Medical problem attributes and information-seeking questions*†

By P. Zoë Stavri, Ph.D.
Fellow in Applied Informatics
National Library of Medicine
Bethesda, Maryland 20894

This study was designed to explore the relationship between certain attributes of a diagnostic problem and a particular information-seeking question. Using case vignettes under experimental conditions, fifty residents in internal medicine and family practice informed the experimenter of the most important thing they would need to know to make a preliminary diagnosis. These data were classified nominally using a preexisting taxonomy. Significant results indicated that Quantification questions tend to be asked under urgent conditions and Verification questions tend to be asked when the least amount of information is presented. While the taxonomy used to classify the questions did not prove rich enough to describe information needs fully, results did suggest some consistency in question asking, a skill that could be developed further during medical education.

Questions have been analyzed in various ways as manifestations of information needs. The number of questions asked has been used as one measure of information need [1-2]. Questions have also been categorized post hoc to create a taxonomy of question attributes with which to measure information need [3-7] or categorized according to a predefined taxonomy [8-9]. Graesser et al. for example, created a taxonomy for categorizing questions by several linguistic attributes: (a) semantic function, or type of information requested; (b) pragmatic function, or purpose for which the question was asked; and (c) communicative function, or properties of the communication [10].

In the medical domain, clinical questions presented to online searchers of MEDLINE have been analyzed [11-12], and verbatim clinical information requests have been categorized according to several attributes, including subject of the request, anticipated source of the response, generality of information sought, and the response required by the information request [13-14]. Thus a number of studies have used information-seeking questions as a behavioral manifestation of the hypothetical construct of information need.

The practice of medicine involves a number of problem-solving activities that require information; these activities include etiologic studies, diagnosis, therapy, prognosis, and work concerning epidemiology [15-21]. Clinicians need to go beyond the information at hand when making a diagnosis because they rarely are presented with all the relevant data [22-24]. While there is evidence that the way clinicians ask for information and generate hypotheses in causal reasoning is heuristically driven, or a form of pattern matching [25], methods of inquiry are just as likely to vary among diagnosticians faced with the same problem [26]. Consequently, a physician's ability to ask information-seeking questions should have an impact on the diagnostic approach [27-28].

Research in the area of medical problem solving suggests that the nature of medical problems, especially their urgency and complexity and the amount of available information, differentiates diagnostic approach more than do the cognitive processes of the individuals themselves [29-30]. For example, there are two conflicting theories about the part played by experience in diagnostic competency [31-36]. One theory is that variability in medical problem solving is a function of the problem rather than the individual diagnostician's information-seeking pattern or problem-solving approach [37]. The second theory is that certain attributes of problems—specifically, the urgency of the problem as presented, the amount of available information directly related to the etiology

* This research was supported in part by a doctoral fellowship from the Medical Library Association, May 1991.
of the presenting condition, or the problem's complexity—exert a greater influence on diagnostic approach than do individual differences [38–39]. However, more empirical evidence is needed to determine whether problem attributes or individual differences contribute more to variations in information-seeking questions used during diagnostic reasoning.

This paper reports on an experiment conducted in 1992 that focused on the relationship between selected attributes of medical problems and information needs as represented by physicians' questions. More specifically, the study explored the relationship between urgency or the presence of etiologic information in experimentally controlled medical problems and the most important piece of information a physician needs to know to make a preliminary diagnosis. This paper describes the background to the study and the method used to collect and analyze the data. It then addresses some unresolved problems and considerations for future research.

Two research questions were designed to address the key issues: First, do certain attributes of medical problems influence information needs as manifested in information-seeking questions? More specifically, is there a relationship between the urgency of the problem or the presence of etiologic information and the critical piece of information defined as the "Most Important Thing," that a subject needs to know to make a preliminary diagnosis?

STUDY DESIGN

Design and materials

When studying the diagnostic process, Elstein et al. compared the results of high-fidelity experiments that used live subjects to low-fidelity experiments that used written case reports. After obtaining similar results, the researchers concluded that using case reports in the laboratory was an effective, time- and cost-efficient method of studying the diagnostic process [40]. Therefore, case reports were used in the present study.

A within-subjects, repeated-measures experiment was used to explore the relationship between medical diagnostic problems and information need. Experimental materials consisted of vignettes, or synopses, of actual cases that varied with respect to two problem attributes or conditions (urgency and etiology). Four presenting conditions were described: B1(Bulimia), B2(Alcoholism), B3(Depression), and B4 (Infarction). The vignettes, created with the assistance of medical experts, described actual cases that were not easily recognized, involved a single office visit, and were deemed equivalent to each other in terms of complexity.

One assumption of the study was that all of the problems used in the experiment met the following criteria: (1) the etiologic information had the same relative relationship to the presenting condition for all problems and (2) the problem was urgent enough to require immediate attention without being life threatening. The independent variables for the experiment were Urgent and Etiology. Base problems were varied by an additional sentence that made the problem Urgent (BU); or a sentence providing etiologic information relevant to the presenting condition (labeled Etiology or BE); or both sentences creating an Urgent/Etiology problem (BEU).

Each problem topic could be permuted in four ways (e.g., B1, BE1, BU1, BEU1), and each subject in the experiment received each topic only once and each problem condition only once. Because the pairing of four problem topics with four problem conditions created 4! or twenty-four combinations of problem sets and 576 possible permutations or orderings, and because a sample of this size was not practical, the problems were presented in a sequence designed to control systematically for first-order effects: Each of the four conditions appeared equally often in each of the four positions (first through last), and each condition appeared equally often after every other condition once in twenty-four trials.

Unit of analysis

The unit of analysis was a specific information-seeking question called the "Most Important Thing" (MIT). The MIT was defined as the information a subject indicated would be most important in making a preliminary diagnosis when given the information presented in each of the vignettes. Each subject generated four MITs (plus one practice MIT) during the course of the experiment.

To arrive at the nominal level value for the dependent variable, each MIT was categorized by its semantic function through Graesser's empirically tested taxonomy, which specified the type of information sought. Another assumption of the study was that the semantic function of a question was the best approximation of information need. Although Graesser suggested that any given question might have more than one semantic function, in the present study questions were assigned only one classification code. The decision to allow only one value for the MIT was pragmatic: The focus on a single value for the dependent variable permitted statistical analysis of the data. In addition, this approach fostered understanding of the relationship between problem and need through inferential statistics, compensating for any loss of information stemming from decision to go with the single value.

Pretesting

Pretesting of the study method proceeded through three distinct phases. The first phase involved seven
neurologists (both residents and faculty) at Boston University Medical School, and helped determine the format of the experiment. The second phase of pre-testing formalized the final version of the vignettes. Finally, the instructions to the subjects were fine-tuned.

Subjects

Subjects for the experiment were volunteers from two residency programs at the University of Wisconsin-Madison Medical School. These two programs were chosen because they were assumed to be equivalent, and because together they offered a subject pool large enough for the study. It was assumed that all subjects were interchangeable. In other words, family practice residents were considered equivalent to internal medicine residents, and all three post-graduate years were deemed equivalent to each other. Residents, having recently completed medical school, do not have a great deal of experience. It was expected that they would ask enough questions to make an investigation of the research question possible.

A unique identification number and four attributes were recorded for each subject:

Residency program: Family Practice (n=25) Internal Medicine (n=25)
Post-graduate year: PG1 (n=14) PG2 (n=14) PG3 (n=22)
Medical School: Wisconsin (n=11) Other U.S. (n=33) Foreign (n=6)
Gender: Female (15) Male (35)

All information except gender was supplied by the departmental offices. The demographic information was used in a secondary analysis of the data to test the homogeneity of the subject pool [41].

METHOD

A faculty member from each department consulted with the investigator regarding the solicitation of subjects. Residents were offered $5.00 for their participation and also entered into a drawing for substantial prizes. Sessions lasted approximately one half hour and were scheduled as early as 7:00 A.M. and as late as midnight at a place of the resident's choosing. Subjects signed informed consent forms and were assured of anonymity.

Subjects were given the opportunity to ask questions both after receiving the instructions and again after a practice session, which was held to clarify instructions and check the tape recorder. After the subject declared a final MIT for each vignette, the investigator confirmed it by restating it and giving the resident an opportunity to change it. At the conclusion of the experimental session, and after the tape recorder was turned off, the investigator explained the study and gave the resident an opportunity to discuss the theory behind it and the anticipated results.

Tapes were transcribed by a professional transcriber. The entire text of each subject's session was examined to extract the single MIT and assign it one semantic function code. Because literal meaning is usually context dependent [42], researchers have often reformulated requests before analysis to incorporate contextual information found in field notes [43-49]. For purposes of the present study, it was deemed necessary to examine the transcripts for contextual information to assign a nominal level value (i.e., semantic function code) to the dependent variable.

To ensure consistency in the extraction of the MITs, the following rules were established:

1. Rule 1: The final MIT, which was confirmed with the subject, was extracted for subsequent analysis.
2. Rule 2: From the final, confirmed MIT, the first information request in a complex request for information was treated as the MIT for subsequent analysis.
3. Rule 3: All extracted MITs were evaluated in context before being coded by semantic function. When a question was reformulated in light of contextual information, it was italicized to indicate that it had been transformed.
4. Rule 4: The essence of the question (the noun or object or action of the question) was determined. The question was then coded with the semantic function category, which corresponded to the type of information being requested [46].

On the first pass, MITs that could easily be classified into one of the categories in Graesser's taxonomy were assigned a code. Through this technique, about 50% of the MITs were classified. On the second pass, MITs that had been classified on the first pass were grouped by question category. Each MIT could then be compared to other MITs in the potential class. It was sometimes necessary to think in terms of the form of the answer expected by the implicit or explicit question when categorizing ambiguous questions.

Three medical school faculty members were consulted to confirm how MITs that required medical knowledge to answer should be coded. In this manner, the remainder of the MITs (for a total of 100%)

‡ While this step had the potential of introducing bias, pre-testing showed it to be necessary to ensure that the MIT had not been misunderstood. The full transcripts indicate a consistency in the restatement procedure.
were classified. The only area of disagreement was an information request for the results of a neurological exam; two out of three medical experts felt that although muscle strength was determined by a physician and not an instrument, relative quantity and not quality was the expected answer. Table 1 shows examples of MITs as classified using Graesser’s taxonomy.

When the coding was completed, the reliability of the measuring instrument was tested. Fifty-three randomly selected MITs were re-coded independently by four information specialists with an average of thirty years’ experience each. After the fourth and final re-coding, Cohen’s Kappa was computed to .74, indicating an acceptable level of agreement among coders after correcting for matches due to chance [46].

DATA ANALYSIS

Three major issues were paramount in deciding upon an appropriate statistical technique for addressing the research question. First, because the dependent variable was set at a nominal level, it was not possible to rely upon well-understood parametric techniques, such as analysis of variance (ANOVA) [47]. Second, most non-parametric multivariate techniques based upon the Chi-square and repeated measures designs violate the assumption of independence of observation central to the use of the Chi-square. Third, the relatively small sample size meant that techniques such as logistic regression could not be used [48]. It was possible to approach the research question indirectly by conducting two sets of analyses using planned Dunn-Bonferroni contrasts [49], based upon the Friedman test. The Friedman test is the nonparametric equivalent of the repeated measures ANOVA [51]. The contrasts were used to test hypotheses about which problem attributes were most likely to produce the most questions in any given question category and which type of question was most likely to be asked given a particular problem attribute. Planned contrasts were performed first, rather than using Scheffe’s method of comparisons (which would not require specifying hypotheses in advance), because it is of most value to test only the comparisons of interest [51].

Hypotheses

The study addressed two sets of hypotheses based on the pretest results. The hypotheses were stated in such a way as to make statistical investigation of the research question possible. The first group of hypotheses was best characterized by the following question: For a given question category, which problem attribute is most likely to have been applied? The second group of hypotheses was best characterized

| Table 1 |
| Semantic function of questions with examples of MITs |

<table>
<thead>
<tr>
<th>Question category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verification:</strong></td>
<td>Does she have chest pain?</td>
</tr>
<tr>
<td>Graesser et al.: implied yes/no/maybe who knows answers</td>
<td>Is he tired or not? Is he tired? [yes/no/maybe]</td>
</tr>
<tr>
<td>Indirect examples: on the surface disjunctive but only one alternative</td>
<td>Is the pain in the joints or the soft tissue? (which is true)</td>
</tr>
<tr>
<td><strong>Disjunctive:</strong></td>
<td>Probably the exact location of the pain.</td>
</tr>
<tr>
<td>Graesser et al.: which of two or more alternatives is true</td>
<td>What is the [place] of the pain?</td>
</tr>
<tr>
<td><strong>Concept completion:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: objects/names/places/times in response to who/what/where/when questions about states, events, and actions;</td>
<td></td>
</tr>
<tr>
<td>List of objects/names describing states, events, and actions</td>
<td>HIV risk factors. What [names] are on a list of states, events, actions related to HIV risks?</td>
</tr>
<tr>
<td><strong>Feature specification:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: request for value of an attribute in an object/attribute/value triplet</td>
<td>The quality features of her pain [in her arm]. What are the features [value] of the pain [attribute] in her arm [object]?</td>
</tr>
<tr>
<td>Request for medical tests in which the features (shape, form, sound, picture) are informing</td>
<td>I would like to look at the chest x-ray first. What does the chest x-ray look like? [x-ray/appearance/value]</td>
</tr>
<tr>
<td><strong>Quantification:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: request for the magnitude (how much) or frequency (how many) value of an attribute measured on a quantitative scale</td>
<td>Her sedimentation rate. What is her sedimentation rate?</td>
</tr>
<tr>
<td>Duration of symptoms (how many days, hours, weeks)</td>
<td>Her amount of purging in the last week. How many times has she purged in the last week? When she took it . . . how long ago? How many hours have passed since she took it?</td>
</tr>
<tr>
<td><strong>Causal antecedent:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: request for prior events and states which causally led to particular events in events in a causal chain</td>
<td>How it started . . . what’s the history of that? What caused him to do that suddenly?</td>
</tr>
<tr>
<td><strong>Causal consequence:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: request for events and states which unfolded causally subsequent to an event in question</td>
<td>[Ipecac] is not a thing people take on a daily basis . . . what are the effects of taking this stuff?</td>
</tr>
<tr>
<td><strong>Goal orientation:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: request for reasons and motives behind an intentional action</td>
<td>I need to know why she’s coming to see me. Why is she coming in to see me?</td>
</tr>
<tr>
<td><strong>Enabling:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: request for objects, abilities, states, and resources needed to perform an action</td>
<td>With regards to the joint pains, what things like aggravating and alleviating. What objects or states aggravate or alleviate the pain?</td>
</tr>
<tr>
<td><strong>Instrumental/procedural:</strong></td>
<td></td>
</tr>
<tr>
<td>Graesser et al.: request for the plan, style, (or instruments used) in performance of an intentional act</td>
<td>How actually is she using her arm?</td>
</tr>
</tbody>
</table>
by the question. For each problem type, which question category is most likely to be asked?

For hypotheses about problem attribute categories, the critical value was \( t(294) = 1.65 \). For hypotheses about question categories involving one contrast, the critical value was \( t(147) = 1.66 \). For hypotheses involving two contrasts per question category, the critical value was \( t(147) = 1.98 \). Alpha was set at .05.

**RESULTS**

Table 2 shows the frequency of each of the question categories under each of the problem conditions. Mean ranks were used for the planned contrasts.

The only planned contrasts that were supported by the data involved Quantification questions and Urgency or Etiology/Urgency conditions. These hypotheses were: (H3) significantly more Quantification questions will be asked under Urgency and Etiology/Urgency conditions than under Etiology and Base conditions (\( t = 1.87 \)); (H8) for Urgency conditions, there will be significantly more Quantification questions than other question categories (\( t = 3.92 \); and (H10) for Etiology/Urgency conditions, there will be significantly more Quantification questions than other question categories (\( t = 4.58 \)).

**DISCUSSION**

In a medical situation that required immediate attention, it is likely that a physician would need objective data to make a preliminary diagnosis. This need would be consistent with the finding of a significant relationship between Urgent conditions and Quantification questions. Other types of questions, such as Feature Specification or Concept Completion questions (see Table 1), rely upon the patient’s response, which in an Urgent condition might not be reliable.

*Post hoc* Scheffé comparisons showed that significantly more Verification questions were asked under Base conditions than under all other conditions (\( F[3, 147] = 2.60; t = 3.12 \). This relationship had not been anticipated. In the pretest, verification questions tended to be very specific in scope and Base problems—which provided the least amount of information—tended to generate Concept Completion questions (who, what, where, when) or Causal Antecedent questions (why), both of which seemed to represent needs for more general information. Therefore, it had been hypothesized that Verification questions would probably be asked under conditions in which more information was present initially (as in Etiology or Etiology/Urgency conditions), in an effort to confirm hunches.

One interpretation of the lack of significance of the majority of the hypotheses is that the pretest data upon which the hypotheses were based differed from the data used in the actual study. An explanation for these differences could be that most of the pretest subjects had been practicing in the field for years (i.e., were experts) or were specialists (e.g., neurologists), as opposed to the subjects in the experiment, who were residents (relative novices) and generalists.

As in previous information-seeking question experiments, relative homogeneity of knowledge was assumed on the parts of the subjects [52-53]. While the subjects in this experiment undoubtedly had roughly equivalent knowledge in a general sense, they may not have been equals in a specific sense. For example, under the Base condition of the Bulimia problem a resident asked, “Is she (the patient) abusing Ipecac?” During debriefing the resident indicated that he once had discussed a similar case with a physician faculty member. Subjects’ prior experience with similar cases may be cause for concern in subsequent studies, especially for a subject pool with varied experience (e.g., experts mixed with novices).

In spite of the investigator’s effort to control the subjects’ interpretation of the instructions, interpretation of the MIT may not have been consistent. An MIT could mean “pursue first” or “most valuable,” and the information the subject would pursue first might not be the most important when considered in a non-experimental setting. Two subjects emphasized that what they needed to know to make a preliminary diagnosis was secondary to what they needed to do to stabilize or treat the patient. While this kind of response supports the theory that the urgency of the problem strongly influences diagnostic approach and information seeking, it also indicates that the interpretation of MITs varied. The tendency to assign the term to a higher-level task was difficult to overcome, especially with subjects for whom English was a second language; these subjects represented 22% of the sample.

The Cohen’s Kappa coefficient was .74 for a random sample of fifty-three MITs, acceptable for the goals of the study. Had most hypotheses been supported, a higher coefficient would have been of concern; conversely, the 74% agreement could make the lack of

---

**Table 2**

<table>
<thead>
<tr>
<th>Semantic function</th>
<th>Base</th>
<th>Etiology</th>
<th>Urgency</th>
<th>ET/UR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verification</td>
<td>18</td>
<td>09</td>
<td>06</td>
<td>04</td>
<td>37</td>
</tr>
<tr>
<td>Disjunctive</td>
<td>05</td>
<td>01</td>
<td>02</td>
<td>02</td>
<td>10</td>
</tr>
<tr>
<td>Concept completion</td>
<td>08</td>
<td>09</td>
<td>10</td>
<td>06</td>
<td>33</td>
</tr>
<tr>
<td>Feature specification</td>
<td>06</td>
<td>10</td>
<td>13</td>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>Quantification</td>
<td>08</td>
<td>15</td>
<td>19</td>
<td>21</td>
<td>63</td>
</tr>
<tr>
<td>Causal antecedent</td>
<td>02</td>
<td>02</td>
<td>00</td>
<td>05</td>
<td>09</td>
</tr>
<tr>
<td>Other</td>
<td>03</td>
<td>04</td>
<td>00</td>
<td>00</td>
<td>07</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>200</td>
</tr>
</tbody>
</table>
significance suspicious. Because there was a 93% agreement among coders when the questions were coded as Verification or Quantification, versus only 84% agreement in coding questions into the other categories, the significant results for the contrasts dealing with Quantification questions and Verification questions were considered relatively reliable.

CONCLUSION

Whether Graesser's taxonomy is the best way to measure the semantics of information-seeking questions remains to be seen. While the taxonomy was used to classify more than 1,000 questions generated under a number of circumstances, and it is comprehensive and mutually exclusive, it may not adequately measure information needs in the medical diagnostic process. After conducting the experiment and coding the data, the investigator felt that some aspect of the information need was lost in the application of the semantic function codes. The semantic function of the question did not seem to describe adequately what the question was about.

For example, the information need in the hypothetical question "Did she try to commit suicide?" differs from the information need in the question "Is the pain localized?" The former question seeks to confirm an act and the latter seeks to confirm a sensation. Yet both of these questions are Verification questions, as defined by Graesser's taxonomy. While the anticipated answer to both questions is yes, no, or maybe, verification does not adequately describe the required information, which would verify either a process or a sensation. Perhaps in the medical domain finer distinctions are needed.

An alternative to rejecting Graesser's taxonomy would be to refine it into an improved measure of information need. To use an example from librarianship: When a particular book is classified, it is never classified in isolation, but rather in conjunction with the collection as a whole, in the context of the particular institution's needs. Classification puts one book in one slot in an abstract hierarchy of knowledge, which is represented, in a specific collection, by other books. The classification code alone, however, does not adequately describe what a book is about. In general, books are also assigned subject headings or descriptors to provide additional information about the intellectual content of the book.

Thus, while Graesser's taxonomy is useful in classifying information-seeking questions, it does not fully describe the dimensions of the need. Perhaps the added dimension needed to describe information need is a descriptor in conjunction with the classification. There are a number of ways to capture the descriptive dimension of information need that may be context dependent. For example, the expected source of the answer could supplement Graesser's taxonomy; this measure has been used in previous studies [54]. The taxonomy of potential answer sources would complement the semantic function of the question. Research in progress by the author will attempt to combine the two taxonomies with the same data set used in the present analysis.

The extent of the variation in what residents thought they needed to know in the experiment surprised senior medical school faculty who were consulted during the study. They had believed that by the residency phase of medical education, the students' diagnostic process would be more consistent in terms of the types of information sought. Yet, unless productive questioning is taught specifically, this skill might not be fully developed by the residency stage. Instruction in this skill is not a new idea; Covell et al. suggested that medical educators need to focus on productive questioning and train students to want what they need to know in making a diagnosis [55]. Other researchers, such as Da Rosa et al., have suggested that physicians' diagnostic skills would increase with improvements in information-seeking skills [56]. Thus, instruction in information need, information seeking, and question asking seems to warrant a more prominent role in medical education. Perhaps problem-based learning will impose a structure for improving students' information-seeking skills.

If the theory is valid that urgency and the presence of etiologic information affect diagnostic approach more than individual differences, perhaps that effect does not extend to information-seeking questions. The theory should not be discarded outright because a previous large-sample study (n=600) of the general population indicated that a significant relationship existed between problems and information-seeking questions [57]. For those findings to be extended into the medical domain, alternative explanations must be considered. One explanation may lie in the failure to validate the assumption that the residents were an homogenous group. Canonical correlations were used with the same data set to look at patterns of relationships among subject attributes, problem attributes, and questions, and the results suggested that homogeneity of subjects may not have been achieved [58].

In conclusion, while it appears that there is some consistency in question asking, and that it is feasible to develop a scientific approach to understanding the relationship between problems and information needs, this area of study is informed by theory more than empirical research, signaling potential investigative difficulties. While this study represents only a step into a relatively unexamined area, it also underscores the feasibility of further empirical investigations. Further progress in this area will depend on success in achieving increased control over experi-
mental conditions, further exploration of the role of subject attributes in defining information need, and further refinements in the taxonomy for classifying information-seeking questions in medicine.

Only with an improved understanding of the relationship between information-seeking questions and the problems that generate them can productive questioning be taught. Medical librarians should contribute their expertise in teaching productive questioning to investigations of this issue as well as to the design of the medical school curriculum.

ACKNOWLEDGMENTS

The author gratefully acknowledges the assistance of her dissertation committee at the School of Library and Information Studies, University of Wisconsin-Madison: Douglas Zweizig, Ph.D., chair; Jane Robbins, Ph.D.; John Boll, Ph.D.; Joan Littlefield, Ph.D.; Art Glenberg, Ph.D.; and Vernon Hunt, M.D.; and of David Lonsdorf, M.D. The author would also like to thank Alexa T. McCray, Ph.D., for her suggestions during the writing of this paper.

REFERENCES

2. MIYAKE N, NORMAN DA. To ask a question, one must know enough to know what is not known. J Verbal Learn Verbal Behav 1979;18:357-64.
12. MCCRAY AT, STAVRI PZ. The language of search requests: implications for direct access to a large bibliographic database. [Presented at the American Society for Information Science Mid-Year Conference, Ann Arbor, MI, 1988.]
33. NORTHRUP, op. cit.
35. BUTCHER, op. cit.
38. Spafaka, op. cit.
40. Graesser, op. cit.
43. Osheroff, op. cit.
44. Forsythe, op. cit.
49. Dunn OJ. Multiple comparisons among means. JASA 1961;56:52-64.
52. Horne, op. cit.
53. Miyake, op. cit.
54. Osheroff, op. cit.
55. Covell, op. cit.
56. DaRosa, op. cit.

Received May 1995; accepted October 1995