Revising the Coherence of Science Texts to Improve Comprehension and Learning I: Traits of Mammals

Eileen Kintsch, Danielle S. McNamara,
Nancy Butler Songer & Walter Kintsch

Institute of Cognitive Science
University of Colorado
Boulder, CO 80309

Technical Report 92-03

Revising the Coherence of Science Texts to Improve Comprehension and Learning I: Traits of Mammals

Eileen Kintsch, Danielle S. McNamara,
Nancy Butler Songer, & Walter Kintsch

University of Colorado

How can information be obtained from a textbook chapter in such a way that it supplements and modifies the reader's already existing knowledge? The study reported here is the initial study in a project designed to explore this question as it applies to middle school students and science texts. The goal of this project is to optimize conditions such that the reader can obtain declarative information from a text. With that we mean that the reader will be able to reproduce the text, say what it is about (e.g., summarize it), and use it together with previous knowledge to infer novel facts and relations in the subject domain. That is, we are interested not only in the ability to reproduce a text, in the original or in some reduced or reorganized form, but also in the integration of the textual information into the reader's knowledge base, and the ability to use this information together with prior knowledge in many situations, including novel situations and new tasks. Our study explores both the reader's mental representation of the text as well as the model of the scientific domain that has been constructed or modified as a result of reading the text. Thus, in our view, comprehension is not restricted to the level of the text proper, but includes the conceptual processing necessary to understand the underlying situation and to acquire expertise in the subject matter. Our goal is to write texts that stimulate this deep, subject matter understanding.

We believe that the ability to obtain information from texts constitutes an important component of literacy. Our culture has traditionally used texts for this purpose, and this use will surely continue. Although it is clear that there is more to literacy than the acquisition of information, the present project is a modest one and strictly limited to this single purpose. A similar limitation of our project also must be recognized with respect to science learning: There is much more to science learning and teaching than giving students a textbook to read, however well written it may be. We in no way claim that improving textbooks is the key to instructing science. In fact, we are convinced that even the best textbook will fall far short of what a good teacher with good facilities can achieve. Indeed, model teachers may use textbooks only rarely. However, since not all schools have ideal science teachers, textbooks, in some form, will always be an important means for transmitting knowledge. Furthermore, texts are well suited for providing a common pool of knowledge to serve as the starting point for meaningful instruction.

Our goal is to explore how to write textbooks that are effective across a broad range of learner skills and knowledge. We have chosen to tackle the difficult problem of trying to accommodate instructional text to the needs of individual learners because we think that research on text comprehension has provided us with some understanding of the issues involved: It has opened up the possibility for principled, theory-based solutions. Experience, ingenuity, and good intuitions will only take us so far in our attempts to improve literacy in our society. We need to develop an educational technology on a scientific basis that can guide us even in those cases where ingenuity is lacking and intuitions don't help. Thus, in this project our intention is to go beyond a demonstration of expert writing of science texts and to derive explicit principles and procedures that would allow even less expert writers to do a better job of writing school texts.

Although at this point, we do not have a very clear picture of what these principles and procedures are, there are some research results to guide our efforts. The following findings serve as our starting point:

(1) A major problem in text comprehension is caused by the lack of coherence of many texts. Coherence on both a local and global level is fundamental to understanding (Kintsch & Vipond, 1979). Where there is a coherence gap in a text, readers must bridge it, which is only possible when they possess the requisite knowledge.

(2) The theory of discourse comprehension specifies the optimal level of coherence for a text such that readers' ability to reproduce a text can be greatly improved (Beck, McKeown, Sinatra, & Loxterman, 1991; Britton & Gulgoz, 1991). Making texts maximally coherent allows readers to construct better textbases (representations of the text in memory), and hence leads to improvements on tests that are based primarily on the quality of the textbase readers construct.

(3) However, a text can be overspecified such that more knowledgeable readers become irritated or bored with what they are reading. This kind of text is not optimal for the construction of mental representations of the scientific domain the text is about (situation models), which are integrated with the reader's prior knowledge (Mannes & Kintsch, 1987).

(4) For the construction of situation models, it is preferable to encourage a more active processing of the text during which the reader is repeatedly required to bridge coherence gaps with his or her own knowledge (E. Kintsch, 1990).

(5) This kind of active processing is possible only if the relevant knowledge is both available and accessible (Perfetti & McCutchen, 1987).

We conclude from these research findings that:

(6) It is not possible to write a single text that is optimal for every reader. Low-knowledge readers require a coherent text, while

high-knowledge readers should be allowed to do their own inferencing whenever that is possible.

(7) Therefore, texts must be adjusted to the reader's level of knowledge, but at the same time require the reader to contribute to the reading process whatever knowledge he or she possesses. Unbridgeable coherence gaps are disasters, whereas bridged ones can become educational opportunities.

(8) We envisage the solution to the problem of how to individualize instructional text in terms of a hypertext system. A long-term goal of this project is a system that evaluates the reader's progress and assigns him or her to a level of text difficulty that is appropriate, in that it encourages inferencing but ensures that the reader is indeed able to do so. Thus, we want to encourage higher-level reasoning, but at the same time ensure that, when left alone with the text without the help of a teacher, the student is capable of reasoning as required by the text. A straight-forward extension of the system would allow students to assess their own comprehension and learning as they read.

These are the long-term goals of our project. The study reported here has a much reduced scope. It does not focus on individual differences; instead, it was designed to investigate three limited, preliminary issues. First, it is necessary to find out whether the research results referred to above on which our arguments are based can be replicated with scientific text and younger students - our target population. Second, the methodology for the evaluation of changes in knowledge as a consequence of reading needs to be developed further. Third, we not only want to explore ways of improving existing science texts, but also to determine what sort of information should optimally be included in such texts. Therefore this study compared three kinds of text: an original middle school science text, a version with essentially the same content which was revised according to theoretical principles to ensure better comprehensibility, and an expanded version concerned with the same scientific topic which emphasized functional knowledge rather than factual knowledge. Thus, our goal was to create texts that would help the reader construct not only an accurate representation of the textbase, but also the conceptually linked knowledge structures that are the basis for domain understanding. In addition, the study attempts to relate the empirical findings, which are described below, to the theoretical model by simulating the changes in knowledge structures that arise from reading each of the texts. The results of this simulation are described in the concluding section of the paper.

The Experiment

The original science text we use is a chapter entitled "Traits of Mammals" from a popular junior-high biology text (Silver Burdett & Ginn, 1987). Thus, we are dealing with a topic about which the students possess a great deal of knowledge already. This is a major limiting condition of this study, as one would expect rather different results with unfamiliar topics in several respects.

The revised version was constructed as in Britton and Gulgoz (1991). Comprehension of the text passage was simulated by a computer program (Miller & Kintsch, 1980; Kintsch, 1988), and explicit bridging text was added whenever a coherence gap was found. However, since we are dealing here with a well-written text at the appropriate grade level, there were, in fact, very few such gaps at the local level (unlike the text Britton & Gulgoz worked with). Most of the coherence gaps existed at the macro-level; that is, explicit statements had to be included in the revision as to how each paragraph related to the rest of the text and to the overall topic. The actual content of the revised version remained the same as the original version, however.

The third, expanded version was constructed because we became dissatisfied with the original text as well as the revised version for reasons having to do with their biology content. These texts emphasized facts - traits of mammals. From a biologist's point of view, such an emphasis is unsatisfactory, for it is not enough to know that mammals have such and such a trait. Rather, the more important point would be to show what role the trait plays in the survival and success of this class of animals. Therefore, a third version with a different emphasis was written, largely including the earlier content but expanded in scope. It was fully coherent at both the local and global level, like the revised version, but considerably longer and much more abstract, because it focused on functional considerations rather than concrete facts.

In the experiment, we first assessed the students' knowledge about traits of mammals (by means of a sorting task and a questionnaire). Then the students read one of the three versions of the text. Finally, changes in knowledge were assessed (by the same tasks as before), as well as memory for the text itself. Thus, we are dealing here only with short-term knowledge changes.

With respect to text recall, we expected the revised version to be recalled substantially better than the original version, as in the Britton and Gulgoz (1991) study, with the improvement to be found primarily at the macro-level - the level at which the text was changed. For the recall of the expanded version, it is difficult to make specific predictions. On the one hand, it should be recalled well because, like the revised version, the expanded version was fully coherent, both at the local and global level. On the other hand, recall might suffer because much of the added content was quite abstract, specifying the relation between general traits and survival. For the knowledge changes, quite specific predictions can be derived from the theory of discourse comprehension (Kintsch, 1988): Concepts that are strongly related to mammal in the text should move closer to the concept mammal in the knowledge structure, and concepts that were discussed in the text but not related to mammal, should not move closer.

Method

Subjects. Thirty-six subjects living in Boulder, Colorado were paid 10 dollars to participate in this study. They were between the ages of 11 and 15 (M=13 years, 2 months) and were entering either 7th (n=15), 8th (n=14), or 9th (n=7) grade. The mean grade point average was approximately 3.4 on a 4-point scale (i.e., a B plus average). Subjects were assigned to one of three conditions based on the order of testing. There were no significant differences between the three groups of subjects in terms of their age, grade level, or grade point average.

General Procedure. Subjects were tested individually, either in their homes or at the university. The subjects first answered questions concerning personal information (e.g., age, grades, personal interests). They were then given two consecutive sorting tasks to complete, followed by a pretest which asked questions concerning mammals and their characteristics. Next they were asked to read silently one of three texts which discussed mammals and their characteristics. After they had read the text twice, they were asked to recall orally as much of the text as they could. Their recall protocols were recorded on a mini-cassette recorder. They then completed the same sorting tasks as they had at the beginning of the session, followed by a posttest which asked the same type of questions as the pretest. The entire session lasted approximately an hour.

Sorting Tasks. The subjects completed the two sorting tasks twice, once before reading the text and once afterward. The subjects were given a shuffled set of cards (3.5" x 1.5"). To familiarize them with the concepts printed on the cards, they were asked to look through the stack of cards and read each card aloud before beginning the two initial sorting tasks (i.e., before reading the text). They were then told to put the cards into piles according to how they thought the concepts should go together. They were told that they could make as few or as many piles or categories as they wished, there needn't be the same number of cards in each pile, they could change their mind and reorganize at any time, and that there was no correct or incorrect way to organize the cards. After stating that they had finished sorting the cards, the subjects were asked why they had put each pile of cards together and how they would label each set of cards.

The first set of cards they were given to sort was comprised of the 21 animal names listed in Appendix A, each of which was printed on a separate card. There were 9 mammals and 12 non-mammals (i.e., amphibians and fish) in this set. After they had finished sorting the animal cards, they were given the second set of cards to sort, which was comprised of the 16 animal characteristics listed in Appendix B. This list included 7 characteristics classified a priori as mammalian traits, 4 as traits that could be characteristic of either mammals or non-mammals, and 5 as non-mammalian traits.

The two sorting tasks were employed in order to assess the changes in the organization of knowledge that occurred as a result of reading the experimental texts. We were primarily interested in the keyword sorts, since that is what was discussed in the experimental texts. The animal name sort was included in order to orient subjects toward the mammal versus non-mammal distinction, so that this distinction would be more likely to play a role in the keyword sorting task, even before reading any of the texts on traits of mammals. In both the animal name and keyword sorting task the items were selected in such a way that they ranged from highly typical for mammals to highly atypical.

<u>Pretest and Posttest Knowledge Questionnaires</u>. After completing the sorting tasks, the subjects were given either one of two questionnaires comprised of multiple choice, true/false, fill-in-the-blank, and short answer questions concerning animal traits and behaviors. The subjects were told that the test was similar to tests in school, but that unlike in school, we were interested in what they thought was the answer to the question, rather than in whether or

not they knew the correct answer. Thus, they were encouraged to do their best, but to guess if they were not sure of an answer to a question.

Altogether there were 32 questions which were classified into three different types: (1) textbased questions for which the necessary information was stated in the original text (n=19); (2) inference questions which required some type of inferencing or analytic reasoning to answer the question (n=9); and (3) unrelated questions which dealt with information that was not included in any of the three texts (n=4). The three different types of questions were included for several reasons. Textbased questions were included to assess prior knowledge of the information presented by the text and also to assess learning of new information. Inference questions were intended to assess differential abilities to make inferences depending on which text was read. That is, we predicted that subjects who read the expanded or revised texts would be more likely to make inferences based on what they read than would subjects who read the original text. Unrelated questions were included to ensure that there were no differences between pretest and posttest and between groups for information that was not stated in the text and that could not be inferred from the text.

The 32 questions were divided pseudorandomly into two sets of 16 questions, with the constraints that each set include approximately half of each of the three types of questions, and that the questions in each set be matched as well as possible for difficulty and content matter. Half of the subjects in each group received one set of questions as the pretest and the other as the posttest, and the remaining subjects were given the tests in the reverse order. For each set, there were two random orders of questions. A random order of one set was paired with a random order of the other set. Half of the subjects were given the tests according to one of the random order pairs, and the other half according to the other random order pair.

The pretest questions and sorting tasks may have biased reading and recall by calling attention to particular points discussed in the text. However, these effects were equally distributed across the three text conditions.

Texts. After completing the initial sorting tasks and the pretest questionnaire, subjects were asked to read one of three texts (i.e., the original text, the revised text, or the expanded text). Subjects were assigned alternately to texts according to the order of testing. An equal number of subjects was assigned to each of the three text conditions (i.e., 12 Ss per text).

The original text was selected from a biology text (Silver Burdett & Ginn, 1987). It discussed a subject - traits of mammals - which was reasonably familiar to students in grades 6-8. The text was 1.5 pages long (590 words). It was completely coherent at the local level, requiring not a single bridging inference to run the simulation. On the other hand, the text had a list-like macrostructure which was less coherent. That is, the various subtopics were not always clearly marked in their relationship to the overall topic and to each other. For example, in the third paragraph, the subtopic "care of young" is not clearly signalled, nor is it specifically identified as one of the traits of mammals under discussion. Additionally, the paragraph structure of the text did not correspond entirely to its macrostructure. For instance, the discussion of instincts is divided over several paragraphs.

The revision consisted primarily in adding material that explicitly identified the major subtopics as traits of mammals and relating these traits to the success of the species. The result was a 2-page text of 821 words in which each paragraph corresponded to one of the traits discussed.

The original and revised texts therefore have the same content. The only difference is that the interconnections between the various content units is made more explicit in the revised version, especially at the global level, but to some extent locally as well. For instance, the original text merely says that "fertilization is internal in mammals", whereas the revision relates this fact to the subtopic of "care and protection".

The expanded version of the text was written because of some obvious shortcomings of both the original and revised versions, not in their structure, but in their content. Identifying various traits that mammals have does not really help a student to understand biology. It is more useful to know how these traits contribute to the success of this group of animals and their ability to survive. Hence, content to this effect was supplied in the expanded version, while at the same time some irrelevant detail was omitted (e.g., the names of the four chambers of the heart). The expanded version was therefore considerably longer, 4 pages or 1167 words. Furthermore, much of the added material was at a relatively high level of abstraction, so that, in effect, a few concrete details were exchanged for some rather lengthy, abstract discussions.

At the local level, all three texts were coherent. At the global level, only the revised and expanded texts were coherent. In terms of content, the situation model underlying the original and the revised version was essentially the same, whereas the expanded version introduced a great deal of new, higher-level discussion about the role which the traits of mammals play in a larger context. The three texts are reprinted in Appendix C.

Results

Reading Time

Approximate reading times were recorded with an accuracy within 1 minute. The average reading times for subjects reading the original, revised and expanded text were 6.83 min, 9.50 min, and 10.87 min, respectively. These differences were statistically significant, $F(2,33) = 7.15$, $p<.01$.

Knowledge Assessment

Question Answering. There were three types of questions, text based questions, inference questions, and general knowledge questions. Figure 1 shows the improvement from pretest to posttest on text based questions for the three different texts. The average percent improvement scores for the original, revised, and expanded text were 18.8, 30.1, and 15.0, respectively. The overall improvement was statistically significant, F(1,33)=30.61. However, the interaction with groups failed to reach statistical significance, F(2,33)=1,56, p=.22, and the predicted superiority of the revised text over the original and expanded versions also was not reliable statistically, F(1,33)=3.10, p=.09.

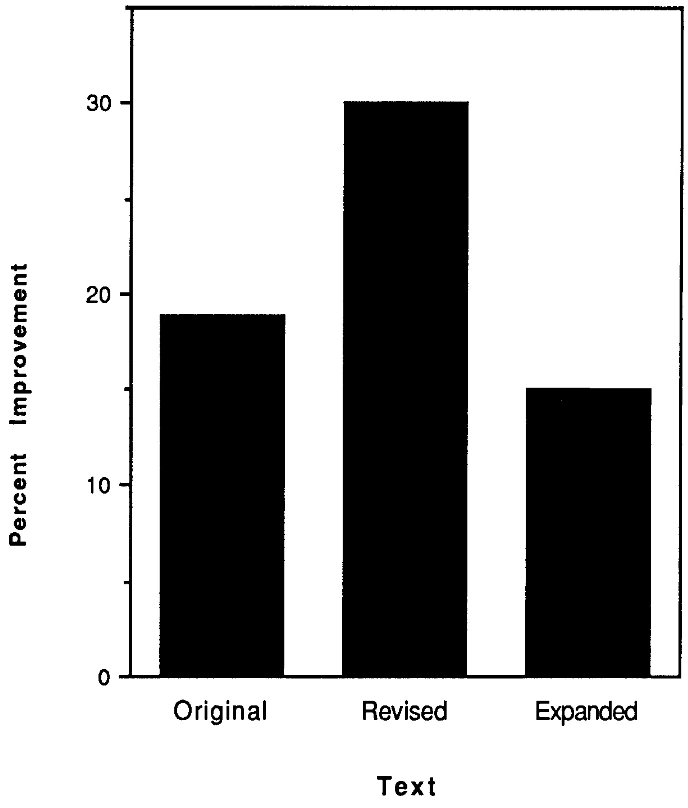

Figure 1. Improvement scores on text based questions.

Inference questions did not improve consistently from pretest to posttest. The average improvement scores for the original, revised, and expanded text were 1.3, -0.4, and 1.1, respectively. Knowledge questions unrelated to the text were answered equally well by all three groups and did not change from pretest to posttest.

In summary, the questions results proved disappointing. They (weakly) confirmed the prediction that subjects who read the revised text should be able to form a good textbase and hence do well on text based questions. The attempt to measure the deeper understanding of the problem domain that these texts communicated to the readers by means of inference questions was a failure, however. Fortunately, we found in the sorting task a more sensitive instrument that yielded more robust data.

Sorting: Animal Names. Since the experimental texts did not discuss most of the animals used in the animal names sorting task, the results of this task are only of ancillary interest as an indication of how well the categories of mammals and non-mammals were differentiated before and after reading.

A similarity matrix was constructed from the sorting task, indicating how often each item was sorted into the same category with every other item. The principal relations in this similarity matrix can be depicted graphically by means of the Pathfinder program of Schvaneveldt, Durso, and Dearholt (1985). Figures 2 and 3 show the structure obtained in this way for the animal names before and after reading the text.

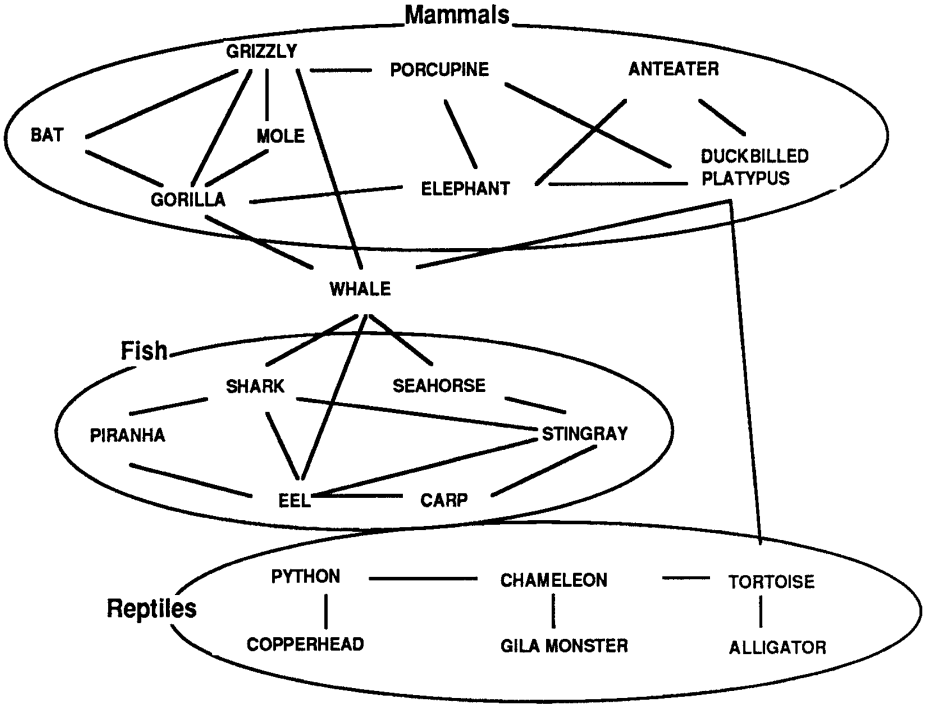

Figure 2. Pathfinder structure depicting pretest sorting of animal names.

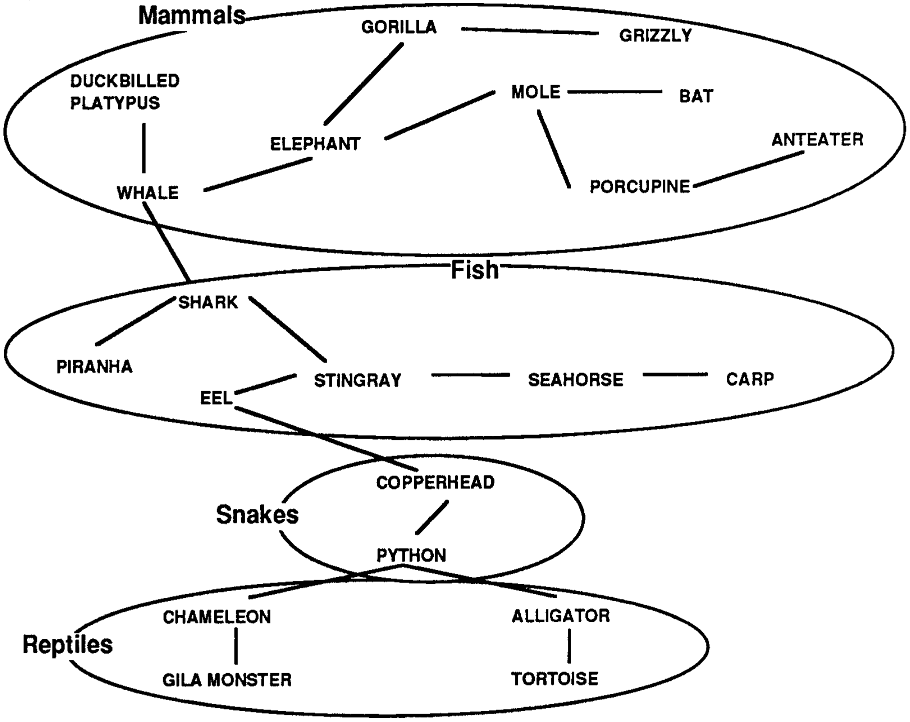

Figure 3. Pathfinder structure depicting posttest sorting of animal names.

Sorting: Keywords. The 16 keywords were analyzed in the same way as the animal names. The resulting Pathfinder structures are shown in Figures 4 and 5. As shown in Figure 4, traits of mammals form a discernible cluster even before reading the text, with non-mammalian traits arranged in a somewhat disorderly manner. A tendency to sort the animal characteristics according to more superficial criteria, such as "behaviors" or "things having to do with babies", is also evident in Figure 4. The distinction between mammalian and reptilian traits is clearly sharpened in Figure 5: The mammalian traits are clearly more organized and integrated, while the two others clusters, reptilian traits and shared behaviors, are more distinct.

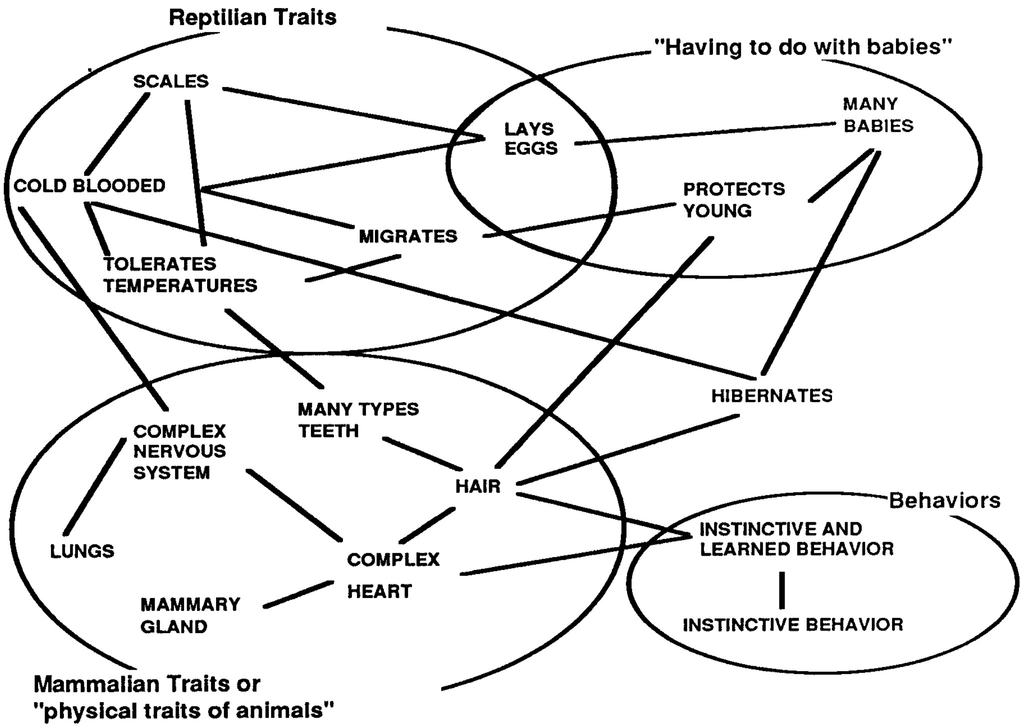

Figure 4. Pathfinder structure depicting pretest sorting of animal characteristics

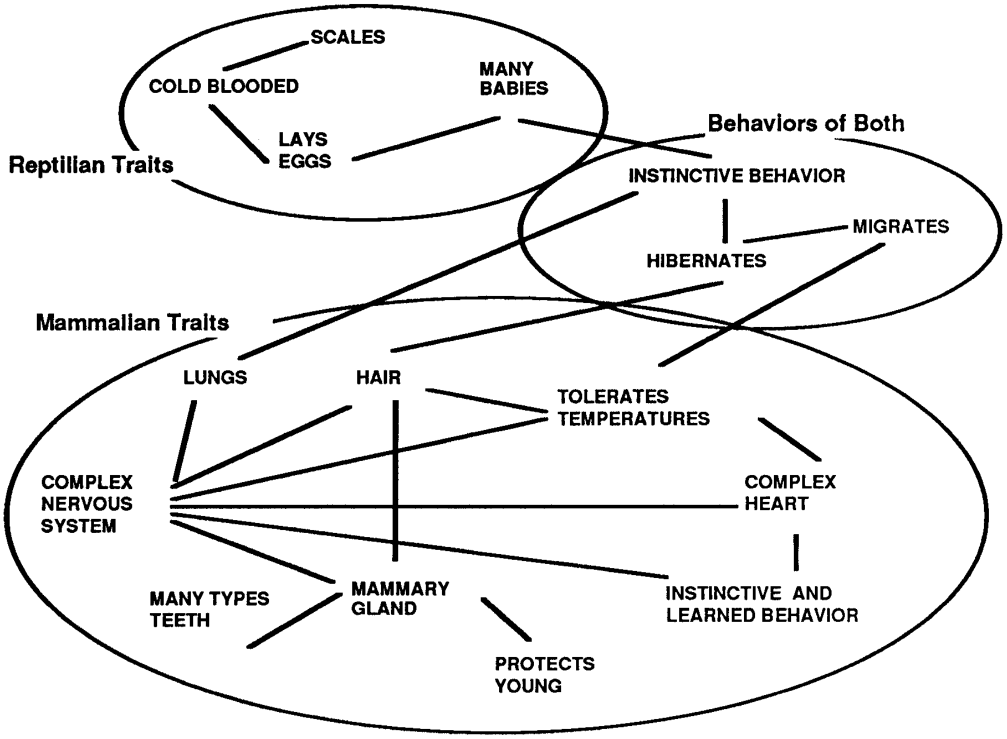

Figure 5. Pathfinder structure depicting posttest sorting of animal characteristics

Since we are particularly interested in how the sorting of the keywords changed as a function of reading the three versions of the experimental text, we need a more detailed, quantitative data analysis than the one afforded by the Pathfinder program. In particular, for each keyword we need a measure of how mammal-like it was perceived, and how this perception changed as a function of reading. Such an index is obtained if one counts for each keyword the frequency with which it was sorted together with other keywords that are characteristics of mammals. The subjects were not instructed to sort these keywords into mammal and non-mammal characteristics, but to the extent that this distinction plays a role in their knowledge structure, one would expect it to be a (partial) determinant of their sorting behavior in a free sorting task.

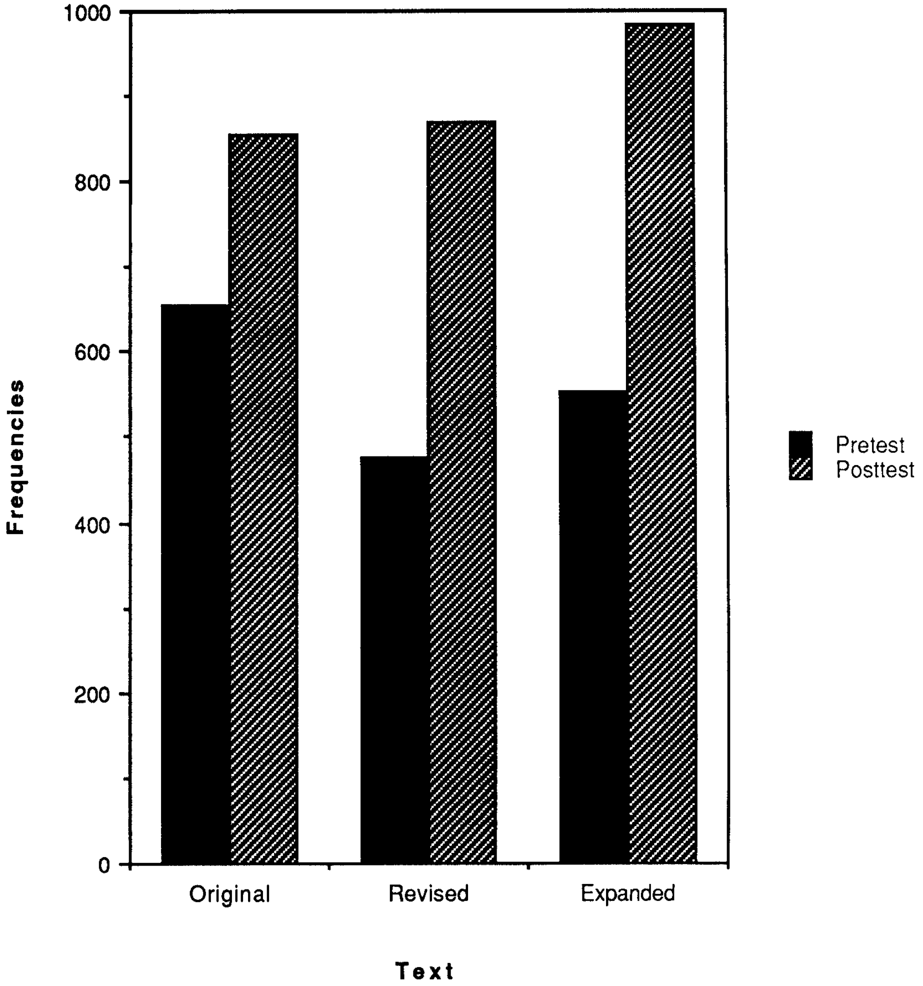

Figure 6. The total number of times items were sorted together with mammal characteristics on the pretest and posttest for three text versions.

Figure 6 shows the average frequency with which items were sorted together with mammal characteristics on the pretest as well as on the posttest. As one would expect for this set of keywords, "mammal" was a significant sorting principle even before reading the text, but it became even more so thereafter, especially for the revised and expanded text. This tendency differed in strength for different items. Correlating sorting scores over the 16 items, we note a fair amount of stability between the pre- and post-reading scores (.79, .76, and .78 for the three texts, respectively). If one compares how much subjects used the "mammal" category as a sorting principle for each item in the three different groups on the pretest, we find considerable variety: The correlation is .54 if one compares the original and revised groups, .58 for the original and expanded group, and .81 for the revised and expanded group. Apparently, subjects employed a variety of sorting principles on the pretest. The text, however, had the effect of strengthening "mammals" as a sorting principle in all groups, so that correlations between groups on the posttest sort are appreciably higher (.93, .89, and .94, respectively).

Instead of looking at all items together as was done in Figure 6, or at each separately, as in the correlations reported above, it is useful to categorize items into three classes: the first class of items, called Mammal items (n=7) are traits unique to mammals (such as "has hair or fur"); the second class of items, Shared items (n=4) are traits of both mammals and non-mammals (such as "hibernates"); the third class of items, Non-Mammal items (n=5) are traits found only in non-mammals (such as, "is cold blooded"). Pre- and post-sorting scores can be calculated for each item in these three classes. From these scores, change scores can be computed. These change scores indicate how much effect the text (as well as other intervening factors) had on the use of the mammal concept as a sorting principle. Positive change scores indicate that an item was perceived as more mammal-like after reading the text, negative that it was perceived less so. Figure 7 shows the average values of these change scores for the three item types for the three text groups. In every case the text had the effect of increasing the tendency to sort items together with mammals most strongly for those items that are in fact characteristic of mammals, somewhat less so for shared traits, while traits which are not characteristic of mammals were sorted less frequently with mammal items after reading any of the texts. The figure also illustrates that this strengthening of the mammal classification by reading the text was stronger for the revised and expanded groups than for the subjects who read the original text. A statistical analysis supports this claim. Before reading the text, the differences in how Mammal, Shared, and Non-Mammal items were sorted were much less strong among the three groups, $X^2(4) = 13.16$, than after reading, $X^2(4) = 162.07$. The stronger tendency to sort mammal traits together among readers of the revised and expanded texts as compared to readers of the original text was also significant, $X^2(2) = 89.2$. The change scores did not differ significantly if only the revised and expanded groups are included in the analysis, $X^2(1) = 1.76$. However, a chi square test of the tendency not to sort Non-Mammal items together with Mammal and Shared items, shows that this change was stronger in the expanded group than in the revised group, X2(1) = 15.07.

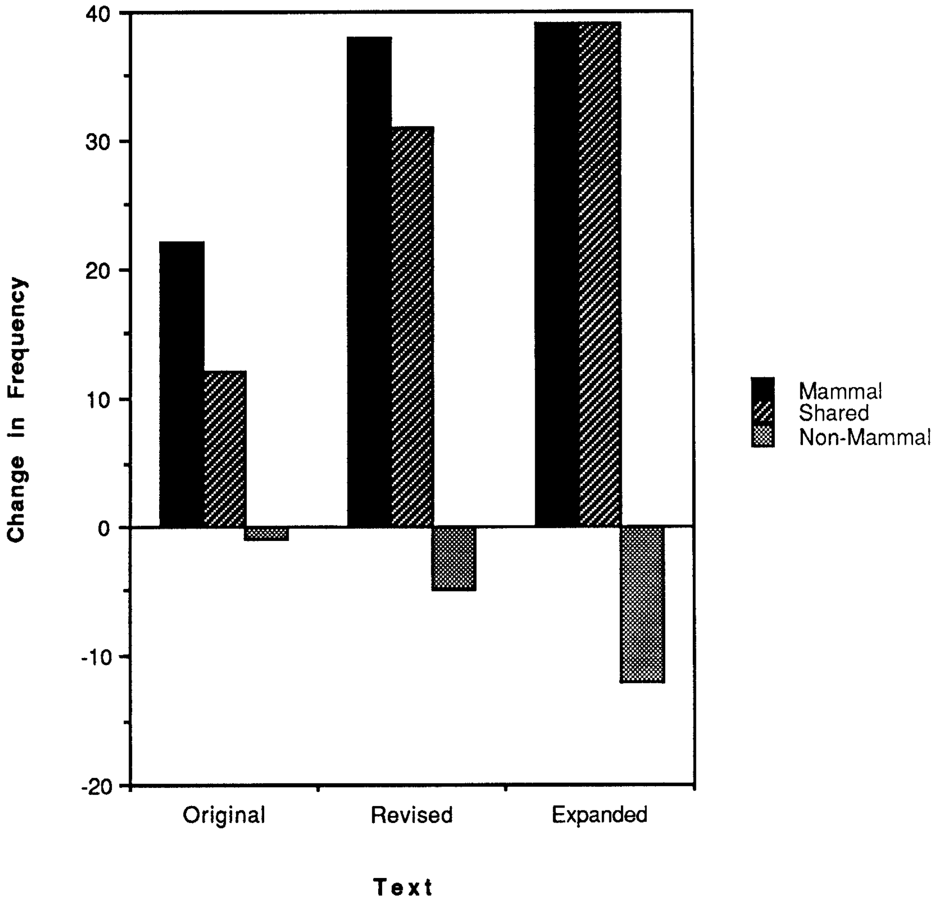

Figure 7. The change in the frequency from pretest to posttest with which three different types of items were sorted together with other mammal characteristics for three texts.

How much an item changes as a function of reading must depend on how strongly it has been classified as a mammal characteristic before. Thus, instead of looking at change frequencies as was done in Figure 7, a better measure of the effects of reading is obtained if one calculates the proportion of the possible change that an item actually achieved. Average change values expressed as proportions of the possible amount of change are shown in Figure 8. The trends are clearly the same as in the frequency analysis, except that the superiority of the Expanded group is more obvious than in Figure 7.

Thus, subjects started out with a weak tendency to sort the keywords into mammal and non-mammal characteristics. Reading the Mammals text greatly strengthened this tendency, more so for the revised and expanded text than for the original text. Indeed, the text effects were weakest for the original text, significantly stronger for both the revised and expanded text, with the expanded text being most effective. The sorting analysis thus confirms the impression gained from the analysis of the questions data that the expanded text had the most profound effect on the knowledge structure of the subjects, and that the revised text had a stronger effect than the original text.

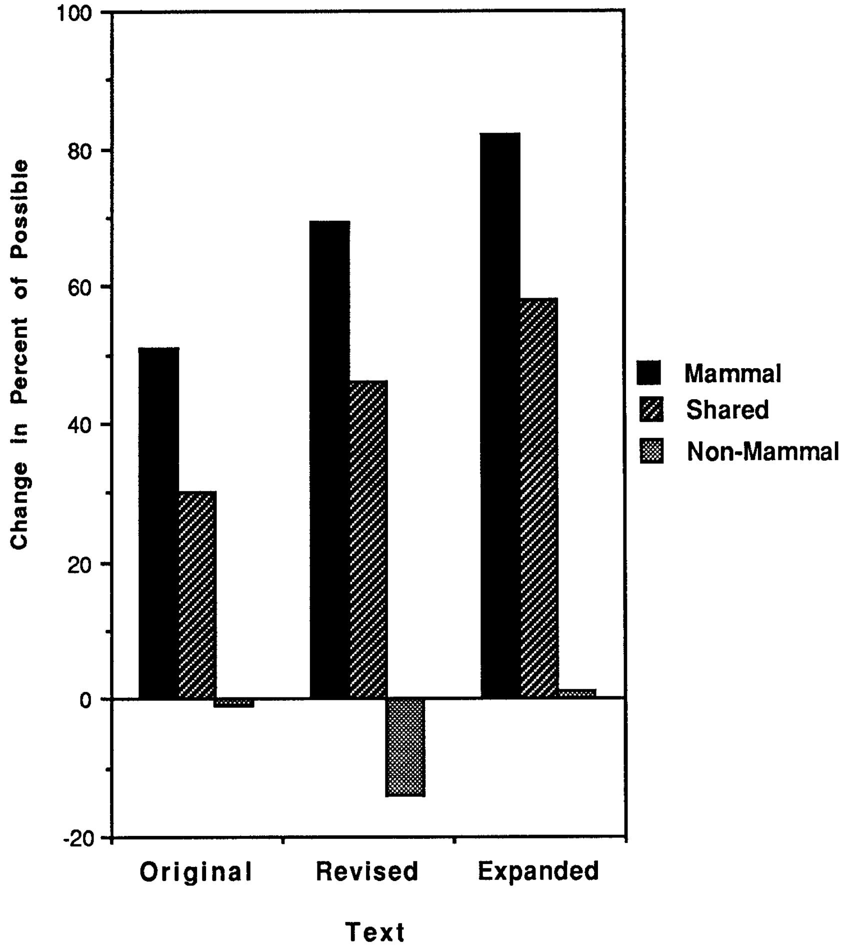

Figure 8. The change in frequency as in Figure 7, expressed as percent of possible change.

Text Recall

Several separate analyses of the recall data were performed. In the first, the number of words reproduced from the text was contrasted with the amount of elaborative, reconstructive recall. The second analysis is based on the same propositional scoring scheme as is used for the simulations reported below. Micropropositions and macropropositions were scored separately. Finally, to gain a clearer picture of how subjects perceived the macrostructure of the three texts, a more qualitative analysis was performed.

Number of Words recalled. In the transcript of each subject the number of words was counted that were derived from the text directly, whether verbatim or paraphrased, and the number of words that were not directly derived from the text. The latter included elaborations, generalizing inferences, knowledge intrusions, and relevant comments, which were all lumped together under the label "elaborations". Repetitions (which were relatively frequent in the spoken protocols) and erroneous statements were excluded. (There were not enough erroneous statements to make a meaningful separate analysis possible).

Figure 9 shows the results of this word count. There were significant differences between the three text versions in the total number of words produced: The revised group generated the longest protocols, and the original group the shortest, with the expanded group in between, F(2,33)=3.32, p<.05. Naturally, there were more text words than elaborations, F(1,33)=78.87, p<.01. Figure 9 also shows a tendency to generate more elaborations in the expanded group and fewer in the original group. In fact, 13% of all words in the original group were elaborations and 16% in the revised group, versus 25% in the expanded group. In spite of this trend in the mean values, there was a great deal of individual variation in this respect - how much students elaborate in recall depends as much on their personal style as on the text - so that this interaction was not statistically reliable, F(2,33)=1.45.

The analysis of the number of words recalled therefore confirms the experimental hypothesis: the revised text produced 41% longer protocols than the original text. It also suggests that the subjects who received the expanded version were more generative than the subjects who received either of the other two versions, although this result was not significant statistically in the word count used here. A more detailed analysis of the nature of the elaborations and inferences the students made in the three conditions is therefore necessary, but first we shall consider the results of the propositional analyses.

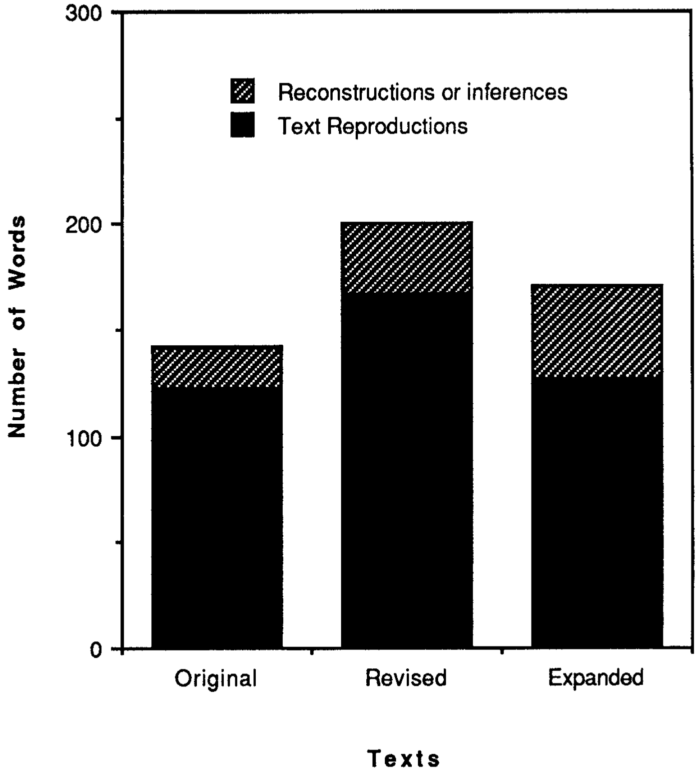

Figure 9. Average number of words recalled: Text words and elaborations.

Propositional Analysis: Microstructure. All three texts were propositionalized according to principles described in Kintsch (1974). For each text proposition it was determined whether a given protocol contained a verbal statement expressing this proposition either verbatim or in paraphrased form. Thus, this manner of scoring is concerned only with reproductive recall. Table 1 shows the results of this analysis of the recall data. If one merely looks at how many propositions are recalled for each text, it is clear that the revised text is better than either of the other two - in agreement with the analysis based upon a word count only. On the other hand, if one considers the different lengths of the three versions, a somewhat different picture emerges. For the original and revised versions the probability that a particular proposition will be reproduced in recall is the same - .20. But it is much reduced for the expanded text, where it only reaches a value of .12. Table 1 also shows an efficiency measure, obtained by dividing the average number of propositions recalled by the average reading time. By this criterion, the groups that read the original and revised versions are once more equivalent, while the group that read the expanded text recalled fewer propositions per second reading time.

Table 1. Number, percentage, and efficiency of recall for subjects reading different text versions.

	Original	Revised	Expanded
Average Number of Propositions Recalled	35.8	47.5	37.3
Percent of Textpropositions Recalled	20	20	12
Average Number of Propositions Recalled per Minute Reading Time	5.24	5.00	3.44

The revised version differed from the original version mainly at the level of the macrostructure. The original version was almost maximally coherent to begin with, so not much could be gained in recall of micropropositions. The failure to find a consistent superiority for the revised version over the original version at the level of microstructure recall is, therefore, just what one would expect. In order to find differences, we have to look at how well subjects recalled the macrostructure of the three texts. We do this in two different ways, one a straightforward extension of the propositional microstructure analysis, and the other a more detailed, qualitative analysis.

<u>Propositional Analysis</u>: <u>Macrostructure</u>. A macrostructure was formulated for each text using the title as the top level macroproposition. The first part of each text consisted of a general discussion of traits and was considered to be directly subordinated to the top level macroproposition. The remainder of the text discussed seven or eight specific traits of mammals, depending on the version. The subtopics "brain" and "behavior", which were discussed separately in the original text, were combined in the revised and expanded versions in order to streamline the basic structure. This subtopic was subdivided into two categories, "instinctual behavior" and "learned behavior" in all versions. In the expanded text "hibernation" formed a separate subtopic, whereas it was merely treated as an example of instinctual behavior in the other texts. Hence, the original and expanded texts each contained a total of nine macropropositions, while the revised text contained eight macropropositions. The macrostructure of the three texts is shown in Appendix D.

The original and revised texts both emphasize specific traits that mammals have, with the main difference being in how explicitly this relationship is stated. Hence, a list of nine macropropositions of the form "X is a trait of mammals" was used to score these two texts ("brain" and "behavior" were scored as separate macropropositions). Since the expanded text had a somewhat different focus (the role these traits played in the animals' survival), it is omitted from this analysis.

Figure 10 shows the probability of recall for the nine macropropositions (eight specific traits of the form "mammals have X", and the general statement "mammals have traits"). It is clear that the revision achieved the goals for which it was designed. Subjects who read the revised version reproduced each specific macroproposition with a probability of .5 or greater, and the general statement with a probability of .33. In contrast, subjects who read the original version successfully reproduced some of these macropropositions (those emphasized in the original text), but did poorly on fully half of the specific macropropositions as well as on the general macroproposition. Thus, while the revision did not do much for the total number of micropropositions recalled, it did achieve its purpose in that it directed subjects' attention more or less evenly to all of the subtopics discussed in the chapter.

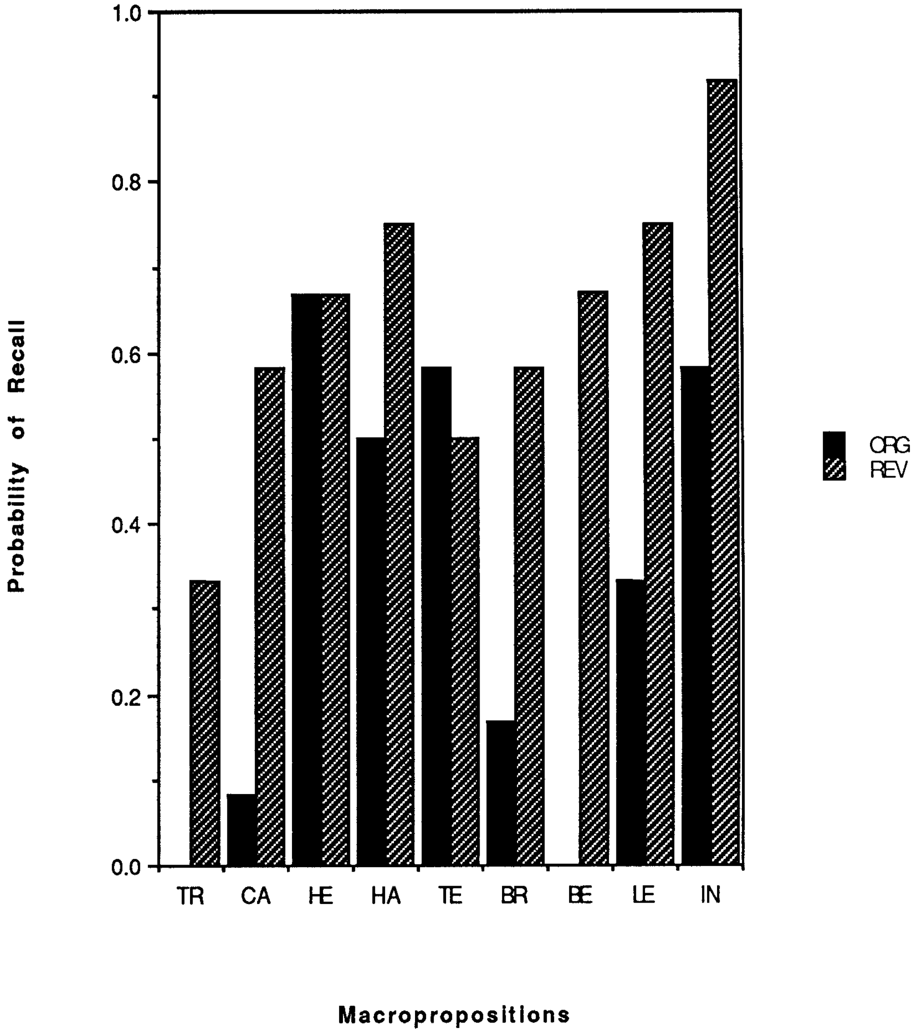

Figure 10. The probability of recalling each macroproposition in the original and revised version. (TR: traits of mammals; CA: care; HE: heart; HA: hair; TE: teeth; BR: brain; BE: behavior; LE: learned; IN: instinct).

Propositional Analysis: Conceptual Levels. In order to analyze macro-level recall further, and include again into our analysis the expanded version of the text, subjects' recall was scored in terms of a qualitative criterion, designed to explore the differences among all three text versions. As already mentioned, the primary difference was that the revised version had a categorical structure, listing various traits of mammals, while the expanded version emphasized the functional role of these traits for survival. To see whether this difference in the nature of the texts is reflected in the subjects' recall, a scoring scheme was developed based on the distinction between three levels of information. This made it possible to determine whether subjects recalled anything at all about specific points discussed in the text, as well as to what extent they recognized the subtopics in the text as traits of mammals, and further, whether they related the specific traits of mammals to their success and adaptability.

We call the three levels of analysis the level of Detail, Trait, and Function. The texts were divided into 16 major points, all concerned with the general concept of traits or specific traits of mammals. Recall is scored at the level of Detail if a protocol contains anything at all about one of these major points. How much is said about a particular point is not taken into account by this way of scoring, nor how well it is said. Recall is scored at the Trait level when a subject specifically mentions that this portion of the text is about the traits of mammals, or that X is a trait of mammals. In other words, we are selecting that subset of the Detail scores for which an explicit connection exists between some fact X and the concept traits of mammals. Finally, at the Function level, we select another subset of the Detail scores where a specific connection has been made between some fact X and its role in the success and survival of this class of animals.

For example, if a subject says that "molars are used for crunching plants" it is scored in the category "teeth" at the Detail level only; if the protocol says "mammals have several types of teeth", it would be scored at that level as well as at the Trait level; if the protocol says "different types of teeth help to use different food sources" it would be scored at the Detail and Function level.

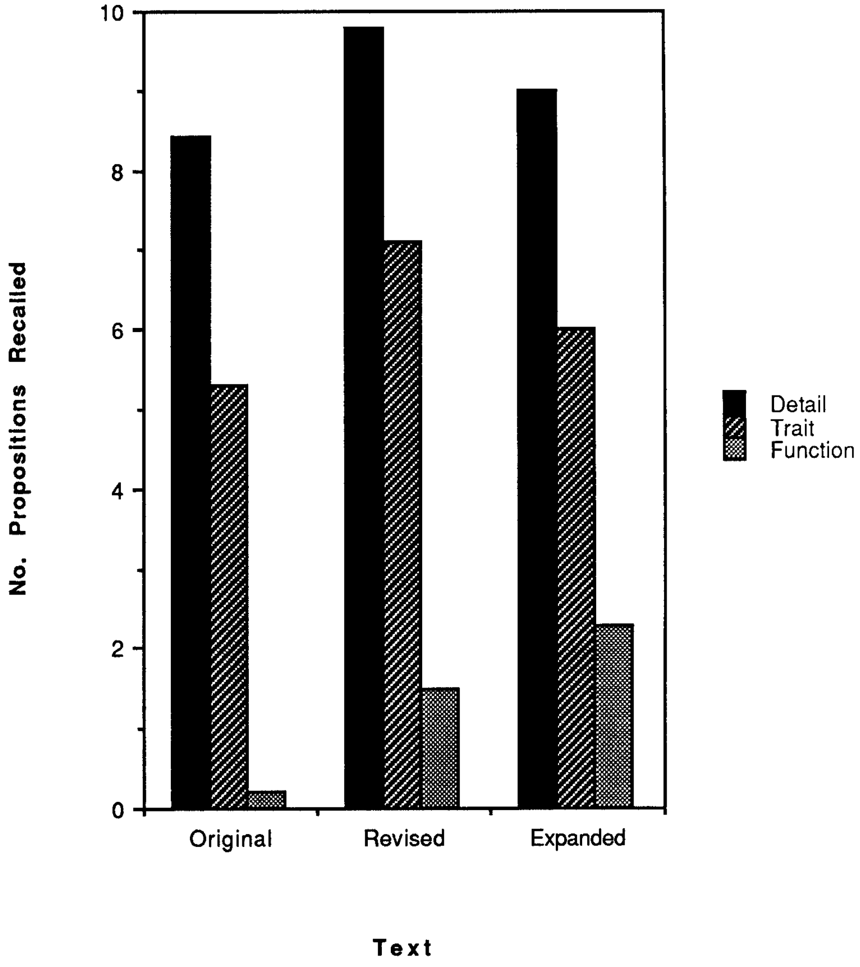

Figure 11. Average number of propositions recalled at three conceptual levels.

Figure 11 shows the results of this analysis. Since the three analyses are completely interdependent, separate analyses of variance were performed at each level of scoring. At the Detail level, Figure 11 confirms earlier results that the revised text, if anything, produced the most recall. However, this result just failed to reach statistical significance, F(2,33)= 3.14, which is perhaps not surprising, given the crude method of scoring employed here. Much more interesting and important are the two analyses concerned with the role subjects assign to the information they recall in the macrostructure. If we focus on the Traits analysis, we see that for all versions subjects managed to relate their recall to "traits of mammals" reasonably well, but that the revised group were best - which was, after all, the major goal of our text revision. These between group differences were statistically reliable, F(2,33)=5.39. $p<.05$. The results are even more striking if one considers the Function analysis. There were hardly any functional relations expressed in the recall protocols of subjects who had read the original text, but this number increased sharply for the revised group, and was highest for the expanded text, F(2,33)=9.99, $p<.01$.

We thus conclude that, in terms of subjects' recall, both the revised and expanded revisions achieved their purpose: The revised version made readers more aware of the significance of the traits categorization in the original text, while the expanded version managed to emphasize not just the categories but also their function.

Discussion

The results described here are complex but quite consistent. Only a few key concepts are necessary to understand their structure.

Of crucial importance is the distinction between situation model and textbase. Measures of performance based primarily on the textbase do not yield the same answers as those based on the level of the situation model. What is an optimal representation in one case is not necessarily optimal in another. The revised text was always

The Simulation

The three experimental texts were propositionalized according to the method described in Kintsch (1974), except that each concept that occurred as an argument of a proposition was also introduced separately by an expression corresponding to an existential proposition. The total number of propositions, as well as the total number of different arguments for each text version is shown in Table 2. The macrostructures for the three texts are as described earlier for the propositional analysis and are presented in Appendix D.

Table 2. Description of the three experimental texts, their textbase, and the simulation of the comprehension process.

	ORIGINAL	REVISED	EXPANDED
TEXT:			
Length (pages)	1.5	2.0	3.0
Length (words)	590	821	1167
TEXTBASE::			
No. of Propositions:	244	309	432
No. of Different Arguments	77	77	105
No. of Macropropositions	9	8	9
Levels in Macrostructure	3	3	3
SIMULATION:			
No. of Processing Cycles	33	41	54
Ave. No. Prop. per Cycle	7.39	7.54	8.00
No. of Links	1,175	1,452	2,221
No. Links per Proposition	4.82	4.70	5.14

Each textbase was divided into processing cycles at sentence boundaries, except that sentences were combined when they contained four or fewer propositions. The current macroproposition was added to each cycle, and the most highly activated proposition from the previous cycle was carried over from cycle to cycle. In addition, concepts introduced in an earlier cycle were reinstated as needed in later cycles. Associative knowledge activation was not included in the simulation.

The simulation process was then performed cycle-by-cycle for all three texts, as described in Kintsch (1988; in press), using the CI-program developed by Mross and Roberts (1991). The calculation of the predicted long-term memory strength values for the various text propositions is discussed in Kintsch and Welsch (1991). The long-term memory strength values calculated in this manner are the basis for the simulation of the sorting results reported below. As Table 1 shows, the number of links created by these simulations increases approximately linearly with the length of the text, so that the number of links per propositional node remains relatively constant.

Modelling the Sorting Data

The goal of this simulation is to predict the pattern of results found in Figures 7 and 8: Reading strengthens the tendency to sort Mammal items together; the same is true for Shared items, but to a lesser degree; and Non-Mammal items are sorted less frequently with mammal characteristics than before.

In order to simulate these results, we need to formulate a theoretical model that specifies how strongly the concept "mammal" is retrieved by each of the 16 retrieval cues in the episodic memory trace of the text. These strength values can be considered as predictions for how strongly the text should influence the sorting of each item as a mammal characteristic. Note that we are not making predictions about knowledge structures since we have no way of knowing what these might be. We are only predicting the change in the knowledge structure that occurs as a consequence of reading one of the experimental texts.

There are many possible retrieval models, but also some constraints, primarily the "fan" effect and the "strength" effect, as described in Anderson (1983). The former arises in the CI-model as a within cycle effect as a consequence of the normalization procedure used: the broader the fan, the higher the activation of the central node, hence the lower the activation of the fan nodes, which are divided by the maximum value. In contrast, the "strength" effect occurs between cycles, when items are repeated in different contexts. As a further, a priori constraint it is assumed that the strength of a retrieval path can never be greater than its weakest link. A retrieval model that satisfies these three constraints is obtained analogously to the computation of joint probabilities by defining the strength of a path between the retrieval cue and the concept "mammal" as the product of the strength of all the links between them, with the maximum strength of a segment being +1 in this computation. (The alternative of first computing probabilities from the strength values via the ratio rule was rejected, since it would yield fan effects but not strength effects - if A has a fan of n concepts, each linked with strength s or with strength k*s, the ratio rule would yield the same probabilities and thus information about absolute strengths would be lost, in contradiction to the Anderson data).

The retrieval strength of a test item is the sum of of the strengths of all paths between the elements of the item and the concept "mammal". Consider, for instance, the item "has several types of teeth", which has the elements (a) TEETH, (b) POSSESS[TEETH], and (c) SEVERAL-TYPES-OF[TEETH]. Since there is the proposition POSSESS[MAMMAL,TEETH] in the textbase, (b) is directly linked to MAMMAL with some value v_1, which can be looked up in the long-term memory matrix obtained from the simulation of the comprehension of the text. Similarly, TEETH is linked in the long-term memory matrix to POSSESS[MAMMAL,TEETH] with some value v_2, so that the strength of the path from TEETH to mammal is v_1*v_2. If either v_1 or v_2 is >1, its value is set equal to 1. There remains the problem what to do about (c), which is not directly contained in the textbase. One could assume that inferences would be made by the reader, for example, identifying "several types" with the text proposition FOUR-TYPES-OF[TEETH]. While such an inference is fairly uncontroversial in this case, determining just where inferences would occur in the text would, in general, introduce an undesirable element of subjectivity into the analyses. We choose, therefore, only to consider the direct overlap between cues and text. The predicted path strength for the three elements of the test item "has several types of teeth" and "mammal" would therefore be $v_1 + v_1 * v_2 + 0$.

Path strength between each of the 16 sorting statements and the concept "mammal" were computed in this manner. Special considerations were necessary concerning item T12, "mostly instinctual", which was changed for the purposes of the simulation to "instinctual rather than learned behavior", to make explicit the contrast with T7, "instinctual and learned behavior". Thus, for T7 the strengths values for instinctual and learned were added to obtain the cue weight, while for T12 the strength of learned was subtracted from instinctual.

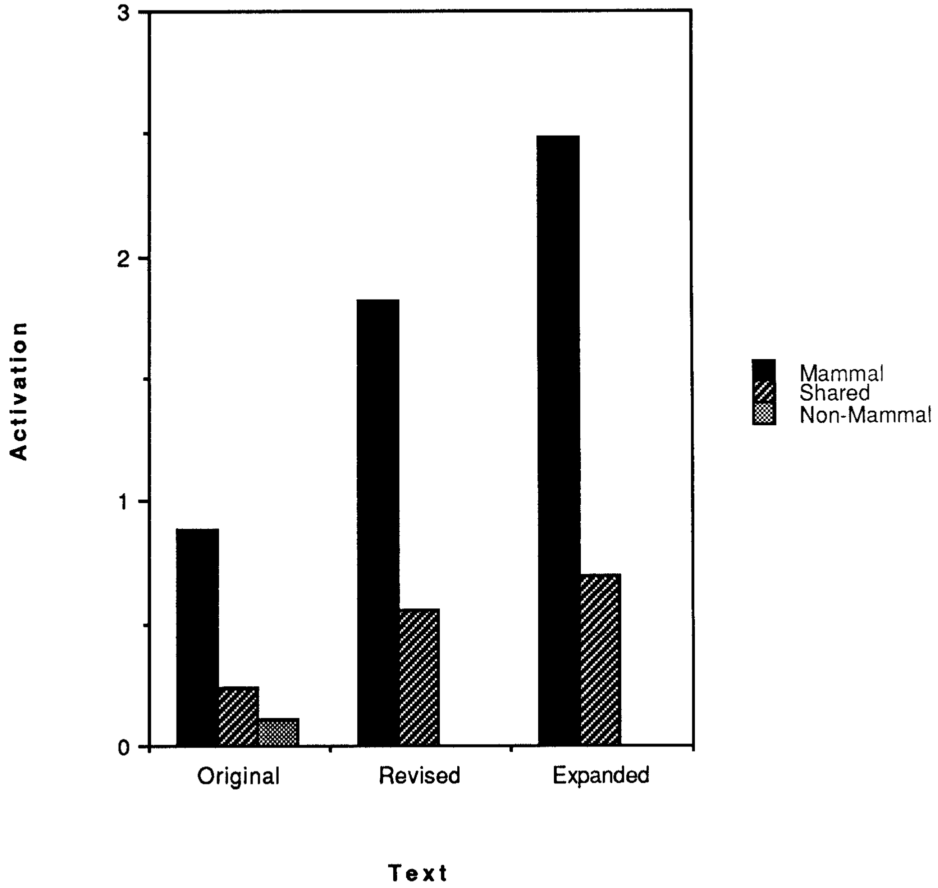

Figure 11. Predicted activation sorting changes for different item types and text versions.

The average cue strengths for the three item types, Mammal, Shared, and Non-Mammal, are shown in Figure 11. Obviously, the pattern of activation values mirrors quite well the pattern of the change in the sorting data (especially Figure 8). For each text, the model predicts the biggest change towards the mammalian classification for the Mammal items, less of a change for Shared items, and even less or no change for Non-Mammal items.

Furthermore, the absolute level of these changes is predicted to be highest for the expanded text, and lowest for the original text, just as required by the data.

A more detailed analysis of the predictions for the individual sorting statements reveals some discrepancies with the empirical results, however. The correlation between the observed change (percent of possible change) and predicted change is reasonably high, Spearman's rho = .84. .78, and .76 for groups original, revised, and expanded, respectively, if all 16 items are included. If sorting items that have no overlap with the text (which are mostly Non-Mammal items) are omitted, these correlations are reduced to .73, .39, and .31, respectively. Thus, the model does not predict with great accuracy which items change how much, though the predictions for groups of items are quite good. The failure to obtain better predictions for individual items may have several sources. First of all, correlations with so few items are unreliable. Second, we have used only direct overlap with the text in our predictions and neglected inferences. Thirdly, there are other features, not just the text that influence sorting. For example, for all three texts the model underpredicts the large observed change in "has mammary gland," because the texts do not say very much about that. But as subjects become more conscious of the mammalian - non-mammalian distinction as a sorting principle on the posttest, the surface features of that item may help to make it a core item for the mammals category.

In summary, the CI model has predicted the observed sorting changes quite well. When subjects read a text, the episodic text memory as postulated in the theory appears to have a clear influence on their post-reading conceptualizations. The model predicted correctly which of the three texts produces the strongest and the weakest influence, it predicted correctly for which item groups these effects should be strongest and weakest, but it did only a fair job in predicting the fate of individual items.

Modelling the Recall Data

Although the focus of the present study is not on how well readers recall a text, but on what they learn from it, it seems worthwhile to at least check whether the model accounts for the recall of the text as well. However, recall protocols cannot be scored in the detail that would be necessary to match them against the predictions of the simulations performed above. For instance, if a subject recalls a predicate with its argument ("X does P"), it is not clear whether the corresponding proposition P[X] as well as the separate concept X, or only the former should be credited. For that reason, separate concept units are usually not included in simulations of recall data (e.g., Kintsch, 1974; Meyer, 1975). Hence, a separate simulation had to be performed in which existential propositions were not included explicitly in the network. This was done only for the revised text. The activation values calculated by the model for each proposition were correlated with the frequency of recall as determined from the subjects' protocols. The resulting correlation was r=.49. This value appears somewhat low when compared with other applications of the model (e.g., Kintsch, 1991), but it should be noted that the present text was considerably longer and more difficult. Typically, simple short narratives were used as the to-be-remembered text in other simulations, as compared with the biology text employed in the present study.

One apparent reason why the present correlation was relatively low was that prior knowledge played a large part in the subjects' recall. Facts that the subjects already knew well were more likely to be recalled than unfamiliar material of comparable importance in the text. More complete simulations of comprehension could, in principle, account for this knowledge effect. If it were known what the subjects knew beforehand, one could estimate what knowledge would be activated during the act of comprehension. For example, if subjects knew that "hair keeps animals from getting too cold", this piece of knowledge would very likely be activated during comprehension, and the corresponding text proposition would be strengthened. On the other hand, the students were probably less likely to know that hair also prevents the animal from getting too hot, so that no strengthening of that text proposition would result. In this way the observed discrepancy in the recall of "becoming too hot" and "becoming too cold" could be accounted for in the model, in spite of the fact that in the text itself these propositions play entirely symmetric roles. Note that these knowledge effects were less disturbing in the sorting analysis, because there we are only concerned with the change in knowledge, not with its overall level.

Conclusions

The two main goals of the present study were to investigate methods for the measurement of changes in the knowledge structure of readers that occur as a function of reading a text and to replicate earlier findings that making texts more coherent improves their recall. The study was successful in both respects.

Sorting a set of keywords or key phrases belonging to a particular knowledge domain that partly overlapped with the text both before and after reading proved to be a sensitive method for the measurement of knowledge changes. The pre-sort yields a map of the knowledge domain before reading, and the post-sort allows us to ascertain the changes that have taken place in that structure as a function of reading. In the present case, items that were discussed in the text as highly typical of mammals were assigned more strongly to the mammal category after reading than before, while items which were less typical moved less strongly in that direction. Items not discussed in the text or not characteristic of mammals showed the least tendency to be categorized with mammal items on the posttest, and indeed were in some cases even more clearly differentiated than before, in spite of the heavy focus on mammals introduced by the text. Thus, pre- and post-sorting of an appropriate set of keywords provides a suitable experimental method for the recording of knowledge changes in reading.

Since we plan to use the construction-integration model as a guide and tool in future work on improving textbooks, the question whether this model could predict the observed knowledge changes in the present experiment was of particular interest. The simulations performed were quite successful in this respect. For the items that moved most strongly towards the mammal category as a function of reading, the links in the textbase between these items and the concept mammal were most strongly activated in the simulation of the comprehension process. For less characteristic items, for which smaller knowledge changes were observed, the activation was weaker. Finally, the predicted activation values were weakest for those items not characteristic of mammals, for which the smallest changes were observed.

The second main question investigated here concerned the differences between the three text versions employed. Here also the simulation correctly predicted that knowledge changes should be larger for the revised than for the original text and largest for the expanded text. The original and revised versions of the text differed in coherence, primarily at the macro-level. As expected, subjects who read the revised version reproduced the macrostructure of the text better than subjects who read the original version. These subjects also showed more differentiated knowledge changes than the readers of the original version: items, for which the text emphasized their mammalian character, moved more strongly into the mammal category, while uncharacteristic items moved slightly away from it.

These results are somewhat different but complementary to those reported by Britton and Gulgoz (1991). These researchers reported a large increase in text recall after a similar revision. However, what they did was to make their text locally coherent, hence achieving a direct effect on the amount recalled. The text used in the present study was already locally coherent to begin with, but not globally coherent. Hence our changes, and their effects on recall, were primarily at the macro-level. Thus, making texts more coherent at either the micro- or macro-level can have beneficial results on recall.

While the recall and sorting data were consistent and generally in line with expectations based on the construction-integration model, the question answering results were disappointing. Our study would have yielded no clear or useful conclusions had we relied solely on question answering. Question answering is the traditional method for assessing learning from text, in school as well as in the laboratory. In the present case, at least, it proved to be an entirely insensitive method. Of course, we probably just asked the wrong questions - but, as in school settings, it is not obvious what the right ones would have been!

So far, this discussion has had a very narrow focus. We have looked at two texts that differed (primarily) in their coherence at the macro-level. This has the advantage that we can make quite unambiguous statements about the effects of that single variable. But there is, of course, much more to good textbook writing than issues of coherence. Beck and McKeown (1991), in their review of textbook analysis methods, are likewise concerned with coherence. Indeed, they identify lack of coherence at the macro-level as one of the major problems, but they raise a number of broader issues. They make the important point that for evaluating a textbook, and presumably for writing one, one must have a clear idea of what kind of information is to be conveyed to the student (the underlying situation model). The problem is not only structural, but also one of content.

It was just such a concern with content that led us to include the expanded version in our study. The improvement made in the revised version over the original version was structural (macro-coherence). The improvement we tried to achieve with the expanded version consisted in providing a more adequate situation model. The situation model on which the original text (and hence also the revised text) was based was found to be inadequate. The text simply listed traits of mammals, without explaining their importance. In the expanded version, we supplied explicit arguments that stated why each of the traits of mammals was being considered. This made the text longer and more abstract (arguments, not concrete facts were added). As a result, recall suffered, at least in terms of the proportion of the total text that was reproduced. But at the same time recall became more constructive, richer in inferences and generalizations than was the case for the other two versions. And, most importantly, the knowledge changes induced by the expanded text were larger and more differentiated than before. The expanded text must therefore be considered a success. It should be noted that this result is also in line with that of Mannes and Kintsch (1987), in which richer constructive processing could be related to the construction of a more elaborated situation model and superior performance on a problem solving task.

Nevertheless, a number of open questions remain. Could we, for instance, improve the recallability of the expanded text if we added not only abstract arguments about survival, but also concrete examples and illustrations? How would efficiency measures look that take into account not only how much a subject recalled, but also study time?

The dominant feature of the present study was the use of a well written, locally coherent text on a topic that was highly familiar to our readers. In further studies we plan to use a more difficult, poorly written, unfamiliar text and to test whether changes in knowledge are maintained over longer periods of time. Furthermore, while there were too few subjects in the present study to examine individual differences in ability and their interaction with text version, such considerations will take precedence in future work.

References

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MS: Harvard University Press.

Beck, I. L. & McKeown, M. G. (1991). Substantive and methodological considerations for productive textbook analysis. In J. P. Shaver (Ed.), Handbook on research on social studies teaching and learning (pp. 496-512). New York: Macmillan.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility Reading Research Quarterly, 27, 251-276.

Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's model to improve instructional text: Effects of inference calls on recall and cognitive structures. Journal of Educational Psychology, 83, 329-345.

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. Cognition and Instruction, 7, 161-195.

Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. Psychological Review, 95, 163-182.

Kintsch, W . (1991). How readers construct situation models for stories: The role of syntactic cues and causal inferences. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes (pp. 261-278). Hillsdale, NJ: Erlbaum.

Kintsch, W. (in press). A cognitive architecture for comprehension. In H. Pick, & P. van den Broek (Eds.), The study of cognition: Conceptual and methodological issues. Washington, DC: American Psychological Association.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), Perspectives of memory research (pp. 325-366). Hillsdale, NJ: Erlbaum.

Kintsch, W., & Welsch, D. M. (1991). The construction-integration model: A framework for studying memory for text. In W. E. Hockley, & S. Lewandowsky (Eds.), Relating theory and data: Essays on human memory in honor of Bennett B. Murdock (pp. 367-385). Hillsdale, NJ: Erlbaum.

Mannes, S., & Kintsch, W. (1987). Knowledge organization and text organization. Cognition and Instruction, 4, 91-115.

Miller, J. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. Journal of Experimental Psychology: Human Learning and Memory, 6, 335-354.

Mross, E. F., & Roberts, J. O. (1992). The construction-integration model: A program and manual. (ICS Tech. Rep. and Program in preparation). University of Colorado, Boulder.

Perfetti, C. A., & McCutchen, D. (1987). Schooled language competence: Linguistic abilities in reading and writing. In S. Rosenberg (Ed.), Advances in applied psycholinguistics (pp. 105-141). New York: Cambridge University Press.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1985). Pathfinder: Scaling with network structures (Tech. Rep. No. MCCS-85-9). Las Cruces: New Mexico State University.

Silver Burdett & Ginn. (1987). Life Science (pp. 300-302). Morristown, NJ: Author.

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

Footnote

This study was supported by a grant from the A. W. Mellon Foundation. The opinions expressed do not necessarily reflect the position or policy of the Mellon Foundation, and no official endorsement should be inferred.

Appendix A

Animal names sorted in Sorting Task 1

Animal type	**Animal sorted**
Mammal	anteater
	bat
	duckbilled platypus
	elephant
	gorilla
	grizzly
	mole
	porcupine
	whale
Non-mammals	alligator
	chameleon
	gila monster
	tortoise
	copperhead
	python
	eel
	carp
	piranha
	seahorse
	shark
	sting ray

Appendix B

Animal characteristics sorted in Sorting Task 2

Animal type	Animal characteristic sorted
Mammals	has hair or fur
	has several types of teeth
	has mammary gland
	instinctive and learned behavior
	tolerates extreme temperatures
	provides a lot of protection for their young
	has complex heart
Either Mammals or Non-mammals	hibernates
	migrates
	has lungs
	has central nervous system
Non-mammals	produces many babies at one time
	is coldblooded
	lays or releases eggs
	mainly instinctive behavior
	has scales or hard shell

Appendix C

Experimental Texts

Traits of Mammals - Original Version

What do a dog, a bat, a monkey, and a whale have in common? They are all mammals. A mammal is a warm-blooded vertebrate that has hair and that feeds milk to its young. Mammals are a successful group of animals. A group of animals is successful if its members are found in many types of environments or in great numbers. Mammals are found on land, in water, and in the air.

Mammals are the only animals that have hair or fur. Though some mammals have little or no hair, most are covered with hair. Hair insulates mammals in much the same way that feathers insulate birds.

Fertilization is internal in mammals. The young of nearly all mammals develop inside the mother's body. The amount of development at birth varies with the type of mammal. A mammal that is not fully developed at birth is dependent on its parents' care. The amount of care given to the young is greater in mammals than in any other class of animal. The mammary gland is found only in mammals and is used in the care of their young. A mammary gland is a structure in female mammals that secretes milk. The milk provides food for the young.

Like birds, mammals have a four-chambered heart. The upper chambers are called atria. The atria receive blood from other parts of the body. The lower chambers, called ventricles, pump blood to other parts of the body. Structures called valves are found between the upper and lower chambers. The valves are flaps that open and close. They control the flow of blood between the upper and lower chambers.

Another important trait of all mammals is a well-developed nervous system that includes a complex brain. In general, the brain of a mammal is larger than that of other vertebrates. Mammals are more intelligent than most other vertebrates.

Mammals have very specialized teeth. There are four types of teeth in mammals: incisors, canines, premolars, and molars. The number and shape of each of these types of teeth are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth that are used for tearing. Their incisors are chisel-shaped and are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials.

The complex behavior of mammals is another trait of this class. Soon after a mammal is born, it finds its mother's mammary gland, or breast, and begins to feed on milk. Without being taught, the newborn seeks its mother's breast. This behavior is inborn. A complex, inborn pattern of behavior is called an instinct. Breast-feeding is one type of instinct. Mammals have many types of instincts. Migration is an instinct of some mammals. The defense of certain territory is another.

Hibernation is another instinct of mammals. Hibernation is a type of deep sleep in which an animal has a lowered body temperature. Many mammals hibernate during winter. Food, needed to supply energy, is scarce in winter. A great amount of energy is needed to maintain a normal body temperature during cold weather. Since a hibernating animal maintains a low body temperature, it uses little energy.

Though many types of animal behavior are instinctive, mammals also learn behavior. An example of learned behavior is seen in bears. Alaskan brown bears teach their offspring to hunt salmon. The young bears carefully watch their mother. After practice, the young bears will become skilled hunters.

Traits of Mammals - Revised Version

What do a dog, a bat, a monkey, and a whale have in common? They are all mammals. Mammals are a special class of vertebrates (animals with backbones) that share certain traits or characteristics. Mammals are found on land, in water, and in the air because they have developed traits that allow them to live in many different environments. Mammals are a successful group of animals. A group of animals is successful (a) if its members are found in many types of environments, (b) if there are many of them or (c) if they have survived for long periods of evolutionary time.

One trait that allows mammals to live in many different environments is that they have developed special ways of protecting their young from the dangers of the environment. One way that the young are protected is that fertilization is internal in mammals, and the young develop inside the mother's body. The amount of development at birth varies with the type of mammal. A mammal that is not well developed at birth is more dependent on its parents' care than one that is fully developed. Mammals care more for their young than other kinds of animals. The mammary gland is found only in mammals and is used in the care of their young. A mammary gland is a structure in female mammals that secretes milk. The milk provides food for the young.

Another trait that allows mammals to live in many different places is that they are warm-blooded. Unlike reptiles, such as snakes or lizards, which need the sun to stay warm, the mammal's body can maintain a nearly constant temperature, regardless of the temperature of the environment. Like birds, they have a four-chambered heart that helps them do this. The heart keeps the body warm by pumping blood to all parts of the body. The upper chambers of the heart are called atria. The atria receive blood from other parts of the body. The lower chambers, called ventricles, pump blood to other parts of the body. Structures called valves are found between the upper and lower chambers. Valves are flaps that open and close. They control the flow of blood between the upper and lower chambers.

A third trait of mammals that allows them to live in many different environments is that they are the only animals that have hair or fur. Though some mammals have little or no hair, most are covered with hair. Hair insulates mammals in much the same way that feathers insulate birds. It keeps them from getting too hot or too cold.

Another physical trait of mammals is that they can eat many different kinds of food because they have very specialized teeth. This trait also helps them to live in different kinds of environments. There are four types of teeth in mammals: incisors, canines, premolars, and molars. The number and shape of each of these types of teeth are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth, that are used for tearing. Their incisors are chisel-shaped and are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials.

Another important reason why mammals are so successful is that they have a well-developed nervous system that includes a complex brain. In general, the brain of a mammal is larger than that of other vertebrates. Mammals are more intelligent than most other vertebrates. Their large and complex brain supports the complex behavior which is another characteristic trait of mammals.

Mammals have two types of behavior: learned and instinctive behaviors. The ability to learn complex behaviors also contributes to the success of mammals. An example of learned behavior is seen in bears. Alaskan brown bears teach their offspring to hunt salmon. The young bears carefully watch their mother. After practice, the young bears will become skilled hunters.

Instincts are complex, inborn patterns of behavior that don't have to be learned. For example, soon after a mammal is born, it finds its mother's mammary gland, or breast, and begins to feed on milk. Without being taught, the newborn seeks its mother's breast. This behavior is inborn. Breast-feeding is one type of instinct that mammals have, but mammals have many other types of instincts. Migration is an instinct of some mammals. The defense of certain territory is another. Hibernation is another instinct of mammals. Hibernation is a type of deep sleep in which an animal has a lowered body temperature. Many mammals hibernate during winter. Food, needed to supply energy, is scarce in winter. A great amount of energy is needed to maintain a normal body temperature during cold weather. Since a hibernating animal maintains a low body temperature, it uses little energy. Thus, mammals who hibernate can live in environments with cold winters.

Traits of Mammals - Expanded Version

What do a dog, a bat, a monkey, and a whale have in common? They are all mammals. Mammals are a special class of vertebrates (animals with backbones) that share certain traits or characteristics. Mammals are a successful group of animals. A group of animals is successful (a) if its members are found in many different types of environments, (b) if there are many of them, or (c) if they are flexible to changes over evolutionary time. These changes can include dramatic shifts in climate, diet, or behavior.

Being flexible to changes over evolutionary time is a very important trait. Natural disasters such as volcanic eruptions or fires can destroy many animals' sources of food, water, and shelter. Animals that are flexible to change are more likely to survive during harsh temperatures or other disasters which destroy some of their food or water sources. Thus, animals, such as mammals, who live in many different types of environments, protect their young, and eat many different foods have a greater chance of surviving over many, many years than animals that are not as flexible.

Mammals include creatures which can fly to great heights, burrow deep in the ground, swim in deep ocean depths, and reach running speeds of 60 miles per hour. Mammals are found in all habitats from oceans to land environments because different species over time have developed traits that allow them to live in many different environments.

One trait that allows mammals to be successful is that they care more for their young than other kinds of animals. In particular, mammals have developed special ways of protecting their young from the dangers of the environment.

One way that the young are protected is that fertilization is internal in mammals. Internal fertilization provides a protective environment for a small number of developing young within the mother's womb. This is very different from external fertilization which is found in many fish and amphibians. In external fertilization, the mother releases millions of eggs into the water. These eggs are unprotected from harsh conditions or enemies.

Mammals also protect their young through a long period of development inside the mother's body. The amount of development at birth varies with the type of mammal. A mammal that is not very developed at birth is more dependent on its parent's care and less likely to survive than one that is fully developed.

The mammalian mother also provides the food needed for early growth of the young mammal. One source of food provided by the mother is milk. The mammary gland is a structure in female mammals that secretes milk. The mammary gland is found only in mammals. It allows the young to get the food they need until they can learn to find food or their own.

Another trait that allows mammals to be successful is that they are warm-blooded. Unlike reptiles, such as snakes or lizards, which need the sun to stay warm, the mammal's body can maintain a nearly constant temperature, regardless of the temperature of the environment. Many of the internal body parts of the mammal work hard to maintain a constant temperature. For example, the mammals' heart has developed four chambers so that it can pump blood to all parts of the body and keep all parts of the body at a constant temperature. When temperatures are very cold, mammals shiver to produce heat. When temperatures are very hot, mammals sweat or pant to release excess heat.

A third trait of mammals that helps them to be successful is that they are the only animals that have hair or fur. Though some mammals, like whales and dolphins, have little or no hair as adult mammals, most are covered with hair. Hair insulates mammals in much the same way that feathers insulate birds. It (hair or fur) keeps them from getting too hot or too cold. Thick hair or fur helps some mammals, like polar bears, to live in very cold climates.

Many mammals that live in very cold climates also hibernate in order to survive cold winters. Hibernation is a type of deep sleep in which an animal has a lowered body temperature. A great amount of energy, such as food, is needed to maintain a normal body temperature during cold weather. Since a hibernating animal maintains a lower body temperature than under normal conditions, the animal uses little energy. Thus, mammals who hibernate can live in environments with cold winters.

Mammals are also successful because over many, many years, they have developed different kinds of specialized teeth. These different teeth allow mammals to eat many different kinds of food. If mammals can eat many different kinds of food, then they will be less likely to die of starvation and become extinct when a change occurs to one of their food sources.

There are four types of teeth in mammals. The number and shape of each of these types of teeth are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth, that are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials. Mammals such as humans have many different kinds of teeth. These help them eat the many different kinds of food in their diets.

Another important reason why mammals are so successful is that they have a well-developed nervous system that includes a complex brain. In general, the brain of a mammal is larger than that of other vertebrates. Mammals are more intelligent than most other vertebrates. Their large and complex brain supports the complex behavior which is another characteristic trait of mammals. The ability to learn complex behaviors also contributes to the success of mammals.

An example of learned behavior is seen in bears. Alaskan brown bears teach their offspring to hunt salmon. The young bears carefully watch their mother. After practice, the young bears will become skilled hunters.

In addition to learned behaviors, mammals also have a second type of behavior called instinctive behaviors. Instincts are complex, inborn patterns of behavior that don't have to be learned. For example, soon after a mammal is born, it finds its mother's mammary gland, or breast, and begins to feed on milk. Without being taught, the newborn seeks its mother's breast. This behavior is inborn. Breast-feeding is one type of instinct. but mammals have many other types of instincts. Migration is an instinct in some mammals. Some mammals, such as whales and seals, migrate at different times of the year to find food or partners for spring mating. The defense of certain territory is another. Hibernation is another instinct of mammals. Young bears know without being taught when and how to begin winter hibernation.

In summary, mammals have evolved many traits over many, many years which help them to be successful in many different environments. These traits will also help mammals to survive for many, many more years.

Appendix D

Macrostructure of Original Text

Level 1: General Topic

traits of mammals

Level 2: Specific Traits

hair or fur
internal fertilization -care
heart
brain
teeth
behavior

Level 3: Subtopics

instincts
learning

Macrostructure of Revised Text

Level 1: General Topic

traits of mammals

Level 2: Specific Topics

care - internal fertilization
warmblooded - heart
hair or fur
teeth
brain and behavior

Level 3: Subtopics

learning
instincts

Macrostructure of Expanded text

Level 1: General Topic

traits of mammals

Level 2: Specific Traits

- care - internal fertilization
- warmblooded
- hair or fur
- hibernation
- teeth
- brain and behavior

Level 3: Subtopics

- learning
- instincts