On the Domain Specificity of Expertise in an Ill-Structured Domain

Colleen M. Zeitz

Learning Research and Development Center
University of Pittsburgh

An important issue concerns the relation between expertise in highly-structured domains and ill-structured domains. This study explored the information processing abilities associated with expertise in literature, an ill-structured domain. Literary experts were superior to novices in gist level recall, the extraction of interpretations and the breadth of aspects addressed of literary texts but not of a scientific text. The results indicate that expertise in literature appears to share features with expertise in highly-structured domains, including domain-specificity and an abstract level of representation.

Our understanding of expertise is based primarily on studies of highly-structured domains. However, research is suggesting that the nature of expertise in ill-structured domains may be different (Glaser, in press; Voss, Green, Post & Penner, 1983). One way to examine this issue is to intensively study experts in a less-structured domain, discovering what characteristics they share with experts in other domains and where they diverge. Literature is a good candidate for such an endeavor, because the domain is a rich, complex one that is neither quantitative nor rule-based. Domains that have been studied by cognitive scientists tend to have mathematical rigor, well-defined rules and clear methods for evaluating the correctness of solutions and distinguishing winners from losers (e.g., chess, Chase & Simon, 1973; physics, Larkin, McDermott, D. Simon, & H. Simon, 1980; computer science, Adelson, 1981). These domains address well-structured problems. In contrast, the domain of literature can be characterized as ill-structured. This is not to say that the field itself is chaotic, but that its practitioners apply themselves to open-ended questions, for which agreed-upon solution evaluation procedures do not exist. While rules or laws that describe the processes a literature expert uses in analyzing a text may exist, they are not transparent. One might assume that findings from well-structured domains could be applied to domains like literature. However, generalizations of this type may be premature. Glaser (1986) has suggested that our understanding of expertise may be biased, because less-structured types of domains have been largely ignored.

In contrast to the literary theorists who describe the importance of literature-specific knowledge and information-processing abilities for processing literature (e.g., Purves, 1984; Mandel, 1976), there are literary theorists who espouse more global, all-encompassing views. Some claim that all of one's experience with the world contributes to the processing of a given literary text (Purves, 1971; Fillion, 1981). Some literary theorists also suggest that the same strategies employed in real life situations are applicable to literary experiences (Fillion, 1981). Others go so far as to claim that the skills gained from the study of literature are sufficiently decontextualized to be applied to all domains (Landow, 1989). This raises the issue of whether expertise in an ill-structured domain is domain-specific, as is expertise in highly-structured domains (e.g., Hatano & Osawa, 1983).

In this study, literary experts and novices were asked to read, perform recall tasks on and analyze both literary and non-literary texts. The results are used to address whether expertise constructs from highly structured domains apply to a less structured domain like literature. Specifically, the analyses consider whether literary experts demonstrate superior memory, abstraction and reasoning with regard to literary texts, but not with regard to non-literary texts. There were three groups of subjects: 24 students in their third year of high school, 13 graduate students who were in their third or higher year in an engineering Ph.D. program, and 16 graduate students who were in their third or higher year in an English Ph.D. program.

Results and Discussion

The most widespread finding in the field of expertise research is experts' superior memory for domain-relevant material. Verbatim, gist and word level free recall of a poem and a scientific text were compared for the three groups. At issue here is whether literature experts have superior recall of specific types of texts, or texts in general, and at what levels of detail any superiority is displayed.

Scoring of free recalls. As the first stage in coding the free recalls, the poem and scientific text were divided into propositional arguments. Each item recalled by subjects was matched to propositions from the appropriate text. A match was categorized as verbatim if it contained the exact wording as a proposition from the text and it was used in an appropriate context. A match was categorized as gist if the wording differed, but the meaning had been retained. Lastly, a match was categorized as being at the word level if some of the content words from a proposition were simply listed, or used in conveying an idea not contained in the text.

Verbatim recall. Subjects' verbatim recall scores for the poem and scientific text were examined first. A 3(group) x 2(text) ANOVA revealed a significant effect for group, $F(2,50) = 9.08$, $p < .0005$. As Figure 1 illustrates, both graduate student groups recalled more at the verbatim level than the high school students. The verbatim scores for the scientific text were significantly better than those for the poem, $F(1,50) = 4.33$, $p < .05$. The interaction was not significant, $F(2,50) = 1.87$, $p > .10$.

It is not possible to conclude that the effects of group in this analysis are due exclusively to the differences in the subjects' experience in the domain of literature. One confounding factor is age. A number of literary theorists claim that age affects literature processing ability (e.g., Purves, 1971, 1984; Fillion, 1981). Another difference may be in intelligence: Perhaps not all the high school students would meet the requirements for entering a doctoral program at Brown University. Thus, the high school students differ from the other groups on a number of important dimensions. The engineering group serves as a control for these factors. Their age, years of education, and ability to meet the University's admission standards were comparable to the English students'. However, their experience with the domain of literature was comparable to that of the high school students. Thus, they provide an important control group whose comparison to the English students can lead more directly to conclusions regarding the effects of experience in the domain of literature. For this reason, each analysis was repeated using only the data from the graduate students. As the purpose of these analyses is to clarify group effects, effects that do not interact with group will not be reported. The analysis of verbatim recall was repeated without the high school students' data. Neither the main effect of group nor the interaction was significant.

Gist recall. The next analysis concerned subjects' recall of both the poem and the scientific prose at the gist level. The effects of group, text and their interaction were significant, $F(2,50) = 7.18$, $p < .005$, $F(1,50) = 5.49$, $p <$

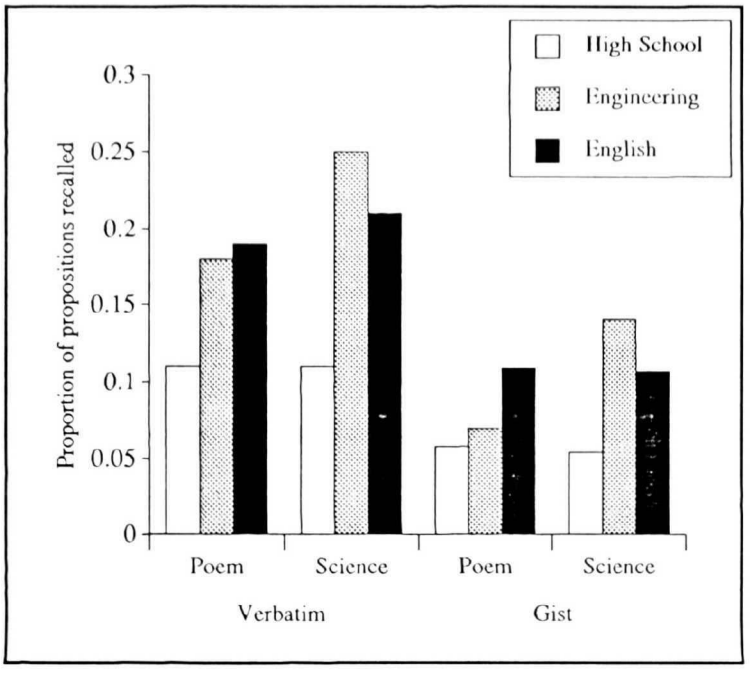

Figure 1. Proportion of items recalled from the poen and scientific text as a function of level of precision of recall.

.05 and F(2,50) = 6.45, p < .005, respectively. In Figure 1 it can be seen that the overall amount recalled by high school students appears to be less than that for either of the graduate student groups. The English students have better gist-level recall of the poem than the other two groups, whereas the engineers recalled the scientific text best.

The same analysis was repeated without the data from the high school students. The effect of group was not significant, $F(1,27) < 1$. At the gist level, the scientific text was recalled better by the engineers, and the poem was recalled better by the English students, as can be seen in Figure 1, $F(1,27) = 6.45$, $p < .05$. Thus, although the two groups of graduate students were equivalent in their overall recall, each group recalled the text within their own area of expertise better. This indicates that neither group had better overall memory for verbal material, but instead that the specialized knowledge they accumulated through experience allowed them to be more efficient at representing information pertaining to their field.

Word recall. An ANOVA for the number of items recalled at the word level did not reveal any significant effects. This category of recall was used very infrequently for all three groups and both texts (less than five percent for each).

The free recall findings suggest that the English students have superior memory for a literary text but not for a scientific text. The fact that the English students' memory was not superior for the scientific text indicates that their expertise applies to a limited set of texts, those that have the characteristics of literature, and not to all written material. Thus, these findings demonstrate the English students' domain-specific skills.

Abstraction

Experts and novices in the domain of literature may differ in the degree of abstraction in their analyses of literary texts. In order to examine these differences, during the second session subjects were asked to list the broadest range of essay ideas that they could generate. I analyzed both the breadth of the topics covered (i.e., the content of the sentences) and the depth to which they were addressed (i.e., the level of abstraction of the sentences). If the English students have deeper, more principle-governed representations of the literary texts, then these should be made manifest in the topics which they can generate regarding the literary texts, a poem and a short story. Their skills with regard to a scientific text should be attenuated because their representations of this text are no more abstract than those of the other groups'.

Scoring of topic sentence data. Subjects were instructed to make a list of the main topics they would include in an analysis of each of the texts, addressing aspects of the structure and content that they thought were important. The instructions directed them to express each of their ideas as the topic sentence of an essay on this theme and to list as many as possible. Fillion's (1981) system for categorizing students' inquiries served as the basis for the coding system. First, the coders classified the content of a given statement. The possible categories were: *events*, for events or plot; *characters*, for characters or relationships; *setting*, for setting, mood or atmosphere; *images*; *themes*, for themes or ideas; or *language*, for language, style or structure. Secondly, the coders classified the level of each sentence. Sentences were categorized as factual if they restated ideas explicitly mentioned in the story. Sentences were classified as interpretive if they offered a conclusion that could be disputed concerning some aspect of the story. The coders developed a number of additional categories for sentences which did not fit into this matrix (e.g., facts that were external to the story). Two coders independently scored all of the protocols. The coders were blind to the level of expertise of the subjects who had produced the topic sentences. All differences of opinion between the two coders were resolved through discussion.

The level of the topic sentences. The number of factual, interpretive and other statements subjects wrote were compared. The data are illustrated in Figure 2. A 3(group) x 3(text) x 3(fact, interpretation or other) ANOVA was performed. There was a significant effect of group, with English graduate students writing the most sentences, and engineers writing the fewest, $F(2,43) = 6.98$, $p < .005$. Subjects wrote the largest number of interpretive sentences and the smallest number of other types of sentences, $F(2,86) = 29.14$, $p < .0001$, but the relative production of sentence types interacted significantly with group, $F(2,43) = 15.63$, $p < .0001$. The tendency to write more interpretive sentences than factual was most exaggerated in the English graduate students and was reversed for the high school students. All of the groups wrote approximately equivalent numbers of factual and other statements for all of the texts. All of the groups wrote the most about the short story and the least about the scientific text, $F(2,86) = 23.08$, $p < .0001$. The amount written for each of the texts interacted marginally with group, $F(4,86) = 2.06$, $p < .10$. The production of different types of sentences did interact with text, $F(4,172) = 22.17$, $p < .0001$. There were more interpretive sentences than factual ones for the poem and short story, but more factual than interpretive ones for the scientific text. The three way interaction did not reach significance.

When this analysis was repeated, omitting the data from the high school students, the same pattern of significance for the results occurred, with one exception. The three-way interaction was marginally significant, $F(2,50) = 2.95$, $p < .10$. This may be due to the extremely large number of interpretive statements the English graduate students made in reference to the short story and poem. As Figure 2 shows, the English students' enormous advantage in producing interpretive statements was attenuated for the scientific text. This is an indication that their

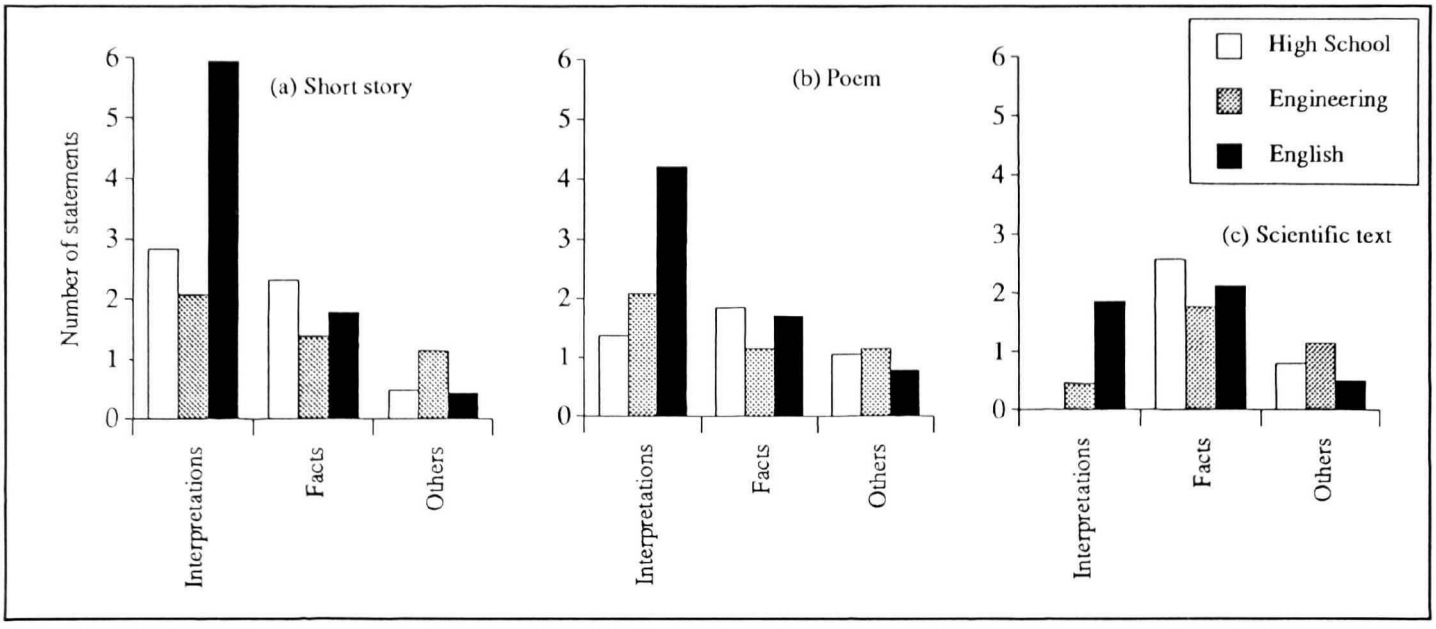

Figure 2. Number of topic sentences as a function of level of abstraction.

analytical skills only apply to a circumscribed set of texts. One possible explanation for this effect is that since the scientific text is written to be taken literally, it only supports a small number of interpretations. If this were the case, one would expect there to be a high degree of overlap between the items that were classified as interpretations. In fact, 22 of the English student groups' 28 interpretive statements were unique. This suggests that although there was a large number of interpretations which could be derived from the scientific text, the English students were not facile at producing them.

These effects can be considered in more depth in the context of each of the texts. For each of the texts, whether the high school students' data is included or not, there is a significant interaction between group and the types of sentences, p's < .05. In all three texts, the English students generated the largest number of interpretations, the high school students generated the largest number of facts, and the engineers generated the largest number of other types of sentences. Both when all three groups, or only the graduate student groups are compared, there is a significant effect of group for both the short story and the poem, p's < .05, but not for the scientific text, p's > .10.

The content of the topic sentences. These analyses consider the content of subjects' proposed topic sentences. There were six possible categories: events, characters, setting, images, themes and language. A 3(text) x 3(group) x 6(content category) ANOVA was performed. Because these data overlap with those of the analyses of factual versus interpretive statements, only the main effect of content category and its interaction with other factors will be considered. These results are illustrated in Figure 3. The number of sentences varied across the content categories, $F(5,215) = 29.10$, $p < .0001$. Overall, the events and characters categories were addressed most frequently, at equivalent levels. Setting and imagery were written about least frequently. The distribution across content areas interacted significantly with group, $F(10,215) = 7.61$, $p < .0001$. The high school students concentrated especially on characters, whereas the English students wrote more about language than any other topic. The English graduate students also heavily emphasized themes, which were largely ignored by the other groups. The distribution across content areas also interacted with the kind of text, $F(10,430) = 35.08$, $p < .0001$. For the short story, the character category was used the most. For the poem, the language category was used the most. Lastly, in

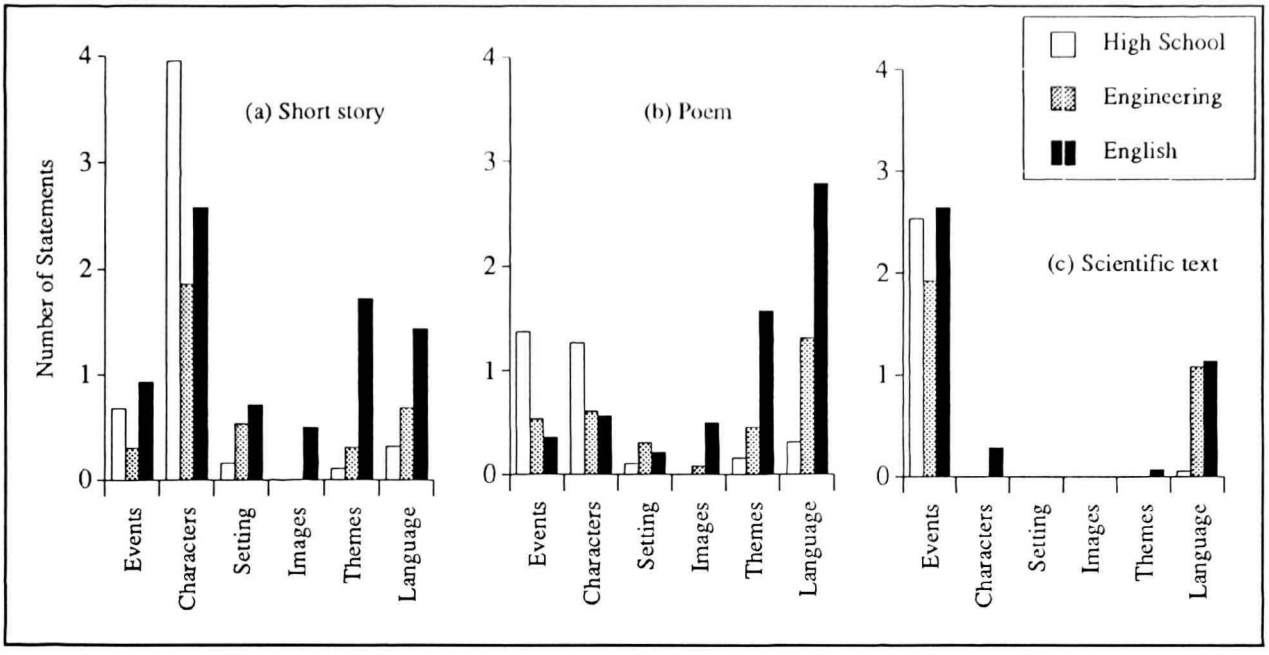

Figure 3. Number of topic sentences as a function of content category.

the scientific text, the events and language categories were used almost exclusively. Thus, on average, the subjects appear to note different aspects of each of the texts. The three-way interaction of content category, group and text also reached significance, $F(20,430) = 3.36$, $p < .0001$. The groups' patterns of category use differed for the poem and short story, but were nearly identical for the scientific text.

The same analysis was repeated without the high school students. The pattern of results was identical, with two exceptions. The interaction of group and content categories used did not reach significance, $F(5,125) = 1.39$, $p > .10$. While the English students wrote much more for each category, the ordering of their usage was similar to that of the engineers. The three-way interaction was marginally significant, $F(10,250) = 1.75$, $p < .10$. The patterns of category usage by the English and engineering students differed somewhat for the poem and short story, but they were practically identical for the scientific text.

When the texts were examined individually and the high school students' data were included, the interaction of group and aspect was significant for all three texts, p's $< .05$. When the graduate student groups are compared, this interaction only reaches significance for the poem, $p < .005$.

It is especially interesting that the English students did not write an exceptionally large number of topic sentences about the scientific text, even though they did so for the two literary texts. Notably, their production of interpretive statements is attenuated for this text. For the literary texts, the English students' extensive production of interpretations is indicative of a deeper level of understanding of these texts. They appeared quite capable of abstracting away from the facts presented in the literary texts. This supports the hypothesis that the English students' expertise is domain-specific. Their domain of expertise is not written language, but specifically literary texts.

Of the two literature texts, the poem is the more specialized genre. Narratives are more common reading fare. In the case of the poem, the aspects mentioned interacted with group, even when only the graduate student groups were considered. There was no significant interaction between content categories and group for the science

text when only the graduate students were considered. For this text the English students were at as much of a loss as the engineers. They were not capable of the more sophisticated and varied analyses that they performed in reference to the literary texts. In addition, this was the only text for which the English students did not produce many more interpretive than factual statements. Again, the domain-specificity of the English graduate students' analytic ability is apparent.

The text processing strategies of the English graduate students, seemed to be the most sensitive to the content and structure of each of the texts. Unlike the high school students, they did not focus on one dimension of each of the texts to the exclusion of all others. Also, the distribution of their comments across categories varies more between texts than the engineers'. They also use the more abstract end of the scale (images, themes and language) much more than the other two groups for the two literary texts. This is a clear example of their deeper analysis of these texts, and that they were able to go beyond the literal information stated. This claim is verified by the fact that the English students generated a higher ratio of interpretations to factual statements for the literary texts.

Conclusions

The results indicate that expertise in literature does exist, and that its general characteristics are like those of expertise in other domains. The English graduate students' performance on a variety of tasks justifies labeling them experts in literature. The implications of the expert-novice differences this study demonstrated are considered below.

These results refute the claims of literary theorists whose ideas are incompatible with the existence of domain-specific expertise in literature. The present findings indicate that experience in literature does lead to measurable advantages in processing literary texts. Some literary theorists would go further, suggesting that all one's experience with the world contributes to the processing of a literary text "by increasing the background knowledge brought to the task" (Fillion, 1981, p. 40). This implies that knowledgeable individuals can process literary texts in a more effective way than people with less knowledge. The results from the engineering graduate students refute this claim. Although they know a large amount about their own disciplines, their performance with respect to literary texts was often comparable to high school students and consistently inferior to the English graduate students'. Thus their large knowledge bases did not seem to aid their processing of literature as much as the literature experts' knowledge bases did. The results also contradict the claim of some literary theorists that experience in literature leads to advantages in other endeavors. The results indicate that the English students' expertise was limited to literary texts. The literature experts performed comparably to or worse than literature novices on a scientific text. The results suggest that the skills associated with expertise, even in ill-structured domains like literature, are domain-specific to a large extent.

Literary experts' knowledge bases. Literature experts' background knowledge allows them to focus on the appropriate level of detail. The memory results for literary texts indicate that the English students' representations differed from those of the other groups primarily at the gist level. Thus, the literature experts' meaning-level representation was more complete and accessible than the novices'. They suggest that literary experts have an additional kind of representation at a higher level of abstraction, that novices lack. This additional level is also evidenced in the experts' text analyses, in the way that they moved beyond the facts to interpretations of the texts.

Taken together, the results of this experiment offer strong evidence for the idea that literary experts have an additional, specialized level of representation, one step removed from concrete literal detail. By this argument, literary experts experience with literature has resulted in the development of a hierarchically organized knowledge

base with at least two levels. The first of these is a mundane literal detailed level of representation, which is comparable to that of novices. The second is a more abstract, principled level, which is the source of the expert-novice differences this study documents. The evidence suggests that experts must also have a certain facility in coordinating information from both levels. It may be the case that an additional, abstract level of representation may account for experts' superior abilities in other domains as well (e.g., Chi, Feltovich & Glaser, 1981; Dawson, Zeitz & Wright, 1989).

The domain-specificity of skills gained from the study of literature bears further investigation. This study demonstrated that literature experts' memory and analytic skills did not extend to another domain. However it remains to be resolved whether their superior analogical reasoning and argumentation skills (Zeitz, 1989) can be applied to other domains. In fact, although literature experts did not display an advantage in processing scientific texts, perhaps they would be superior to literature novices in processing information from domains more closely related to their own, such as history or another art form, such as painting. Clearly, further study of literature and other types of less structured domains can lead to a more complete understanding of the information processing abilities that develop with extensive experience in these types of domains.

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. Memory and Cognition, 9, 422-433.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

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

Dawson, V. L., Zeitz, C. M., & Wright, J. C. (1989). Expert-novice differences in person perception: Evidence of experts' sensitivities to the organization of behavior. Social Cognition, 7, 1-30.

Fillion, B. (1981). Reading as inquiry: An approach to literature learning. English Journal, 70, 39-45.

Glaser, R. (in press). Expert knowledge and processes of thinking. Chemtech.

Glaser, R. (1986). On the nature of expertise. In F. Klix, & H. Hagendorf (Eds.), Human memory and cognitive capabilities. Amsterdam: Elsevier Science Publishers.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. Cognition, 15, 95-110.

Landow, G. P. (1989). Hypertext in literary education, criticism and scholarship. Computers and the Humanities, 23, 173-198.

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

Mandel, B. J. (1976). What's at the bottom of literature? College English, 38, 250-262.

Purves, A. (1971). Evaluation of learning in literature. In B. S. Bloom, J. T. Hastings, & G. Madaus (Eds.), Handbook on formative and summative evaluation of student learning. New York: McGraw Hill.

Purves, A. (1984). The potential and real achievement of U.S. students in school reading. American Journal of Education, 93, 82-106.

Voss, J. F., Green, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skills in the social sciences. The Psychology of Learning and Motivation: Advances in Research and Theory, 17, 165-213.

Zeitz, C. M. (1989). *Expert-novice differences in memory and analysis of literary and non-literary texts.* Unpublished doctoral dissertation, Brown University, Providence, RI.

This research was supported by the James S. McDonnell Foundation Program in Cognitive Studies for Educational Practice grant number 87-42 to Kathryn T. Spoehr and post-doctoral fellowship to Colleen Zeitz. I am grateful to Robert Glaser, Robert Hoffman, Gregory Murphy, Clark Quinn, Kathryn Spoehr, and Jacque Wright for helpful comments and suggestions on earlier drafts of this paper.