DOCUMENT RESUME

ED 291 752 TM 011 037

AUTHOR Risner, Gregory P.
TITLE Cognitive Levels of Questioning Demonstrated by Test Items That Accompany Selected Fifth-Grade Science Textbooks.
PUB DATE May 87
NOTE 46p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Measurement; Cognitive Tests; Comprehension; Critical Thinking; *Elementary School Science; *Grade 5; Intermediate Grades; *Item Analysis; Knowledge Level; Questioning Techniques; *Science Tests; Test Items; *Textbook Evaluation; Textbooks
IDENTIFIERS *Blooms Taxonomy; Higher Order Skills

ABSTRACT

This study sought to determine the cognitive levels of questioning of end-of-chapter test items accompanying selected fifth-grade science textbooks. The extent to which test items emphasized high level questioning and significant differences among texts in the distribution of knowledge and above-knowledge-level test items were also investigated. Data were derived from 500 test items randomly selected from five most widely used fifth-grade science textbooks in Alabama, California, Tennessee, and Texas. Three raters classified the items according to Bloom's six major taxonomic categories. Collectively, 304 items were judged as assessing knowledge, 169 as assessing comprehension, 26 as application items, and one as an evaluation item. Implications derived from the findings of this research are based on two assumptions that: (1) the use of higher order questions stimulates critical thinking; and (2) tests that accompany textbooks are important guides to instruction and evaluation. In general, it appears that high-order thinking skills are not elicited by end-of-chapter tests in science textbooks and that ready-made tests, which accompany science textbooks, tend to become guides for instruction as well as evaluation. An 87-item reference list concludes the document. (TJH)

ED291752

COGNITIVE LEVELS OF QUESTIONING DEMONSTRATED BY TEST ITEMS THAT ACCOMPANY SELECTED FIFTH-GRADE SCIENCE TEXTBOOKS

by
Gregory P. Risner, Ed.D.
George Peabody College for Teachers
of Vanderbilt University
May 1987

The elementary science curriculum must include real-life experiences that enhance development of problem-solving skills and provide students opportunities to apply the scientific method. Because the ability to think critically is a widely held goal of elementary science, and both textbooks and questioning are such prominent tools in the instructional process, elementary science textbooks and accompanying resource materials must be analyzed to determine their ability to elicit higher cognitive processes.

This study sought to determine the cognitive levels of questioning of test items accompanying selected fifth-grade science textbooks. Also investigated were the extent to which test items emphasized higher level questioning and whether significant differences existed among the texts in distribution of knowledge- and above-knowledge-level test items.

Data were derived from 500 test items, 100 randomly selected from each of the five most widely used fifth-grade science textbooks in Alabama, California, Tennessee, and Texas. Three raters classified the items according to Bloom's (1956) six major taxonomic categories.

TM011 037

BEST COPY AVAILABLE

Collectively, 304 (61%) of the 500 test items were judged as knowledge, 169 (34%) as comprehension, and only 26 (5%) as application items. No test item represented the analysis and synthesis categories; only one evaluation item was found. The textbooks contained significantly more knowledge- than above-knowledge-level test items ($\underline{X}^2$ = 23.3287; $\underline{p} < .001$).

When data were dichotomized into low and higher levels of questioning, two textbooks had slightly greater percentages of higher level questions than lower. However, the overwhelming majority of higher level questions were comprehension, the lowest level of understanding.

Some textbooks contained significantly greater percentages of higher level test items than others ($\underline{X}^2$ = 41.5716; $\underline{p} < .001$). Individual post hoc comparisons indicated that two textbooks were significantly superior to the others in their emphasis of above-knowledge-level items.

Because these tests focused on memorization and not the development of higher cognitive skills, it was concluded that students could master these tests with little or no true understanding of science processes. It was recommended that publishers include a more logical progression of test items, reflective of all six levels of Bloom's taxonomy. Additional recommendations addressed publishers, textbook selection committees, preservice teacher educators, and classroom teachers.

INTRODUCTION

Background and Statement of the Problem

The emergence of elementary science textbooks and supplementary text materials has paralleled closely the development of science education in the public schools. In the pioneer days of the American graded school, the standard curriculum was limited to reading, spelling, arithmetic, and writing. Although science was not among the basics, children's literature often emphasized science themes. The McGuffey Reader, for example, included such topics as plants, weather, and animals (Harbeck & Marcuccio, 1978).

Prior to the development of public schools in the United States, noted scholars Jean Jacque Rousseau (1712-1788) and Johann Pestalozzi (1746-1827) had advocated the study of science through sensory activities and individual examination of natural objects. By the late 1800s, American grammar schools utilized object teaching activities in which students composed descriptions of animate and inanimate objects. The focus of such exercises was to develop the mental processes of the student in preparation of obtaining knowledge for reasoning in later years. Because students did not use textbooks with the object exercises, the amount of science knowledge possessed by the teacher was a critical factor in the success of the lessons. Teacher manuals were published to provide science background; however, most were so laden with incorrect science information that they were of little value (Underhill, 1941).

As the population of the United States migrated from rural to urban communities during the industrial revolution, educators began to stress the study of nature. The period between 1890 and 1920 became known as the

Nature Study movement and was characterized by an enormous increase in the production of teacher manuals and supplementary materials. By 1918, science methods courses were introduced by many teacher-training institutions. As a result, numerous texts of elementary science methodology and curriculum were published (Underhill, 1941).

In 1932, the 31st Yearbook of the National Society for the Study of Science Education emphasized a teaching methodology based on conceptual learning rather than memorization of scientific facts. In highlighting the necessity of science education, the Yearbook defined the need for a specific scope and sequence throughout the entire school curriculum. The landmark publication greatly contributed to the advancement of science as a regular component of the elementary curriculum (Smith, 1975).

As science progressed rapidly at the elementary school level during the 1930s, numerous science textbook series were published. Between 1940 and 1950 most elementary schools began using science textbooks with standard content (Harbeck & Marcuccio, 1978). The dominant purpose of these texts was to transmit factual knowledge. Underhill (1941) discussed the weakness of these early science textbook publications:

> The emphasis is almost wholly on the knowledge of facts. Very little is suggested by the material itself in the way of discipline or mind training of any sort, or the development of powers or faculties, although of course, educators who support the use of such materials do so in terms of such theories of mental development. Even the question-and-answer form degenerates into a catechism type of questioning which leads almost inevitably to a process of memorization. (p. 50)

Although experimental activities were included in the first science textbook publications, the exercises merely required students to replicate textbook procedures rather than use investigative inquiry (Harbeck & Marcuccio, 1978).

New emphasis was placed on the higher levels of cognition when Bloom (1956) published a taxonomy for classifying educational goals and objectives. His hierarchical scheme was based on the premise that simple concepts become incorporated with other equally simple concepts to form more complex concepts. The resulting concrete-to-abstract progression in cognition found support in Piaget's (1970) psychological principles of cognitive development.

Concurrently with Bloom's publication, the launching of Sputnik I by the Soviet Union in 1957 initiated the shift of emphasis from learning content to learning process. Because science education was considered imperative for our national security, federal funds were appropriated through the National Science Foundation to revise elementary science curricula. Pedagogical methodology in science shifted from traditional memorization of factual information to an inquiry/discovery mode of learning (Hunkins, 1972). The process skills of science were taught through manipulative materials. The 59th Yearbook of the National Society for the Study of Science Education strongly advocated teaching process skills through inquiry or discovery lessons (Smith, 1975). Science was regarded as a method of problem solving that could be learned without textbooks.

The activity programs that emerged during the National Projects Era of the late 1960s and 1970s represented a departure from the traditional drill-and-practice procedure in science education. Programs such as Science--A Process Approach (SAPA), Elementary Science Study (ESS), and the Science Curriculum Improvement Study (SCIS) were nontraditional in the sense that they: (a) were activity-based, utilizing lab materials rather than textbooks; and (b) were process-oriented, emphasizing the application of science knowledge (Shymansky, Kyle, & Alport, 1983).

For a brief period following the introduction of new science curricula, critical thinking received prominent emphasis in the professional literature. However, a limited amount of research has been conducted to determine the influence of Bloom's (1956) cognitive model on elementary science textbook content. No published research has been done specifically for determining the cognitive levels of test items which accompany elementary textbooks. This study will present an analysis of the cognitive levels of test items that accompany selected elementary science textbooks.

Rationale for the Study

The textbook remains the most widely used teaching tool in elementary science. Wright (cited by Mechling & Oliver, 1983) reported that an overwhelming majority of all science classes use the basic text for instruction. The National Science Foundation conducted an extensive study of science classes and found that reading assignments and discussion information were limited to the science text (Smith, 1980, cited by Mechling & Oliver, 1983). Textbook instruction is so predominant at the elementary level that most of the allocated science time is devoted to reading and answering questions from the science book. According to Mechling and Oliver (1983), the most important decision made by school districts is the choice of the science book; curriculum and instructional methods are dictated by the text.

The dependence on textbooks for science instruction presents an educational problem. Teaching and learning strategies which foster critical thinking skills often are neglected in the traditional textbook approach. Objectives which facilitate higher order thought processes are not utilized by writers in creating textbook questions, student activities, and tests (Marksberry, McCarter, & Noyce, 1969). Concerning the prefe[illegible] state of elementary science, Johnson (1982) wrote: "While it is desired that elementary school science be taught with real alternatives, real problems, and concrete materials, there is evidence that science is taught in a didactic manner using a single text" (p. 7).

During this century science teaching has reflected one of two instructional approaches: the product or the process orientation. The product approach encompasses teacher-centered lecture and recitation of factual knowledge (Lapp, 1980). Textbooks are the core of this approach; learning is equated with memorization of knowledge contained in the science book rather than participation in experiences which involve analysis, synthesis, and evaluation (Mechling & Oliver, 1983). Except for a brief encounter with the innovative curricula of the 1960s, the product approach has been the predominant practice in elementary schools.

The process-oriented approach encourages problem solving through critical and divergent thinking. Activity-based science, inquiry, experimentation, and process science are central to this method. In this approach, science is considered a method of solving problems rather than a body of knowledge (Horak & Roubinek, 1981). This approach was conceptualized by Dewey (1916) and later accentuated by prominent educators such as Bruner (1962), Gagne (1970), and Piaget (1970). ESS, SCIS, and SAPA are innovative science curricula projects that reflect the process-oriented theme. A survey of the research literature (Bredderman, 1983; Shymansky et al., 1983) reveals the positive effect of these programs in facilitating critical thinking in science.

The differences between these two approaches reflect the discrepancies that exist between the goals of elementary science and the instructional practices utilized to attain them (Smith, 1963). It is clear that there are inconsistencies between actual and desired states in elementary science (Johnson, 1982). The gap between theory and practice deserves attention, and necessitates an examination of the questioning strategies included in science textbooks.

Questioning is one of the most widely used teaching and learning techniques. Several studies (Floyd, 1968; Hoetker, 1968; Stevens, 1912) have shown that a tremendous portion of the school day is spent in question and answer recitation. It was not until publication of Bloom's (1956) taxonomy and the launching of Sputnik that researchers and educators became concerned about higher level questioning skills. Crump (1970) commented on the poor quality of teacher questioning: "Yet is the potential of the question being fully exploited? Although they have the capacity of initiating critical and creative thinking, many questions focus on memory of specific facts" (p. 657).

Research evidence (Haynes, 1935; Stevens, 1912) clearly indicates that teachers have had a long history of posing a substantial amount of factual-type questions. Studies conducted after the publication of Bloom's (1956) taxonomy reveal that the overwhelming majority of teachers' questions require memory-convergent thinking rather than evaluative-divergent thinking (Davis & Hunkins, 1966; Davis & Tinsley, 1967; Floyd, 1968; Gall, 1970). It must be recognized, however, that acquisition of specific facts serve a valid educational purpose. To suggest that teachers use only higher cognitive questions would diametrically oppose the theoretical perspectives asserted by Bloom (1956) and Taba, Levine, & Elzey (1964). Inherent in both cognitive models is the premise that low-level, factual thinking provides the foundation which allows students to draw inferences, form generalizations, and make judgments. Facts, therefore, should serve as a vehicle to enable students to engage in critical thinking. Sanders (1966) elaborated on the purpose of facts when he stated that teachers, in conceptionalizing a new unit of study, should ask: "What are the most important generalizations that deserve to be emphasized? What facts are necessary to develop these generalizations?"

(p. 21). Factual knowledge represents a vital building block in the problem-solving process; unfortunately, many teachers' questions never go beyond this inital step in thinking.

Bloom's (1956) taxonomy prompted researchers to conduct experimental studies to determine the effects of higher cognitive questions. Taba et al. (1964) found an extremely high correlation between the levels of thinking demonstrated by pupils' answers and the levels of questions asked by their teachers. Additional studies (Andre, 1979; Doak, 1970; Hunkins, 1969; Ladd & Anderson, 1970; Redfield & Rousseau, 1981) provide evidence that the use of higher level questions promotes higher level thinking.

Because one of the major goals in elementary science education is to foster problem-solving capacities through independent critical thinking, the concern for stimulating higher cognitive thought through appropriate questioning strategies is justified. The literature suggests a definite need to examine the content of elementary science textbooks with regard to the types of questions included by writers for pupil evaluation.

Purpose of the Study

The purpose of this study will be to determine the levels of questioning, according to Bloom's (1956) taxonomy, generated by test items that accompany selected fifth-grade science textbooks.

Research Questions

1. What levels of questioning are generated by test items included in the end-of-chapter tests?

2. To what extent do the end-of-chapter test items in the basic science books emphasize the higher levels of questioning, as defined by Bloom's '56) taxonomy?

3. Are there significant differences among the basic science texts studied in the number of knowledge-level and above-knowledge-level test items included in the end-of-chapter tests?

Need for the Study

A study to determine the levels of questioning generated by end-of-chapter test items that accompany science textbooks is justified for the following reasons:

1. The science textbook approach is the most commonly used approach for the teaching of elementary science.

2. End-of-chapter tests accompany most basic science series and are available for teacher usage.

3. Authorities in the field of education agree that the use of higher level questions is essential to the development of higher level thinking skills or critical thinking.

Assumptions

The following assumptions are basic to this research study. First, the use of higher cognitive questions facilitates the development of higher level thinking skills (Hunkins, 1970). Second, end-of-chapter test items that accompany science textbooks are important guides for instruction and evaluation (Sanders, 1966).

Limitations

Limitations of the study relate to specific aspects of questioning not analyzed in this research. First, only end-of-chapter test items that accompany science textbooks will be studied. End-of-chapter questions contained in students' textbooks or manuals and questions contained in teachers' editions for classroom use will not be included in this research. Second, this study will be restricted to fifth-grade science textbooks. Third, this research will be limited to state-adopted textbooks for the states of Alabama, California, Tennessee, and Texas.

Definition of Terms

For the purpose of this study, the following definitions will apply:

Basic science textbook - A systematically organized presentation of science concepts and skills presented in a series of books written by an author or group of authors and presented by a single publisher.

End-of-chapter test - A science evaluation of accompanying textbook material which is designed for systematic administration following each chapter of instruction, devised by the author(s) of the corresponding text and presented by the corresponding publisher.

Bloom's (1956) Taxonomy of Educational Objectives: Cognitive Domain - A hierarchically ordered classification scheme for six levels of questioning: knowledge, comprehension, application, analysis, synthesis, and evaluation.

Definitions of Bloom's six levels of questioning are provided as adapted by Grorlund (1978).

Low Level of Questioning (Gronlund, 1978):

> Knowledge. Knowledge is defined as the remembering of previously learned material. This may involve the recall of a wide range of material, from specific facts to complete theories, but all that is required is the bringing to mind of the appropriate information. Knowledge represents the lowest level of learning outcomes in the cognitive domain. (p.28)

Higher Levels of Questioning (Gronlund, 1978):

> Comprehension. Comprehension is defined as the ability to grasp the meaning of material. This may be shown by translating material from one form to another (words to numbers), by interpreting material (explaining or summarizing), and by estimating future trends (predicting consequences or effects). These learning outcomes go one step beyond the simple remembering of material, and represent the lowest level of understanding.

Application. Application refers to the ability to use learned material in new and concrete situations. This may include the application of such things as rules, methods, concepts, principles, laws and theories. Learning outcomes in this area require a higher level of understanding than those under comprehension.

Analysis. Analysis refers to the ability to break down material into its component parts so that its organizational structure may be understood. This may include the identification of the parts, analysis of the relationship between parts, and recognition of the organizational principles involved. Learning outcomes here represent a higher intellectual level than comprehension and application because they require an understanding of both the content and the structural form of the material.

Synthesis. Synthesis refers to the ability to put parts together to form a new whole. This may involve the production of a unique communication (theme or speech), a plan of operations (research proposal), or a set of abstract relations (scheme for classifying information). Learning outcomes in this area stress creative behaviors, with major emphasis on the formulation of new patterns or structures.

Evaluation. Evaluation is concerned with the ability to judge the value of material (statement, novel, poem, research report) for a given purpose. The judgments are to be based on definite criteria. These may be internal criteria (organization) or external criteria (relevance to the purpose) and the student may determine the criteria or be given them. Learning outcomes in this area are highest in the cognitive hierarchy because they contain elements of all of the other categories, plus conscious value judgments based on clearly defined criteria. (p. 28)

ANALYSIS OF DATA AND RESULTS

The major purpose of this study was to determine the levels of questioning generated by test items included in the end-of-chapter tests that accompany selected fifth-grade science textbooks. Classification of the test items was based on the six major categories included in Bloom's (1956) taxonomy: knowledge, comprehension, application, analysis, synthesis, and evaluation.

Textbook information was obtained from the state departments of education of Alabama, California, Tennessee, and Texas to obtain a rank ordering of the fifth-grade science textbooks used most frequently in each state. From that list, the five series used most frequently for all states combined were selected for ana'ysis. The five series studied were: Silver Burdett Science (1984), published by Silver Burdett Company; Holt Science (1984), published by Holt, Ri ehart, and Winston Company; Scott, Foresman Science (1984), published by Scott, Foresman and Company; Accent on Science (1983), published by Charles E. Merrill Company; and HBJ Science (1985), published by Harcourt Brace Jovanovich Company.

Training of Raters

Prior to the actual classification of data, three raters were prepared for the task through three training sessions. The training involved a review of Bloom's (1956) taxonomy and practice classification exercises included in Sander's (1966) handbook, Classroom Questions: What Kinds? Each rater was presented with a training packet which included the following: (a) Taxonomy of Educational Objectives Handbook I: Cognitive Domain (Bloom, 1956), (b) a list of key words used as criteria for item classification, (c) a sample test item classification form, and (d) pertinent notes from Classroom Questions: What Kinds? (Sanders, 1966). The scope of the initial training session focused on developing a common framework of terminology and criteria needed for reliable classification.

The second training session involved classifying written questions from the chapter-end practice exercises in Sander's (1966) Classroom Questions: What Kinds? After independent classifications were made, the raters compared their results and discussed the rationale for each classification. Discrepancies among the three judges were noted and discussed; agreements were reached after further deliberation and additional review of Bloom's (1956) taxonomy and training notes. At the close of this session, the raters reached 86% agreement on a random sample of 50 science test items (10 from each textbook series) not included in the actual classification study. Although this level met the predetermined level of acceptable agreement, an additional training session was held to provide more experience with science test items from the five textbook series.

During the third session, the raters discussed each of the 50 test items classified in the previous session and the rationale for the category chosen. Of particular interest in this final session were discrepant test items revealed in the prior classification. After discussion, concensus was achieved on all test items in question. Following this session, an additional pool of 50 science test items, not included in the previous training session or actual classification study, was randomly selected for further analysis. Interrater agreement reached 96% on this sample, which exceeded the specified acceptable level of 85%. Upon obtaining 96% interrater agreement, the classification of actual data began.

Data Treatment

The sample consisted of a total of 500 science test items, 100 from each of the five science textbook series. These items were randomly selected prior to the training sessions and extracted from all preclassification training exercises and discussions. In categorizing the data, the three raters used training packets and the key word list to independently classify each of the 500 test items.

The interrater agreement achieved during the actual classification study reached 88%. The remaining test items (12%) were classified according to the assignment of the majority of raters. In all cases where two-out-of-three agreement was used, the discrepant rater had assigned the test item to a category adjacent to the one chosen by the majority of raters.

To represent the data; answer the research questions; and test the proposed hypotheses, the study utilized bar-graph comparisons, frequency distribution tables, and a series of chi-square statistical analyses. The chi-square procedure was employed to determine whether the distribution of observed frequencies differed from that expected by chance.

Findings

Research Question 1

The first research question asked: What levels of questioning are generated by test items included in the end-of-chapter tests? In order to answer this question, the following null hypothesis was tested: There is no significant difference between the frequencies of end-of-chapter test items classified at the knowledge-level and above-knowledge-level of questioning for all textbooks combined.

Each test item chosen for classification was rated independently by the three trained analyzers. The results of the ratings were compared and the test item was placed in the category ascribed by the majority of raters. Table 1 reports a frequency distribution of the results of this classification. These results are represented in a bar graph in Figure 1.

Inspection of these data indicate that the majority of test items that accompany these five series were classified in the knowledge catgegory, the lowest of Bloom's (1956) cognitive domains. The overall frequency distribution and percentage tabulation reveals that of all 500 items randomly selected for analysis, 61% dealt with knowledge. The second largest percentage of items was categorized as comprehension (34%). Only 5% of the test items fell into the application category. Neither the analysis nor synthesis questioning level was represented in the five textbooks: only one evaluation test item existed. Collectively, results of the classification demonstrate that, with the exception of the one evaluation test item, only the first three levels in Bloom's (1956) cognitive domain were represented by the textbooks studied.

Although these findings indicate that the total observed frequencies were largely in the knowledge category, comparisons between individual textbook distributions yield important differences which should be discerned. The Scott, Foresman series, for example, contained the largest percentage of knowledge-level test items. The next largest percentage of knowledge-level test items was demonstrated by the Silver Burdett and Harcourt Brace Jovanovich series, respectively.

The series analyzed in this study differed vastly with respect to the comprehension category. The data show that the Holt Science series contained the largest percentage of comprehension items. Merrill possessed a similar distribution with the next highest percentage of comprehension items and the lowest percentage of knowledge-level items. In comparison with the other texts, Merrill's emphasis was different in the area of application; 14% of its items fell into that category. Merrill is the only series that had any of its test items rated as evaluation; one item necessitated an evaluation or judgment on the part of student based on external criteria. Of major note is the finding that among the five series studied, no analysis or synthesis test items existed, and only one item was classified under evaluation.

A chi-square goodness of fit analysis was calculated to determine if the observed differences between the total frequencies of the combined textbooks' test items (knowledge-level vs. above-knowledge-level) departed significantly from those expected by chance. The following results were obtained:

	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Knowledge	304	250	54	2916	11.664
Above-Knowledge	196	250	-54	2916	11.664
					$X^2 = 23.328$

$X^2(1) = 23.328$; $p < .001$

Because the calculated chi-square value of 23.328 exceeds the critical chi-square value beyond the .001 level, the null hypothesis is rejected. This statistic indicates that the distribution of total frequencies for knowledge-level and above-knowledge-level test items for all series combined departed significantly from a distribution based on chance.

Research Question 2

The second research question asked: To what extent do the end-of-chapter test items in the basic science books emphasize the higher levels of questioning, as defined by Bloom's (1956) taxonomy? To determine the extent to which the test items in the selected science books emphasize the higher levels of questioning, results of the data were dichotomized into two major categories: lower level and higher level. Thus, lower level test items refer to Bloom's knowledge level, and higher level items represent levels two through six in the hierarchy.

By collapsing the comprehension, application, analysis, synthesis, and evaluation levels into one major category, it is possible to analyze the emphasis of each science series on understanding and using science information (higher level) as compared to rote memorization (lower level). The resulting data were tabulated and percentages were calculated (see Table 2 and figures 2 - 6). These results indicate that, even when Bloom's cognitive levels two through six were combined for examination, the lower level of questioning represented the largest percentage of test items. It should be noted that the overwhelming majority of higher level test items were categorized as comprehension, the lowest level of understanding.

The relatively low emphasis placed on higher level questioning (39% overall) reveals interesting similarities and differences among the series studied. The Holt and the Merrill series had a generally even balance of lower level and higher level test items. Both series had slightly larger percentages of higher level than lower level items; however, there were differences in the distribution of items in the higher level category. As shown in Table 2, Merrill placed a heavy emphasis on higher level test items (58%). Close inspection of the distribution of Merrill's test items reveals that 14 out of 58 higher level questions (24%) represented the application category. In contrast, Holt, the series with the next largest percentage of higher level items, had only two application questions which represented 3% of the total; the remaining questions in the higher level category were comprehension.

The Harcourt Brace Jovanovich and Silver Burdett series demonstrated the next largest percentages of higher level questions, respectively. An analysis of the higher level questioning categories for these series shows that the clear majority of the items were classified under comprehension. Of the 36 higher level items in the Harcourt Brace Jovanovich series, 8 (22%) were judged as application items. Silver Burdett had only one application item. The Scott, Foresman series contained the lowest percentage of higher level questions. The higher level category in this series was comprised of 17 comprehension items and one application item.

When considering the overall percentages of higher level questions, the Merrill and Holt series gave greater emphasis to this category than did the other series studied. It is worth noting again, however, that in these five textbooks the majority of higher level questions were classified as comprehension. Relatively few higher level questions were application and only one was classified in the three highest levels of Bloom's (1956) taxonomy: analysis, synthesis, and evaluation.

Research Question 3

The third research question asked: Are there significant differences among the basic science textbooks studied in the distribution of knowledge-level and above-knowledge-level test items included in the end-of-chapter tests? The second null hypothesis was tested to answer research question 3: There are no significant differences among the basic science text-books studied in the distribution of knowledge-level and above-knowledge-level test items included in the end-of-chapter tests.

To determine if there were significant differences among the basic science textbooks studied in the number of knowledge-level and above-knowledge-level test items, a chi-square test of independence was performed. The calculation consisted of two steps. First, a 5 x 2 matrix

(Textbook Series x Knowledge-Level and Above-Knowledge-Level Questions) was established to determine whether the distribution of one variable (textbook series) was independent of the second variable (test items). The overall 5 x 2 chi-square test yielded the following results:

	O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Silver Burdett KL	67	60.8	6.2	38.44	.632
Silver Burdett AKL	33	39.2	-6.2	38.44	.980
Holt KL	49	60.8	-11.8	139.24	2.290
Holt AKL	51	39.2	11.8	139.24	3.552
Scott, Foresman KL	82	60.8	21.2	449.44	7.392
Scott, Foresman AKL	18	39.2	-21.2	449.44	11.465
Merrill KL	42	60.8	-18.8	353.44	5.813
Merrill AKL	58	39.2	18.8	353.44	9.016
Harcourt Brace KL	64	60.8	3.2	10.24	.168
Harcourt Brace AKL	36	39.2	-3.2	10.24	.261
					$\underline{X}^2 = 41.5716$

Note: KL = Knowledge-Level
AKL = Above-Knowledge-Level

$\underline{X}^2(4) = 41.57$; $\underline{p} < .001$

Since the calculated chi-square value (41.5716) exceeds the critical chi-square value beyond the .001 level, the two variables (textbook series and test items) are not independent (i.e., a significant relationship is indicated). This reveals that the distribution of test items (knowledge-level and above-knowledge-level) for the five textbooks series depart significantly from that expected by chance, indicating that some textbooks had a significantly greater percentage of higher level test items than others. On the basis of this calculation, the null hypothesis is rejected.

Second, a series of post hoc analyses were necessary to determine where these differences in frequencies of the levels of test items existed; therefore, ten 2 x 2 chi-square calculations tested whether the difference in the distributions of knowledge-level and above-knowledge-level test items was statistically significant. The results of these comparisons are presented in Table 3.

As Table 3 evidences, the series of 2 x 2 (Textbook Series x Knowledge-Level and Above-Knowledge-Level Items) analyses produced several significant results concerning the number of above-knowledge-level test items. A statistically significant difference was indicated between the Merrill series, which had the most above-knowledge-level items, and three of the series analyzed. A highly significant difference was observed between the Merrill and Scott, Foresman series ($\underline{X} = 39.95$; $\underline{p} < .001$). Also, a significant difference beyond the .001 level was noted between the Merrill and Silver Burdett series ($X = 12.60$). The comparison between the Merrill series and Harcourt Brace Jovanovich series yielded a chi-square value of 9.71, which was significant beyond the .01 level. No significant difference was found between the Merrill and Holt series.

Interpretation of Table 3 further indicates that a significant difference existed between the Holt science series, which had the second largest number of above-knowledge-level items, and three series studied. The comparison of Holt's distribution of test items to Scott, Foresman's distribution resulted in a significant ($p < .001$) chi-square value of 24.09. When compared to the Silver Burdett series, the Holt series had a significantly ($p < .01$) greater number of above-knowledge-level test items, as evidenced by the chi-square value of 6.65. A significant difference beyond the .05 level was noted between the Holt and the Harcourt Brace Jovanovich series.

In addition to the aforementioned results, Table 3 reveals two additional chi-square comparisons which reached significance. The Harcourt Brace Jovanovich series had significantly ($p < .01$) more above-knowledge-level items than the Scott, Foresman series ($\underline{X}^2 = 8.21$). When the above-knowledge-level items were compared between the Silver Burdett and Scott, Foresman series, a significant difference beyond the .05 level was found.

Summary

The results of the classification of 500 science test items according to Bloom's (1956) taxonomy indicate that a preponderance of knowledge-level items existed across the five series studied. The combined frequency distribution and percentage tabulation demonstrate that 304 (61%) test items emphasized knowledge, the lowest level of Bloom's hierarchy. The textbooks included more knowledge-level test items than all other items combined. Bloom's comprehension category accounted for 169 (34%) of the items under study. A total frequency of 26 (5%) application items existed. Noticeably absent from the frequency distribution were Bloom's three highest levels of cognition; no analysis or synthesis items were found and only one evaluaion item existed. The chi-square goodness of fit statistic, which compared the frequencies of knowledge-level and above-knowledge-level items for all textbooks combined, resulted in a significant ($p < .001$) chi-square value of 23.3287. On the basis of this finding, the first null hypothesis was rejected.

By dichotomizing the resulting data into lower and higher levels of questioning, it was determined that the Merrill and Holt series possessed slightly greater percentages of higher level questions. Of the 58 higher level questions from the Merrill series, 14 (24%) were application; the remaining higher level questions represented the comprehension category. Two of the 51 higher level questions from the Holt series were application; all remaining questions fell into the comprehension category. In the other three textbooks analyzed, the lower level of questioning represented the largest percentage of test items.

To answer the third research question, two steps were necessary. First, a chi-square test of independence of two variables (Textbook Series vs. Test Items) was calculated. In the overall 5 x 2 test, a chi-square value of 41.5716 was obtained, which was significant beyond the .001 level. Second, a series of 10 post hoc analyses were needed to compare all possible combinations of textbook series to determine where the differences in distributions existed. These comparisons showed that the Merrill and Holt textbook series possessed the greatest number of above-knowledge-level items as compared with the other series studied. When these two series were compared, no statistically significant difference was indicated. In all other 2 x 2 analyses, however, the Merrill and Holt series included significantly more above-knowledge-level test items than did the other science series. The Scott, Foresman series was found to contain the fewest above-knowledge-level items as compared with the other series. In all cases, the Merrill, Holt, HBJ, and Silver Burdett series contained significantly more above-knowledge-level test items than the Scott, Foresman text. On the basis of these findings, the second null hypothesis, which stated that there would be no significant differences among the basic science textbooks studied in the distribution of knowledge-level and above-knowledge-level test items, was rejected.

SUMMARY, FINDINGS AND CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS

Summary

The elementary science curriculum must address real-life experiences that foster the development of problem-solving skills and provide students with opportunities to apply the scientific method. Since the ability to think critically is a desired outcome of the science program, and elementary teachers rely heavily on textbooks as the primary source of science instruction, elementary science textbooks and accompanying resource materials, such as tests, must be analyzed to determine their ability to stimulate higher levels of thought.

This study sought to determine the cognitive levels of questioning displayed by test items contained in the chapter-end tests that accompany selected fifth-grade science textbooks. Additionally, this research examined the extent to which each basic series emphasized the higher levels of questioning and what significant differences, if any, existed among the texts in the number of knowledge-level and above-knowledge-level test items.

The sample of materials examined in this study included the five most widely used fifth-grade science textbooks from the states of Alabama, California, Tennessee, and Texas. Data were derived from 500 test items, 100 randomly selected from each textbook. After obtaining a 96% interrater agreement on training materials, three raters independently classified the 500 items according to the major categories included in Bloom's (1956) taxonomy. Results of the classifications were analyzed and presented in frequency distribution tables and a series of bar-graph comparisons. The chi-square statistic was selected to test the proposed hypotheses.

Findings and Conclusions

The first research question sought to determine the cognitive levels of questioning revealed by test items included in the end-of-chapter tests. Corresponding to this question, the following null hypothesis was tested: There is no significant difference between the frequencies of knowledge-level and above-knowledge-level test items for all textbooks combined.

A total of 304 (61%) test items in the textbooks studied was judged as knowledge. The textbook series contained 169 (34%) comprehension items and only 26 (5%) application items. Items representing the analysis and synthesis categories were nonexistent in the distribution; only one evaluation item was found. The goodness of fit comparison of knowledge- and above-knowledge-level test items yielded a significant ($p < .001$) chi-square value of 23.3287; therefore, the first null hypothesis was rejected.

When examined collectively, the five textbooks present significantly more knowledge-level than above-knowledge-level test items. Therefore, it is concluded that end-of-chapter tests that accompany selected science textbook series predominantly emphasize mere recall or recognition of isolated bits of factual information. Because these findings parallel those from earlier studies (Davis & Hunkins, 1966; Pfeiffer & Davis, 1965) conducted during the post-Sputnik period of the 1960s, it is concluded that little or no change has occurred over time concerning the cognitive emphasis of textbook questioning. Furthermore, the heavy concentration of knowledge-level questions corresponds with other research conducted in elementary science (e.g., Arnold et al., 1973), thus indicating that science questioning remains factually oriented. The cognitive range of questioning displayed by the five textbook series is restricted to the lowest three categories of Bloom's (1956) hierarchy. This finding demonstrates a weighting of items toward the base of the taxonomy and a decreasing quantitative ordering of knowledge, comprehension, and application items. The distribution further reveals that the textbooks analyzed in this research place no emphasis on Bloom's upper three taxonomic categories; therefore, it is concluded that these textbook publications display a uniform neglect of the highest questioning levels: analysis, synthesis, and evaluation.

The second research question focused on the extent to which the test items in the basic texts emphasized the higher levels of questioning. Levels two through six of Bloom's (1956) taxonomy were combined and percentages were calculated. By collapsing the upper five levels into one major category, textbooks were examined according to their emphasis on understanding science information (higher level) as compared to memorization (lower level). Overall, the five series overwhelmingly stress lower level questions. However, this analysis revealed that two series, Merrill and Holt, inc de slightly greater percentages of higher level questions than lower.

It is concluded that the Merrill and Holt series possess generally balanced disbursements of lower and higher level questions on end-of-chapter examinations. Based on these results, it is surmised that the Merrill and Holt series place far greater emphasis on higher level questioning than do the other textbooks studied, thereby exposing students to cognitive stages other than mere recall. When these two textbook series were compared, however, one notable difference emerged. Merrill includes more opportunities than Holt for students to respond to test items requiring the application of scientific concepts to real-life situations. In contrast, only three of Holt's higher level questions fall into the application category. Since the clear majority of higher level questions demonstrated by Merrill and Holt tests are comprehension, it is concluded that both series, although equal in their representation of lower and higher level questioning, focused primarily on the lowest level of understanding.

The third research question compared all possible combinations of textbooks series in order to determine whether significant differences existed in the distributions of knowledge- and above-knowledge-level test items. The following null hypothesis was tested: There are no significant differences among the basic science textbooks studied in the distribution of knowledge-level and above-knowledge-level test items included in the end-of-chapter tests.

The chi-square test of independence of two variables (Textbook Series x Test Items) resulted in a significant ($p < .001$) chi-square value of 41.5716. Because this finding reveals that some textbooks possess significantly greater percentages of higher level test items than others, the null hypothesis was rejected.

Interpretation of each post hoc 2 x 2 analysis revealed significant differences between the Merrill and Holt series and all other series examined in the study. No significant difference was evidenced, however, between Merrill and Holt. Additionally, significant differences were found in two other comparisons: (a) Harcourt Brace Jovanovich and Scott, Foresman; and (b) Silver Burdett and Scott, Foresman.

Based on the results of these data, it is concluded that both the Merrill and Holt series contain significantly more above-knowledge-level items than Harcourt Brace Jovanovich, Silver Burdett, and Scott, Foresman. The greatest discrepancies exist in the Merrill and Scott, Foresman comparison and the Holt and Scott, Foresman comparison. Also, rather large discrepancies occur in the analyses comparing Merrill to Silver Burdett and Holt to Silver Burdett.

An overall conclusion from these post hoc analyses is that the Merrill and Holt textbooks are significantly superior to the other series studied in their emphasis of above-knowledge-level test items. Since every comparison with Scott, Foresman reached significance, a rather obvious conclusion is that Scott, Foresman possesses significantly fewer above-knowledge-level items that the other four textbook series examined.

Implications

The implications derived from the findings of this research are based on two fundamental assumptions. First, the use of higher order questions stimulates critical thinking. As noted by Taba (1964), questions are central to the instructional process and can be arranged in a simple-to-complex progression for the purpose of elevating student thought to higher levels. Indeed, much research evidence has veried Taba's premise. Since the inception of Bloom's (1956) taxonomy, much scholarly literature has supported its use as a means to plan purposely and incorpoprate questions representing a variety of cognitive levels (Gall, 1970). Abundant research evidence has highlighted the need for teachers to include Bloom's upper three taxonomic levels to enhance critical thinking. Second, tests that accompany textbooks are important guides to instruction and evaluation (Billings, 1970; Sanders, 1966).

Based on the assumption that the test items analyzed in this research are representative of all items in the five textbooks studied, and the five science textbooks selected for study are representative of all science textbooks, one rather obvious implication is that higher order thinking skills are not elicted by end-of-chapter tests that accompany science textbooks. It is recognized that the acquisition of knowledge precedes functioning at higher cognitive levels; however, the emphasis on knowledge demonstrated by these tests is excessive and indicates that elementary science students have little or no encouragement, at least on published examinations, to understand and apply information. If exposing students to higher order thinking skills through a balance of higher cognitive questions is an integral part of elementary science, none of the end-of-chapter tests are desirable. A lesser goal might be to provide students with cognitive experiences other than rote memorization, while not necessarily concentrating on the upper three levels in Bloom's (1956) taxonomy. In this case, all textbooks examined in this study rate extremely low except Merrill and Holt. While tests from these series fail to emphasize analysis, synthesis, and evaluation, they do focus primarily on some form of understanding (comprehension and application). Though significant differences exist between the Merrill and Holt series and all other series studied, these are attributed to differences shown in the comprehension and knowledge levels, the lowest two categories in the hierarchy.

These findings have broader implications concerning the instructional process in general and, more specifically, elementary science teaching in the 1980s. It is recognized that ready-made tests that accompany newly acquired science textbooks tend to become guides for instruction as well as evaluation. Designed to be administered after each chapter of instruction, these tests quickly become the focus of the science experience. Some teachers, particularly those with little confidence in their science instruction, refuse to deviate from a textbook-approach and feel compelled to adhere strictly to the lock-step instruction and evaluation techniques that accompany recent publications. Similarly, the rigid testing movement currently underway in our nation's schools encourages the use of traditional methods in science. Concern for science achievement as measured by standardized achievement tests limits the majority of elementary science classes to the textbook and accompanying materials (Smith, 1980, cited by Mechling & Oliver). Therefore, based on the assumption that evaluation is representative of instruction, the findings of this research infer that science instruction as well as evaluation focus on little more than remembering facts.

Also, the findings from this study should be considered in view of the commonly held objectives of elementary science. Although it was beyond the scope of this research to analyze the goals and objectives inherent in the science series, the findings regarding the cognitive levels of science textbook examinations suggest a vital dimension relative to science textbook publishers. The five series selected for this study include scope and sequence charts that list the basic science concepts covered in each textbook. A review of the teacher's edition of each series reveals a listing of higher order skills such as making inferences, forming hypotheses, drawing conclusions, and the like. Silver Burdett Science (Ackerman et al.,1984), the most widely used series in each state included in this study, makes the following claim: "The basic thinking skills emphasized in the Silver Burdett Science program are summarized in the chart below. The skills are divided into three major areas, based on Bloom's taxonomy of thinking skills" (p. 29). In addition to a thorough explanation of the process skills of science, the Scott, Foresman teacher's edition includes the following statement in describing the series: "Effective questioning is essential . . . Open ended questions stimulate a creative exchange of ideas. Asking questions that require little reasoning or a simple yes or no answer is not a true evaluation of understanding" (p. 28).

The goals and objectives outlined by the aforementioned textbook series are ironic in view of the evidence gathered in this study, particularly because these two textbook series demonstrate the highest percentages of knowledge-level items. The other series studied make similar proclamations concerning their respective emphasis on critical thought processes and thought-provoking questioning.

The five textbook series, in varying degrees, attempt to include some experiences representative of the upper three levels in Bloom's hierarchy. Examples of such experiences include teaching activities, experiments, chapter inserts for independent enrichment, and even questions in the teacher's edition for classroom discussion. Likewise, the charts of science objectives corresponding to each of the five science series contain terminology related to complex process skills and Bloom's application, analysis, synthesis, and evaluation categories. Since the purpose of published examinations is to assess the degree to which stated goals and objectives are met, it is expected that test items would correspond to the varying cognitive levels exhibited in textbook skill charts. However, the tests measure primarily the recall of facts. This finding reveals that a discrepancy exists between publishers' preferred goals of elementary science and the instruments that measure the accomplishment of science goals. The results of the data imply that there is a disunified approach to include the higher cognitive processes in science textbooks, thus supporting the premise that a gap exists between science theory and actual practice.

Finally, it must be noted that these textbook examinations reflect, to some extent, the preferences of the majority of classroom teachers. Most textbook adoption presentations begin with survey data reflecting teachers' opinions concerning the desired format and accompanying materials for science textbooks. Based on the assumption that science textbooks adequately reflect teachers' opinions, it is implied that factually oriented tests are preferred by the majority of elementary teachers. Perhaps this inference further highlights the insecurities elementary teachers possess about science instruction. It is more likely, however, that the strong dependence upon supplementary textbook materials, such as tests, stems from the fact that they are scored easily and quickly, thereby reducing the teacher's workload. In either case, students are the beneficiaries of an unstimulating science experience absent of any real-life application.

It must be remembered that tests and grades greatly influence the image of science that students take with them from elementary school. If successful science is equated with the ability to memorize a given body of knowledge, pupils are deprived of a challenging intellectual climate. The data obtained in this study imply that students can master these tests without experiencing any true understanding of the scientific principles involved. Developing higher levels of understanding requires a more logical progression through all taxonomic levels and more attention to the levels of application, analysis, synthesis, and evaluation.

Recommendations

The results of this study should be particularly beneficial to science publishers in planning and developing textbook questions, especially those designed for student evaluation. Also, this study should aid those who serve on state and local textbook selection committees in critically assessing science texts and accompanying materials during the adoption process. Further, preservice teacher education programs as well as practicing classroom teachers should find the data useful in examining current questioning strategies in elementary science and developing teaching and evaluation techniques sensitive to all cognitive levels in Bloom's (1956) taxonomy. Recommendations from the results of this research address: textbook publishers, textbook selection committees, preservice teacher educators, and classroom teachers, respectively:

1. Science textbook publishers should attempt to include a logical progression of science test items reflective of all six levels in Bloom's (1956) hierarchy.

2. Publishers should make a deliberate effort within each unit of study to correlate the varying cognitive levels of textbook objectives, activities, questions, and test items for the purpose of creating unified experiences that represent a balance of cognitive levels.

3. In addition to tests that emphasize all cognitive levels, science textbook publishers should include corresponding resource materials, relative to each unit of study, that engage students in substantial practice with comprehension, application, analysis, synthesis, and evaluation questions.

4. Textbook selection committees should become familiar with research concerning the cognitive levels of questions used by classroom teachers, questions included in textbooks, and questions included on examinations, as well as the research related to the effective and logical

use of questioning.

5. During the examination process, textbook selection committees need to analyze the questioning style of each textbook series in light of the previous research findings; a well-balanced cognitive approach to questioning, in textbooks and accompanying materials, should be included in the criteria for adoption.

6. Textbook selection committees should analyze critically the claims proposed by textbook representatives concerning the higher levels of questioning included in their respective series.

7. Preservice teacher educators need to emphasize research findings and include instruction on effective questioning techniques in methods courses.

8. Activity-based programs such as SCIS, SAPA, and ESS should be included in the science methods courses as a means to promote student awareness concerning effective science questioning.

9. Preservice teacher educators should provide instruction and practical experiences in supplementing science textbook material, including tests, with higher level questions.

10. Preservice teacher educators should include micro-teaching sessions that promote self-assessment and improvement of both oral and written questioning techniques.

11. Teachers should become familiar with the various cognitive levels of questioning and the research findings concerning the use of higher level questions.

12. Teachers need to critically assess the level of questioning

demonstrated by recent textbooks and accompanying tests and devise supplementary questions, both written and oral, that facilitate the development of the higher cognitive processes.

13. Teachers should consider the need for higher level questions in planning science lessons; objectives and evaluation should promote critical thinking through purposely planned discussion, activities, and evaluation that include higher level questions.

Suggestions for Additional Research

Because questioning is such a vital aspect of the instructional process, and higher level questions are essential in developing critical thinking skills, further research needs to be conducted related to questioning. The results of this study indicate a need for the following research:

1. A similar study needs to be conducted in elementary science to determine the cognitive levels of questions included in teacher's editions for classroom discussion and pupils' textbooks for chapter-end review.

2. A study in elementary science should be conducted to determine the relationships among the cognitive levels of questions demonstrated in teachers' editions, textbook activities, and end-of-chapter tests.

3. Additional research is needed to determine the relationship between question level and response level in elementary science.

4. A similar study should be conducted to determine the oral questioning levels used by elementary science teachers.

5. More experimental studies are needed to determine the effects of

higher level questions on student achievement, attitude, and interest in elementary science.

References

Adams, T.H. (1964). The development of a method for analysis of questions asked by teachers in c'assroom discussion. Unpublished doctoral thesis, Rutgers State University, 1964.

Andre, D.T. (1979). Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 49, 280-318.

Arnold, D.S., Atwood, R.K., & Rogers, V.M. (1973). An investigation of relationships among question level, response level, and lapse time. School Science and Mathematics, 73, 591-594.

Aschner, M.J. (1961). Ask'ng questions to trigger thinking. National Education Association Journal, 50, 44-47.

Atwood, R.K., & Stevens, J.T. (1974). Relationships among question level, response level, and lapse time: secondary science. School Science and Mathematics, 74, 249-254.

Billings, G. (1971). Cognitive levels of elementary science tests. School Science and Mathematics, 71, 824-830.

Black, T.R. (1980). An analysis of levels of thinking in Nigerian science teachers' examinations. Journal of Research in Science Teaching, 17, 301-306.

Bloom, B.S. (1956). Taxonomy of educational objectives: the classification of educational goals, handbook I: the cognitive domain. New York: McKay.

Bolden, J. (1984). Statistics with finesse [Computer program]. Grennville, IL.

Borg, W.R., & Gall, J.D. (1983). Educational research: an introduction. New York & London: Longman.

Bredderman, T. (1983). Effects of activity-based science on student outcomes: a qualitative synthesis. Review of Educational Research, 53, (4), 499-518.

Bredderman, T. (1984). The influence of activity-based elementary science programs on classroom practices: a quantitative synthesis. Journal of Research in Science Teaching, 21, (3),. 289-303.

Bruce, L.R. (1971). A study of the relationship between the SCIS teachers' attitude toward the teacher-student relationship and question types. Journal of Research in Science Teaching, 8, (2), 157-164.

Bruner, J.S. (1960). The process of education, Cambridge, Massachusettes: Harvard University Press.

Bruner, J.S. (1960). The process of education, Cambridge, Massachusettes: Harvard University Press.

Brunner, J.F. (1983). Some comments on reading comprehension instruction from a taxonomic point of view. Reading Improvement, 20, (1), 2-6.

Buggey, L.J. (1972). A study of the relationship of classroom questions and social studies achievement of second grade children (Report No. SO 003 731). Chicago, IL: Educational Research Association (ERIC Document Reproduction Service No. ED 066 391).

Carner, R.L. (1963). Levels of questioning. Education, 83, 546-550.

Cole, R.A., & Williams, D.M. (1973). Pupil response to teacher questions: cognitive level, length and syntax. Educational Leadership, 31, 142-145.

Corey, S.M. (1940). The teachers out-talk the pupils, The School Review, 48, 745-752.

Crump, C. (1970). Teachers, questions, and cognition. Educational Leadership, 27, 657-660.

Cunningham, H.A. (1925). Types of thought questions in general science textbooks and laboratory manuals. General Science Quarterly, 9, 91-95.

Davis, O.L., & Hunkins, F.P. (1966). Textbook questions: what thinking processes do they foster? Peabody Journal of Education, 43, 285-292.

Davis, O.L., & Tinsley, D.C. (1967). Cognitive questions revealed by classroom questions asked by social studies student teachers. Peabody Journal of Education, 45, 21-26.

Dewey, J. (1916). Democracy and education. New York: Macmillan.

Dewey, J. (1933). How we think: a restatement of the relation of reflective thinking to educative process. Washington: D.C.: Heath and Company.

Doak, D.E. (1970). Evaluating levels of thinking. School and Society, 98, 177-178.

Ferguson, G.A. (1981). Statistical analysis in psychology and education. New York: McGraw-Hill.

Floyd, W. (1960). An analysis of the oral questioning activity in selected Colorado primary classrooms. Unpublished doctoral dissertation, Colorado State College.

Floyd, W. (1968). Do teachers talk too much? Instructor, 9, 53-57.

Funk, H.J. (1977). An examination of the levels of questions on standardized tests for elementary science. Science and Children, 14, (4), 30-32.

Gagne, R.M. (1970). The conditions of learning. New York: Holt, Rinehart and Winston.

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

Gall, M.D., Ward, B.A., Berliner, D., Cahen, L.S., Winne, P.H., Elashoff, J.D., & Stanton, G.C. (1978). Effects of questioning techniques and recitation on student learning. American Education Research Journal, 15, 175-199.

Gallagher, J.J. (1965). Expressive thought by gifted children in the classroom. Elementary English, 42, 559-568.

Glass, G.V. (1978). Integrated findings: The meta-analysis of research. Illinois: Peacock.

Gronlund, N.E. (1978). Stating objectives for classroom instruction. New York: Macmillan.

Guilford, J.P. (1956). Structure of the intellect. Psychological Bulletin, 53, (4), 267-297.

Gusak, F.J. (1967). Teacher questioning and reading. Reading Teacher, 21, 227-334.

Harbeck, M.B., & Marcuccio, P. (1978). Science in the lives of children. Childhood Education, 55, (2), 94-98.

Haynes, H. (1935). The relationship of teacher intelligence, teacher experience and type of school to types of questions. Doctoral dissertation, George Peabody College, Nashville, TN

Heller, M.O. (1978). Asking questions - a basic skill. Curriculum Review, 17, 115-117.

Hoetker, J. (1968). Teaching questioning behavior in nine junior high school english classes. Research in the Teaching of English, 2, 106-109.

Hoetker, J., & Ahlbrand, W.P. (1969). The persistence of the recitation. American Educational Research Journal, 6, (2), 99-106.

Horak, W., & Roubinek, D.L. (1981). The social studies and science curriculum. Contemporary Educational Psychology, 6, 287-301.

Hunkins, F.P. (1968). The influence of analysis and evaluation questions on achievement in sixth-grade social studies. Educational Leadership, 25, 326-330.

Hunkins, F.P. (1970). Analysis and evaluation questions: their effects upon critical thinking. Educational Leadership, 3, 697-705.

Hunkins, F.P. (1972). Involving students in questioning. Boston: Allyn and Bacon.

Jarolimek, J. (1962). The taxonomy: quide to differentiated instruction. Social Education, 26, 445-447.

Johnson, R.T. (1983). Elementary science. In Focus on excellence (Vol. 1, no. 2, pp. 4-10). Washington, DC: National Science Teachers Association.

Kniep, W.M. (1977). Some effects of high level questions with systematic positive reinforcement on the achievement of selected social studies concepts. Social Education, 41, 340-344.

Konya, B.A. (1972). The effects of higher and lower order teacher questions on the frequency and type of student verbalizations. Dissertation Abstracts International, 33, 4177A. (University Microfilms No. 73-2466)

Kropp, R.P., & Stoker, H.W. (1966). The construction and validation of tests of the cognitive processes as described in the taxonomy of educational objectives (Report No. CRP-2117). Tallahassee, FL: Florida State University, Institute of Human Learning. (ERIC Document Reproduction Service No. ED 010 044)

Ladd, G.T., & Anderson, H.O. (1970). Determining the level of inquiry in teachers' questions. Journal of Research in Science Teaching, 7, 389-400.

Lapp, D.M. (1980). Helping teachers teach science: the need for support systems. National Elementary Principal, 59, (2), 62-64.

Lombard, J.W. (1965). Preparing better classroom tests. The Science Teacher, 32, 33-38.

Manson G., & Clegg, A.A. (1970). Classroom questions: keys to children's thinking? Peabody Journal of Education, 47, 302-307.

Marksberry, M.L., McCarter, M., & Noyce, R. (1969). Relation between cognitive objectives from selected texts recommendations of national committees. Journal of Educational Research, 62, (9), 422-429.

Martikean, A. (1973). The levels of questioning and their effects upon student performance above the knowledge level of Bloom's taxonomy of educational objectives. (Report No. SE 017 956). Indiana University Northwest. (ERIC Document Reproduction Service No. ED 091 248)

Measel, W., & Wood, D.E. (1972). Teaching verbal behavior and teacher and pupil thinking in elementary school. Journal of Educational

Research, 66, 991-03.

Mechling, K.R., & Oliver, D.L. (1983). Activities, not textbooks: what research says about science programs. Principal, 62, (4), 41-43.

Mills, S.R., Rice, C.T., Berliner, D.C., & Rosseau, E.W. (1979). The correspondence between teacher questions and student answers in classroom discourse. The Journal of Experimental Education, 48, 194-204.

Moon, T.C. (1969). A study of verbal behavior patterns in primary grade classrooms during science activities. Dissertation Abstracts International, 30, 5325A. (University of Microfilms No. 70-9607)

Piaget, J. (1970). Science of education and the psychology of the child. New York: Orion Press.

Pfeiffer, I., & Davis, O.L. (1965). Teacher-made examinations: what kinds of thinking do they demand? National Association of Secondary School Principals, 49, 1-10.

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

Reed, R.L. (1977). Questioning and its implications for educational research. (Report No. SO 011 332). (ERIC Document Reproduction Service No. ED 164 374).

Riley, J.P. (1981). The effects of preservice teacher's cognitive questioning level and redirection on student science achievement. Journal of Research in Science Teaching, 18, 303-309.

Rogers, V.A. (1969). Varying the cognitive levels of classroom questions in elementary social studies: an analysis of the use of questions by student-teachers. Dissertation Abstract International, 30, 1459A-1460A. (University Microfilms No. 69-15, 862)

Ryan, F.L. (1973). Differentiated effects on levels of questioning on student achievement. The Journal of Experimental Education, 41, 63-67.

Sanders, N.M. (1966). Classroom questions: what kinds? New York: Harper and Row.

Savage, T.V. (1972). A study of the relationship of classroom questions and social studies achievement of fifth-grade children. Dissertation Abstracts International, 33, 2245A. (University Microfilms No. 72-78, 661).

Schreiber, J.E. (1967). Teachers' question-asking techniques in social studies. Dissertation Abstracts International, 31, 522A-523A. (University Microfilms No. 67-9099).

Shymansky, J.A., Kyle, W.C., & Alport, J.M. (1983). The effects of new science curricula on student performance. Journal of Research in Science Teaching, 20, (5), 387-404.

Simmons, B. (1976). A linguistic analysis of disadvantaged kindergarten children's verbal responses to questions. Journal of Educational Research, 69, 253-255.

Smith, C.T. (1967). Improving comprehension? That's a good question (Report No. CS 003 559). San Leandro, CA: Livermore California Schools. (ERIC Document Reproduction Service No. ED 142 944)

Smith, C.T. (1974). The relationship between the type of questions, stimuli, and the oral language production of children. Unpublished doctoral dissertation, University of California, Berkeley.

Smith, H.A. (1963). Educational research related to science instruction for the elementary and junior high school: a review and commentary. Journal of Research in Science Teaching, 1, 199-225.

Smith, H.A. (1975). Historical background of elementary science. In E.Victor & M.S. Learner, Readings in Science Education for the Elementary School (pp. 38-44). New York: Macmillan.

Stevens, R. (1912). The question as a measure of efficiency in instruction: a critical study of classroom practice. Teachers College Contributions in Education, 48, 123-127.

Suchman, J.R. (1966). Developing inquiry. Chicago: Science Research Associates.

Taba, H., Levine, S., & Elzey, F. (1964). Thinking in elementary school children (Report No. CRP-1574). San Francisco, CA: San Francisco State College. (ERIC Document Reproduction Service No. ED 003 285)

Underhill, O.E. (1941). The origins and development of elementary school science. Chicago, Atlanta, Dallas: Macmillan.

Wildman, P.R. (1967). The fallacy of facts. Peabody Journal of Education, 44, 177- 180.

Willson, I.A. (1973). Changes in mean levels of thinking in grades 1-8 through use of an interaction analysis system based on Bloom's Taxonomy. The Journal of Educational Research, 66, 423-429.

Wilson, J.H. (1967). Differences between the inquiry-discovery and the traditional approaches to teaching science in elementary schools. Dissertation Abstracts International, 28, 887A. (University Microfilms No. 67-11, 480)

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