ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

University of Illinois at
Urbana-Champaign Library
Large-scale Digitization Project, 2007.

Technical Report No. 463

QUESTIONS IN ELEMENTARY SCIENCE AND SOCIAL STUDIES TEXTBOOKS

Bonnie B. Armbruster
University of Illinois at Urbana-Champaign

Joyce Ostertag
Benchmark School - Media, Pennsylvania

March 1989

Center for the Study of Reading

TECHNICAL REPORTS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
174 Children's Research Center
51 Gerty Drive
Champaign, Illinois 61820

CENTER FOR THE STUDY OF READING

Technical Report No. 463

QUESTIONS IN ELEMENTARY SCIENCE AND SOCIAL STUDIES TEXTBOOKS

Bonnie B. Armbruster
University of Illinois at Urbana-Champaign

Joyce Ostertag
Benchmark School - Media, Pennsylvania

March 1989

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. OEG 0087-C1001, with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

Abstract

In this study, approximately 7,500 questions from fourth and fifth grade science and social studies textbooks and teacher's manuals were analyzed and classified according to several characteristics that affect learning: (a) Type of cognitive demand (similar to the levels of Bloom's taxonomy), (b) Source of Answer (where the information to answer the question would most likely be found), (c) Target Relationship (the kind of information required in the response, such as definition or cause/effect), (d) Form of Question (short answer, multiple choice, etc.) and (e) Question Purpose (i.e., instruction or assessment). The density of questions (number of words of text per question) was also calculated. The results are discussed in terms of implications for learning, and recommendations are offered for teachers and publishers.

QUESTIONS IN ELEMENTARY SCIENCE AND SOCIAL STUDIES TEXTBOOKS

It is now widely acknowledged that textbooks exert a powerful influence on American education. From his earlier study of American schools, Goodlad (1976) concluded that "the textbook predominated throughout as the medium of instruction, except in kindergarten" (p. 14). Goodlad's more recent study (1984) corroborates the earlier finding. Another study (EPIE, 1974) concluded that students spend 75% of their classroom time and 90% of their homework time using textbooks and related materials. Finally, Tyson-Bernstein (1988), in her recent treatise on textbooks, observes that "according to virtually all studies of the matter, textbooks have become the *de facto* curriculum of the public schools . . ."(p. 11).

Not only instruction but also assessment is strongly influenced by commercially published materials. Researchers at the Center for the Study of Evaluation (Burry, Catterall, Choppin, & Dorr-Bremme, 1982; Burry, Dorr-Bremme, Herman, Lazar-Morrison, Lehman, & Yeh, 1981) found that teachers tend to use assessment techniques that they feel are closely linked to their instruction. Among the favored assessment techniques are "curriculum embedded tests," including tests that accompany commercially published textbooks. Teachers frequently use curriculum embedded tests for placement, tracking student progress, assessing performance, and grading.

One significant component of textbooks is questions. Most of the instructional recommendations in teachers' editions revolve around questions. Questions abound in student textbooks, especially at the ends of lessons, chapters, and units. Questions are the heart of supplemental materials such as workbooks and ditto masters. And, of course, the tests that accompany the textbooks are comprised of questions.

Since textbooks have such a powerful influence on classroom instruction, it is important for educators to be informed about the questioning practices in commercially published materials. With this knowledge, educators can evaluate the instructional philosophy implicit in the publishers' questions. The purpose of this study was to investigate the characteristics of the questions asked in commercially published materials. Specifically, the study addressed the issue of how many and what kinds of questions appear in elementary science and social studies textbooks and teachers' manuals.

Before turning to the study itself, we will very briefly review some of what is known about the effect of questions on learning and the type of questions used by teachers and textbooks.

What Effect do Questions Have on Learning?

One area of research on questions has investigated the effect of frequency of teacher questions on learning. The results of this line of research are quite consistent: the frequency of academic questions is positively correlated with learning outcomes (Brophy & Evertson, 1976; Soar, 1973; Stallings & Kaskowitz, 1974). Good and Brophy (1978) offer two possible explanations for this finding. One explanation is that teachers with high rates of academic questioning usually have well-managed, well-organized classrooms and spend most of their time on academic activities. Thus, the high frequency of academic questions is an index of a focus on academic learning. A second explanation is that a high rate of academic questioning provides opportunities for student participation and involvement with the content.

Most of the research on teacher questions, however, has looked beyond the frequency of questions to the *type* of questions asked. This research has involved coding questions according to some classification scheme and relating the resulting question types to student achievement. Many classification schemes for categorizing questions have been developed. Often, such schemes sort questions into some kind of hierarchy of cognitive demands. As an example, probably the most

influential of such classification schemes in research on questioning is Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). This six-level scheme assumes a continuum of cognitive demands, ranging from relatively simple (e.g., knowledge of facts) to more complex (e.g., synthesis and evaluation).

Two main areas of research on the effect of question type on learning involve *teacher questions* (oral questions teachers ask as a part of classroom instruction) and *text questions* (written questions accompanying text, often called "adjunct questions"). Both areas of research are relevant to the present study, since commercial materials contain both suggested teacher questions (in the teacher's edition) and text questions (in the pupil's edition).

Research on Teacher Questions

Research on the effect of teacher questions has been rather thoroughly reviewed over the past decade. In a review of 18 experimental and quasi-experimental studies, Winne (1979) concluded that there is no substantive relationship between the use of higher cognitive questions and enhanced student achievement. On the other hand, Redfield and Rousseau's (1981) later meta-analysis of the results of the same studies Winne reviewed plus one additional study found a significant, large, positive effect for teacher use of higher cognitive questions on student achievement. Later, another research team (Samson, Strykowski, Weinstein, & Walberg, 1987) analyzed the same studies included in the previous reviews plus three newer studies and concluded that higher cognitive questioning has a small positive effect on student achievement. In sum, although reviews of research on teacher questions have failed to demonstrate a strong, consistent relationship between types of questions and student achievement, two out of three reviews show some support for a positive relationship between higher cognitive questions and enhanced student achievement.

Research on Text Questions

Research on the effect of text questions on student achievement has been reviewed in two recent studies. Hamaker (1986), in a meta-analysis of research on text (adjunct) questions, concluded that "higher order questions may have a somewhat broader general facilitative effect than factual adjunct questions" (p. 237). In another review of the text question literature, Andre (1987) concluded that "higher level adjunct questions facilitate the learning of factual information from text and increase the amount of attention readers devote to processing text" (p. 81). Thus, these two recent reviews of text questions tend to support the use of higher order adjunct questions.

Another conclusion from research on text questions is that such questions have a so-called "forward effect" (Rickards, 1979). That is, students form expectations based on the type of question they receive, and these expectations affect learning from reading subsequent material. In other words, students' interactions with questions directly influence future learning outcomes. As Wixson put it, "What you ask about is what children learn" (Wixson, 1983, p. 287). The implication is that higher order questions would promote higher order processing of the text.

In sum, assuming that higher student achievement and higher order processing of text are worthy instructional goals, the research on both teacher questions and text questions tends to give a nod toward higher order questions.

What Type of Questions are Used by Teachers and Textbooks?

A classic research finding is that teachers tend to rely on lower cognitive level questions. For example, in a review of research on teachers' questioning practices from 1912 to 1967, Gall (1970) concluded that "about 60% of teachers' questions require students to recall facts; about 20% require students to think; and the remaining 20% are procedural" (p. 713). The preponderance of lower level teacher questions has been verified at all grade levels in a variety of subject areas (Wilen, 1982).

The tests teachers give also reflect a preference for lower level questions. Pfeiffer and Davis (1965) analyzed ninth grade teacher-made semester exams and concluded that the exams "clearly emphasized the objective of knowledge acquisition and the mental process of memory" (p. 10). More recently, in his extensive study of American schools, Goodlad (1984) observed of social studies instruction that "The tests we examined rarely required other than the recall and feedback of memorized information . . ." (p. 212). The situation was similar for science instruction: "Once more, the tests used emphasized heavily the recall of specific information rather than the exercise of higher intellectual functions" (Goodlad, 1984, p. 215).

Research suggests that textbooks, too, tend to emphasize lower level questions. Davis and Hunkins (1965) analyzed all questions from a sample of chapters from three fifth grade social studies textbooks, using Bloom's taxonomy. They concluded that about 87% of the questions required knowledge of specifics, while only about 9% required comprehension. Trachtenberg (1974), also using Bloom's taxonomy, analyzed all the study questions, exercises, activities, and test items in nine sets of commercially published world history materials. Of the almost 62,000 items analyzed, an average of about 63% were "knowledge" and about 36% were "comprehension" items, with a negligible number of items representing the higher cognitive levels.

In sum, it appears that the majority of questions asked in American classrooms--both by teachers and by textbooks--are lower cognitive level questions.

The purpose of this study was to update the research on questions appearing in commercially published textbooks and teachers' manuals. The Davis and Hunkins and Trachtenberg studies are now dated. Much has changed since those studies were completed. First, recent theory and research in cognitive psychology has revealed a great deal about learning from text and the processes which support such learning. Second, research on questions, as just reviewed, has advanced understanding of the effect of questions on learning. Third, educational policy currently favors an emphasis on critical thinking skills across the curriculum. Are these recent trends reflected in the questioning practices of textbooks?

The study reported here is a descriptive study of questions in fourth and fifth grade science and social studies materials. We selected science and social studies because these are core content areas in the academic curriculum. The purpose of reading in these subjects is to learn content. Therefore, the questions should reflect the publisher's curriculum policy concerning the content to be learned. We selected grades 4 and 5 because it is at these grades that students are beginning to learn how to learn from reading. Therefore, content area textbooks play an especially critical role in these grades in developing and shaping students' ability to learn from reading.

Method

Materials

The materials were fourth and fifth grade science and social studies pupils' and teachers' editions of three major publishers who publish both science and social studies programs: Harcourt Brace Jovanovich; Scott Foresman; and Silver Burdett. These programs command a respectable portion of the market and are representative of commercially published materials. We used the most recent versions of the textbooks available to us when we began the study. The specific textbooks included in the study are listed in Table 1.

[Insert Table 1 about here.]

Procedure

Development of a question classification system. Because of the research on the effect of type of questions on learning, we wanted to use some kind of hierarchical classification scheme in our

research. Several schemes have been developed. As previously mentioned, Bloom's taxonomy of educational objectives for the cognitive domain has undoubtedly been the most influential so far in research on questioning. Bloom's hierarchical system includes the six levels of knowledge, comprehension, application, analysis, synthesis, and evaluation.

Another popular taxonomy influenced by Bloom's taxonomy is that of Barrett (1976). Barrett's four levels of questions are literal recognition or recall, inference, evaluation, and appreciation. A major addition made by Barrett is the subdivision of these levels according to the specific information targeted by the question (e.g., recognizing and recalling main ideas, inferring cause and effect relationships, identification with characters and incidents).

A third taxonomy, which has grown in popularity over the last decade, is Pearson and Johnson's (1978) taxonomy of question-answer relationships. This taxonomy is based on recent theories of reading, which stress that reading is not simply a text-based activity, but an interactive process in which meaning evolves from an interaction of reader with text. The Pearson and Johnson comprehension categories (textually explicit, textually implicit, and scriptally implicit) are identified by the data source that must be used by the reader to answer the question.

We were not totally satisfied with any of these three classification systems. While Pearson and Johnson's taxonomy is most responsive to current reading theory, we found that the three categories, particularly the "textually implicit" category, were too broad for our purposes. "Textually implicit" means "there is at least one step of logical or pragmatic inferring necessary to get from the question to the response . . ." (Pearson & Johnson, 1978, p. 161). "At least one step" implies that the "steps" of inferring could range from one to many. The inferences required by textually-implicit questions include everything from a relatively simple inference about the referent for a pronoun to a relatively complex inference involving the synthesis of information to generate an unstated main idea of an entire passage. We were dissatisfied with the Pearson and Johnson taxonomy precisely because we thought a question classification system should reflect the differential cognitive processing required of different numbers or types of inferences.

Bloom's and Barrett's taxonomies came closer to reflecting the variety of cognitive demands that we thought were required by different types of questions. But these taxonomies also had their weaknesses. Neither taxonomy recognizes the interactive nature of reading and the need to consider question-answer relationships. We found Bloom's taxonomy difficult to use. Perhaps because it is a taxonomy of educational objectives rather than questions *per se*, we were often puzzled about how to classify particular questions. Barrett's taxonomy was somewhat easier to use, but it appeared to be more appropriate for narrative than for expository text. For example, categories such as "inferring character traits" and "judgments of reality and fantasy" were not represented in the social studies and science textbooks of our sample.

We finally decided to compromise by including aspects of all three taxonomies in our own classification system. Our system includes five dimensions. The first dimension, *Type of Question*, uses a modification of Bloom's and Barrett's taxonomies in an attempt to index the type of cognitive process students would probably use to answer the question. Five types of cognitive processes are captured in this dimension. The five types are defined and illustrated in Table 2.

[Insert Table 2 about here.]

The second dimension, *Source of Answer*, is included in response to Pearson and Johnson's concern about the need to consider question-answer relationships. We determined four possible sources of answers for textbook questions: the prose itself, graphic aids, activities or demonstrations, and prior knowledge. These sources are defined and exemplified in Table 3. In coding questions in this dimension, we tried to determine the single most likely source of the answer, even though students *may* have been able to use a different source or more than one source to generate the answer. Also, "the

answer" refers to the answer given in the teacher's edition. Where answers were missing or obviously wrong or inappropriate, we determined an appropriate answer.

[Insert Table 3 about here.]

The third dimension is *Target Relationship*. In including this dimension, we were influenced by Barrett's taxonomy, with its attention to the particular type of relationship (e.g., sequence, cause-effect, comparison) asked for in the question. As previously mentioned, we included this dimension because it seems particularly relevant for questions about content. Table 4 defines and gives examples of the types of relationships reflected in the Target Relationship dimension.

[Insert Table 4 about here.]

A fourth dimension of our classification system is **Form of Question**. The Forms of Questions include multiple-choice; true/false; matching; fill-in; short answer (requiring an answer consisting of anything from a single word to two sentences); essay (requiring an answer of three or more sentences); and graphic (requiring an answer in the form of a picture, diagram, chart, map, etc.).

Finally, we included a fifth dimension called *Question Purpose*. This dimension was included because it provides additional information about the cognitive demands of the question. Beck and McKeown (1981) suggest that questions are asked for two purposes: to assess comprehension and to aid in the development of comprehension. Therefore, we decided to differentiate between questions that were included primarily for assessment purposes and those included primarily for instructional purposes. *Assessment questions* included only items on chapter and unit tests. *Instructional questions* included questions in: (a) the teacher's edition (except tests), (b) the pupils' edition at the ends of lessons and/or chapters, (c) activities, such as science investigations, and (d) chapter or unit reviews.

Question analysis. First, we conducted a small pilot study to test our classification system. Both authors worked together to code the questions from four randomly selected chapters from the textbook sample previously described. This exercise enabled us to reach consensus on what counted as a question and how various representative question types should best be coded. We decided to include the following types of questions in the analysis: (a) standard wh- and how questions, (b) "pseudoquestions" (Pearson & Johnson, 1978), that is, questions stated in the imperative rather than the interrogative form (e.g., Explain what the Continental Divide is), (c) indirect or implicit questions in the teacher's edition (e.g., Encourage students to talk about the hardships and dangers explorers would have had to face in the early 1800s). Excluded from the analysis were questions embedded within the prose of the pupils' edition (because these are often rhetorical questions, with no apparent expectation of a response) and questions within special "skills lessons" appearing within the unit (because these questions appeared to be about skills rather than content). Having decided which questions to include, we developed Tables 2-4 as a guide for classifying questions.

In order to establish interrater reliability, all questions (n = 196) from one fourth grade science unit were coded independently by the first author and a graduate assistant not otherwise involved with the project. The graduate assistant used the definitions and examples of questions exactly as they appear in Tables 2 through 4. The resulting percentages of agreement for the four categories of questions so analyzed were: Type of Question, 79%; Source of Answer, 96%; Target Relationship, 86%; and Form of Question, 99%. The overall agreement (pooling across the four categories) was 90%.

Next, we began the main study. We randomly sampled two units from each textbook in our sample. (Two units comprised between one-quarter and one-third of the total content of each textbook.) Then we identified and numbered each candidate question. Both authors independently coded every question. The analysis required us to consider each question in relationship to both the instructional context (the text as well as the instruction recommended in the teacher's edition) and the expected

answer (as given in the teacher's edition). Therefore, we had to read carefully the pupils' edition as well as the teacher's edition in order to complete the analysis. After coding the questions in each chapter, we compared our results. Every discrepancy was resolved through discussion. Thus, we reached 100% consensus on the final coding. In all, 7,463 questions were coded in this manner.

We used this tedious coding and checking procedure in order to ensure the highest possible reliability of the data. Although the previously determined overall reliability of 90% is certainly acceptable, the lower reliability of the Type of Question category was less than desirable. Through discussion, we felt we could eliminate discrepancies due to carelessness, misreadings, oversights, etc.

The data were analyzed by computer using an SPSS program, Crosstabs, which yields frequencies and percentages of questions in the various categories of our system. (Note that because we were interested general trends across publishers rather than differences among publishers, we did not include publishers as a factor in our analysis.)

In addition to investigating question characteristics, we were also interested in determining question density. We calculated question density as the number of words of text divided by the number of questions.

Results and Discussion

Question Characteristics

Tables 5 through 8 present the percentages of total instructional and assessment questions in the sampled social studies and science materials for the dimensions of Type of Question, Target Relationship, Source of Answer, and Form of Question, respectively.

Type of question. As Table 5 shows, almost half of instructional questions and two-thirds of assessment questions are Type 1 questions--questions involving little or no inference. The next most frequent question type is Type 2 questions--questions that require students to make some degree of inference. About one-half of instructional questions in social studies and between one-quarter and one-third of instructional questions in science are questions of this type. For both social studies and science, however, only about one-quarter of the assessment questions require some degree of inference. In both content areas, Type 3, 4, and 5 questions are relatively sparse in either instruction or assessment.

[Insert Table 5 about here].

The profile of question types we found differs somewhat from the findings of earlier studies on questions in commercially published materials (Davis & Hunkins, 1965; Trachtenberg, 1974). Assuming that our Type 1 and 2 categories are similar to Bloom's "knowledge" and "comprehension" levels, respectively, our results show a profile quite different from Davis and Hunkin's and somewhat different from Trachtenberg's results. At least for instructional questions, we found fewer Type 1 questions, more Type 2 questions, and more questions in the other categories. For assessment questions, however, our results look more similar to Trachtenberg's findings.

How might the profile of question types we found affect student learning? Since we found more Type 1 (little or no inference) questions than any other kind, it appears that the questions in the materials we analyzed do not conform to the research-based recommendations about higher level questions. However, this seeming indictment needs some qualifications.

First, the picture is not as bleak as we thought it might be, based on previous research. We were encouraged by the higher proportion of higher cognitive level questions than had been previously

found. Thus, it appears that publishers may be more responsive now to increasing the cognitive level of at least the questions used in instruction.

Second, we were impressed by the relatively high degree of cognitive processing required by *some* of the Type 2 questions. Our Type 2 category includes questions that require rather complex inferences, such as pulling together information from several locations in the text to answer a "main idea" question. Indeed, we felt that some Type 2 questions require higher levels of cognitive processing than did some Type 3 through 5 questions. For example, the Type 5 question "If you were a farmer, which would you prefer to use--horses or a tractor? Why?" may well promote less sophisticated thinking than the Type 2 question "How do the juices given off by the mouth, stomach, and small intestine help in digestion?" given that the information required to answer the question comes from several different sections. A similar point has been made by Beck and McKeown (1987).

Third, the research does not suggest what an appropriate distribution of questions *should* be. Redfield and Rousseau (1981), for example, conclude that achievement is facilitated by a "predominant" use of higher level questions during instruction, but they do not define "predominant." Certainly no one suggests that *all* questions should be high level. Indeed, the use of factual questions can be defended on the grounds that students need to know certain basic information before they can engage in higher order thinking.

The preceding points notwithstanding, we believe that too high a proportion of the questions in these science and social studies materials are questions requiring little or no inference from single text sentences. If the goal of instruction is to promote higher order thinking skills, as it seems to be with the current emphasis on critical thinking, then the pattern of questions we observed is probably inappropriate. When half of the questions asked by the teacher or textbook and two-thirds of the test questions ask for facts, students are getting a clear message about educational priorities. Furthermore, they are learning to engage in minimal cognitive processing and to rely on copying or rote memorization.

Target relationships. Table 6 shows the distribution of the Target Relationships asked in questions. The most frequent instructional questions in both social studies and science are cause/effect questions. This result seems reasonable; after all, explanations of causality are the heart of science and social studies. The prevalence of cause/effect questions is consistent with a finding in a previous study of history questions (Armbruster, Anderson, Bruning, & Meyer, 1984), where "explanation" questions predominated. The next most frequent target relationships asked in instructional questions are identity (asking for the name of a described concept or object) and example (asking for examples of a concept).

For assessment questions, the profile is different. About half of the total questions are identity and definition questions. These two question types are closely related: Identity questions define or describe a concept or object and ask for a name, while definition questions provide the name of a concept or object and ask for a definition. In other words, this finding shows that students are being held accountable largely for names and definitions. Knowledge of cause/effect relationships receives lower priority in assessment than in instruction.

[Insert Table 6 about here.]

Source of answer. The distribution of Source of Answer for questions is shown in Table 7. Most of the instructional and, especially, assessment questions require answers based on reading the prose itself. Of course, we expected that most of the questions would require text-based answers, since these are questions accompanying a textbook series. Graphics are the second greatest source of instructional questions in social studies but the least source in science instruction. Again, it is not surprising that many questions in social studies refer to maps, charts, and tables. The fact that a greater proportion of instructional questions in science are based on activities is also not surprising, since science instruction typically includes many "hands-on" activities or demonstrations. The finding that science instruction

includes more prior knowledge items than social studies is consistent with the finding that science materials have more Type 4 questions (see Table 2). Questions that ask for predictions, hypotheses, and analogies obviously require the reader to draw on prior knowledge.

[Insert Table 7 about here.]

Form of question. Table 8 shows the distribution of questions by Form. About five-sixths of instructional questions are short answer questions. This is hardly surprising, since it is difficult for teachers to ask questions orally in any other form. For assessment questions, about half are multiple choice; matching is the second most popular form of question. True/false, fill-in, and short answer questions round out the profile of assessment questions. This situation is probably attributable to (a) the belief that teachers prefer easily scorable, objective tests, and (b) the pervasive influence of standardized tests and standardized test formats.

The commercially published materials offer very few opportunities for students to synthesize and integrate extended information in an essay. We think this is unfortunate, given the current emphasis on writing and the relationship between reading and writing (see, for example, Squire, 1983).

[Insert Table 8 about here.]

Question Density

Density of questions was calculated by dividing total number of words of text by total number of instructional and assessment questions. The results are shown in Table 9.

[Insert Table 9 about here.]

As Table 9 reveals, there is an instructional question for about every 30 words of text. While research does support frequent questioning, a question for every 30 words of text seems excessive to us. First, it ensures that many questions will be lower level, since small units of text typically yield only factual or detail kinds of questions. Second, such a high frequency of questions dictates frequent interruptions of reading. Frequent interruptions make it difficult for students to "keep track" of the main ideas, to form a coherent interpretation of the information, and to engage in higher order cognition. Overly frequent questioning, in combination with a high proportion of Type 1 questions, may encourage students to be locators of information rather than learners of content.

It may not be the publishers' intent that teachers ask *all* the questions in the teacher's edition, but rather to ask questions selectively as time permits. Indeed, given the large number of questions, teachers probably would not have time to ask all of them. Yet no guidelines for question priorities are offered in the teacher's editions. The undifferentiated listing of questions gives the impression that the questions are equal in value.

Summary and Recommendations

Although the situation seems to have improved since earlier studies, textbooks continue to make heavy use of lower cognitive level questions that require little or no inference. This situation is especially true of questions used to evaluate student achievement. Assessment questions, more than instructional questions, tend to be at the lowest cognitive level, to ask mostly for names and definitions, and to appear in a multiple-choice or other objective item format. The bottom line of the *de facto* curriculum of textbooks is clear: What *really* counts is names, definitions, and other facts that can be assessed "objectively," and *not* meaningful learning of "big ideas," or higher order, critical thinking, or the ability to write extended answers.

The profile for assessment items may exist because of a conscious plan to emulate items on standardized tests. We suspect, however, that most of the questions in commercially published materials are produced somewhat willy nilly, not according to any informed, systematic, well-articulated plan. Of course, it would be difficult for publishers to work from a systematic plan because there is not yet a science of questioning.

An important part of creating a science of questioning will involve the development of alternative methods of assessment. There is currently a serious discrepancy between what we know about learning from text and how we measure that learning. Fortunately, this urgent problem is beginning to be addressed (see, for example, Valencia & Pearson, 1986).

One probable outcome of work on new methods of assessment is the development of new taxonomies, or at least typologies, of questions. The available taxonomies (e.g., Bloom's, Barrett's, and Pearson & Johnson's) seem to fall short of capturing all the important dimensions of questions as recognized by current cognitive psychology. We feel our own effort at a multidimensional taxonomy is an improvement, but it, too, has its weaknesses. For example, our Type 2 category (and Bloom's comprehension level, Barrett's inferential level, and Pearson & Johnson's text implicit category) is still too broad to reflect important differences in cognitive processing required of different questions. We believe a question taxonomy should reflect finer distinctions in the requisite cognitive processing.

Of course, hand in hand with the development of new assessment methods, there must be further research on issues such as the optimal type, number, and sequencing of questions to use in furthering various desired educational outcomes.

Meanwhile, however, as we await the development of a science of questioning, we offer the following recommendations to publishers of science and social studies materials and to the teachers who use them. These recommendations reflect our considered judgment, based on the results of this study, current reading theory and research, and common sense.

1. Reduce the number of instructional questions by weeding out questions that do not reinforce or assess important content. Include a higher proportion of questions that ask for "main ideas."

2. Assuming that higher order cognition is a desirable educational objective, we recommend turning many Type 1 questions into higher order questions. In many Type 1 questions, we saw a missed opportunity to get students to think rather than copy or memorize. With very little effort, most Type 1 questions could be transformed into higher order questions. With a paraphrase here and a request for support or evidence there, questions could tap the same basic information, but demand considerably more cognitive processing (see, for example, Anderson, 1972).

3. Use alternative instructional and assessment methods where currently available. For example, in the area of vocabulary, which seems to be very important in content area learning, many alternative instructional and assessment techniques have been proposed (see, for example, Beck, Perfetti, & McKeown, 1982, and Chapter 7 in McNeil, 1987).

4. Provide more opportunities for students to write extended answers that require them to synthesize and integrate information across extended text.

5. Ensure that questions are worded clearly and that the recommended answer is reasonable and complete.

We believe that even these few, relatively simple recommendations would do much to improve the questioning practices in elementary science and social studies materials.

References

Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research, 42*, 145-170.

Andre, T. (1987). Questions and learning from reading. *Questioning Exchange, 1*, 47-86.

Armbruster, B. B., Anderson, T. H., Bruning, R. H., & Meyer, L. A. (1984). *What did you mean by that question?: A taxonomy of American history questions* (Tech. Rep. No. 308). Urbana: University of Illinois, Center for the Study of Reading.

Barrett, T. (1976). A taxonomy of reading comprehension. In R. Smith & T. C. Barrett (Eds.), *Teaching reading in the middle grades*. Reading, MA: Addison-Wesley Publishing Co.

Beck, I. L., & McKeown, M. G. (1987). Getting the most from basal reading selections. *The Elementary School Journal, 87*, 343-356.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506-521.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, W. H. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: Longman.

Brophy, J. E., & Evertson, C. M. (1976). *Learning from teaching: A developmental perspective*. Boston: Allyn & Bacon.

Burry, J., Catterall, J., Choppin, B., & Dorr-Bremme, D. (1982). *Testing in the nation's schools and districts: How much? What kinds? To what ends? At what costs?* (CSE Rep. No. 194). Los Angeles, CA: University of California, Center for the Study of Evaluation.

Burry, J., Dorr-Bremme, D., Herman, J. L., Lazar-Morrison, C. M., Lehman, J. D., & Yeh, J. P. (1981). *Teaching and testing: Allies or adversaries* (CSE Rep. No. 165). Los Angeles, CA: University of California, Center for the Study of Evaluation.

Davis, O. L., Jr., & Hunkins, F. P. (1965). Textbook questions: What thinking processes do they foster? *Peabody Journal of Education, 43*, 285-292.

EPIE Institute (1974). *Fits and misfits: What you should know about your child's learning materials*. Columbia, MD: National Committee for Citizens in Education.

Gall, M. D. (1970). The use of questions in teaching. *Review of Educational Research, 40*, 707-721.

Good, T. L., & Brophy, J. E. (1978). *Looking in classrooms*. New York: Harper and Row.

Goodlad, J. I. (1976). *Facing the future: Issues in education and schooling*. New York: McGraw-Hill.

Goodlad, J. I. (1984). *A place called school*. New York: McGraw-Hill.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*, 212-242.

McNeil, J. D. (1987). *Reading comprehension: New directions for classroom practice* (2nd ed.). Glenview, IL: Scott, Foresman and Company.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension*. New york: Holt, Rinehart, & Winston.

Pfeiffer, I., & Davis, D. L., Jr. (1965). Teacher-made examinations: What kinds of thinking do they demand? *NASSP Bulletin, 49,* 1-10.

Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research, 51,* 237-245.

Rickards, J. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research, 49,* 181-196.

Samson, G. E., Strykowski, B., Weinstein, T., & Walberg, H. J. (1987). The effects of teacher questioning levels on student achievement: A quantitative synthesis. *Journal of Educational Research, 80,* 290-295.

Soar, R. S. (1973). *Follow-through classroom process measurement and pupil growth (1970-71): Final report*. Gainesville: University of Florida, College of Education.

Stallings, J. A., & Kaskowitz, D. (1974). *Follow-through classroom observation evaluation, 1972-73.* Menlo Park, CA: Stanford Research Institute.

Trachtenberg, D. (1974). Student tasks in text material: What cognitive skills do they tap? Peabody *Journal of Education, 52,* 54-57.

Tyson-Bernstein, H. (1988). *A conspiracy of good intentions: America's textbook fiasco*. Washington, DC: The Council for Basic Education.

Valencia, S. V., & Pearson, P. D. (1986). *New models for reading assessment.* (Reading Education Report No. 71). Urbana: University of Illinois, Center for the Study of Reading.

Wilen, W. W. (1982). *What research says to the teachers: Questioning skills for teachers*. Washington, DC: National Education Association.

Winne, P. H. (1979). Experiments relating teacher's use of higher cognitive questions to student achievement. *Review of Educational Research, 49,* 13-50.

Wixson, K. K. (1983). Questions about a text: What you ask about is what children learn. *The Reading Teacher, 36,* 287-293.

Table 1

Textbooks Used in Analysis

Fourth grade

Social Studies

States and regions. Orlando, FL: Harcourt Brace Jovanovich Publishing Co., 1985.

Regions of our country and our world. Glenview, IL: Scott, Foresman Publishing Co., 1983.

States and regions. Morristown, NJ: Silver Burdett Co., 1986.

Science

HBJ Science. Orlando, FL: Harcourt Brace Jovanovich Publishing Co., 1985.

Scott, Foresman Science. Glenview, IL: Scott, Foresman Publishing Co., 1984.

Silver Burdett Science. Morristown, NJ: Silver Burdett Co., 1986.

Fifth grade

Social Studies

The United States: Its history and neighbors. Orlando, FL: Harcourt Brace Jovanovich Publishing Co., 1985.

America past and present. Glenview, IL: Scott, Foresman Publishing Co., 1983.

The United States and its neighbors. Morristown, NJ: Silver Burdett Co., 1986.

Science

HBJ Science. Orlando, FL: Harcourt Brace Jovanovich Publishing Co., 1985.

Scott, Foresman Science. Glenview, IL: Scott, Foresman Publishing Co., 1984.

Silver Burdett Science. Morristown, NJ: Silver Burdett Co., 1985.

Table 2

Definitions and Examples of "Type of Question" Dimension

Type 1. Answering a Type 1 question involves little or no inferencing. The answer is either (a) explicitly stated in the text within a single sentence, or (b) assumed to be an intact part of memory (i.e., stored in exactly the required form). This level is similar to the "knowledge" level of Bloom's taxonomy, the "literal recognition and recall" level of Barrett's taxonomy, and the "textually explicit" level of the Pearson-Johnson taxonomy.

Examples: (Example 1 illustrates point a; Example 2 illustrates point b)

1. The frontier was the imaginary dividing line between (a) the North and South (b) California gold mining camps (c) settled and unsettled land. (*America past and present*, Teacher's Edition, 1983, p. 123)

(The text says: "The frontier was the imaginary dividing line between settled and unsettled land." *America past and present*, 1983, p. 194)

2. What is this a statue of? (*The United States: Its history and neighbors*, 1985, p. 338)

(There is an uncaptioned photograph of a statue of Abraham Lincoln. Students either know the answer or not; there is no possibility of inferring a correct answer.)

Type 2. Answering a Type 2 question involves some inferencing. The answer is either (a) in the text but must be inferred across sentences or paragraphs, or (b) assumed to be part of prior knowledge, but probably not stored in exactly the required form. (This level is similar to the "comprehension" level of Bloom's taxonomy, the "inferential" level of Barrett's taxonomy, and the "textually implicit" level of the Pearson-Johnson taxonomy.)

Examples: (Example 1 illustrates point a; Example 2 illustrates point b)

1. How did cities grow upward and outward?
(*America past and present*, 1983, p. 318)

(This example requires students to pull together information over several pages of text.)

2. What is going on in the picture? (*The United States: Its history and neighbors*, 1985, p. 358)

(This example requires students to make an inference from an uncaptioned picture of: soldiers fighting, a cannon, and a flag.)

Type 3. Answering a Type 3 question involves applying information from the text or prior knowledge to a novel situation. (This level is similar to the "application" level of Bloom's taxonomy and is included in the "inferential" level of Barrett's taxonomy and the "scriptally implicit" level of the Pearson-Johnson taxonomy.)

Examples:

1. Did Andrew Carnegie's company provide goods or services?
(*The United States: Its history and neighbors*, Teacher's Edition, 1985, p. 408)

Table 2 (continued)

This question requires students to apply new knowledge about the distinction between goods and services to recently read information about Andrew Carnegie.)

2. What simple machines are found in this compound machine?
(*Silver Burdett Science*, Teacher's Edition, Grade 4, 1985, p. 66.)

(This question requires students to find examples of the simple machines they have read about in a simple line drawing of a wheelbarrow.)

Type 4. Answering a Type 4 question involves putting together information in order to make predictions, generate hypotheses, and form analogies. (This level is included in the "comprehension" and "synthesis" levels in Bloom's taxonomy, the "inferential" level of Barrett's taxonomy, and the "scriptally implicit" level of the Pearson-Johnson taxonomy.)

Examples:

1. Imagine the United States with no Mississippi River. How would the country be different? What towns might be different? How might people's lives be different?
Regions of our country and our world, 1983, p. 315)

2. What would happen if the lens in your eye was not able to change its shape?
Scott, Foresman Science, Grade 5, 1984, p. 279)

(These two questions require students to generate hypotheses or make predictions based on information they have just read.)

3. Have the students compare a bus system with the systems in the body.
(*Silver Burdett Science*, Teacher's Edition, 1985, p. 273)

(These questions require students to form an analogy between information they have just read and prior knowledge.)

Type 5. Answering a Type 5 question involves making value judgments about events or situations based on information presented in the text and internal critieria. (This level is similar to the "evaluation" levels in Bloom's and Barrett's taxonomies and is included in the "scriptally implicit" level of the Pearson-Johnson taxonomy.)

Examples:

1. If you were a farmer, which would you prefer to use--horses or a tractor? Why?
(*Regions of our country and the world*, Teacher's Edition, 1983, p. 397)

2. Where would you rather live--in a cold, dry place or in a cold, wet place? Why?
(*States and regions*, 1985, p. 258)

Table 3

Definitions and Examples of Questions in "Source of Answer" Dimension

---

Prose. The answer is primarily derivable from the prose itself.

Examples:

1. How did the coming of the first white settlers to America change the forests?
(*States and regions*, 1986, p. 35)

The text says: "The first white settlers cut down many trees. They built their homes of wood and cleared fields for farming. They used wood as a fuel for heating their houses and for cooking. As people moved west more trees were cut. Forests were destroyed to make farmland and to make room for highways and railroads." (*States and regions*, 1986, p. 34)

2. The surface of a mirror is always
a. very smooth b. very rough
c. made of silver
(*HBJ Science*, Grade 4, 1984, p. 91)

[The text says: "The reflecting surface of a mirror is very smooth."
(*HBJ Science*, Grade 4, p. 90)]

Graphic. The answer is primarily derivable from a picture, photograph, diagram, chart, table, time line, map, or globe.

Examples:

1. Reading a Time Line. Have pupils use the time line on this page to answer the following questions . . .
(*The United States and its neighbors*, Grade 5, 1986, p. 108)

2. (Refers to unit opening photographs) How are these animals different?
(*HBJ Science*, Teacher's Edition, Grade 4, 1985, TM1)

Activity. The answer is primarily derivable from an activity demonstrated by the teacher or completed by the students.

Examples:

1. What happened when you let a lot of water fall quickly on the sand in the same place?
(*Scott, Foresman Science*, Grade 4, 1984, p. 115).

(This question follows an activity about water as an agent of change.)

2. Which leaf is wilted?
(*HBJ Science*, Grade 5, 1985, p. 277)

(This question follows an activity comparing plants in salt and fresh water environments.)

Table 3 (continued)

Prior Knowledge. The answer depends primarily on what the reader already knows, rather than on information presented in the text.

Examples:

1. Ask the class to describe what can happen to metal objects when they are left outside for many weeks.
(*Scott, Foresman Science*, Grade 4, 1984, p. 117)

(This question precedes text discussing rust)

2. Is there a river near our community?
(*The United States: Its history and neighbors*, 1985 p. 54)

Table 4

Definitions and Examples of Questions in "Target Relationship" Dimension

---

Process. Process questions ask about a sequence of events or series of steps that are not necessarily causally connected.

Examples:

1. Describe, in sequence, the events that occur in pond succession.
(*Silver Burdett Science*, Teacher's Edition, Grade 5, p. 100)

2. Number the events below from 1 to 7 in the order that they happened.
(Seven events from the Civil War follow.)
(*America past and present*, Teacher's Edition, Grade 5, 1983, p. 127)

Cause-Effect. Cause-effect questions ask about a causal relationship between an antecedent and a consequent.

Examples:

1. Near the mid-ocean ridges, magma breaks through easily because the
   a. Earth's crust is very thick there.
   b. Earth's plates are rigid there.
   c. Earth's crust is very thin there.
   d. crust is hot enough to melt rock.

(*HBJ Science*, Teacher's Edition, Grade 5, p. 27)

2. Why did the government keep troops in the South for several years after the Civil War?
(*America past and present*, Teacher's Edition, 1983, p. 161)

Definition. Definition questions ask for a definition of a concept.

Examples:

1. What is refraction?
(*Scott, Foresman Science*, Grade 5, 1984, p. 275)

2. Capital is (a) saved up wealth, (b) natural resources.
(*The United States and its neighbors*, 1986, p. 218)

Property. Property questions ask about properties or characteristics of objects or events.

Examples:

1. When it hatches, a salmon is usually
   a. about 2 meters long.
   b. about 2 centimeters long.
   c. about 20 centimeters long.
   d. about 1/2 meter long.

(*HBJ Science*, Teacher's Edition, 1985, p. 45P)

Table 4 (continued)

2. Describe a cool, fairly dry climate.
(*Regions of our country and our world*, 1983, p. 263)

Example. Example questions ask for one or more examples of a concept.

Examples:

1. Name a bird that cannot fly.
(*Silver Burdett Science*, Teacher's Edition, Grade 5, p. 67)

2. What were some of the problems the homesteaders faced?
(*The United States: Its history and neighbors,* 1985, p. 430)

Identity. Identity questions ask for the name of a person, object, event, or concept described in the question.

Examples:

1. It is a cold-blooded vertebrate. Its body is covered with scales. It can change color for protection. What is it?
(*Silver Burdett Science*, Grade 5, 1985, p. 78)

2. What is the highest mountain in the world?
(*States and regions*, 1986, p. 371)

Compare/contrast. Compare/contrast questions ask for similarities and/or differences between two or more concepts or events.

Examples:

1. Compare conduction and convection.
(*Silver Burdett Science*, Grade 4, 1985, p. 76)

2. The Quechuas must depend on themselves to meet their needs. How is their life in the Andes different from life in the city?
(*States and regions*, 1986, p. 371)

No Relationship. No relationship questions do not ask for a relationship between ideas. This category includes questions that ask for only Yes/No answers, True/False questions, and questions for which it was unclear what relationship was being assessed.

Examples:

1. If you lived on Pluto or Neptune, would you be able to celebrate your birthday frequently?
(*HBJ Science*, Teacher's Edition, Grade 5, 1985, p. 102)

2. True/False Slavery continued in the northern states for many years after the Civil War.
(*America past and present*, 1983, p. 322)

Table 5

Distribution of Questions by Type

		Type of Question				
Content Area	Question Purpose	Type 1	Type 2	Type 3	Type 4	Type 5
Social Studies	Instructional (n=3,529)	45.6%	45.0%	3.7%	4.0%	1.7%
	Assessment (n=866)	68.7%	28.6%	2.5%	0.0%	0.1%
Science	Instructional (n=2,609)	46.0%	28.9%	16.4%	8.4%	0.3%
	Assessment (n=459)	66.7%	26.4%	6.8%	0.2%	0.0%

Table 6

Distribution of Questions by Target Relationships

Content Area	Question Purpose	Target Relationship: Process	Cause/ Effect	Definition	Property	Example	Identity	Compare Contrast	None
Social Studies	Instructional (n=3,529)	1.0%	27.9%	11.6%	5.7%	18.9%	22.4%	8.1%	4.2%
	Assessment (n=866)	0.8%	14.2%	29.7%	4.8%	9.1%	35.0%	3.0%	3.3%
Science	Instructional (n=2,609)	3.4%	26.0%	10.3%	10.2%	13.3%	17.4%	11.0%	8.4%
	Assessment (n=459)	0.7%	20.0%	23.7%	13.3%	9.8%	26.6%	3.7%	2.2%

Table 7

Distribution of Questions by Source of Answer

Content Area	Question Purpose	Source of Answer: Text	Graphic	Activity	Prior Knowledge
Social Studies	Instructional (n=3,529)	76.7%	13.8%	0.5%	8.9%
	Assessment (n=866)	94.1%	4.8%	0.0%	1.0%
Science	Instructional (n=2,609)	55.5%	10.0%	17.9%	16.6%
	Assessment (n=459)	92.4%	6.3%	0.2%	1.1%

Table 8

Distribution of Questions by Form

Content Area	Question Purpose	Form of Question						
		Multiple Choice	True/ False	Matching	Fill-In	Short Answer	Essay	Graphic
Social Studies	Instructional (n=3,529)	2.9%	4.3%	5.6%	0.7%	84.2%	1.9%	0.4%
	Assessment (n=866)	49.7%	9.2%	16.4%	9.6%	10.2%	3.3%	1.6%
Science	Instructional (n=2,609)	6.6%	0.9%	5.4%	2.4%	82.7%	0.7%	1.3%
	Assessment (n=459)	48.8%	11.3%	15.7%	10.7%	10.7%	1.7%	1.1%

Table 9

Question Density: Mean number of Words of Text Per Question

	Question Purpose	
Content Area	Instructional	Assessment
Social Studies	36.4	148.4
Science	26.5	144.7