15	-0.18	0.15
dict	[75,]	0.94	0.07	0.88	0.15	0.89	-0.08	0.85	0.10	0.87	0.04
dict%s	[76,]	-0.01	0.40	-0.10	-0.55	-0.12	-0.47	-0.13	0.27	-0.11	0.76
tt.con	[77,]	-0.66	-0.08	-0.83	-0.25	-0.76	-0.13	-0.77	-0.16	-0.84	0.17
tt.fun	[78,]	-0.67	0.03	-0.82	-0.44	-0.78	-0.19	-0.79	-0.17	-0.84	0.28
var	[79,]	-0.01	-0.48	-0.12	0.47	-0.02	0.46	-0.03	-0.45	-0.10	-0.64
sylrep	[80,]	-0.05	-0.53	-0.11	0.28	-0.02	0.30	-0.07	-0.59	-0.10	-0.53
ch.n	[81,]	0.90	-0.01	0.99	0.31	0.94	0.0	0.97	0.04	0.99	-0.08
ch.m	[82,]	-0.07	0.18	0.04	-0.14	0.07	-0.15	-0.01	0.28	0.05	0.34
ch.sd	[83,]	0.09	0.57	0.10	0.0	0.04	-0.20	-0.02	0.53	0.08	-0.11
qualty	[84,]	0.43	-0.11	0.49	0.55	0.51	0.34	0.52	-0.34	0.49	-0.36

CORRELATIONS AMONG VARIABLES

		adjno [,51]	pron% [,52]	pronn [,53]	nom% [,54]	nomno [,55]	nouno [,56]	prono [,57]	posop [,58]	adjop [,59]	artop [,60]
kin	[1,]	0.13	-0.34	0.0	-0. [illegible]	0.01	-0.05	-0.26	NA	0.14	0.08
aut	[2,]	0.21	-0.26	0.14	-0.07	0.11	0.06	-0.26	NA	0.19	0.17
c-l	[3,]	0.10	-0.42	-0.01	0.23	0.05	0.10	-0 18	NA	0.06	0.08
fle1	[4,]	0.24	-0.42	0.08	0.11	0.11	0.10	-0.19	NA	0.22	0.20
fle2	[5,]	-0.08	0.46	0.09	-0.10	0.02	0.04	0.24	NA	-0.09	-0.04
nosnt	[6,]	0.96	0.07	0.90	-0.03	0.75	0.89	0.30	NA	0.80	0.76
nowds	[7,]	0.98	0.09	0.93	-0.04	0.75	0.84	0.23	NA	0.86	0.72
avslen	[8,]	0.21	0.12	0.23	-0.38	0.11	0.01	-0.14	NA	0.21	0.16
avwlen	[9,]	0.03	-0.45	-0.08	0.33	0.01	0.06	-0.17	NA	0.01	0.02
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	0.99	0.07	0.92	-0.05	0.74	0.84	0.24	NA	0.87	0.72
%cwds	[13,]	-0.01	-0.62	-0.19	0.35	-0.08	-0.03	-0.17	NA	0.05	-0.07
avlencw	[14,]	0.08	-0.40	-0.02	0.27	0.07	0.07	-0.19	NA	0.04	0.16
shtsent	[15,]	0.21	0.11	0.23	-0.39	0.11	0.01	-0.15	NA	0.21	0.17
%shts	[16,]	0.12	0.38	0.18	-0.30	0.02	0.06	0.12	NA	0.08	0.15
noshts	[17,]	0.90	0.11	0.89	-0.11	0.71	0.82	0.15	NA	0.77	0.68
ings	[18,]	0.21	0.11	0.23	-0.39	0.11	0.01	-0.15	NA	0.21	0.17
%ings	[19,]	0.02	0.43	0.10	-0.34	-0.05	-0.05	0.16	NA	0.03	0.02
nolngs	[20,]	0.78	0.15	0.82	-0.14	0.65	0.66	-0.03	NA	0.75	0.48
lgests	[21,]	0.81	0.23	0.82	-0.26	0.60	0.62	0.05	NA	0.75	0.56
where	[22,]	0.46	0.04	0.50	-0.06	0.41	0.46	-0.08	NA	0.30	0.76
shtsts	[23,]	-0.26	-0.31	-0.30	-0.13	-0.20	-0.35	-0.28	NA	-0.19	-0.30
where	[24,]	0.91	0.01	0.75	-0.14	0.48	0.63	0.19	NA	0.86	0.68
smpl%	[25,]	-0.28	-0.31	-0.27	0.26	-0.08	-0.17	0.09	NA	-0.24	-0.25
simpno	[26,]	0.75	0.05	0.75	0.11	0.77	0.89	0.50	NA	0.51	0.69
cplex%	[27,]	0.12	0.29	0.08	-0.21	-0.06	0.02	0.02	NA	0.12	0.15
plexno	[28,]	0.94	0.07	0.79	-0.13	0.52	0.67	0.21	NA	0.87	0.72
pound%	[29,]	0.44	0.14	0.44	-0.05	0.38	0.45	0.13	NA	0.36	0.33
poundno	[30,]	0.82	0.07	0.81	-0.02	0.74	0.80	0.15	NA	0.72	0.49
c-c%	[31,]	0.10	0.04	0.14	-0.10	0.06	0.06	-0.23	NA	0.08	0.05
c-cno	[32,]	0.77	0.07	0.81	-0.07	0.68	0.74	-0.06	NA	0.67	0.57
verbt	[33,]	0.99	0.09	0.92	-0.07	0.71	0.81	0.20	NA	0.90	0.67
tobe%	[34,]	0.03	-0.31	-0.10	-0.05	-0.05	-0.17	-0.21	NA	0.09	0.02
tobeno	[35,]	0.97	0.01	0.86	-0.08	0.64	0.70	0.10	NA	0.93	0.65
aux%	[36,]	-0.17	0.09	-0.13	-0.32	-0.17	-0.15	-0.09	NA	-0.14	-0.24
auxno	[37,]	0.82	0.13	0.86	-0.10	0.71	0.73	0.05	NA	0.77	0.46
inf%	[38,]	0.0	0.15	0.01	0.11	0.01	-0.05	0.23	NA	0.04	-0.15
infno	[39,]	0.96	0.10	0.88	-0.05	0.68	0.73	0.23	NA	0.92	0.52
pass%	[40,]	0.17	0.50	0.27	-0.10	0.20	0.24	0.42	NA	0.08	0.14
passno	[41,]	0.82	0.10	0.88	-0.03	0.78	0.82	0.12	NA	0.69	0.60
prep%	[42,]	-0.03	-0.32	-0.03	0.30	0.16	0.12	0.05	NA	-0.09	-0.07
prepno	[43,]	0.96	0.08	0.93	0.01	0.81	0.88	0.30	NA	0.81	0.69
conj%	[44,]	0.24	0 52	0.40	-0.29	0.28	0.37	0.37	NA	0.08	0.31
conjno	[45,]	0.89	0.13	0.91	-0.03	0.81	0.91	0.25	NA	0.70	0.70
adv%	[46,]	-0.02	0.05	-0.04	-0.57	-0.18	-0.10	-0.10	NA	0.08	0.04
advno	[47,]	0.95	0.05	0.89	-0.16	0.66	0.75	0.09	NA	0.89	0.71
noun%	[48,]	0.0	-0.48	-0.02	0.45	0.23	0.18	-0.08	NA	-0.10	0.11
nounno	[49,]	0.96	0.08	0.92	-0.01	0.79	0.88	0.27	NA	0.81	0.74
adj%	[50,]	0.02	-0.57	-0.22	0.16	-0.16	-0.20	-0.26	NA	0.12	-0.10
adjno	[51,]	1.00	0.03	0.87	-0.07	0.66	0.76	0.21	NA	0.92	0.69
pron%	[52,]	0.03	1.00	0.35	-0.16	0.15	0.15	0.46	NA	-0.06	-0.01
pronno	[53,]	0.87	0.35	1.00	0.02	0.84	0.85	0.28	NA	0.74	0.56
nom%	[54,]	-0.07	-0.16	0.02	1.00	0.35	0.11	0.09	NA	-0.15	-0.12
nomno	[55,]	0.66	0.15	0.84	0.35	1.00	0.34	0.36	NA	0.47	0.39
nounc	[56,]	0.76	0.15	0.85	0.11	0.84	1.00	0.42	NA	0.51	0.56
pronop	[57,]	0.21	0.46	0.28	0.09	0.36	0.42	1.00	NA	-0.01	0.12
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	0.92	-0.06	0.74	-0.15	0.47	0.51	-0.01	NA	1.00	0.49
artop	[60,]	0.69	-0.01	0.56	-0.12	0.39	0.56	0.12	NA	0.49	1.00
totop	[61,]	-0.05	0.29	0.03	-0.07	-0.03	0.06	0.14	NA	-0.09	-0.10
prep%	[62,]	-0.08	-0.23	-0.11	0.11	-0.01	-0.10	-0.06	NA	-0.04	-0.13
prepno	[63,]	0.32	-0.07	0.39	0.13	0.55	0.54	0.27	NA	0.20	0.16
adv%	[64,]	-0.05	-0.17	-0.12	0.04	-0.06	-0.09	-0.09	NA	0.01	-0.03
advno	[65,]	0.18	-0.17	0.0	-0.02	-0.06	-0.03	-0.02	NA	0.27	0.06
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	0.37	-0.11	0.33	-0 07	0.22	0.20	-0.19	NA	0.33	0.56
sub-cno	[69,]	0.65	-0.04	0.62	-0.12	0.41	0.40	-0.16	NA	0.60	0.72
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	0.24	-0.02	0.19	-0.14	0 03	0.02	-0.11	NA	0.16	0.65
expno	[73,]	0.24	-0.02	0.19	-0.14	0.03	0.02	-0.11	NA	0.16	0.65
abst	[74,]	-0.17	-0.14	-0.18	0.53	-0.17	-0.19	-0.15	NA	-0.14	-0.02
dict	[75,]	0.82	-0.06	0.83	0.10	0.81	0.82	0.09	NA	0.73	0.47
dict%s	[76,]	-0.08	-0.54	-0.18	0.23	-0.02	-0.11	-0.23	NA	0.01	-0.22
tt.con	[77,]	-0.80	-0.06	-0.76	0.14	-0 66	-0.81	-0.35	NA	-0.66	-0.57
tt.fun	[78,]	-0.80	-0.17	-0.78	0.23	-0.63	-0.72	-0.27	NA	-0.66	-0.73
ver	[79,]	-0.13	0.59	0.04	-0.16	-0.06	-0.01	0.37	NA	-0.13	-0.16
sylrep	[80,]	-0.11	0.49	0.05	-0.17	0.0	-0.04	0.15	NA	-0.10	-0.16
ch.n	[81,]	0.98	0.09	0.93	-0.04	0.76	0.84	0.23	NA	0.86	0.71
ch.m	[82,]	0.06	-0.41	-0.08	0.07	0.01	0.13	-0.15	NA	-0.02	0.15
ch.sd	[83,]	0.0	-0.11	0.04	0.04	0.19	0.12	0.06	NA	-0.07	-0.01
qualty	[84,]	0.48	0.43	0.54	-0.35	0.29	0.43	0.16	NA	0.37	0.44

CORRELATIONS AMONG VARIABLES

		totop [,61]	prep% [,62]	prepn [,63]	adv% [,64]	advno [,65]	verb% [,66]	verbn [,67]	sub-c [,68]	sub-c [,69]	conj% [,70]
win	[1,]	0.23	-0.18	-0.16	-0.15	-0.09	NA	NA	-0.09	0.01	NA
aut	[2,]	0.34	-0.26	-0.14	-0.27	-0.19	NA	NA	-0 09	0.08	NA
c-i	[3,]	0.39	-0.45	-0.23	-0.08	-0.04	NA	NA	-0.02	0.04	NA
fiel	[4,]	0.20	-0.30	-0.18	0.02	0.10	NA	NA	0.06	0.10	NA
fle2	[5,]	-0.24	0.27	0.21	0.04	-0.01	NA	NA	0.05	0.01	NA
nosnt	[6,]	-0.08	-0.08	0.43	0.0	0.16	NA	NA	0.43	0.67	NA
nowds	[7,]	-0.06	-0.07	0.40	-0.05	0.13	NA	NA	0.41	0.69	NA
avslen	[8,]	0.07	0.12	0.06	-0.29	-0.22	NA	NA	-0.11	0.08	NA
avwlen	[9,]	0.35	-0.45	-0.24	-0.01	0.0	NA	NA	-0.01	0.01	NA
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	-0.06	-0.08	0.39	-0.04	0.14	NA	NA	0.41	0.68	NA
%cwds	[13,]	0.09	-0.15	-0.08	0.07	0.10	NA	NA	0.0	-0.04	NA
avlencw	[14,]	0.24	-0.38	-0.20	0.05	0.05	NA	NA	0.10	0.10	NA
shtsent	[15,]	0.06	0.11	0.07	-0.28	-0.21	NA	NA	-0.09	0.10	NA
%shts	[16,]	0.16	-0.18	0.04	-0.07	-0.06	NA	NA	0.05	0.13	NA
noshts	[17,]	-0.07	-0.06	0.47	-0.05	0.06	NA	NA	0.43	0.71	NA
lngs	[18,]	0.06	0.11	0.07	-0.28	-0.21	NA	NA	-0.09	0.10	NA
%ings	[19,]	0.13	-0.09	-0.02	-0.11	-0.10	NA	NA	-0.02	0.06	NA
nolngs	[20,]	-0.04	-0.05	0.37	-0.09	0.0	NA	NA	0.37	0.68	NA
lgests	[21,]	0.02	-0.09	0.29	-0.09	0.04	NA	NA	0.30	0.57	NA
where	[22,]	-0.17	-0.09	0.23	0.01	-0.05	NA	NA	0.63	0.74	NA
shtsts	[23,]	-0.01	0.30	-0.09	-0.27	-0.26	NA	NA	-0.34	-0.27	NA
where	[24,]	-0.17	-0.09	0.21	0.13	0.35	NA	NA	0.53	0.74	NA
smpl%	[25,]	-0.08	0.04	0.06	0.16	0.09	NA	NA	-0.04	-0.17	NA
simpno	[26,]	-0.13	-0.04	0.59	0.08	0.15	NA	NA	0.37	0.50	NA
cplex%	[27,]	0.27	-0.35	-0.28	-0.06	0.02	NA	NA	0.05	0.09	NA
plexno	[28,]	-0.04	-0.15	0.16	-0.01	0.24	NA	NA	0.44	0.67	NA
pound%	[29,]	0.12	-0.06	0.22	-0.14	-0.11	NA	NA	-0.01	0.12	NA
poundno	[30,]	0.03	-0.05	0.44	-0.11	-0.02	NA	NA	0.20	0.46	NA
c-c%	[31,]	-0.32	0.49	0.23	-0.13	-0.13	NA	NA	0.0	0.09	NA
c-cno	[32,]	-0.09	0.02	0.40	-0.10	-0.05	NA	NA	0.37	0.66	NA
verbt	[33,]	-0.06	-0.08	0.35	-0.04	0.16	NA	NA	0.40	0.68	NA
tobe%	[34,]	-0.14	0.17	0.03	-0.01	0.01	NA	NA	0.03	0.05	NA
tobeno	[35,]	-0.10	-0.05	0.32	-0.03	0.18	NA	NA	0.42	0.71	NA
aux%	[36,]	0.07	0.15	0.03	-0.25	-0.27	NA	NA	-0.23	-0.15	NA
auxno	[37,]	-0.01	-0.04	0.36	-0.14	-0.04	NA	NA	0.32	0.62	NA
inf%	[38,]	0.20	-0.08	-0.08	-0.18	-0.13	NA	NA	-0.18	-0.10	NA
infno	[39,]	-0.01	-0.09	0.29	-0.06	0.18	NA	NA	0.29	0.57	NA
pass%	[40,]	-0.08	-0.08	0.20	0.15	0.12	NA	NA	0.26	0.24	NA
passno	[41,]	-0.11	-0.04	0.49	-0.03	0.02	NA	NA	0.48	0.74	NA
prep%	[42,]	-0.09	0.02	0.13	0.17	0.13	NA	NA	0.02	-0.07	NA
prepno	[43,]	-0.06	-0.06	0.47	-0.03	0.12	NA	NA	0.38	0.64	NA
conj%	[44,]	-0.08	0.20	0.30	-0.18	-0.18	NA	NA	0.0	0.11	NA
conjno	[45,]	-0.04	-0.04	0.49	-0.11	-0.01	NA	NA	0.35	0.62	NA
adv%	[46,]	-0.19	0.20	0.08	0.02	0.02	NA	NA	0.07	0.08	NA
advno	[47,]	-0.11	-0.05	0.37	-0.04	0.14	NA	NA	0.48	0.77	NA
noun%	[48,]	-0.19	-0.04	0.14	0.35	0.27	NA	NA	0.22	0.04	NA
nounno	[49,]	-0.07	-0.07	0.44	-0.03	0.12	NA	NA	0.42	0.68	NA
adj%	[50,]	0.17	-0.04	-0.15	-0.19	-0.10	NA	NA	-0.18	-0.09	NA
adjno	[51,]	-0.05	-0.08	0.32	-0.05	0.18	NA	NA	0.37	0.65	NA
pron%	[52,]	0.29	-0.23	-0.07	-0.17	-0.17	NA	NA	-0.11	-0.04	NA
pronno	[53,]	0.03	-0.11	0.39	-0.12	0.0	NA	NA	0.33	0.62	NA
nom%	[54,]	-0.07	0.11	0.13	0.04	-0.02	NA	NA	-0.0[illegible]	-0.12	NA
nomno	[55,]	-0.03	-0.01	0.55	-0.06	-0.06	NA	NA	0.22	0.41	NA
nouno	[56,]	0.06	-0.10	0.54	-0.09	-0.03	NA	NA	0.20	0.40	NA
pronop	[57,]	0.14	-0.06	0.27	-0.09	-0.02	NA	NA	-0.19	-0.16	NA
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	-0.09	-0.04	0.20	0.01	0.27	NA	NA	0.33	0.60	NA
artop	[60,]	-0.10	-0.13	0.16	-0.03	0.06	NA	NA	0.56	0.72	NA
totop	[61,]	1 00	-0.71	-0.39	-0.59	-0.54	NA	NA	-0.54	-0.33	NA
prep%	[62,]	-0.71	1.00	0.61	-0.07	-0.08	NA	NA	-0.11	-0.10	NA
prepno	[63,]	-0.39	0.61	1.00	-0.11	-0.14	NA	NA	-0.07	0.05	NA
adv%	[64,]	-0.59	-0.07	-0.11	1.00	0.93	NA	NA	0.63	0.21	NA
advno	[65,]	-0.54	-0.08	-0.14	0.93	1.00	NA	NA	0.60	0.26	NA
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	-0.54	-0.11	-0.07	0.63	0.60	NA	NA	1.00	0.86	NA
sub-cno	[69,]	-0.33	-0.10	0.05	0.21	0.26	NA	NA	0.86	1.00	NA
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	-0.19	-0.06	-0.10	-0.05	-0.06	NA	NA	0.63	0.75	NA
expno	[73,]	-0.19	-0.06	-0.10	-0.05	-0.06	NA	NA	0.63	0.75	NA
ebst	[74,]	-0.03	-0.03	-0.17	0.01	-0.02	NA	NA	0.09	0.07	NA
dict	[75,]	-0. 1	-0.02	0.48	0.02	0.09	NA	NA	0.37	0.60	NA
dict%s	[76,]	-0.11	0.15	0.04	0.02	-0.01	NA	NA	-0.03	-0.03	NA
tt.con	[77,]	0.03	0.14	-0.35	-0.07	-0.20	NA	NA	-0.31	-0.45	NA
tt.fun	[78,]	-0.01	0.17	-0.29	0.01	-0.11	NA	NA	-0.35	-0.52	NA
var	[79,]	0..7	-0.18	-0.12	-0.01	-0.04	NA	NA	-0.05	-0.07	NA
sylrep	[80,]	0.0	0.15	-0.03	-0.15	-0.18	NA	NA	-0.15	-0.10	NA
ch.n	[81,]	-0.06	-0.07	0.41	-0.05	0.12	NA	NA	0.42	0.70	NA
ch.m	[82,]	0.19	-0.35	-0.19	0.15	0.15	NA	NA	0.05	-0.03	NA
ch.sd	[83,]	-0.22	0.04	0.15	0.29	0.23	NA	NA	0.19	0.06	NA
qualty	[84,]	0.13	0.07	0.24	-0.42	-0.31	NA	NA	-0.03	0.25	NA

BEST COPY AVAILABLE

CORRELATIONS AMONG VARIABLES

		conjn [,71]	exp% [,72]	expno [,73]	abst [,74]	dict [,75]	dict% [,76]	tt.co [,77]	tt.fu [,78]	ver [,79]	sylre [,80]
kin	[1,]	NA	0.02	0.02	-0.08	0.11	0.43	0.02	-0.07	-0.59	-0.57
aut	[2,]	NA	0.09	0.09	-0.05	0.16	0.27	-0.06	-0.18	-0.50	-0.52
--l	[3,]	NA	0.06	0.06	0.29	0.15	0.52	-0.07	0.01	-0.45	-0.70
fle1	[4,]	NA	0.04	0.04	0.11	0.19	0.40	-0.20	-0.21	-0.57	-0.78
fle2	[5,]	NA	0.0	0.0	-0.10	-0.10	-0.54	-0.02	-0.01	0.59	0.72
nosnt	[6,]	NA	0.27	0.27	-0.16	0.86	-0.12	-0.86	-0.84	-0.08	-0.10
nowds	[7,]	NA	0.27	0.27	-0.17	0.87	-0.11	-0.82	-0.83	-0.09	-0.08
avslen	[8,]	NA	0 06	0.06	-0.45	0.07	-0.17	-0.05	-0.30	-0.24	0.06
avwlen	[9,]	NA	0.04	0.04	0.43	0.11	0.52	-0.02	0.11	-0.36	-0.70
noqst	[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
nocwds	[12,]	NA	0.26	0.26	-0.17	0.87	-0.11	-0.83	-0.83	-0.09	-0.10
%cwds	[13,]	NA	-0.05	-0.05	0.57	0.05	0.36	-0.07	0.10	-0.10	-0.71
avlencw	[14,]	NA	0.15	0.15	0.26	0.15	0.58	0.0	0.01	-0.57	-0.63
shtsent	[15,]	NA	0.07	0.07	-0.44	0.08	-0.17	-0.05	-0.29	-0.24	0.03
%shts	[16,]	NA	0.08	0.08	-0.08	0.09	-0.20	0.09	-0.07	0.20	0.13
noshts	[17,]	NA	0.30	0.30	-0.13	0.89	-0.08	-0.70	-0.73	-0.06	-0.07
lngs	[18,]	NA	0.07	0.07	-0.44	0.08	-0.17	-0.05	-0.29	-0.24	0.03
%lngs	[19,]	NA	0.02	0.02	0.12	0.0	-0.33	0.17	-0.08	0.60	0.13
nolngs	[20,]	NA	0.21	0.21	-0.04	0.86	-0.07	-0.51	-0.59	0.09	0.02
lgests	[21,]	NA	0.17	0.17	-0.29	0.72	-0.23	-0.59	-0.75	-0.03	0.0
where	[22,]	NA	0.75	0.75	0.08	0.44	-0.13	-0.48	-0.61	-0.05	-0.09
shtsts	[23,]	NA	-0.14	-0.14	-0.33	-0.24	0.21	0.34	0.29	-0.42	0.10
where	[24,]	NA	0.37	0.37	-0.08	0.71	-0.09	-0.72	-0.69	-0.07	-0.16
smpl%	[25,]	NA	-0.09	-0.09	0.38	-0.10	0.58	0.28	0.27	-0.06	-0.42
simpno	[26,]	NA	0.22	0.22	-0.12	0.74	-0.07	-0.78	-0.74	-0.06	-0.17
cplex%	[27,]	NA	0.07	0.07	-0.34	-0.06	-0.53	-0.12	-0.15	0.18	0.39
plexno	[28,]	NA	0.31	0.31	-0.17	0.68	-0.17	-0.79	-0.78	-0.06	-0.07
pound%	[29,]	NA	0.0	0.0	-0.16	0.31	-0.25	-0.50	-0.58	-0.17	-0.06
poundno	[30,]	NA	0.04	0.04	-0.18	0.83	-0.09	-0.73	-0.73	-0.10	-0.02
c-c%	[31,]	NA	0.06	0.06	-0.07	0.11	-0.08	-0.08	0.02	-0.10	0.16
c-cno	[32,]	NA	0.23	0.23	-0.08	0.88	-0.01	-0.58	-0.60	-0.06	0.04
verbt	[33,]	NA	0.23	0.23	-0.17	0.87	-0.10	-0.81	-0.79	-0.06	-0.05
tobe%	[34,]	NA	0.06	0.06	-0.36	0.08	0.60	0.23	0.13	-0.72	-0.07
tobeno	[35,]	NA	0.28	0.28	-0.16	0.86	-0.02	-0.71	-0.71	-0.16	-0.06
aux%	[36,]	NA	-0.05	-0.05	-0.27	-0.20	-0.32	0.13	0.26	0.21	0.52
auxno	[37,]	NA	0.19	0.19	-0.18	0.82	-0.14	-0.68	-0.65	0.02	0.13
inf%	[38,]	NA	-0.11	-0.11	0.34	-0.01	-0.24	0.02	0.11	0.67	0.08
infno	[39,]	NA	0.09	0.09	-0.19	0.83	-0.08	-0.78	-0.73	0.02	-0.02
pass%	[40,]	NA	0.13	0.13	-0.09	0.26	-0.08	-0.13	-0.16	0.44	0.06
passno	[41,]	NA	0.30	0.30	-0.12	0.94	-0.01	-0.66	-0.67	-0.01	-0.05
prep%	[42,]	NA	-0.09	-0.09	0.02	0.07	0.40	-0.08	0.03	-0.48	-0.53
prepno	[43,]	NA	0.22	0.22	-0.19	0.88	-0.10	-0.83	-0.82	-0.12	-0.11
conj%	[44,]	NA	0.05	0.05	-0.08	0.15	-0.55	-0.25	-0.44	0.47	0.28
conjno	[45,]	NA	0.20	0.20	-0.15	0.89	-0.12	-0.76	-0.78	-0.02	-0.02
adv%	[46,]	NA	0.08	0.08	-0.17	-0.08	-0.47	-0.13	-0.19	0.46	0.30
advno	[47,]	NA	0.34	0.34	-0.14	0.85	-0.13	-0.77	-0.79	-0.03	-0.07
noun%	[48,]	NA	-0.01	-0.01	0.15	0.10	0.27	-0.16	-0.17	-0.45	-0.59
nounno	[49,]	NA	0.27	0.27	-0.18	0.87	-0.11	-0.84	-0.84	-0.10	-0.10
adj%	[50,]	NA	-0.03	-0.03	0.15	0.04	0.76	0.17	0.28	-0.64	-0.53
adjno	[51,]	NA	0.24	0.24	-0.17	0.82	-0.08	-0.80	-0.80	-0.13	-0.11
pron%	[52,]	NA	-0.02	-0.02	-0.14	-0.06	-0.54	-0.06	-0.17	0.59	0.49
pronno	[53,]	NA	0.19	0.19	-0.18	0.83	-0.18	-0.76	-0.78	0.04	0.05
nom%	[54,]	NA	-0.14	-0.14	0.53	0.10	0.23	0.14	0.23	-0.16	-0.17
nomno	[55,]	NA	0.03	0.03	-0.17	0.81	-0.02	-0.66	-0.63	-0.06	0.0
nouno	[56,]	NA	0.02	0.02	-0.19	0.82	-0.11	-0.81	-0.72	-0.01	-0.04
pronop	[57,]	NA	-0.11	-0.11	-0.15	0.09	-0.23	-0.35	-0.27	0.37	0.15
posop	[58,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
adjop	[59,]	NA	0.16	0.16	-0.14	0.73	0.01	-0.66	-0.66	-0.13	-0.10
artop	[60,]	NA	0.65	0.65	-0.02	0.47	-0.22	-0.57	-0.73	-0.16	-0.16
totop	[61,]	NA	-0.19	-0.19	-0.03	-0.11	-0.11	0.03	-0.01	0.17	0.0
prep%	[62,]	NA	-0.06	-0.06	-0.03	-0.02	0.15	0.14	0.17	-0.18	0.15
prepno	[63,]	NA	-0.10	-0.10	-0.17	0.48	0.04	-0.35	-0.29	-0.12	-0.03
adv%	[64,]	NA	-0.05	-0.05	0.01	0.02	0.02	-0.07	0.01	-0.01	-0.15
advno	[65,]	NA	-0.06	-0.06	-0.02	0.09	-0.01	-0.20	-0.11	-0.04	-0.18
verb%	[66,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
sub-c%	[68,]	NA	0.63	0.63	0.09	0.37	-0.03	-0.31	-0.35	-0.05	-0.15
sub-cno	[69,]	NA	0.75	0.75	0.07	0.60	-0.03	-0.45	-0.52	-0.07	-0.10
conj%	[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
exp%	[72,]	NA	1.00	1.00	0.21	0.07	-0.08	-0.13	-0.27	-0.08	-0.09
expno	[73,]	NA	1.00	1.00	0.21	0.07	-0.08	-0.13	-0.27	-0.08	-0.09
abst	[74,]	NA	0.21	0.21	1.00	-0.14	0.12	0.35	0.34	0.19	-0.32
dict	[75,]	NA	0.07	0.07	-0.14	1.00	0.19	-0.67	-0.61	-0.13	-0.11
dict%s	[76,]	NA	-0.08	-0.08	0.12	0.19	1.00	0.22	0.30	-0.53	-0.45
tt.con	[77,]	NA	-0.13	-0.13	0.35	-0.67	0.22	1.00	0.86	0.05	0.12
tt.fun	[78,]	NA	-0.27	-0.27	0.34	-0.61	0.30	0.86	1.00	0.05	0.09
ver	[79,]	NA	-0.08	-0.08	0.19	-0.13	-0.53	0.05	0.05	1.00	0.45
sylrep	[80,]	NA	-0.09	-0.09	-0.32	-0.11	-0.45	0.12	0.09	0.45	1.00
ch.n	[81,]	NA	0.28	0.28	-0.17	0.87	-0.11	-0.82	-0.83	-0.09	-0.08
ch.m	[82,]	NA	-0.07	-0.07	0.13	0.08	0.21	-0.14	-0.04	-0.35	-0.48
ch.sd	[83,]	NA	-0.01	-0.01	-0.29	0.10	0.22	-0.04	-0.12	-0.32	-0.17
quality	[84,]	NA	0.20	0.20	-0.36	0.29	-0.50	-0.50	-0.51	0.14	0.22

BEST COPY AVAILABLE

CORRELATIONS AMONG VARIABLES

		ch.n [,81]	ch.m [,82]	ch.sd [,83]	qual.t [,84]
kin	[1,]	0.07	0.61	0.33	0.01
aut	[2,]	0.18	0.57	0.23	0.12
c-i	[3,]	0.06	0.68	0.22	-0.16
fla1	[4,]	0.18	0.76	0.27	-0.08
fla2	[5,]	-0.02	-0.71	-0.37	0.16
nosnt	[6,]	0.98	0.07	0.05	0.48
nowds	[7,]	1.00	0.04	0.03	0.51
avslen	[8,]	0.20	0.05	0.11	0.37
avwlen	[9,]	0.0	0.64	0.15	-0.26
noqst	[10,]	NA	NA	NA	NA
noimp	[11,]	NA	NA	NA	NA
nocwds	[12,]	1.00	0.05	0.03	0.50
%cwds	[13,]	-0.05	0.38	-0.15	-0.40
avlencw	[14,]	0.06	0.58	0.29	-0.21
shtsent	[15,]	0.20	0.07	0.10	0.37
%snts	[16,]	0.14	-0.16	-0.17	0.24
noshts	[17,]	0.94	-0.02	0.01	0.49
lngs	[18,]	0.20	0.07	0.10	0.37
%lngs	[19,]	0.05	-0.13	-0.03	0.18
nolngs	[20,]	0.83	-0.08	-0.03	0.44
lgests	[21,]	0.83	0.0	0.02	0.56
where	[22,]	0.55	-0.03	0.0	0.34
shtsts	[23,]	-0.30	0.13	0.18	-0.10
where	[24,]	0.89	0.0	-0.02	0.42
smpl%	[25,]	-0.26	0.01	0.11	-0.58
simpno	[26,]	0.82	0.12	0.17	0.32
cplex%	[27,]	0.10	-0.04	0.01	0.29
plexno	[28,]	0.91	0.03	-0.01	0.48
pound%	[29,]	0.45	-0.17	0.0	0.40
poundno	[30,]	0.85	-0.04	0.0	0.44
c-c%	[31,]	0.09	0.13	-0.19	0.38
c-cno	[32,]	0.82	0.08	-0.06	0.47
verbt	[33,]	0.99	0.03	-0.01	0.49
tobe%	[34,]	-0.02	-0.07	0.31	-0.21
tobeno	[35,]	0.96	0.0	0.01	0.44
aux%	[36,]	-0.16	-0.25	-0.33	0.31
auxno	[37,]	0.87	-0.13	-0.09	0.50
inf%	[38,]	-0.01	-0.06	-0.59	0.06
infno	[39,]	0.94	0.01	-0.06	0.44
pass%	[40,]	0.23	-0.39	0.28	0.21
passno	[41,]	0.90	-0.07	0.09	0.43
prep%	[42,]	-0.01	0.18	0.57	-0.11
prepno	[43,]	0.99	0.04	0.10	0.49
conj%	[44,]	0.31	-0.14	0.0	0.55
conjno	[45,]	0.94	0.07	0.04	0.51
adv%	[46,]	0.0	-0.15	-0.20	0.34
advno	[47,]	0.97	-0.01	-0.02	0.52
noun%	[48,]	0.04	0.28	0.53	-0.34
nounno	[49,]	0.99	0.05	0.08	0.49
adj%	[50,]	-0.08	0.34	-0.11	-0.36
adjno	[51,]	0.98	0.06	0.0	0.48
pron%	[52,]	0.09	-0.41	-0.11	0.43
pronno	[53,]	0.93	-0.08	0.04	0.54
nom%	[54,]	-0.04	0.07	0.04	-0.35
nomno	[55,]	0.76	0.01	0.19	0.29
nouno	[56,]	0.84	0.13	0.12	0.43
pronop	[57,]	0.23	-0.15	0.06	0.16
posop	[58,]	NA	NA	NA	NA
adjop	[59,]	0.86	-0.02	-0.07	0.37
artop	[60,]	0.71	0.15	-0.01	0.44
totop	[61,]	-0.06	0.19	-0.22	0.13
prep%	[62,]	-0.07	-0.35	0.04	0.07
propno	[63,]	0.41	-0.19	0.15	0.24
adv%	[64,]	-0.05	0.15	0.29	-0.42
advno	[65,]	0.12	0.15	0.23	-0.31
verb%	[66,]	NA	NA	NA	NA
verbno	[67,]	NA	NA	NA	NA
sub-c%	[68,]	0.42	0.05	0.19	-0.03
sub-cno	[69,]	0.70	-0.03	0.06	0.25
conj%	[70,]	NA	NA	NA	NA
conjno	[71,]	NA	NA	NA	NA
exp%	[72,]	0.28	-0.07	-0.01	0.20
expno	[73,]	0.28	-0.07	-0.01	0.20
abst	[74,]	-0.17	0.13	-0.29	-0.36
dict	[75,]	0.87	0.08	0.10	0.29
dict%s	[76,]	-0.11	0.21	0.22	-0.50
tt.con	[77,]	-0.82	-0.14	-0.04	-0.50
tt.fun	[78,]	-0.83	-0.04	-0.12	-0.51
var	[79,]	-0.09	-0.35	-0.32	0.14
sylrep	[80,]	-0.08	-0.48	-0.17	0.22
ch.n	[81,]	1.00	0.01	0.04	0.51
ch.m	[82,]	0.01	1.00	0.14	-0.20
ch.sd	[83,]	0.04	0.14	1.00	-0.15
qualty	[84,]	0.51	-0.20	-0.15	1.00

BEST COPY AVAILABLE

DATA FOR ALL RESPONSE CATEGORIES

category-->		[03]	[11]	[12]	[13]	[15]	[16]	[17]	[18]	[19]
1	win	9.70	10.00	10.00	15.70	11.10	11.90	11.40	18.50	12.80
2	eut	10.60	9.00	11.20	15.00	11.60	12.50	11.90	18.70	12.30
3	c-i	9.70	12.40	16.30	14.70	12.50	13.70	13.00	14 90	12.80
4	fle1	8.70	14.00	15.00	17.00	13.30	14.20	13.80	17.00	14.90
5	fle2	62.60	43.60	36.90	21.60	48.20	41.80	44.90	15.20	37.50
6	nosnt	2.00	7.00	2.00	1.00	17.00	11.00	16.00	2.00	11.00
7	nowds	42.00	81.00	15.00	22.00	310.00	199.00	287.00	59.00	211.00
8	evslen	21.00	11.60	7.50	22.00	18.20	18.10	17.90	29.50	19.20
9	evwlen	4.57	5.22	6.13	5.41	5.08	5.29	5.17	5.39	5 12
10	noqst	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	noimp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	nocwds	21.00	51.00	12.00	14.00	188.00	123.00	176.00	38.00	132.00
13	%cwds	50.00	63.00	80.00	63.60	60.60	61.80	61.30	64.40	62.60
14	evlencw	6.00	6.75	7.08	7.07	6.19	6.63	6.43	6.50	6.52
15	shtsent	16.00	7.00	3.00	17.00	13.00	13.00	13.00	25.00	14.00
16	%shts	50.00	14.00	0.0	0.0	18.00	9.00	13.00	50.00	18.00
17	noshts	1.00	1.00	0.0	0.0	3.00	1.00	2.00	1.00	2.00
18	lngs	31.00	22.00	18.00	32.00	28.00	28.00	28.00	40.00	29.00
19	%lngs	0.0	0.0	0.0	0.0	6.00	0.0	0.0	50.00	9.00
20	nolngs	0.0	0.0	0.0	0.0	1.00	0.0	0.0	1.00	1.00
21	lgests	28.00	22.00	10.00	22.00	35.00	28.00	28.00	41.00	37.00
22	where	2.00	5.00	2.00	1.00	7.00	4.00	5.00	1.00	9.00
23	shtsts	14.00	4.00	5.00	22.00	7.00	12.00	8.00	18.00	7.00
24	where	1.00	7.00	1.00	1.00	5.00	8.00	1.00	2.00	2.00
25	smpl%	0.0	71.00	100.00	100.00	29.00	55.00	63.00	50.00	9.00
26	simpno	0.0	5.00	2.00	1.00	5.00	6.00	10.00	1.00	1.00
27	cplex%	100.00	29.00	0.0	0.0	53.00	36.00	31.00	50.00	45.00
28	plexno	2.00	2.00	0.0	0.0	9.00	4.00	5.00	1.00	5.00
29	pound%	0.0	0.0	0.0	0.0	6.00	9.00	6.00	0.0	27.00
30	poundno	0.0	0.0	0.0	0.0	1.00	1.00	1.00	0.0	3.00
31	c-c%	0.0	0.0	0.0	0.0	12.00	0.0	0.0	0.0	18.00
32	c-cno	0.0	0.0	0.0	0.0	2.00	0.0	0.0	0.0	2.00
33	verbt	6.00	9.00	2.00	2.00	40.00	21.00	32.00	5.00	30.00
34	tobe%	67.00	33.00	0.0	100.00	23.00	14.00	25.00	0.0	30.00
35	tobeno	4.00	3.00	0.0	2.00	9.00	3.00	8.00	0.0	9.00
36	aux%	33.00	0.0	0.0	0.0	10.00	14.00	9.00	20.00	20.00
37	auxno	2.00	0.0	0.0	0.0	4.00	3.00	3.00	1.00	6.00
38	inf%	17.00	11.00	50.00	0.0	25.00	29.00	31.00	40.00	23.00
39	infno	1.00	1.00	1.00	0.0	10.00	6.00	10.00	2.00	7.00
40	pass%	0.0	13.00	0.0	0.0	10.00	7.00	9.00	0.0	4.00
41	passno	0.0	1.00	0.0	0.0	3.00	1.00	2.00	0.0	1.00
42	prep%	4.80	14.80	13.30	13.60	10.30	14.60	13.90	10.20	8.50
43	prepno	2.00	12.00	2.00	3.00	32.00	29.00	40.00	6.00	18.00
44	conj%	0.0	1.20	0.0	0.0	4.80	3.50	3.50	6.80	3.80
45	conjno	0.0	1.00	0.0	0.0	15.00	7.00	10.00	4.00	8.00
46	adv%	2.40	4.90	0.0	0.0	4.50	2.50	3.10	10.20	3.80
47	advno	1.00	4.00	0.0	0.0	14.00	5.00	9.00	6.00	8.00
48	noun%	21.40	35.80	33.30	31.80	28.70	32.70	31.70	30.50	25.60
49	nounno	9.00	29.00	5.00	7.00	89.00	65.00	91.00	18.00	54.00
50	adj%	19.00	13.60	26.70	31.80	16.10	16.10	16.40	13.60	22.30
51	adjno	8.00	11.00	4.00	7.00	50.00	32.00	47.00	8.00	47.00
52	pron%	4.80	1.20	0.0	0.0	4.50	6.00	5.60	1.70	2.80
53	pronno	2.00	1.00	.0	0.0	14.00	12.00	16.00	1.00	6.00
54	nom%	5.00	4.00	13.00	0.0	3.00	8.00	7.00	0.0	2.00
55	nomno	2.00	3.00	2.00	0.0	9.00	15.00	19.00	0.0	4.00
56	nouno	1.00	2.00	2.00	0.0	10.00	8.00	10.00	1.00	5.00
57	pronop	0.0	0.0	0.0	0.0	0.0	2.00	2.00	0.0	0.0
58	posop	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
59	adjop	0.0	1.00	0.0	1.00	2.00	0.0	2.00	0.0	4.00
60	artop	1.00	1.00	0.0	0.0	4.00	1.00	1.00	1.00	2.00
61	totop	100.00	57.00	100.00	100.00	94.00	100.00	94.00	100.00	100.00
62	prep%	0.0	0.0	0.0	0.0	0.0	0.0	6.00	0.0	0.0
63	prepno	0.0	0.0	0.0	0.0	0.0	0.0	1.00	0.0	0.0
64	adv%	0.0	29.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0
65	advno	0.0	2.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0
66	verb%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
67	verbno	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
68	sub-c%	0.0	14.00	0.0	0.0	6.00	0.0	0.0	0.0	0.0
69	sub-cno	0.0	1.00	0.0	0.0	1.00	0.0	0.0	0.0	0.0
70	conj%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
71	conjno	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
72	exp%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
73	expno	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
74	abst	0.0	2.50	13.30	0.0	1.30	1.00	0.70	3.30	2.40
75	dict	0.0	2.00	1.00	1.00	4.00	2.00	3.00	0.0	2.00
76	dict%s	0.0	28.57	50.00	100.00	23.52	18.18	13.75	0.0	18.18
77	tt.con	1.00	0.73	1.00	0.93	0.64	0.57	0.55	0.91	0.63
78	tt.fun	0.86	0.63	1.00	0.67	0.41	0.44	0.37	0.57	0.37
79	var	0.50	0.73	1.00	0.0	0.90	0.84	0.80	1.00	0.61
80	sylrep	0.52	0.35	0 29	0.32	0.39	0.39	0.39	0.34	0.41
81	ch.n	7.00	14.00	3.00	4.00	53.00	38.00	55.00	10.00	37.00
82	ch.m	33.00	36.07	35.33	34.75	35.08	32.53	35.13	37.50	35.16
83	ch.sd	9.32	19.28	6.80	17.28	13.60	13.44	18.66	16.87	11.10
84	qualty	15.80	4.60	4.40	2.10	26.20	25.20	17.00	19.20	22.00

BEST COPY AVAILABLE

DATA FOR ALL RESPONSE CATEGORIES

category-->		[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]
1	win	12.00	13.20	13.00	12.60	14.50	13 60	14.80	9.40	10.40
2	out	12.20	14.30	14.00	13.50	14.30	13.60	14.40	10.10	11.00
3	c-	14.50	13.60	13.60	14.10	14.00	13.10	14.00	12.40	11.90
4	fle1	14.90	14.40	14.40	14.60	17.00	15.00	17.00	11.20	11.60
5	fle2	37.20	40.50	40.40	39.30	29.50	36.40	28.10	53.90	52.00
6	nosnt	26.00	5.00	36.00	22.00	2.00	5.00	6.00	7.00	8.00
7	nowds	409.00	112.00	777.00	430.00	43.00	108.00	131.00	102.00	142.00
8	evslen	15.70	22.40	21.60	19.50	21.50	21.60	21.80	14.60	17.80
9	evwlen	5.47	5.21	5.23	5.34	5.30	5.15	5.30	5.14	4.99
10	noqst	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	noimp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	nocwds	251.00	59.00	464.00	258.00	27.00	66.00	78.00	59.00	89.00
13	%cwds	61.40	52.70	59.70	60.00	62.80	61.10	59.50	57.80	62.70
14	ev.encw	7.02	6.86	6.67	7.02	6.33	6.47	7.13	6.41	6.24
15	shtsent	11.00	17.00	17.00	15.00	17.00	17.00	17.00	10.00	13.00
16	%shts	23.00	20.00	33.00	27.00	0.0	20.00	17.00	29.00	38.00
17	noshts	6.00	1.00	12.00	6.00	0.0	1.00	1.00	2.00	3.00
18	ings	26.00	32.00	32.00	30.00	32.00	32.00	32.00	25.00	28.00
19	%ings	4.00	20.00	17.00	9.00	0.0	0.0	0.0	0.0	13.00
20	noings	1.00	1.00	6.00	2.00	0.0	0.0	0.0	0.0	1.00
21	igests	32.00	33.00	63.00	28.00	24.00	30.00	30.00	21.00	32.00
22	where	8.00	5.00	10.00	20.00	1.00	5.00	2.00	1.00	8.00
23	shtsts	7.00	12.00	9.00	7.00	19.00	11.00	14.00	8.00	3.00
24	where	11.00	2.00	20.00	17.00	2.00	1.00	1.00	4.00	6.00
25	smpl%	58.00	40.00	33.00	36.00	0.0	40.00	50.00	43.00	50.00
26	simpno	15.00	2.00	12.00	8.00	0.0	2.00	3.00	3.00	4.00
27	cplex%	23.00	20.00	31.00	45.00	50.00	40.00	33.00	43.00	38.00
28	plexno	6.00	1.00	11.00	10.00	1.00	2.00	2.00	3.00	3.00
29	pound%	8.00	0.0	14.00	5.00	0.0	0.0	17.00	14.00	13.00
30	poundno	2.00	0.0	5.00	1.00	0.0	0.0	1.00	1.00	1.00
31	c-c%	12.00	40.00	22.00	14.00	50.00	20.00	0.0	0.0	0.0
32	c-cno	3.00	2.00	8.00	3.00	1.00	1.00	0.0	0.0	0.0
33	verbt	45.00	13.00	101.00	51.00	6.00	14.00	10.00	14.00	14.00
34	tobe%	31.00	8.00	40.00	41.00	17.00	36.00	70.00	29.00	21.00
35	tobeno	14.00	1.00	40.00	21.00	1.00	5.00	7.00	4.00	3.00
36	aux%	4.00	15.00	13.00	10.00	33.00	0.0	0.0	14.00	21.00
37	auxno	2.00	2.00	13.00	5.00	2.00	0.0	0.0	2.00	3.00
38	inf%	16.00	15.00	25.00	16.00	33.00	36.00	10.00	7.00	14.00
39	infno	7.00	2.00	25.00	8.00	2.00	5.00	1.00	1.00	2.00
40	pass%	11.00	0.0	14.00	12.00	0.0	11.00	0.0	8.00	17.00
41	passno	4.00	0.0	11.00	5.00	0.0	1.00	0.0	1.00	2.00
42	prep%	13.70	9.80	12.10	10.70	11.60	10.20	17.60	15.70	12.70
43	prepno	56.00	11.00	94.00	46.00	5.00	11.00	23.00	16.00	18.00
44	conj%	5.60	6.30	4.60	3.50	2.30	1.90	3.10	3.90	2.80
45	conjno	23.00	7.00	36 00	15.00	1.00	2.00	4.00	4.00	4.00
46	adv%	3.70	3.60	5.00	5.60	4.70	5.60	3.10	4.90	6.30
47	advno	15.00	4.00	39.00	24.00	2.00	6.00	4.00	5.00	9.00
48	noun%	31.50	25.90	27.20	27.90	23.30	26.90	32.80	26.50	30.30
49	nounno	129.00	29.00	211.00	120.00	10.00	29.00	43.00	27.00	43.00
50	adj%	17.40	12.50	17.60	18.10	23.30	19.40	20.60	14.70	16.20
51	adjno	71.00	14.00	137.00	78.00	10.00	21.00	27.00	15.00	23.00
52	pron%	2.70	10.70	4.50	3.50	2.30	6.50	3.80	8.80	4.20
53	pronno	11.00	12.00	35.00	15.00	1.00	7 00	5.00	9.00	6.00
54	nom%	3.00	4.00	4.00	1.00	0.0	2.00	5.00	0.0	1.00
55	nomno	11.00	5.00	28.00	6.00	0.0	2.00	6.00	0.0	2.00
56	nouno	16.00	4.00	18.00	5.00	2.00	2.00	1.00	6.00	4.00
57	pronop	2.00	0.0	0.0	0.0	0.0	0.0	0.0	1.00	0.0
58	posop	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
59	adjop	1.00	0.0	12.00	5.00	0.0	2.00	2.00	0.0	2.00
60	artop	6.00	1.00	3.00	8.00	0.0	1.00	3.00	0.0	1.00
61	totop	96.00	100.00	92.00	82.00	100.00	100.00	100.00	100.00	88.00
62	prep%	4.00	0.0	3.00	0.0	0.0	0.0	0.0	0.0	13.00
63	prepno	1 00	0.0	1.00	0.0	0.0	0.0	0.0	0.0	1.00
64	adv%	0 0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
65	advno	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
66	verb%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
67	verbno	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
68	sub-c%	0.0	0.0	6.00	14.00	0.0	0.0	0.0	0.0	0.0
69	sub-cno	0.0	0.0	2.00	3.00	0.0	0.0	0.0	0.0	0.0
70	conj%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
71	conjno	0 0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
72	exp%	0.0	0.0	0.0	5.00	0.0	0.0	0.0	0.0	0.0
73	expno	0.0	0.0	0.0	1.00	0.0	0.0	0.0	0.0	0.0
74	abst	1.20	4.60	0.90	5.00	0.0	0.0	1.50	1.90	1.40
75	dict	5.00	0.0	14.00	3.00	0.0	2.00	0.0	0.0	1.00
76	dict%s	19.23	0.0	38.88	13.63	0.0	40.00	0.0	0.0	12.50
77	tt.con	0.51	0.88	0.44	0.64	0.71	0.73	0.86	0.72	0.72
78	tt.fun	0.30	0.60	0.20	0.29	0.82	0.53	0.47	0.64	0.52
79	var	0.61	1.00	0.74	0.56	0.70	0.71	0.19	0.87	0.70
80	sylrep	0.36	0.44	0.41	0.36	0.36	0.36	0.35	0.43	0.31
81	ch.n	68.00	19.00	140.00	79.00	7.00	19.00	24.00	19.00	28.00
82	ch.m	37.97	35.89	33.56	33.72	38.14	34.42	33.96	32.74	30.14
83	ch sd	14.29	13.09	14.35	13.46	10.63	9.98	15.44	15.13	13.06
84	qualty	27.40	26.80	30.60	29.20	28.80	26.40	26.60	25.00	30.00

DATA FOR ALL RESPONSE CATEGORIES

category-->		[29]	[30]	[31]	[32]	[33]	[34]	[35]
1	win	14.40	7.60	10.40	7.10	10.50	17.90	16.20
2	aut	14.20	7.50	9.60	9.20	10.00	15 90	14.30
3	c-i	14.20	8.50	8.40	11.60	13.70	19.30	17.30
4	fle1	17.00	8.10	10.40	8.10	14.40	17.00	17.00
5	fle2	29.20	69.20	58.00	68.90	40.70	5.90	6.40
6	nosnt	30.00	2.00	2.00	3.00	5.00	2.00	3.00
7	nowds	632.00	32.00	42.00	42.00	58.00	31.00	46.00
8	avslen	21.10	16.00	21.00	14.00	11.60	15.50	15.30
9	avwlen	5.33	4.44	4.36	5.02	5.45	6.29	5.96
10	noqst	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	noimp	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	nocwds	402.00	17.00	24.00	26.00	39.00	20.00	28.00
13	%cwds	63.60	53.10	57.10	61.90	67.20	64.50	60.90
14	avlencw	6.64	5.53	5.67	6.12	6.59	8.10	7.54
15	shtsent	16.00	11.00	16.00	9.00	7.00	11.00	10.00
16	%shts	20.00	50.00	0.0	0.0	0)	50.00	0.0
17	noshts	6.00	1.00	0.0	0.0	( )	1.00	0.0
18	lngs	31.00	26.00	31.00	24.00	22.00	26.00	25.00
19	%lngs	7.00	50.00	0.0	0.0	0.0	0.0	0.0
20	nolngs	2.00	1.00	0.0	0.0	0.0	0.0	0.0
21	lgests	48.00	28.00	21.00	16.00	18.00	21.00	19.00
22	where	1.00	1.00	1.00	3.00	5.00	1.00	3.00
23	shtsts	9.00	4.00	21.00	13.00	7.00	10.00	13.00
24	where	26 00	2.00	1.00	1.00	3.00	2.00	1.00
25	smpl%	27.00	50.00	50.00	67.00	40.00	100.00	33.00
26	simpno	8.00	1.00	1.00	2.00	2.00	2.00	1.00
27	cplex%	60.00	50.00	0.0	33.00	60.00	0.0	67.00
28	plexno	18.00	1.00	0.0	1.00	3.00	0.0	2.00
29	pound%	7.00	0.0	0.0	0.0	0.0	0.0	0.0
30	poundno	2.00	0.0	0.0	0.0	0.0	0.0	0.0
31	c-c%	7.00	0.0	50.00	0.0	0.0	0.0	0.0
32	c-cno	2.00	0.0	1.00	0.0	0.0	0.0	0.0
33	verbt	90.00	6.00	5.00	6.00	8.00	4.00	5.00
34	tobe%	39.00	17.00	60.00	17.00	0.0	75.00	60.00
35	tobeno	35.00	1.00	3.00	1.00	0.0	3.00	3.00
36	aux%	7.00	17.00	20.00	33.00	0.0	25.00	0.0
37	auxno	6.00	1.00	1.00	2.00	0.0	1.00	0.0
38	inf%	29.00	50.00	20.00	33.00	25.00	25.00	0.0
39	infno	26.00	3.00	1.00	2.00	2.00	1.00	0.0
40	pass%	5.00	33.00	0.0	0.0	0.0	0.0	20.00
41	passno	3.00	1.00	0.0	0.0	0.0	0.0	1.00
42	prep%	10.90	6.30	11.90	11.90	8.60	16.10	17.40
43	prepno	69.00	2.00	5.00	5.00	5.00	5.00	8.00
44	conj%	2.70	6.30	4.80	2.40	1.70	0.0	0.0
45	conjno	17.00	2.00	2.00	1.00	1.00	0.0	0.0
46	adv%	4.70	6.30	7.10	9.50	10.30	3.20	0.0
47	advno	30.00	2.00	3.00	4.00	6.00	1.00	0.0
48	noun%	25.50	15.60	26.20	28.60	25.90	25.80	32.60
49	nounno	161.00	5.00	11.00	12.00	15.00	8.00	15.00
50	adj%	23.60	12.50	19.00	11.90	17.20	32.30	21.70
51	adjno	149.00	4.00	8.00	5.00	10.00	10.00	10.00
52	pron%	3.20	12.50	0.0	2.40	1.70	0.0	2.20
53	pronno	20.00	4.00	0.0	1.00	1.00	0.0	1.00
54	nom%	1.00	0.0	5.00	0.0	2.00	3.00	7.00
55	nomno	5.00	0.0	2.00	0.0	1.00	1.00	3.00
56	nouno	8.00	0.0	0.0	2.00	2.00	1.00	3.00
57	pronop	1.00	2.00	0.0	0.0	0.0	0.0	0.0
58	posop	0.0	0.0	0.0	0.0	0.0	0.0	0.0
59	adjop	15.00	0.0	1.00	1.00	2.00	1.00	0.0
60	artop	4.00	0.0	0.0	0.0	1.00	0.0	0.0
61	totop	93.00	100.00	50.00	100.00	100.00	100.00	100.00
62	prep%	0.0	0.0	50.00	0.0	0.0	0.0	0.0
63	preono	0.0	0.0	1.00	0.0	0.0	0.0	0.0
64	adv%	3.00	0.0	0.0	0.0	0.0	0.0	0.0
65	advno	1.00	0.0	0.0	0.0	0.0	0.0	0.0
66	verb%	0.0	0.0	0.0	0.0	0.0	0.0	0.0
67	verbno	0.0	0.0	0.0	0.0	0.0	0.0	0.0
68	sub-c%	3.00	0.0	0.0	0.0	0.0	0.0	0.0
69	sub-cno	1.00	0 0	0.0	0.0	0.0	0.0	0.0
70	conj%	0.0	0.0	0.0	0.0	0.0	0.0	0.0
71	conjno	0.0	0.0	0.0	0.0	0.0	0.0	0.0
72	exp%	0.0	0.0	0.0	0.0	0.0	0.0	0.0
73	expno	0.0	0.0	0.0	0.0	0.0	0.0	0.0
74	abst	1.10	3.10	2.50	0.0	3.50	3.70	2.20
75	dict	5.00	0.0	1.00	0.0	1.00	2.00	2.00
76	dict%s	16.66	0.0	50.00	0.0	20.00	100.00	66.66
77	tt.con	0.42	1.00	0.95	0.77	0.71	0.90	0.93
78	tt.fun	0.22	0.79	0.85	0.65	0.55	0.90	0.83
79	ver	0.59	2.25	0.38	1.40	1.00	0.20	0.30
80	sylrep	0.35	0.48	0.45	0.51	0.40	0.32	0.36
81	ch.n	112.00	6.00	7.00	8.00	10.00	6.00	.9.00
82	ch.m	35.07	29.67	31.29	31.25	36.90	37.17	35.89
83	ch.sd	12.01	10.75	13.89	12.44	10.23	11.60	25.76
84	qualty	29.60	22.60	21.60	24.20	17.00	18.20	19.00

Appendix C

Redesigned FH Instructions and Items Used in Small Pilot Test

Generating Hypotheses

The purpose of this test is to measure your ability to think of hypotheses that might explain a social phenomenon or the findings from a research study. The ability to think of possible explanations is important for problem solving in any field of study.

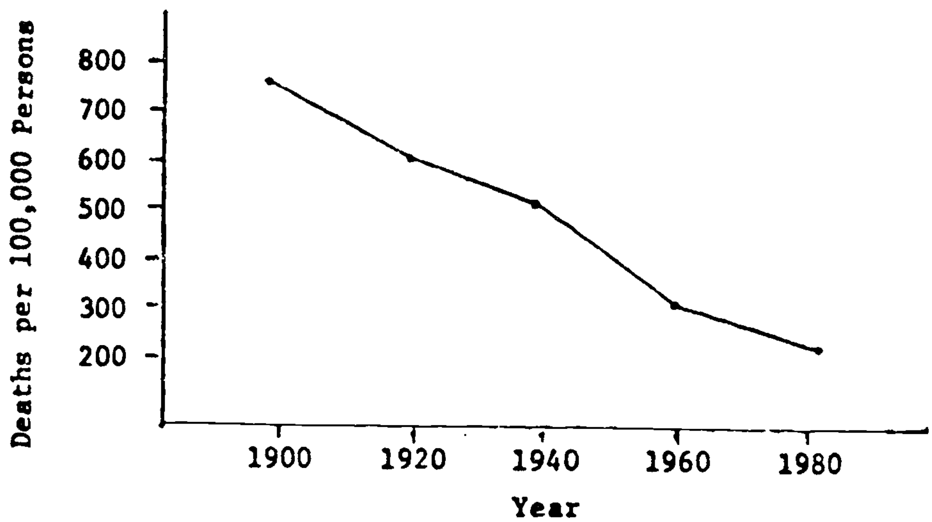
The problems presented here do not require any special or technical knowledge. They involve situations or results similar to ones you might read about in a newspaper and want to explain.

Your task is to think of as many possible interpretations or factors that might contribute to an explanation as you can. You are not looking for one right answer, but for many answers that might be considered, whether or not they prove to be correct.

The next page shows a sample problem and examples of hypotheses you might think of to explain the result.

Sample Problem

Rate of Death from Infectious Diseases in Alcadia



Finding: In Alcadia, a small country in central America, the rate of death from infectious diseases declined steadily from 1900 to 1980. What factors might account for the decrease?

Examples of Hypotheses:

Disease-causing organisms have gradually been eliminated by improved sanitation.

Better nutrition has resulted in a healthier population, better able to resist diseases.

more widespread innoculation against diseases

better medical treatment for those who become sick

Dissemination of health information has improved people's ability to avoid diseases.

Many who were susceptible to infectious diseases died before producing children, so that the percent of the population that is genetically resistant to the diseases has gradually increased.

The population has gradually build up immunity to the diseases that used to result in many deaths.

This list contains hypotheses that deal with environmental/living conditions, informed health and medical care, and biological/genetic factors.

The list is not a complete list of possible factors, but presents a sufficient number of good responses to this problem.

The responses are of high quality because they deal specifically with the data, and require only general knowledge. They did not require any specific knowledge about Alcadia or infectious disease.

Unless otherwise instructed, you should assume that the data are correct and that the study did not have any methodological problems.

Now go on to answer the four test questions. Take your time to think through the problems; there is no time limit.

Family Situation of Juvenile Delinquents

The family situation of children charged with juvenile delinquency in New York City in 1984 was investigated. The table compares the rate of juvenile delinquency for children from intact, two-parent homes with that for children from disrupted families.

Family Situations of Children Aged 10-17
Charged with Juvenile Delinquency

Family Situations	Percent Charged
Two-parent homes	4
One-parent homes	13

Finding: Proportionately more children who were charged with delinquency came from disrupted, single-parent families than from families in which both parents were present.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Family Situation of Juvenile Delinquents

The family situation of children charged with juvenile delinquency in New York City in 1984 was investigated. The table compares the rate of juvenile delinquency for children from intact, two-parent homes with that for children from disrupted families.

Family Situations of Children Aged 10-17
Charged with Juvenile Delinquency

Family Situations	Percent Charged
Two-parent homes	4
One-parent homes	13

Finding: Proportionately more children who were charged with delinquency came from disrupted, single-parent families than from families in which both parents were present.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Begin each hypothesis with the phrase, "Children from broken homes..."

Family Situation of Juvenile Delinquents

The family situation of children charged with juvenile delinquency in New York City in 1984 was investigated. The table compares the rate of juvenile delinquency for children from intact, two-parent homes with that for children from disrupted families.

Family Situations of Children Aged 10-17
Charged with Juvenile Delinquency

Family Situations	Percent Charged
Two-parent homes	4
One-parent homes	13

Finding: Proportionately more children who were charged with delinquency came from disrupted, single-parent families than from families in which both parents were present.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Begin each hypothesis with one of the following phrases:

"Children from broken homes..."

"Children from intact homes..."

"Families that are not intact..."

"Families that are intact..."

Family Situation of Juvenile Delinquents

The family situation of children charged with juvenile delinquency in New York City in 1984 was investigated. The table compares the rate of juvenile delinquency for children from intact, two-parent homes with that for children from disrupted families.

Family Situations of Children Aged 10-17
Charged with Juvenile Delinquency

Family Situations	Percent Charged
Two-parent homes	4
One-parent homes	13

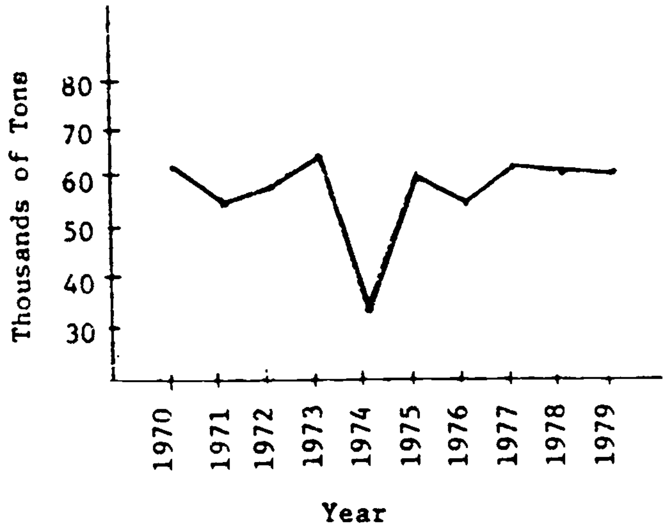
Finding: Proportionately more children who were charged with delinquency came from disrupted, single-parent families than from families in which both parents were present.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

One hypothesis might state, "Children from broken homes are delinquent because they are trying to get attention." This hypothesis could be stated as a phrase, "Delinquency gets attention." List your hypotheses in the form of phrases like this one.

Annual Mackerel Catch by Fleet Sailing from Port Byardia

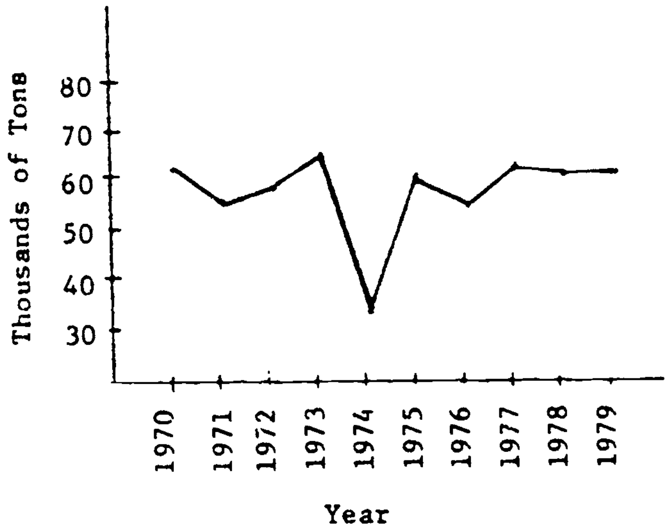


Finding: The Port Byardia fleet had a mackerel catch that was relatively constant year-to-year during the 1970's, except for a sharp drop in 1974.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Annual Mackerel Catch by Fleet Sailing from Port Byardia



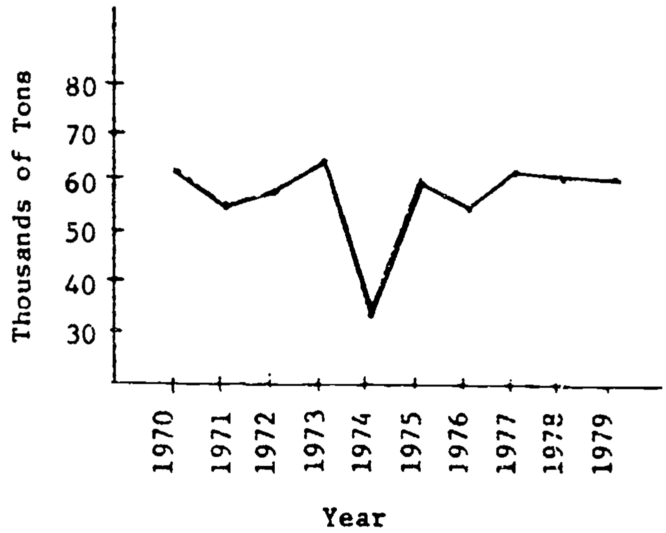
Finding: The Port Byardia fleet had a mackerel catch that was relatively constant year-to-year during the 1970's, except for a sharp drop in 1974.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Begin each hypothesis with the phrase, "During 1974..."

Annual Mackerel Catch by Fleet Sailing from Port Byardia



<u>Finding</u>: The Port Byardia fleet had a mackerel catch that was relatively constant year-to-year during the 1970's, except for a sharp drop in 1974.

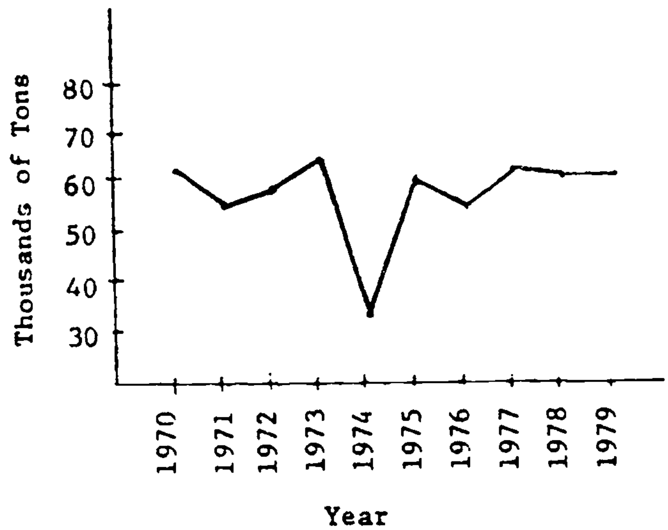
Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Begin each hypothesis with either one of two phrases:

"During 1974..."
or
"In every year except 1974..."

Annual Mackerel Catch by Fleet Sailing from Port Byardia



Finding: The Port Byardia fleet had a mackerel catch that was relatively constant year-to-year during the 1970's, except for a sharp drop in 1974.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Some of the factors that influenced the finding might involve weather. What other factors might have contributed to this outcome?

Violence in Family Relationships

The high level of violence in family relationships is a growing social problem. A study was conducted to determine if the rate of violence was as great for unmarried couples living together (cohabiting) as it was for married couples. The amount of interpersonal violence reported in a survey by male and female respondents is presented in the following table.

Interpersonal Violence Rates for Married and Cohabiting Couples

	Married	Cohabiting
Severe Violence	5.6	27.0
Overall Violence	15.1	37.8

Finding: Cohabitors report a higher rate of violence than their married counterparts.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Violence in Family Relationships

The high level of violence in family relationships is a growing social problem. A study was conducted to determine if the rate of violence was as great for unmarried couples living together (cohabiting) as it was for married couples. The amount of interpersonal violence reported in a survey by married and cohabiting respondents is presented in the following table.

Interpersonal Violence Rates for Married and Cohabiting Couples

	Married	Cohabiting
Severe Violence	5.6	27.0
Overall Violence	15.1	37.8

Finding: Cohabitors report a higher rate of violence than their married counterparts.

Think of hypotheses (possible explanations) to account for the finding. Write each hypothesis as a separate answer. Begin each hypothesis with the phrase, "Cohabiting couples..."

Violence in Family Relationships

The high level of violence in family relationships is a growing social problem. A study was conducted to determine if the rate of violence was as great for unmarried couples living together (cohabiting) as it was for married couples. The amount of interpersonal violence reported in a survey by married and cohabiting respondents is presented in the following table.

Interpersonal Violence Rates for Married and Cohabiting Couples

	Married	Cohabiting
Severe Violence	5.6	27.0
Overall Violence	15.1	37.8

Finding: Cohabitors report a higher rate of violence than their married counterparts.

Think of hypotheses (possible explanations) to account for the finding. Write each hypothesis as a separate answer. Begin each hypothesis with one of two phrases:

"Married couples..."
or
Cohabiting couples..."

Violence in Family Relationships

The high level of violence in family relationships is a growing social problem. A study was conducted to determine if the rate of violence was as great for unmarried couples living together (cohabiting) as it was for married couples. The amount of interpersonal violence reported in a survey by married and cohabiting respondents is presented in the following table.

Interpersonal Violence Rates for Married and Cohabiting Couples

	Married	Cohabiting
Severe Violence	5.6	27.0
Overall Violence	15.1	37.8

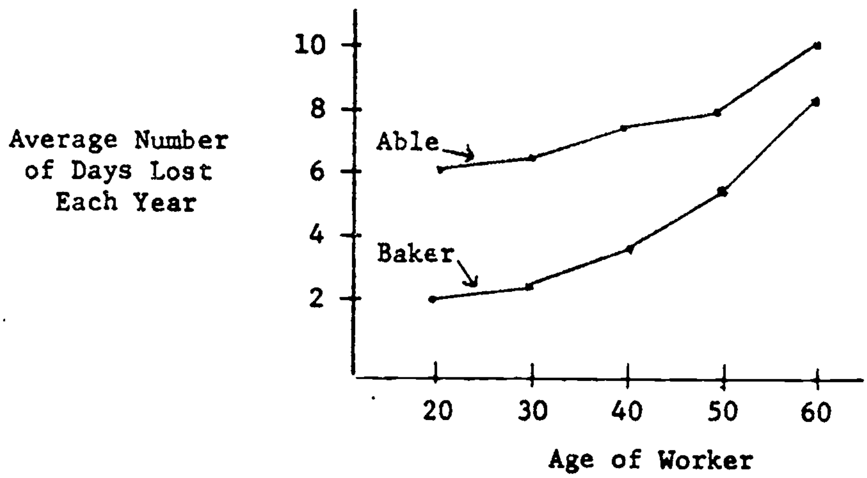
Finding: Cohabitors report a higher rate of violence than their married counterparts.

Think of hypotheses (possible explanations) to account for the finding. Write each hypothesis as a separate answer.

One hypothesis might state, "Cohabiting couples have psychological problems that lead to violent behavior." This hypothesis could be stated as a phrase, "Cohabiting couples have psychological problems." List your hypotheses in the form of phrases like this one.

Time Lost from Work Due to Illness or Injury

For two manufacturing companies, a study was made of the average number of days lost from work by assembly line workers because of illness or injury. The results were as follows:



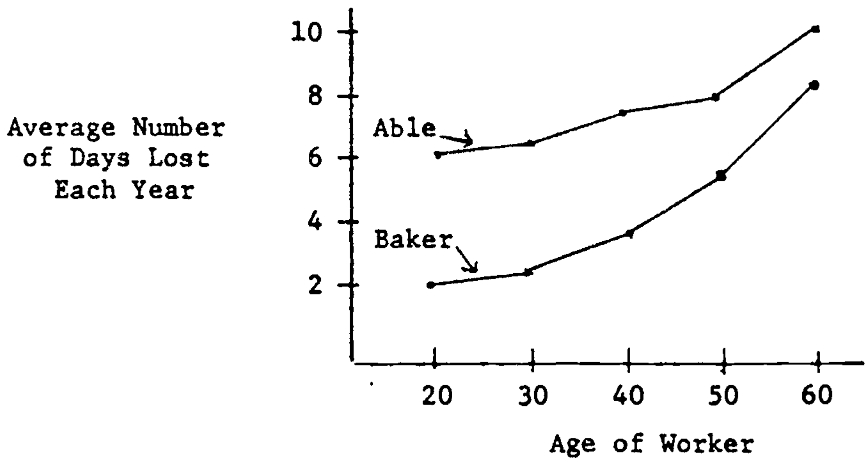
Finding: The average number of days lost each year was greater at Able Corporation than at Baker Corporation, especially among younger workers.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Time Lost from Work Due to Illness or Injury

For two manufacturing companies, a study was made of the average number of days lost from work by assembly line workers because of illness or injury. The results were as follows:



Finding: The average number of days lost each year was greater at Able Corporation than at Baker Corporation, especially among younger workers.

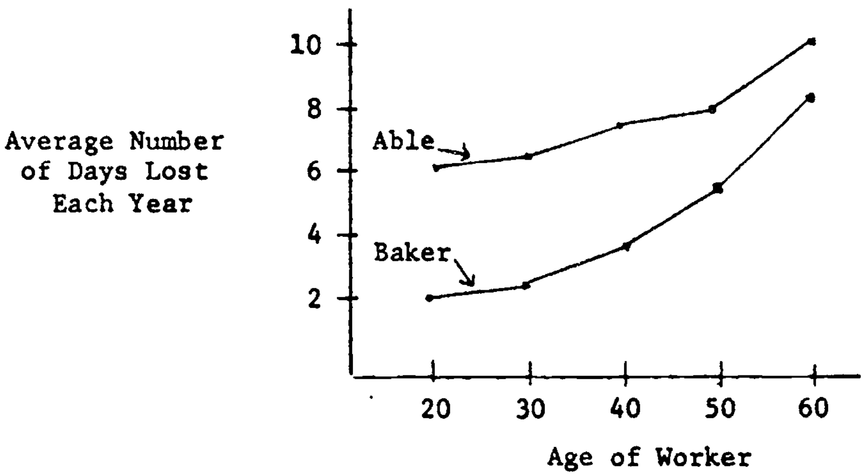
Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

Begin each hypothesis with the phrase, "Able Corporation..."

Time Lost from Work Due to Illness or Injury

For two manufacturing companies, a study was made of the average number of days lost from work by assembly line workers because of illness or injury. The results were as follows:



<u>Finding</u>: The average number of days lost each year was greater at Able Corporation than at Baker Corporation, especially among younger workers.

Think of hypotheses (possible explanations) to account for the finding.

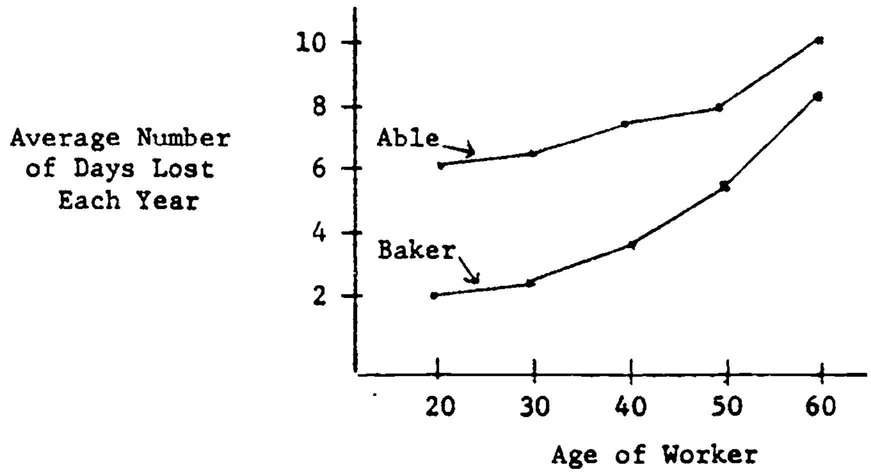
Write each hypothesis as a separate answer.

Begin each hypothsis with either one of two phrases:

"Able Corporation..."
or
"Baker Corporation..."

Time Lost from Work Due to Illness or Injury

For two manufacturing companies, a study was made of the average number of days lost from work by assembly line workers because of illness or injury. The results were as follows:



Finding: The average number of days lost each year was greater at Able Corporation than at Baker Corporation, especially among younger workers.

Think of hypotheses (possible explanations) to account for the finding.

Write each hypothesis as a separate answer.

One of the factors that influenced the number of days lost at Able Corporation might involve the problem of malingering--workers pretending to be sick or injured when in fact they were not. What other factors might have contributed to this finding?