1988

Linguistic content analysis of the Holt, Rinehart and Winston series of high school biology textbooks: a longitudinal study focusing on the use of inquiry

Elizabeth Miles Eltinge
Iowa State University

Order Number 8825392

Linguistic content analysis of the Holt, Rinehart and Winston series of high school biology textbooks: A longitudinal study focusing on the use of inquiry

Eltinge, Elizabeth Miles, Ph.D.

Iowa State University, 1988

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Linguistic content analysis of the Holt, Rinehart and Winston series of high school biology textbooks: A longitudinal study focusing on the use of inquiry

by

Elizabeth Miles Eltinge

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Professional Studies in Education

Major: Education (Curriculum and Instructional Technology)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1988

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Statement of Research Problem

For many years the most dominant influence over what is taught in the high school science classroom has been the science textbook. The content of the science taught in the science classroom is predominantly guided, organized and restricted to what is contained within the textbook (Harms and Yager, 1981; Helgeson, Blosser and Howe, 1977; Nelkin, 1977; Stake and Easly, 1978). It is therefore of critical importance to science educators to carefully investigate what is contained in science textbooks.

The ways in which science teachers use textbooks help to determine the order, the examples, and the applications of science topics which are presented to the students. Teacher's questions generally focus on the information contained in the textbook, and students are trained to seek the "right" answers in textbooks. "Science," as presented in classrooms, is quite often limited to the facts and concepts included in the textbook. Science textbooks also tend to emphasize the learning and memorization of a highly specialized terminology. In brief, what a student knows, does or thinks regarding science is limited to two or three commonly used textbooks in a given area of science. In biology, these textbooks have been the yellow and green versions of the Biological Sciences Curriculum Studies (BSCS) and the Holt Modern Biology series. The Holt textbook was the best selling high school textbook from 1945 until the late 1970s (Grabiner and Miller, 1974; Helgeson, Blosser and Howe, 1977; Lowery and Leonard, 1978a; Weiss, 1978). Since the late 1970s the Holt textbook has remained one of the three best selling high school biology textbooks (Joy, 1988).

Since textbooks play such a critical role in the science classroom, it is important to the science educator to try to ascertain exactly what is being presented in them. Are science textbooks accurately representing the goals of science education? Are they consistent with the research and recommendations of science educators?

One of the goals of science education which has been predominant over the past several years is bringing students to a state of scientific inquiry (Herron, 1971). It is argued by many that teaching science as inquiry presents a more realistic view of science (see for example Herron, 1971; Tamir, 1985; Welch, Klopfer, Aikenhead and Robinson, 1981). One of the arguments put forth for teaching science as inquiry is that it is an important content objective. It helps to explain how science functions. Also, students could become frustrated and confused if they view science as a stable, unchanging body of facts and absolute truths, and then later in life they encounter revised versions of these "absolute truths." In addition to being frustrated and confused, students could worse yet become cynical about science, and lose confidence in it. Understanding the nature of science as inquiry can help prepare students to understand critical societal issues involving science and technology. If they are taught not to see science as a massive body of knowledge, too large for anyone to ever memorize, but rather to see it as a changing, growing, developing system of interrelated events, it could help them to better understand critical science and technology issues in todays society. This better understanding could lead to better decision making concerning these issues.

The number of studies focusing on the nature of science as inquiry in textbooks is limited. Of the work that has been done, a few of the more useful studies have used content analysis. Content analysis can be viewed as a systematic quantification, description, and statistical analysis of text, with possible reference to other text-related data (Weber, 1986). The analysis has a particular focus, dependent on the intent of the researcher. Content analysis of textbooks can not be used to describe the actual experiences of learners in a classroom, but rather the opportunities for learning (Tamir and Lunetta, 1981). One method of content analysis, linguistic content analysis, is especially useful in analyzing the inter-relationships of words in a text (Roberts, 1988). This type of study can help to determine the treatment of a topic, such as science shown as a process of inquiry, within a text.

The idea of looking at how science is expressed in textbooks is certainly not new. The ideas expressed by Crowell in 1937 very accurately reflect the nature of this study.

> It is generally agreed that science consists not only of facts which have been verified and classified, but that it is also a method of thinking. It is the belief of the writer that science teachers have concerned themselves unduly with the teaching of facts, to the neglect of science as a method by which the pupil can solve his own problems. Most persons will probably agree that facts have a place in any scheme of education if those facts serve as vehicles for the development of individually and socially worth-while attitudes, appreciations, habits and skills, and are not considered solely as ends in themselves. The writer believes that many of the facts which are at present taught as isolated bits of knowledge might well be taught for the purpose of developing attitudes and skills associated with the scientific method of thinking....It is the conviction of the writer that the topic of scientific method (of inquiry) will not be adequately treated in the classroom until the authors of science textbooks recognize the importance of the subject and build the organization of their books around it. This is true because science textbooks probably determine to a large degree what topics shall be taught and the method by which they are taught (Crowell, 1937, p. 525, parentheses mine).

Over the past three decades, the image of science as inquiry has been a popular theme among experts in the field of science education. Because textbooks play a very critical role in science education it is of interest to investigate whether this image has been incorporated into textbooks during this time period.

Purpose of Study

The purpose of this study was to determine the presence of science as inquiry in high school science textbooks over the past three decades. The science textbook used was a biology textbook. Biology was selected because it is the most frequently taken high school science course (Helgeson, Blosser and Howe, 1977; National Science Board Committee on Science Indicators, 1985; Office of Educational Research and Improvement, 1986). Also, for most students, biology, taken in the tenth grade, is often the last science course they take in high school (Stuart, 1982). Thus, the high school biology classroom is often the last place students have the opportunity to acquire scientific literacy.

This study was a longitudinal study, which allowed for an investigation of how the presence of science as inquiry has changed over time. The data collection phase employed linguistic content analysis techniques. The resulting categorical data were analyzed using logistic regression.

To limit the introduction of extraneous variables, only one series of textbooks was studied. One criterion of selection for the series of textbooks studied was the distribution and use of the textbooks in that series. Assuming that a textbook series with a large market share in high school biology textbooks would have greater impact on opportunities for classroom learning of biology in the United States, it was important to select a series with a strong sales record. Since 1945, the Modern Biology series by Holt, Rinehart and Winston has been a leader in the market in sales of high school biology textbooks (Grabinar and Miller, 1974; Helgeson, Blosser and Howe, 1977; Joy, 1988; Lowery and Leonard, 1978a). A second criterion for selection of the textbook series studied was whether or not the years of publication of textbook editions spanned the time period of interest. To determine the impact of the curriculum movements in science education during the 1960s, a series which had editions published prior to and after the 1960s had to be used. The first publication of the Holt Modern Biology series was in 1921. Holt has continued to publish textbooks for high school biology since that first edition, with the newest edition planned for 1989. Thus, the textbook series which best fit the selection criteria was the Modern Biology series by Holt, Rinehart and Winston.

Limitations of the Study

1. The study is limited to textbooks only. The study does not address how a textbook is actually used in the classroom. The use of the textbook by the teacher, and the relative emphasis on various portions of the textbook could be very important in the portrayal of the image of science.
2. The study only looks at one series of textbooks. Although other materials developed during the 1960s were designed to stress inquiry, it is not within the scope of this study to examine all high school biology textbooks over the past 30 years.
3. The study only looks at three subject areas. Inferences drawn from the study can only apply to the treatment of these three subject areas in the Holt *Modern Biology* textbooks.

Definitions

1. *Content Analysis:* Content analysis is a systematic, quantitative analysis of written material which allows the researcher to make inferences about the text.

2. Linguistic Content Analysis: Linguistic content analysis is a form of content analysis which allows for the study of inter-relationships among words.
3. Science as Inquiry: Science seen as a process of inquiry which demonstrates the methods by which scientific knowledge is obtained.

REVIEW OF THE LITERATURE

Textbooks have played an important role in biology education. To fully appreciate this role, a brief historical review of biology courses and textbooks will be helpful. Therefore, this review of the literature includes a history of biology courses and textbooks, and a discussion of the important role of textbooks in biology education. Following the discussion of the role of textbooks in biology education is a treatment of some of the studies which have been reported on analyzing textbooks. The next topics covered will be the role of inquiry in science education, and previous studies on the treatment of inquiry in science curriculum. The last part of this section will be devoted to a review of content analysis.

History of Biology Textbooks in American Secondary Schools

A major factor influencing high school science curriculum in the last half of the twentieth century was the launching of the satellite Sputnik by the Russians on October 4, 1957. This event resulted in a fear that the Russians would become superior to the Americans in science and technology. This fear produced a demand for more rigorous school curriculum, particularly in the areas of mathematics and science. This demand resulted in the development of several federally funded national curriculum projects in mathematics and science. Three programs were developed in high school biology, all by the Biological Sciences Curriculum Study (BSCS). The three series of textbooks which arose from these programs are the BSCS blue, green and yellow versions. This curriculum reform program of the 1960s has had such a major impact on high school biological science that the following review of biology courses and textbooks will in part be organized around the dates of its development. Specifically, the review will be divided into the two time periods of before 1960 and from 1960 until the present time.

Biology courses and textbooks before 1960

General biology has not always existed as a specific course in the secondary schools. General biology evolved from a combination of three separate, older biological science courses; botany, zoology and physiology (Reynolds, 1967). These three specialized courses in turn emerged from a course known as "natural history". The course, natural history, usually covered all objects which could be classified as animal, vegetable or mineral.

Of the three early science courses, botany was offered the most extensively. Anatomy and classification of flowering plants received a great deal of emphasis in these early classes. Memorization of botanical terms was the principle method of study. During the middle and latter part of the nineteenth century, the emphasis was mainly on description and classification of plants. At the end of the nineteenth century, there was a demand for a more practical approach to the subject. The emphasis shifted from classification to a study of morphology. The emphasis on practical application was also incorporated into the laboratory (Reynolds, 1967).

Zoology started out in the secondary curriculum as a form of natural history. For the most part, zoology in the nineteenth century was a less technical subject than botany. There was emphasis on classification in secondary zoology teaching, as well as comparative anatomy studies. Zoology was quickly absorbed by general biology after the turn of the nineteenth century (Reynolds, 1967).

Human physiology was taught in both the grammar and the secondary schools. Most textbooks were written by physicians, and as a result emphasis was placed on the practical value of the science. It was more common to have separate botany and zoology courses, with physiology being added later. All three of the courses were patterned after college courses (Reynolds, 1967).

By the end of the nineteenth century, the new subject of general biology was in the American secondary schools, and, to a large extent, it took the place of the older biological sciences. Eventually, by 1930, general biology was established as the science course most commonly studied during the tenth grade.

The early biology text materials were almost entirely descriptive in nature. The texts were largely informational and utilitarian. There were virtually no suggestions for pupil activities in the textbooks until after 1850. These early textbooks were generally very small, with small print and few pictures. The writing style followed the deductive approach, and memorization of the material was often implied. In general, the early texts required little pupil activity other than memorization and formal application of rules (Nietz, 1966).

The early general biology textbook evolved from botany, zoology and physiology textbooks, much as the course, general biology, evolved from early botany, zoology and physiology courses. The organization of general biology textbooks contained separate, abbreviated courses in these three areas. The first biology textbook to present a blend of these three areas was probably Needham's <u>General Biology</u> published in 1910. In other instances, the tripartite arrangement of biology textbooks lasted until the 1920s. During the period of 1907 to 1935, many general biology texts were organized around everyday life, and moved away from this tripartite arrangement toward a unified arrangement. Some of the issues addressed by books of that time dealt with improvement of crops, conservation of biological resources and other agricultural problems. A second type of organization which emerged was of a physiological background. With this arrangement the human body was stressed, and organization centered around human welfare (Reynolds, 1967).

The first publication of what was to become one of the most widely used high school biology textbook series in America appeared in 1921. This was Truman Moon's Biology for Beginners published by Holt. The Holt series did not immediately dominate the market however. The most widely used text in the late 1920s appears to have been Smallwood, Reveley and Bailey's New General Biology (New York: Allyn and Bacon, 1929) (Grabiner and Miller, 1974). The most widely used text in the 1930s was Arthur O. Baker and Lewis H. Mills' Dynamic Biology (New York: Rand McNally, 1933). Dynamic Biology remained the leading book in the field until 1945. During the period from 1945 to 1960 the most widely used text was Truman Moon and Paul Mann's Biology: A

Revision of Biology for Beginners (New York: Holt, 1941) and later Moon, Mann and Otto's Modern Biology (New York: Holt, 1947, 1951, 1956, 1958) (Grabiner and Miller, 1974). Subsequent editions of the Holt textbook remained the dominant biology textbook on the market from until the late 1970s (Helgeson, Blosser and Howe, 1977; Lowery and Leonard, 1978a; Weiss, 1978), and one of the three leading textbooks in the 1980s (Joy, 1988).

Biology courses and textbooks after 1960

The time following and including 1960 can be divided into two broad periods. The first being the early sixties, immediately post-Sputnik, from 1960 to about 1975. The second period would be from 1975 to the current time.

The early 1960s was a very important time for all of science education. During this time period a great deal of time, effort and money was spent on improving science education. New curricula were developed for chemistry (CBA, CHEMS), physics (PSSC) and biology (BSCS). The impact of the BSCS materials on biology education has been very broad (Grobman, 1969; Hurd, Bybee, Kahle and Yager, 1980). The development of the BSCS courses brought together a mixture of scientists and educators with the common goal of improving biology education in the United States. The start of the study involved a careful study of the current state of education and a restructuring of the principle areas of focus. One of the central parts of the BSCS program was an emphasis on inquiry (Grobman, 1969).

Since the development of these curricula, there have been no movements of a similar magnitude in biology education. Some contend that the goals and objectives which were established for the 1960s are no longer suitable for the 1980s (Hickman and Kahle, 1982; Hofstein and Yager, 1982). These authors state that during the early 1960s the goals of science education were oriented to the immediate past and present. Americans wanted to increase the supply of highly trained scientists and engineers. They state that one of the goals of science education at that time was to increase the number of students planning to become scientists and engineers. The courses were designed to make students in the classroom and laboratory act much as a scientist was perceived to act at work. The criterion for selection of material to be included in the program was what was perceived as being basic to the discipline. By the early 1970s, it appeared as if some of the objectives of the 1960s had been met. There was a surplus of scientists and engineers in this country, and the United States had made major advances past the Russians in various space projects, including landing the first man on the moon. However, dissatisfaction with societal issues, such as the Vietnam War, led to the notion that a more rigorous science curriculum was not helping to solve the present societal issues and concerns (Hofstein and Yager, 1982).

The start of the final period, 1975 to the present, was marked with growing dissatisfaction with the discipline-centered science curricula of the 1960s. Science and technology are viewed as means for improving society. There are those that believe science should be taught with the purpose of helping students to cope in an increasingly technological society, not with the sole purpose of creating more scientists and engineers. Science is not viewed as being value-free, but rather as being value-laden. The interface of science, technology and society has become a focus of science education in the 1980s, along with all of the moral and ethical issues it includes (Hofstein and Yager, 1982).

Perhaps one way to view the historical changes in biology education during the last half of this century is to consider the goals of biology education. Biology programs have been structured on basically five goals. These goals can by roughly described as biological knowledge, scientific methods, social issues, personal needs and career preparation. Statements which might better describe what has historically been incorporated into these goals are as follows:

- Biology education should develop a fundamental understanding of biological systems.
- Biology education should develop a fundamental understanding of, and ability to use, the methods of scientific investigation.
- Biology education should prepare citizens to make responsible decisions regarding science-related social issues.
- Biology education should contribute to an understanding and fulfillment of personal needs and thus contribute to the development of individuals.
- Biological education should inform students about careers in the biological sciences (Hurd, Bybee, Kahle and Yager, 1980, p. 391).

Changes which have occurred in science education can be viewed not so much as changes in these five goals, but as changes in the relative emphasis placed on these five goals. The goal of achieving biological knowledge has traditionally received the greatest emphasis. The goal of scientific methods, which includes inquiry processes, was emphasized in the 1960s, but has received less emphasis in the 1980s. In its place, the goal of social issues has received greater attention (Hickman and Kahle, 1982; Hurd, Bybee, Kahle and Yager, 1980).

The role of textbooks in biology education

The textbook is the most widely used of all teaching aids. Few teachers would attempt to teach a course without one. In science courses, the textbook has historically been closely related to both the content and the methods of the secondary school course (Reynolds, 1967; Hurd, Bybee, Kahle and Yager, 1980; Helgeson, Blosser and Howe, 1977; Nelkin, 1977; Stake and Easley, 1978; Weiss, 1978). Over 90% of all science teachers use a textbook 90 to 95% of the time (Stake and Easley, 1978). This close relationship between the textbook and the course is particularly true in biology courses.

Three major National Science Foundation (NSF) studies conducted during the 1970s all indicated the great dependence of science teachers on the textbook (Helgeson, Blosser and Howe, 1977; Stake and Easley, 1978; and Weiss, 1978). Yager (1983) presents a consolidation of some of the facts and insights on textbooks gleaned from these massive reports. Among these findings are that textbooks determine the order, the examples, and the applications of science topics that students experience. Science, as presented to the students, is limited to the facts and concepts included in the textbook. Yager also concluded that a relatively small number of textbooks actually determine the science curriculum in the nation's schools, and that most instruction within a single classroom is based on the information in a single textbook. As mentioned previously, one of the most frequently used single textbooks has been the Holt Modern Biology series in high school biology classes.

Teacher's questions, both for discussion and for testing, focus on information in the textbook. Students are trained to seek the "right" answers in textbooks. Other teaching aids, such as audio-visual materials supplement the textbook, and at times merely prove to be a means to present the same information as that in the textbook. Textbooks tend to focus on the disciplines of science. Science textbooks also place a great deal of emphasis on words and specialized terminology. In brief, what a student knows, does, or thinks regarding science in an American classroom can be approached by reviewing two or three commonly used textbooks in a given discipline at a given level of instruction (Yager, 1983).

The content of the textbook obviously has a great impact on what is taught in the science classroom. However, despite the obvious importance of science textbooks in the classroom, comparatively little work has been done on analyzing their content (Reynolds, 1967; Tamir, 1985). It will be helpful now to turn to a review of previous work in the field of textbook analysis.

Analysis of Textbooks

Many different attributes can be considered in an analysis of textbooks. Attributes of texts which are considered for analysis depend on the purpose for performing the textbook analysis. Very often textbooks are analyzed as part of a process of selection or adoption for use in the classroom. Attributes which might be considered for such purposes are durability, cost, appearance, organization, content, readability, and even weight of the book (Doran and Sheard, 1974; St. Laurence, 1951; Vogel, 1951). Other issues of concern prior to textbook adoption might include treatment of women and minorities, treatment of controversial social issues, such as evolution, quality of teacher aids, and quality and use of illustrations (Neie, 1979; Newton, 1984; Thomas, 1978).

Other reasons for analyzing textbooks could be to focus on a particular research interest. Attributes selected for analysis based on research interests may or may not be included in an analysis for the purpose of textbook adoption. Some of these attributes could include cognitive demands of textbooks (Hartford and Good, 1976), misconceptions presented in textbooks (Cho, Kahle and Nordland, 1985), analysis of science concepts (Vachon and Haney, 1983; Yost, 1973), use of mathematics (Pratt, 1985), use of analogies (Curtis and Reigeluth, 1984), as well as treatment of controversial social issues (Levin and Lindbeck, 1979; Rosenthal, 1984).

A review of previous work on science textbook analysis follows. It is broken into the following categories: analyzing terminology and readability; analyzing concepts; analyzing objectives and questions; and the treatment of science as inquiry.

Analyzing terminology and readability in textbooks

Textbooks are the central feature in most science classrooms, and terminology is the central feature in most science textbooks. Yager (1983) counted the number of special/technical words on every tenth page of 25 widely used K-12 science textbooks. Based in these data, he concluded that a typical science course requires students to master more new words than would typically be required in learning a foreign language. The number of new words encountered in science alone often approaches the total number of new words that a student could be expected to learn in a given year. Science textbooks then seem to demand more than a fair-share in terms of vocabulary development. In addition, most of the terms in a science textbook have little meaning in terms of meeting personal needs of students. Yager and Yager (1984) studied the effect of schooling on mastery of eight selected science vocabulary terms. They randomly sampled students in grades 3, 7, and 11 from a single school district in the midwest. When questioned on the meanings of the selected science terms, the third grade students performed as well as the eleventh grade students on three of the eight terms. The seventh grade students performed almost equally to the eleventh grade students on all eight of the terms. Yager and Yager suggest that the current emphasis upon knowledge acquisition and vocabulary development is ineffective. This study suggests that the use of curricula emphasizing term mastery is meeting with limited success.

Textbooks are also analyzed for their levels of readability. Readability is a measure of the style of writing of a text. It is generally intended to be used as a measure of the difficulty in comprehending a given piece of text. There are currently about 100 different readability formulas, but most of the popular formulas have a few common elements. These common elements are: average number of syllables per word; average number of words per sentence; percentage of short words; percentage of long words; and percentage of words which are found on a standard word list (Dukes and Kelly, 1979).

Readability formulas often assign grade levels as indicators of the level of difficulty, however, some authors caution that readability is a continuous variable, and as such, assigning specific cut-off points for grade levels is not entirely meaningful (Dukes and Kelly, 1979). Readability studies have been performed on physics textbooks (Dukes and Kelly, 1979), biology textbooks (Gould, 1977; Kennedy, 1979; Wright, 1982), and elementary science textbooks (Shymansky and Yore, 1979). Some studies find a given reading level to be more or less appropriate for an intended grade level (Dukes and Kelly, 1979; Kennedy, 1979; and Shymansky and Yore, 1979), while other studies contend that the reading levels are too high for an intended grade level (see for example Wright, 1982). Quite often it is difficult to draw meaningful conclusions from a readability study because the variation within the sample is high (Shymansky and Yore, 1979).

One study investigated the effect of lowering the readability level of a high school biology text on comprehension and biology achievement (Wright, 1982). A portion of a high school biology textbook was rewritten to a lower readability level. Students were measured both on comprehension of the text and on achievement. Comprehension was measured by the cloze test, a technique in which every fifth word of a piece of written material is deleted, and subjects are instructed to supply the missing word. Achievement was measured with end-of-unit exams. The results showed that reducing the readability of the text material increased comprehension, but had virtually no effect on achievement. Lack of effect was partially attributed to the technical difficulty of the material (genetics) and to the fact that while comprehension was raised by a statistically significant amount, it still did not reach a very high overall level.

In some cases, efforts to make textbooks obtain a lower readability score result in confusing and sometimes inaccurate texts (Fiske, 1984). This confusion in the text results from creating short, choppy sentences, without important casual connectives between main ideas. To date, the studies on readability seem to have many problems, and the results are far from conclusive.

Analyzing concepts in textbooks

Analyzing the content of textbooks can be very helpful to educational researchers. It can help to determine the relative emphasis given to various areas of interest. It can also assist people in determining which textbooks to adopt, and for which grade level a given textbook might be appropriate. One useful way to analyze content is to divide it into concepts. This can provide a more economical means of handling the information. However, there is disagreement about how a concept is to be defined (Finley, 1981). To add to the confusion, many other terms are used synonymously with concept (e.g., theme, major idea, unifying thread, etc.). Despite these drawbacks, efforts have been made to analyze biology textbooks by concepts, and to study treatment of specific concepts, particularly evolution, in textbooks.

Stuart (1982) analyzed the content of the fifteen most widely used biology textbooks for grades 7-12. The purpose of his study was to survey the content of the textbooks, to identify basic concepts by textbook source, and to determine the overall extent to which each concept was covered. Stuart used a classification scheme developed and validated in two earlier studies. The scheme divided concepts into six broad areas: physiology, morphology, genetics, evolution, ecology, and applied biology. Each page of the 15 textbooks was read. Written material outside of the main body of the text, such as captions and special laboratory instructions, were excluded from analysis. The results show that in grades 7-12, the concepts of ecology and physiology received the most emphasis, and evolution received the least.

Cho and Kahle (1984) went beyond content analysis, and tried to relate textbook content to student achievement, as well as investigating the impact of national recommendations on textbook content. The measurement of student achievement was taken from the 1977 survey of the National Assessment of Educational Progress. Textbook content was assessed by grouping objectives from the three most widely used high school biology textbooks in 1973 into ten conceptual areas. The same grouping of concepts was performed in 1983 textbooks to determine the impact of the recommendations from Project Synthesis, a major national study on science education whose results and recommendations were published in 1981. Results showed a direct linear relationship between achievement level in a concept and emphasis of that concept in biology textbooks. It was also concluded that in general, the recommendations of Project Synthesis were not reflected in the newer textbooks. This lack of evidence of the 1981 recommendations in 1983 textbooks could have been due to the relative short time span between when the recommendations were released and the textbooks were published. This could be a key factor, especially considering the long time period required for preparation of a textbook.

The contents of science textbooks have also been analyzed as to their coverage of controversial issues. With the rising importance of science and technology in our society, many feel it is important that societal issues involving science be included in textbooks. Levin and Lindbeck (1979) conducted a study whose purpose was to determine the coverage of controversial issues and biosocial problems in selected biology textbooks. Five high school biology textbooks published in 1973 were chosen for the study. The researchers then generated a list of 11 controversial issues and biosocial problems and counted the number of pages, to the nearest tenth of a page, devoted to these issues. These page counts were then expressed as a percentage of the total pages in the textbook. The authors also gave qualitative ratings to the quality of the treatment given to controversial issues. Conclusions were that textbooks differ considerably in their treatments of controversial issues, and no one textbook gave adequate treatment to all issues and problems.

In an extension to the work by Levin and Lindbeck, Rosenthal (1984) studied the treatment of controversial issues in high school biology textbooks published between 1963 and 1983. She developed a list of 12 social issues, based on a study of the literature. She also developed reliability and validity measures on her classification scheme. Both her reliability and validity values were within an acceptable range. Using her classification scheme, Rosenthal tabulated amount of space, to the nearest tenth of a page, devoted to social issues, and expressed it as a percentage of the total length of the text. Her results demonstrated that attention to social issues in high school biology textbooks has decreased between 1963 and 1983.

Several studies have focused on the treatment of one particular controversial issue, that of evolution. It has been hypothesized that pressure from vocal, anti-evolution groups have caused publishers to avoid this controversial topic in textbooks (Skoog, 1979, 1984).

Grabiner and Miller (1974) studied the effects of the Scopes trial in 1925 on treatment of evolution in high school biology textbooks. The Scopes trial involved the prosecution of John Thomas Scopes for violating Tennessee's law against teaching the theory of evolution. Scopes was found guilty, but the verdict was later reversed on a technicality. It is usually held that the pro-evolution forces were favored in the popular public opinion. Grabiner and Miller examined the treatment of evolution in high school biology textbooks before and after the Scopes trial. They concluded that the treatment of evolution in high school biology textbooks declined after the Scopes trial. The treatment continued to decline until the new Biological Sciences Curriculum Study (BSCS) texts were published in the 1960s.

Skoog performed a massive study of treatment of evolution in high school biology textbooks (Skoog 1979, 1984). He studied 105 high school biology textbooks published between 1900 and 1983. He determined the number of words devoted to evolution for each textbook. Skoog concluded that prior to 1960, evolution was treated in textbooks in a cursory and generally noncontroversial manner. He saw an increase in treatment from 1900 to 1950, then a decrease in treatment in the 1950s. In the 1960s, under the influence of the Biological Sciences Curriculum Studies, there was a great increase in the treatment of evolution. However, the trend has decreased again during the 1970s and 1980s. Skoog attributes part of the current de-emphasis to the persistent anti-evolutionist pressures.

Rosenthal (1985) also performed a longitudinal study on the treatment of evolution in high school biology textbooks (see also Beard, 1986). She studied 22 textbooks published between 1963 and 1983. Rosenthal calculated the percentage of text devoted to evolution, rather than the word counts used by Skoog. She also found a decrease in attention to evolution, though not as dramatic in certain textbooks as in Skoog's studies.

Analyzing objectives and questions in textbooks

Various researchers have worked on categorizing objectives and questions found in text materials. The purpose of such categorizations is generally to try to ascertain the cognitive level of activities in text materials. Some researchers have used the "Taxonomy of Educational Objectives: Cognitive Domain" developed by Bloom and others (Reynolds, 1967; Scott, 1972).

Bloom and his colleagues developed the taxonomy of educational objectives to help curriculum builders specify objectives, in order to make it easier to plan diverse learning experiences and to prepare evaluation devices. The five levels of the taxonomy, ranging from lowest cognitive order to highest cognitive order are: knowledge; comprehension; application; analysis; and evaluation. The levels are meant to represent increasing amounts of cognitive activity for the student, with each subsequent level building on the previous one. The taxonomy thus allows the curriculum builder a classification scheme of objectives based on perceived cognitive demands on the student (Bloom, Engelhart, Furst, Hill and Krathwohl, 1956).

The Bloom taxonomy, which has the advantage of being well known and recognized among educational researchers, also has several disadvantages when applied to analyses of written materials. One disadvantage is that the taxonomy was originally designed for use in the construction of test questions and objectives; it was not designed for analyzing written material. Another problem in using the taxonomy with written text materials is the unknown variable of students' past experience. The level of knowledge a student has prior to answering a question would determine their level of cognitive activity. What could be an activity at the "synthesis" level for one student could merely be an activity at the "knowledge" level for another student. A third problem with using Bloom's taxonomy is the variable of teacher technique. The amount of prompting, guidance and direction given by a teacher could alter the intended level of cognitive activity. Scott (1972) has proposed a method of helping to address some of these problems. He proposed rating activities with a span of levels, from the lowest clearly intended cognitive level required to a higher level which, given the circumstances, could be required. Scott has used the technique in a curriculum study to help alleviate the afore mentioned problems.

The Bloom taxonomy was also used by Reynolds (1967) in a study of end-of-chapter questions in high school biology textbooks. Reynolds studied four textbooks in popular use at the time. He classified questions at the end of each chapter according to Bloom's taxonomy. He also categorized questions according to the form of the question. His results show that knowledge level questions accounted for 77% of all review questions. Of the remaining five categories, 1% of the questions were comprehension, and 8% of the questions were either application, analysis, synthesis or evaluation. The form of most of the questions was that of short answer.

Other scales have also been developed for ranking the cognitive level of text materials (Nicely, 1980-81, 1985a, 1985b; Lowery and Leonard, 1978a, 1978b; Stuart and Burns, 1984; Feezel, 1985). Most of these scales follow the same theory as that upon which Bloom's taxonomy was developed, in that they assign an ordered ranking to the objectives or question, ranging from lower level thinking skills, to higher level thinking skills.

Lowery and Leonard have presented the argument that inquiry behavior is very desirable in children, and that questions can be effective in eliciting inquiry behavior (Lowery and Leonard, 1978a, 1978b; Leonard and Lowery, 1984). They further argue that because inquiry is central to science education and the textbook is the basis for most science instruction, textbooks should be examined for their ability to stimulate students to inquiry by examining the questions they contain. They developed an instrument called "Textbook Questioning Strategies Assessment Instrument" (or TQSAI) to categorize questions into specified types. They also recorded the frequency of the various types of questions and the placement of the questions. Lowery and Leonard then used this instrument to perform an analysis of the four most popular high school biology textbooks of 1976. They found significant differences in the number of questions asked, types of questions asked, and placement of questions within the book. The meaning of these differences in not entirely clear. For example, research has not shown which placement of questions is the best, or what frequency of questions is most desirable. However, one conclusion which can be drawn is that most of the questions were "lower cognitive order" questions. If one values either more "higher cognitive order" questions, or a balance of both higher and lower order questions, all of the textbooks studied could be improved.

Nicely (1980-81, 1985a, 1985b) developed an instrument for analyzing mathematics textbooks. The major purpose of his instrument is to aid in decision making about textbook selection. His instrument has four categories of classification; type of content, level of cognitive activity, stage of mastery and mode of response. Perhaps the most generalizable of these four scales to science textbooks is level of cognitive activity. This scale contains 27 verbs grouped into nine categories and arranged in an ordinal scale. End-of-unit activities in mathematics textbooks were coded according to this scale. Nicely has used the instrument on mathematics textbooks published from 1961 to 1980. His results indicate that in general, textbooks do not stress problem solving and higher order skills. Most responses were coded at a very low level of cognitive activity, and no books contained responses at the very highest level of behavior, that of evaluation. He further reports that from 1961 to 1980, there was an increase in lower order skills, and a decrease in higher order skills.

Science as inquiry

Terms such as "scientific inquiry," "the nature of inquiry" and "inquiry skills" have been prominent in science education literature, particularly during the curriculum movement of the early 1960s. Goals for science education have often included representing science as a process of inquiry, and encouraging students to obtain those skills deemed necessary to participate in science as a process of inquiry. Yet, despite the obvious importance science educators have placed on science as inquiry, little evidence exists that this importance has been incorporated into teaching practices (Tamir, 1983, 1985; Herron, 1971; Hurd, Bybee, Kahle and Yager, 1980). Most science goal statements issued by state education agencies, those which teachers must follow, are more content-oriented rather than inquiry-related (Welch, Klopfer, Aikenhead and Robinson, 1981). Teachers often feel more responsibility for teaching facts. They also feel much more prepared to teach facts rather than to teach inquiry (Duschl, 1986).

Part of this reluctance to address the issue of inquiry could be a confusion as to what is actually meant by inquiry. In an effort to help alleviate this confusion, Tamir (1983, 1985) falls back on the dual nature of "teaching science as inquiry" as described originally by

Schwab. "Teaching science as inquiry" is seen as consisting of two separate, identifiable parts. One part is "teaching by inquiry" and the other is "science as inquiry." These are best viewed as the process and the product of what might occur in a science classroom.

Teaching by inquiry involves the means by which the students gain knowledge. It includes the development of the so-called inquiry skills, such as the abilities to identify and define a problem, to formulate a hypothesis, to design an experiment, to collect, analyze and interpret data and to draw meaningful conclusions. Hands-on, experiential laboratory activities which encourage creative student activities are examples of what could be teaching by inquiry.

Science as inquiry deals with the image of science. The image of science should be put forth as realistically as possible. Science is both a collection of facts, and a method by which those facts are obtained. The scientific method of inquiry shows science as being fallible, self-corrective and progressive, rather than infallible and conservative. It is skeptical rather than dogmatic, and constantly undergoing change (Lampkin, 1951). When viewed as a process of inquiry, science is seen not as the final truth about the world, but rather as the most adequate account of the world at any given time (Connelly et al. as found in Tamir, 1985).

Scientific knowledge is a dynamic, not a stable entity. If students' exposure to science is primarily as a body of static truths, their image of science will be distorted. Disagreement among scientists and conflicts of "scientific truths" could lead to confusion. For example, confusion could result if what students learned about atomic structure in the fifth grade became obsolete by the time they graduated from high school. Students could become cynical and skeptical about science if they are only presented with facts about scientific explanations, such as those describing atomic structure. Distorted views of what science is would still arise even if the facts learned at different times were never in conflict with one another. Science is both a way of obtaining information as well as a body of information. A representation of science as simply a body of information is a distortion of the true image of science. In contrast, students could be taught about the inquiry processes which led to one set of conclusions at one point in time, and possibly to a revised set of conclusions as more knowledge and information are gathered and analyzed. This would make the changing of theories and ideas to accommodate new knowledge a reasonable and understandable process, rather than a sudden shift in absolute truths.

Assessing inquiry in written materials

Despite the importance of both inquiry and textbooks in science education, very little work has been done on assessing the inquiry in textbooks (Tamir, 1985; Herron, 1971). Of the work which has been done, most of it can be categorized as "teaching by inquiry" rather than "science as inquiry." Yet, Schwab has stated that of the two, science as inquiry is the most important in science education (Tamir, 1985). The following four studies are examples of research done in the category teaching (or learning) by inquiry.

Tamir and Lunetta (1978) developed and used an instrument for analyzing inquiry in laboratory investigations. They argue that the laboratory is one of the easiest, most obvious places to convey inquiry. The instrument used in this study contains 16 items based on task analysis, and 4 items to indicate the degree of integration of the laboratory work with the rest of the course. The first 16 items listed activities which would be associated with learning by inquiry, such as recognizing and defining problems, formulating hypotheses and recording results. They used this instrument to analyze the 70 laboratory activities contained in the third edition of the BSCS yellow series. For a given laboratory activity, a category in the 16 item task analysis was either recorded as being present or absent. The frequency of occurrence was not recorded. The researchers concluded that the BSCS yellow version had a suitable amount of laboratory work integrated into the text. However, they did not feel students were given enough opportunities to identify and formulate problems or to design experiments. In addition, they noted a lack of explicit instructions to teachers for post-laboratory discussions.

In an extension of their previous work, Tamir and Lunetta (1981) used a modified version of their previous instrument for analyzing inquiry in laboratory activities. The new version, called the Laboratory Structures and Task Analysis Inventory (or LAI) is longer than the previous form. It has two main categories, one devoted to laboratory organization and one to laboratory tasks. The section on laboratory organization has 14 categories divided into four sub-sections, while the section on laboratory tasks has 24 categories divided into four sub-sections. As before, only the presence or absence of an item was recorded, not the frequency. The authors did a comparison of the LAI scoring on the laboratory activities in one biology textbook (BSCS yellow), two physics textbooks and five chemistry textbooks. Their results indicated a stronger orientation for inquiry in the biology textbook. However, it should be noted that they only used one biology textbook, and that one was written to encourage inquiry activities. Also, the instrument used to analyze the biology textbook was an earlier form of the LAI, they did not repeat the analysis with the newer version. In general however, they concluded that most of the laboratory investigations in all of the textbooks examined were highly structured. The students were not provided with opportunities to investigate and inquire. These findings are in contrast with the goals put forth by science educators and curriculum developers.

Tamir, Nussinovitz and Friedler (1982) developed an instrument called the Practical Tests Assessment Inventory (or PTAI). This instrument is designed to assess student responses on inquiry-oriented practical laboratory examinations. There are 16 items, and student responses are ranked from high to low within each item. For example, one category is "formulating problems." Student responses are given a score of "one" if no problem is formulated, a score of "two" if a nonrelevant problem is stated, and so forth, on up to a score of "six" if a relevant problem is formulated as a question. The test has been used in Israel with twelfth grade students who take the biology matriculation examination. Each school then obtains a detailed report of the results in terms of the PTAI categories and ranks. These reports provide helpful feedback for the teachers trying to stress inquiry skills in their science teaching.

Tafoya, Sunal and Knecht (1980) developed an instrument to assess inquiry potential in elementary curriculum. Their purpose was to develop a tool to aid curriculum decision makers in selecting textbooks which promote inquiry skills. The Assessment of Inquiry Potential (or AIP) is an analytic tool which is claimed to give reproducible results in classifying science materials on the basis of their potential for inquiry learning. Sentences are randomly selected from written material and analyzed with the two step process of the AIP. The first step classifies sentences according to the function they serve. In the second step, knowledge claims are categorized according to the manner in which they are verified. The AIP is not intended to be used for making value judgements, but to rank curriculum materials as to their potential contribution to obtaining inquiry skills.

One final study on assessing inquiry in written material can be classified as assessing science as inquiry (Tamir, 1985). In this study five paragraphs were randomly selected from a chapter in a science textbook. Each sentence in a paragraph was categorized as either "rhetoric of conclusions" or as "narrative of inquiry." A sentence classified as "rhetoric of conclusions" is one which presents information as final and absolute without referring to the ways, procedures and ideas which lead to the establishment of this information. A sentence classified as "narrative of inquiry" is one which presents conclusions in the framework of which they arise and are tested. It would include explanation of the problems posed, experiments performed, difficulties encountered, rival interpretations and doubts and limitations of the conclusions. Once the sentences have been classified in one of the two categories, a ratio called the index of "science as inquiry" is calculated by dividing the number of "narrative of inquiry" statements by the number of "rhetoric of conclusions" statements.

To provide more detail about the specific components of the narrative of inquiry, a 23 item categorization scheme was developed. The 23 items, shown in Table 1, list examples of what might appear in a narrative of inquiry, such as information being presented as being tentative and incomplete, names of researchers being mentioned, problems and hypotheses being formulated and experiments being described. In Tamir's study, chapters were the units of analysis, and tally marks were recorded on the scale for each of the 23 items found in the chapter. The maximum number of marks used was two. That is, once an item had appeared twice in a chapter, any further appearances were not recorded. Tamir's instrument allows for a more detailed report of the ways in which the concept of "science as inquiry" is being presented in a textbook.

Content Analysis

Content analysis is a recently developed research technique. The first work in content analysis was conducted in the late 1930s by Harold Lasswell and Paul Lazarsfeld. The purpose of their work was primarily to determine the effects of radio on public opinion. Later, during World War II, Lasswell, Lazarsfeld and others set up various research organizations which worked on refining the techniques of content analysis. Their studies typically involved the analysis of World War II propaganda. By the late 1940s and early 1950s texts on content analysis were starting to appear, mainly as a result of the work on war propaganda projects (Roberts, 1988).

The object of content analysis is that of description. In particular, content analyses of textbooks have not provided direct data on student growth. What they do allow,

TABLE 1. Indicators of science expressed as inquiry (Tamir, 1985, p. 91)

1. Information is presented as tentative and incomplete.
2. Doubts are raised, validity is tested.
3. Controversial opinions are presented.
4. History of ideas and of discoveries is described.
5. One scientific method is implied.
6. Different scientists use different methods.
7. Facts depend on guiding conceptions of scientists.
8. Names of researchers are mentioned.
9. Personal and social backgrounds of researchers are described.
10. Contribution of technology to research is described.
11. Questions are raised.
12. Problems are formulated.
13. Hypotheses are formulated.
14. Predictions are made.
15. Observations and measurements are described.
16. Data is (sic) presented.
17. Data is (sic) interpreted, conclusions are drawn.
18. Experiments are described.
19. Explanations are presented.
20. Assumptions and limitations are mentioned.
21. Tables are presented and interpreted.
22. Graphs are presented and interpreted.
23. Pictures/Drawings are presented and interpreted.

however, is a comparison of the actual text materials and the stated goals of curriculum developers and science educators (Tamir and Lunetta, 1981). Content analysis can also provide an indicator of the state of beliefs, values, or ideologies expressed in written text (Weber, 1986).

Content analysis has several advantages as a data gathering technique. Communication is a central aspect of social interaction and content analysis can operate directly on the written forms of communication. Also, since written documents can be preserved for long periods of time, the researcher is afforded ready access to data for longitudinal (and other) studies. Content analysis is an unobtrusive measure, eliminating the possibility of unintended "instrument effects." That is, performing content analysis on a written document will have no effect on its original author or its intended audience; there is not a problem of interaction with the data gathering device and the data source (Weber, 1986).

One of the central ideas of content analysis is reducing the many words of the original text into fewer categories (Weber, 1986). Of course some information will inevitably be lost from the original text with this process. The researcher must decide how to perform this reduction, and choose the criteria for deciding what is important to preserve versus what is not critical to the research questions.

Content analyses which are highly quantitative have the advantage of high reliability. An example would be counting the occurrences of a certain word, such as "evolution" to determine the extent of the treatment of the topic of evolution in a textbook. This type of analysis could have high inter-coder reliability and could be performed fairly easily with a computer. However, trying to draw inferences on the treatment of the topic of evolution in a textbook based on word counts alone could be difficult. For example, materials attacking the "evils of evolution" in favor of a creationist approach could have more occurrences of the word "evolution" than materials describing a particular detail of the theory of evolution. Trying to infer meaning on word counts taken out of context could lead to erroneous conclusions.

A highly qualitative form of content analysis could help to avoid some of the problems of highly quantitative studies. Using the illustration above, a more qualitative content analysis study would analyze the settings in which the term "evolution" was found. The researcher would attempt to get a broader, overall perspective of the intended meaning of the writer. Highly qualitative studies would avoid extensive, over-simplification of the data. However, highly qualitative studies run the risk of being too subjective, and having very low replicability. The research reported here uses a linguistic content analysis technique which maintains a balance in both quantitative and qualitative aspects. The technique affords data that have high inter-rater agreement as well as contextual integrity.

Program for linguistic content analysis

"Program for Linguistic Content Analysis" (PLCA) is the software tool used in generating the data for this study. The program allows for classification of words into meaning categories, as well as a means for representing the interrelations among words (Roberts, 1988). Individual clauses of a text are manually coded into a microcomputer, generating a numerical data matrix. (See Appendix B for an illustration of a PLCA data matrix.) The data matrix is then used for subsequent statistical analyses.

One of the first steps in using PLCA is the development of a dictionary to be used by the program. The dictionary consists of numerical codes for frequently used verbs and nouns. The user creates the dictionary based on the subject matter to be coded and the purposes of the study. Appendix A contains the dictionary created for this study.

The unique feature of this program is not the generation of a data matrix (which could be developed almost as easily without a computer), but the ability of the program to allow for a check of the face validity of the coded text. The program provides the user with an immediate "translation" of the clause being coded. If the clause has been coded correctly, the translation, though different in style, should reflect the original relationships of the words in the text being coded. For example, a PLCA translation of the sentence, "Science is not limited by national boundaries," is "The NATIONAL BOUNDARIES do(es) not DISCOURAGE/IMPEDE a(n) SCIENCE." In the translation, the words in all upper case letters are from the dictionary, the rest of the words are generated by the program. Appendix C contains more examples of translated clauses generated from the matrix in Appendix B. If the coder is not satisfied with the translation, the phrase can be recoded until an accurate representation is achieved.

Thus PLCA allows for a quantitative description of text, which identifies relations among words and their classification into meaning categories. It is also qualitative in nature in that coders subjectively evaluate the relations among words as they are believed to have been intended by their authors. Finally, it allows for a check of face validity by providing a translation of the coded material.

Summary

Textbooks have played a vital role in biology classrooms in the past. They play a vital role in current biology classrooms, and one can anticipate that they will continue to determine a large part of what will occur in future biology classrooms. For many, the science textbook is the science curriculum. Whether or not it is the most desirable situation, teacher dependence on textbooks is strong and not likely to change in the near future. It is therefore critical to determine what exactly is contained within the pages of this paramount of teaching aids, the science textbook.

A variety of studies have been performed on science textbooks. Studies on terminology indicate that textbooks are too dependent on vocabulary. Readability research draws conflicting conclusions. Some studies conclude textbooks have appropriate levels of readability, while others conclude that the levels of readability are too high. However, most studies also have the problem of a wide variance within sampling units, which may cause problems in drawing conclusive results.

Treatment of various concepts in science textbooks has also been examined. In general these studies demonstrate that societal issues receive sparse attention in textbooks. Studies focussing on the treatment of evolution note fluctuations over time in the amount of space given to evolution in biology textbooks. These fluctuations seem to have been associated with various social attitudes on the appropriateness of teaching evolution in the schools.

Other researchers have tried to measure the degree of higher order thinking skills present in textbooks by ranking end-of-chapter questions on hierarchical scales. All of the studies discussed here report a disturbingly low number of higher order thinking skill questions. These results do not argue well for the textbook being a good source for developing higher order thinking skills, a topic which has been prevalent in the science education literature for several years.

Another prevalent topic in science education literature has been teaching science as a process of inquiry. This concept can be divided into the two sub-concepts of teaching/learning by inquiry and science expressed as a method of inquiry, the process and product respectively of teaching science as a process of inquiry.. Few studies have been done on analyzing this important concept in textbooks, and of those few, most have focussed on teaching/learning by inquiry and not on science expressed as a method of inquiry. This second sub-concept, science as a method of inquiry, is viewed by some as being the more important of the two for science teaching. There is thus a need for more work in analyzing the presence of science expressed as a method of inquiry in science textbooks.

Content analysis is one approach which can be used to assess the treatment of science as inquiry in textbooks. Content analysis allows for a systematic, quantitative analysis of written material. One form of content analysis, linguistic content analysis, is especially suited for detecting the image of a concept, such as science as inquiry, in written material.

METHODOLOGY AND RESEARCH DESIGN

The object of this study was to perform a longitudinal analysis of the treatment of science as inquiry in high school biology textbooks. The series of textbooks studied was the Modern Biology series published by Holt, Rinehart and Winston. The Modern Biology series was selected because it has been one of the best selling high school biology textbook during the later half of the century, and because the years of publication of editions in this series span the time period of interest.

Research Design

Nine editions of Modern Biology were published by Holt from 1956 to 1985. The years of publication for these editions are: 1956, 1960, 1963, 1965, 1969, 1973, 1977, 1981 and 1985. It was not feasible to study carefully all nine editions, therefore four editions were selected to best represent intervals of roughly 10 years in the approximately 30 year time span. The textbook editions selected were: 1956, 1965, 1977, and 1985. Of the remaining editions, the 1981 edition was selected for reliability analyses.

A blocking design was used within the textbooks. Three subject areas were strategically selected for blocks. The first subject area, which was addressed in Chapter One, entitled "The Science of Life" in all four editions, was the general description of what the subject of biology is, and how it fits into the realm of science. This subject was selected because it provides a description of how the authors view biology and science, and could provide an overall view of their approach to the image of science as inquiry.

The second subject area selected, genetics, was one in which several major scientific advances have been made in the last several years. It was hypothesized that this area would have the highest potential for expressing science as an active, ongoing process of inquiry. The text material dealing with the genetic material itself was selected. In the 1956 edition, this material was contained in Chapter 48 entitled "The Principles of Heredity." In the subsequent editions, the treatment of the material was expanded, divided into two chapters, and moved closer to the front of the book. The two chapters which dealt with this material in the 1965, 1977 and 1985 editions were Chapters 9 and 10, entitled "Principles of Heredity" and "The Genetic Material" respectively. Chapters 9 and 10 of the last three editions were treated as a single block for the random selection of the sentence sample, and for the subsequent statistical analysis.

The final subject area selected, leaf structure and function, was one which was hypothesized to be highly descriptive. It was hypothesized that this subject area would be less likely to demonstrate an image of science as a process of inquiry and more likely to portray science as a body of knowledge. The chapter number and titles devoted to leaf structure in the four textbook editions studied are as follows: 1956, Chapter 14 "Leaves and Their Activities;" 1965, Chapter 25 "Leaf Structure and Function;" 1977 and 1985, Chapter 23 "The Leaf and Its Function."

Sentences within blocks were randomly sampled using the random number generator of the SAS statistical package. Only sentences in the main body of the written text were sampled. Excluded from analysis were chapter introductions, figure and picture captions, marginal notes, chapter summaries, and end-of-chapter questions. A twenty percent sample of sentences was selected. Dukes and Kelly (1979) recommend sampling ten percent of a text for readability studies, so it was assumed that a twenty percent sample of sentences would be more than adequate. Also, a twenty percent sample allowed for reasonably large expected cell frequencies. Sampled sentences were assigned identification codes which included information on the year the edition was published, the subject area, the page number in the book on which the sentence was located, the sentence number within the chapter and the position of the sentence within a paragraph (i.e., first, middle, last or only sentence within a paragraph). Sampled sentences were then copied from the textbooks, along with their identification numbers using the Apple Writer II word processing program. Contextual clues which would be helpful with the coding were also copied with the sentences. For example, in the phrase "It was a major discovery," the event being referred to by the word "It" was copied and flagged to indicate it was a contextual clue and not part of the actual sentence to be coded. Printouts were made of the sentences, allowing space between sentences for notes made by the coder. In order to allow for the coder to be unaware of the publication year of the material, the portions of the identification codes indicating year of publication were masked. It was not possible for the coder to also be unaware of the subject area, as the sentence topics provide obvious clues to their chapter source. The 15 chapters were coded in a random order. Sentences within the blocks were coded in the order in which they appeared in the text. This nonrandom order within chapters was used because of the important role contextual clues play in the coding process. It was believed to be easier to make an accurate judgement on the intent of the authors if the sentences were coded in order of appearance in text, rather than random order.

This study utilizes content analysis. As used here, content analysis will be defined as the systematic analysis of data derived from the coding of textual material. A computer-aided, linguistic approach was used for this study.

Using PLCA

The program accesses a dictionary, which consists of a list of numerical codes and the words to which they correspond (see Appendix A). Due to limited storage capacity in the memory of the computer, the dictionary must also be limited. The dictionary has to be modified to fit the needs of the particular analysis being done. To develop the dictionary, portions of text from Modern Biology were coded on paper. Frequently used nouns and verbs were noted. The dictionary was then altered to include vocabulary appropriate for this analysis. The dictionary was revised and re-edited until it reached a form which allows for coding of text material in high school textbooks. The final form of the dictionary contains 60 codes for verbs, 37 for objects and 6 for people.

The unit of analysis for PLCA is a clause. A clause is defined as any sentence or portion of a sentence that includes a predicate and a subject (or implied subject). With the PLCA, clauses can be classified into one of four types. This classification of clauses into types is perhaps one of the most critical steps in the coding process.

The first type of clause is "perception." This is the most common type of clause in written material. It describes an activity. Examples of perception clauses are: A scientist performs experiments; refer to Figure 3; a green plant grows toward the sunlight.

The second type of clause is "recognition." This is the second most common type of clause. This type of clause classifies a phenomenon as being a member (or not a member) of a category. Examples of recognition clauses are: Biology is a science; a starfish is actually not a type of fish; glucose is a sugar.

The third and fourth types of clauses are very rare in descriptive types of writing. The third type of clause is "justification." It judges the goodness or badness of an activity. Examples of justification clauses are: Drunk driving is wrong; it is good to clean your microscope after it comes into contact with salt water.

The last type of clause is "evaluation." This judges how well a phenomenon fits into a category. Examples of evaluation clauses are: Pasteur is recognized as being a brilliant scientist; when considering observable genetic traits, blood type is a good example.

The selection of the type of clause is very important in the coding. It is also difficult to determine at times. The coder must rely on the context, and on subject matter knowledge to try to determine the intent of the writer.

To help illustrate the importance of determining clause type, an example taken from Roberts (1988) will be used. The example shows that the clause, "John bought that book," can be classified into each of the four clause types in four different contexts.

An example of a perception context would be: "We both went to the store. John bought that book." This is an accounting of an activity which occurred.

An example of a context for a recognition clause would be: "I'm trying to figure out what he spent his allowance on. Aaah, John bought that book!" Here, the purchasing of the book falls into the category of "what John spent his allowance on." A PLCA translation of the clause would be: "The BOOK is JOHN's PURCHASE."

An example of a context for a justification clause would be: "I pointed out that he might be arrested if he stole it. Fortunately, John bought that book." Here, the phrase is intended to communicate that the purchasing of the book was the right or appropriate thing to do. A PLCA translation of the clause would be: "JOHN's BUYing a(n) BOOK was appropriate."

Finally, an example of a context for a evaluation clause would be: "John is an excellent cook. From all the cookbooks in the store, John bought that book." Here the phrase communicates that the book is a very good cookbook. A PLCA translation of the clause would be: "The COOKBOOK is good."

Clearly, the type of clause is important in coding. Practice with the coding technique, subject matter knowledge, and unambiguous text can all help in selecting clause types. Fortunately for this study, ambiguities are less common in descriptive texts or in texts intended to convey specific information than in other forms of written text, such as transcripts of conversations or poetry (Roberts, 1988).

In addition to coding clause types, the valance of the clause (positive or negative) and whether or not the clause is in the form of a question are also recorded. Direct quotations can be coded to indicate the speaker of a clause and the audience of the speaker. For the current analysis, this feature was used only when the text contained a direct command to the reader, when a term such as "we," "you," or "our" was used or when the clause was a question. After a clause type has been selected and the information on valence, question, speaker and audience has been entered, the program runs through a series of prompts that correspond to the syntax of that clause type. The user then enters appropriate numerical codes from the dictionary. For example, the prompt sequence for a perception clause is: semantic subject, modifier of the subject, verb, modal auxiliary of the verb, semantic object, and modifier of the object. If a clause does not contain one of these syntactic elements, such as a modifier of the subject, the user simply strikes the "return" key and passes on to the next prompt. The program then enters a code of "888" (or "88" depending on the column width) for that particular element.

Prior to the actual data collection, sentences from the introduction, the genetics and the leaf structure chapters of the 1965 edition of Modern Biology and sentences from the introduction of the 1981 edition of Modern Biology were coded to verify that the dictionary contained sufficient vocabulary for use in coding high school biology textbooks and to develop a set of coding rules. As a result of this process the following coding rules were established.

The first rule dealt with the use of the variables called "speaker" and "audience." If a clause was phrased as a question, was a direct command, or contained one or more of the words "we," "you" or "us," then the "speaker" of the clause was coded as author (code = 135) and the "audience" was coded as reader (code = 136). Otherwise the speaker and audience codes were "888," as described above. If audience and speaker codes were used, they were collapsed into a category called "involve" for involvement of reader. The category "involve" was used in the creation of the dependent variable, as described in the following chapter.

The second rule addressed the treatment of verbs. Verbs are a pivotal element in the coding process. Because a clause must contain an inflected verb, and the unit of analysis is a clause, locating the inflected verbs in a sentence was the first step in identifying its clause(s). It was found that the easiest way to begin coding a sentence was to first circle all the inflected verbs in the sentence. The second rule applied in the event of multiple verbs for a single subject. Each verb was coded as a separate clause if the verbs denoted very different actions. If the actions were very similar, the verbs were collapsed into a single verb code. For example, in the phrase "the scientist tested, then reported her theory," the two verbs "tested" and "reported" denote two separate types of action. This would have been coded as two separate clauses. In the phrase "the child hopped and skipped around the room," the two verbs "hopped" and "skipped" denote rather similar actions. This phrase would have been coded as a single clause with a more generic single verb such as "moved" for the two verbs in the original form.

A third rule concerned the use of gerunds, such as "testing," "looking," etc. Gerunds were not coded at all. For example, the phrase "the researcher investigated the effects of the virus while controlling for environmental factors" would be coded as a single clause "the researcher investigated the effects of the virus." There is only one inflected verb in the phrase, so there is only one clause. There is some loss of information due to this rule, but the justification for not coding gerunds is that if the author(s) did not consider the information important enough to devote a inflected verb to it, the information is not important enough to code. This particular phrase also provides an example of a modifier, in this case a modifier of the direct object. The object of the verb "investigate" is "effect." The modifier of "effect" is "virus."

A fourth rule had to be developed due to the limited memory of the computer. Biology textbooks have a very large vocabulary, particularly nouns, and it was impossible to provide a separate code for every noun which might appear.

Instead, certain noun groupings were established for the many technical and scientific nouns which appeared in the text. Initially the groupings were for "plant science word," "animal science word," "genetic science word," "mineral science word," and "science word." All of these "science word" categories were later merged into one category called "technical word" during the data analysis. Technical word was thus an investigator designed variable. Basically, a technical word was a word which one might find listed in a glossary. Examples of words classified as technical words are: photosynthesis, epidermis, genotype, parenchyma and heterozygous. Examples of words not classified as technical words are: plant, skin, guinea pig, leaf and disease.

To summarize the coding process, there were a total of 17 variables recorded for each clause. The Program for Linguistic Content Analysis records 12 variables; 5 additional identifying variables were also recorded. The 12 variables recorded with the program were: type of clause; tense of clause (past, present, or future); whether or not the clause was a question (the program allows for more detailed information on type of question, but for this analysis, this variable was simply treated as a dichotomous variable); valence (positive or negative); audience;

speaker; semantic subject of the verb; modifier of the subject; verb; modal auxiliary of the verb; semantic object ot the verb; and modifier of the object. The 5 identifying variables recorded were: year of publication (1956, 1965, 1977, 1985); subject area (introduction, genetics, leaf structure); page number on which the sentence appeared; the sequence number of the sentence within a chapter (i.e. sentence number 17 out of 256 total sentences in the chapter); and the position of the sentence within the paragraph (first, middle, or last). There were a few cases when a sentence was the only sentence in a paragraph. In these instances, the sentence was treated as if it were the first sentence in a paragraph.

One issue of concern with any coding technique is the reliability of that technique. The efforts to determine reliability will be described in the next section.

Reliability

Reliability can be defined as the level of consistency or stability of an instrument (Borg and Gall, 1983). Two types of reliability were evaluated: self-reliability (or intrarater agreement) and cross-reliability (or interrater agreement). Self-reliability refers to the extent that coding by the same person is identical when performed at different times. Cross-reliability refers to the extent of identical coding across raters.

To determine self-reliability, one of the 15 chapters in the study was coded twice. The chapter selected for double coding was the middle, or eighth chapter in the random order of the 15 chapters. The middle chapter was selected to help control for any possible effects due to changes in the coding process in the beginning phases of the data collection. It was also chosen to allow more time between coding than would occur if one of the later chapters were coded twice. There was a time span of around two weeks between the first and the second coding of the chapter. Approximately 700 other clauses were coded during these two weeks, reducing the chance of recall of any particular clause from the first to the second coding.

Cross-reliability was evaluated by having two people code the same group of sentences. The sentences were randomly selected from the three subject areas. All of the sentences for the cross-reliability study came from the 1981 edition of Modern Biology.

A statistic which can be used for calculating reliability of categorical ratings is Kappa (Fleiss, 1973; Cohen, 1960). In the coding of categorical data, there is a certain amount of agreement which can occur by chance.

Because of this chance agreement, simply reporting percent or proportion of agreement could be biased slightly upward. The Kappa statistic is designed to correct for this chance agreement, yielding a more conservative estimate of reliability. It was decided before hand that a Kappa of at least 0.60 would be acceptable. This is based on the fact that the coding technique has never been applied to science textbooks. It is a rather subjective technique, and hence responses can vary across raters, lowering reliability. Also, the number of choices for certain categories, such as verbs, was very large, allowing for even more variance among raters. Fuller (1987) reports on the difficulty of obtaining high reliability values (though not Kappa statistics) in rather simple, straight-forward items such as age, gender, level of education and occupation. These are traits which should be rather objective, and have limited numbers of possible choices. One can see then, how difficult it is to achieve high reliability on something which is much more subjective, with many more possible choices.

Statistical Analysis

All of the variables in the study were categorical with the exception of one. The variable which was not categorical was sequence number of sentence within the chapter. This variable was converted from an interval to a continuous scale by dividing the sequence number by the total number of sentences in the chapter. The resulting variable was a measure of the position of the sentence within the chapter. This variable ranged from zero to one, with zero not inclusive. Values close to zero represented sentences at the beginning of the chapter, while values close to one represented sentences at the end of the chapter.

The dependent variable was a dichotomous variable representing whether or not inquiry was present. Creation of this variable from the linguistic content analysis data is described in the following chapter. A useful technique to use when a dependent variable is dichotomous is logistic regression analysis. Logistic regression analysis can be interpreted in much the same manner as linear regression analysis, in that expected values of a dependent variable are modeled as a linear combination of one or more independent variables. Contributions of the independent variables to the model are tested for their statistical significance by evaluating the decrement in the model likelihood ratio chi-square attributable to each variable. Logistic regression models estimate the natural logarithm of the odds of the occurrence (versus nonoccurrence) of the dependent variable (Knoke and Burke, 1980).

Rather than a test of the overall R-Square or change in R-square as in linear regression, the test for significance of a logistic regression model is a test of the change in the likelihood ratio chi-square statistic. When testing models, one looks for a large decrement in the "model chi-square," where "model chi-square" is calculated as:

$$[-2 * \ln(\text{model likelihood ratio})].$$

In other words, the statistical dependence between a dependent variable and a set of independent variables is evaluated by calculating a chi-square value for the model with the dependent variable as a function of the intercept only, then for the model with the dependent variable as a function of the intercept plus the independent variables. The addition of the extra variables should model expected frequencies closer to observed frequencies, and hence lower the overall chi-square. The decrement in chi-square is the amount that the overall chi-square is lowered. In a similar manner, model fitting can be performed on the significance of "r" independent variables by testing the decrement in chi-square values from a model with "x" independent variables to a model with "x + r" independent variables. This decrement has a chi-square distribution with "r" degrees of freedom and can be tested for statistical significance by referencing any standard chi-square table.

For this study a logistic regression model was developed and tested. Details of the development of the model, which used analysis of contingency tables and log-linear analysis, will be explained in the following chapter.

Research Questions

The six proposed research questions can be grouped into two categories. The first two questions are the major questions of the study, and stem from rationale developed in Chapter One, Chapter Two and earlier in this chapter. The remaining four questions are exploratory in nature and hence are not grounded in previous studies.

1. Does level of inquiry vary across years? In particular, one could anticipate a peak in level of inquiry shortly after 1960. This could be anticipated because of the high level of attention given to inquiry during the early 1960s.
2. Does level of inquiry vary across subject areas studied? The subject of genetics was specifically selected to be high in inquiry, while the subject of leaf structure was selected

to be low in inquiry. The introduction material was selected to represent an overall approach to treatment of inquiry.

3. Does level of inquiry vary with respect to the position of the sentence within a chapter? Often chapters start by giving an historical background of the subject matter in the chapter. According to Tamir (1985), description of the history of ideas and discoveries is an indicator of inquiry. One could anticipate that inquiry would be higher at the beginning of a chapter, and lower at the end of a chapter.

4. Does level of inquiry vary with respect to the position of the sentence within a paragraph? One could anticipate the measure of inquiry could be higher at the beginnings and ends of paragraphs because often questions are placed at these points in a paragraph. For example, questions may be used at the beginnings of paragraphs as introductions to the material in the paragraph and may be used at the ends of paragraphs as transition statements to the next topic. (It should be noted that any such trend could be a factor of writing style as well as a possible measure of inquiry.)

5. Does level of inquiry vary with the use of technical words? Clauses which have technical words could tend to be highly descriptive, and thus less likely to exhibit much inquiry.
6. Are any interactions among the five main effect variables associated with level of inquiry? No specific interactions will be listed at this point. In the following section potentially interesting interactions are identified with the use of log-linear analysis, and modeled using logistic regression.

RESULTS

The results will be reported separately for the reliability studies and for the model development and testing. First however, results from the coding process will be discussed.

Results of Coding Process

As mentioned above, the unit of analysis for the PLCA is a clause. The number of clauses sampled is shown in Table 2. As can be seen from Table 2, the total number of clauses sampled was 1,350. The smallest cell size within Table 2 was 58, the number of clauses sampled from the leaf structure chapter in the 1985 edition. The largest cell size was 233, the number of clauses sampled from the genetics material in 1965. It should be noted at this point that the clauses for the treatment of the genetic material came from two chapters in 1965, 1977 and 1985, hence the large sample sizes for these blocks. The average number of clauses per sentence for the entire sample was 1.7. It is also necessary to note that only the first two types of clauses, perception and recognition, were observed in the sentences. The occurrence of only these two types of clauses might be expected in material which is highly descriptive, such as textbooks.

TABLE 2. Number of clauses per block

Subject	1956	1965	1977	1985	Total
Introduction	69	138	67	75	349
Genetics	75	233	219	142	669
Leaf Structure	117	66	91	58	332
Total	261	437	377	275	1,350

Reliability

The Kappa statistic was calculated from two-dimensional contingency tables in which clauses were classified according to values from the first coding and values from the second coding (in the case of self-reliability) or values from the coding of one rater and values from the coding of the second rater (in the case of cross-reliability). Separate contingency tables were created for each of the twelve variables in PLCA. The formula for Kappa (as found in Fleiss, 1973) is:

$$\text{Kappa} = (P_o - P_c) / (1 - P_c).$$

P_o is the sum of the diagonal elements of the contingency table divided by the total sample size. P_o is thus the proportion of agreement. P_c is the sum of the products of the row and column totals divided by the square of the total sample size. P_c is the correction factor for chance agreement. Kappa has a range of zero to one, with values close to zero indicating low reliability and values close to one indicating high reliability. As can be seen from the formula, when P_o is equal to one, Kappa is also equal to one.

The results of the self-reliability studies, based on a sample size of 82 clauses are shown in Table 3. The values for P_o, or proportion agreement, are shown in the first column, the values for Kappa are shown in the second column, and the standard deviations for Kappa are shown in the last column. (Standard deviations for Kappa were calculated according to Fleiss, Nee and Landis, 1979.) The means and standard deviations reported at the bottom of Table 3 are based on a sample size of twelve. In other words, the mean of the twelve values of Kappa is 0.86, and these twelve values have a standard deviation of 0.10. The value 0.10 is not the standard deviation of an estimate of Kappa, these values are reported in the last column. As expected, when P_o has a value of one, Kappa also has a value of one. It is interesting to note that the correction due to chance on variables with large contingency tables is smaller than the correction due to chance on variables with small contingency tables. For example, the variable "tense" has a three-by-three contingency table, and a change of eleven percent from P_o to Kappa. The variable "verb" has a roughly fifty-by-fifty contingency table, and only a two percent change from P_o to Kappa. This would be expected with a statistic that corrects for chance agreement, since chance agreement is greater with a smaller number of choices.

The results of the cross-reliability studies, based on a sample size of 46 clauses are shown in Table 4. The format for Table 4 is the same as that for Table 3. As might be expected due to a slightly smaller sample size, and greater potential for diversity across occurrences of ratings, the standard deviations of Kappa are slightly higher for the cross-reliability study than for the self-reliability study. The lowest values of Kappa are for the variables "Modifier of Subject" and "Modifier of Object." These very low values (0.15 and 0.35 respectively) are the result of highly skewed marginal distributions. The distributions were skewed because the modifiers were infrequently used, and hence the majority of codes for these variables were "888." One of the coders tended to be more

TABLE 3. Results of self-reliability analysis

N = 82

Variable	P_o	Kappa	Standard Deviation
Type	0.96	0.92	0.08
Tense	0.90	0.80	0.07
Question	1.00	1.00	0.08
Valence	1.00	1.00	0.08
Audience	0.99	0.74	0.08
Speaker	0.99	0.74	0.08
Semantic Subject	0.94	0.93	0.02
Modifier of Subject	0.99	0.92	0.05
Verb	0.89	0.87	0.03
Modal Auxiliary	0.96	0.69	0.06
Semantic Object	0.86	0.85	0.02
Modifier of Object	0.98	0.87	0.05
mean	0.95	0.86	
standard deviation	0.05	0.10	

liberal when using modifiers, and this difference is, in effect, magnified by the skewed marginal distributions. The average value of Kappa across all twelve variables for cross-reliability is 0.71. Based on arguments put forth in the previous chapter, the values of Kappa for both the self- and the cross-reliability were judged as being sufficiently large.

TABLE 4. Results of cross-reliability analysis

N = 46

Variable	P_o	Kappa	Standard Deviation
Type	0.95	0.81	0.10
Tense	0.85	0.68	0.09
Question	1.00	1.00	0.08
Valence	1.00	1.00	0.10
Audience	0.93	0.69	0.10
Speaker	0.93	0.69	0.10
Semantic Subject	0.78	0.76	0.03
Modifier of Subject	0.78	0.15	0.06
Verb	0.76	0.74	0.03
Modal Auxiliary	0.95	0.78	0.07
Semantic Object	0.85	0.84	0.03
Modifier of Object	0.85	0.35	0.06
mean	0.88	0.71	
standard deviation	0.09	0.24	

Developing a Model

After the reliability studies were competed, work was begun to develop a logistic regression model for the data. The first step was to study contingency tables of relevant variables. For example, Tamir (1985) states that giving historical backgrounds of discoveries is an indicator of inquiry (see Table 1). An easy way to detect such an item is to look at use of verbs in the past tense. Tables of use of past tense crossing year of publication, subject area and position in paragraph were thus studied. Another item mentioned by Tamir is references to figures, so the occurrence of references to figures was studied with similar types of contingency tables.

One approach to analyzing data for linguistic content analysis is to focus on particular verbs, and to identify the actors and objects associated with them. To this end, a group of verbs, called "inquiry verbs" was created (see Appendix A). Inquiry verbs were verbs which showed actions associated with inquiry, such as test, experiment, research, investigate, etc. The distribution of these inquiry verbs was also studied with contingency tables. Two other variables, called "inquiry scientist" and "inquiry reader" were also created by combining certain actors of the clauses with the inquiry verbs. Inquiry scientist was a variable in which a scientist was the semantic subject of an inquiry verb. Likewise, inquiry reader was a variable in which a reader was the semantic subject of an inquiry verb. The distributions of these variables were also studied with contingency tables.

The occurrence of any one of the aforementioned variables was not, by itself, a meaningful measure of inquiry. Each variable might serve as an indicator of inquiry, but inquiry involves more than any one of these indicators. For this reason, a dependent variable called "inquiry" was created by merging together several indicators of inquiry. The variable "inquiry" was a dichotomous variable which had a value of one in the presence of one or more of the following traits: a scientist as the semantic subject of an inquiry verb; a question; the use of a scientist's name; reference to a figure; involvement of reader; or use of past tense. In the absence of any of these traits, the variable inquiry had a value of zero.

The trait of a scientist as the semantic subject of an inquiry verb is described above. Involvement of the reader is indicated by whether or not the speaker/audience codes were used (as described in the previous chapter). This was included as an indicator of inquiry because a reader must be involved in the learning process to develop any sense of inquiry. It was reasoned that use of questions, commands, and words such as "you" and "we," creates more of a sense of involvement, and hence greater potential for inquiry by the reader. The speaker/audience codes were used to indicate these traits associated with involvement. The use of questions, use of a scientist's name, reference to a figure, and use of past tense as indicators of inquiry are based on Tamir's list of indicators of inquiry (Tamir, 1985) (see Table 1). It should be noted that the use of questions is subsumed under the category "involve." The trait of use of a questions was left in the list of indicators of inquiry for descriptive purposes.

The independent main effect terms in the model were derived from the research questions. These terms were: year of publication (1956, 1965, 1977, 1985); subject area (introduction, genetics, leaf structure); position in chapter (a continuous variable ranging from zero to one); position in paragraph (first, middle, last); and use of technical words (yes, no).

To identify potentially statistically significant two-way interaction terms, a log-linear analysis was performed with the data. There are no dependent or independent variables with log-linear analysis, but rather the cell frequencies within a contingency table are modeled as functions of all the variables in the model. To compare log-linear analysis with logistic regression, we must consider the interaction terms in log-linear analysis. In all following discussions, any references to dependent and independent variables are in reference to the dependent and independent variables of a logistic regression model, not a log-linear model. With log-linear analysis, a significant two-way interaction term involving the dependent variable (in this case inquiry) and one other variable is comparable to a significant main effect term with that variable in logistic regression. Likewise, a three-way interaction term involving inquiry and two other variables is comparable to a two-way interaction between those two variables in a logistic regression model.

The results of the log-linear analysis showed that all two-way interactions of the variable "inquiry" with the five independent variables were statistically significant. This was an indication that these five independent variables would be statistically significant main effect terms in a logistic regression model. The results also indicated five potentially significant two-way interaction terms for a logistic regression model. These terms were chosen from the log-linear analysis as three-way interactions involving the variable "inquiry" which had p-values of less than 0.1. Possible interactions involving three or more independent variables (four-way or greater than four-way interactions in log-linear analysis) were not considered due to the difficulty of applying meaningful interpretations to interactions involving greater than two variables.

Fitting models

The five potentially statistically significant two-way interaction terms for logistic regression modeling indicated by the log-linear analysis were: subject-by-position in chapter; position in paragraph-by-position in chapter; year of publication-by-position in chapter; subject-by-position in paragraph; and year of publication-by-subject. Model fitting was performed using logistic regression. The full model for the five main effect terms and the five two-way interaction terms was tested, and a model chi-square obtained. Then the least statistically significant interaction term was dropped from the model. A new model chi-square for the reduced model was obtained, and the difference in the two model chi-square values was calculated. If the difference, or decrement, in the chi-square value was statistically significant, it indicated the term deleted from the model significantly improved the fit of the model to the observed occurrences of inquiry. In other words, a significant decrement in the model chi-square value caused by deleting a term from the model is evidence that the term should be left in the model. If a significant decrement in the model chi-square did not occur with the deletion of a term, the deleted term was left out of the model. If a significant decrement did occur with the deletion of a term, the deleted term was returned to the model and other interaction terms were evaluated for deletion in the same manner.

The results of sequentially deleting the least statistically significant interaction terms are shown in Table 5. The variables in the model are shown in the left-hand column. The next column shows the model chi-square value for each model, with the appropriate degrees of freedom in parentheses underneath. The third column indicates the interaction term which was deleted for statistical significance testing. The last column shows the change, or decrement, in the model chi-square value resulting from deletion of the indicated term. Degrees of freedom associated with the deleted term are shown in parentheses underneath the change in the chi-square value.

As can be seen from Table 5, the deletion of the first two interaction terms (subject-by-position in chapter and position in paragraph-by-position in chapter) did not cause a statistically significant change in the value of the model chi-square. However, deletion of the remaining three interaction terms (year of publication-by-subject, subject-by-position in paragraph, and year of publication-by-position in chapter) did cause statistically significant decrements in the model chi-square values. Based on these

TABLE 5. Fitting logistic regression models

Variables in Model	Model Chi-Square (df)	Variable Removed	Change in Model Chi-Square (df)	
Y, S, PC, PP, TW Y*S, Y*PC, S*PP, S*PC, PP*PC	399.52 (26)	-	-	
Y, S, PC, PP, TW Y*S, Y*PC, S*PP, PP*PC	398.81 (24)	S*PC	0.71 (2)	
Y, S, PC, PP, TW Y*S, Y*PC, S*PP	396.59 (22)	PP*PC	2.22 (2)	
Y, S, PC, PP, TW Y*PC, S*PP	375.70 (16)	Y*S	20.89 (6)	**
Y, S, PC, PP, TW Y*S, Y*PC	382.68 (18)	S*PP	13.91 (4)	**
Y, S, PC, PP, TW Y*S, S*PP	383.68 (19)	Y*PC	12.91 (3)	**

Y - year of publication; S - subject area; PC - position in chapter; PP - position in paragraph; TW - technical word

**$p < 0.01$.

results, the model adopted for further study and testing was the model with the five main effect terms and the three statistically significant interaction terms listed above.

The model was thus arrived at by means of both subject matter consideration (for the main effect terms) and statistical model fitting (for the interaction terms). The proposed model will be examined in greater detail in the following section.

Analysis of the Proposed Model

Before describing the analysis of the model, it is necessary to consider the composition of the data set in greater detail. In particular, the data set consists of two types of clauses, the perception clauses and the recognition clauses. These two types of clauses can be viewed as two separate populations within the data set. For example, one of the indicators of inquiry developed for the dependent variable was the presence of a particular groups of verbs called "inquiry verbs." Recognition clauses can never have "inquiry verbs" because the verb in a recognition clause is always a form of the verb "to be." If the two clause types are indeed separate populations, these two clause types should be analyzed separately. On the other hand, both types of clauses can contain indicators of inquiry. The argument can be made that because the purpose of the study is to assess the level of inquiry in the total body of written material, and both types of clauses carry indicators of inquiry, the two types of clauses should be analyzed together. For purposes of comparison, both approaches to the analysis were used, first with the perception and recognition clauses combined, and then with the perception clauses only. An analysis of recognition clauses only was not performed because this type of analysis would require a different methodology. Logistic regression techniques require large sample sizes to obtain reliable results, and the sample size of recognition clauses (n=262) was not viewed as being large enough.

<u>Analysis of perception and recognition clauses combined</u>

The results of fitting the model to the data set consisting of both types of clauses is shown in Table 6. The decrement in chi-square for this model is 396.59 with 22 degrees of freedom. This chi-square value is based on a sample size of 1,350 and is statistically significant ($p < 0.0001$). The statistic C, reported at the bottom of Table 6, can be viewed as being somewhat analogous to a correlation coefficient of the observed values of inquiry with the values of inquiry predicted by the model. This statistic is considered to be a good measure of the predictive ability of the model (Harrell, 1986). (See Appendix D for a description of how C is calculated.) For this model the value of C was 0.794, indicating fairly good predictive ability. The terms in the model, both main effect and interaction, are listed in the left most column. The chi-square value associated with each term is listed in the second column. The chi-square values associated with each term were calculated by deleting a given term from the model and calculating the decrement in the model chi-square for the reduced model. The difference between the calculated value and the model chi-square value associated with the full model was then recorded as the chi-square value for the deleted term. The third column indicates the degrees of freedom associated with each term, while the last column reports the probability value of the chi-square statistic. The intercept term is comparable to the intercept term in linear regression, and is of no real substantive interest. As can be seen from Table 6, all of the terms in the model are statistically significant at at least the 0.01 level of significance.

Analysis of perception clauses only

The results of fitting the same model to a data set containing perception clauses only is shown in Table 7. The model chi-square is 335.68 with 22 degrees of freedom. This chi-square value is based on a sample size of 1,088 and is statistically significant ($p < 0.0001$). The statistic C, a measure of the predictive ability of the model, is 0.802. This value of C indicates fairly good predictive ability.

TABLE 6. Logistic regression model for perception and recognition clauses combined

Variable	Chi-Square	d.f.	p
Intercept	15.51	1	<0.0001 **
Year	27.94	3	<0.0001 **
Subject	156.96	2	<0.0001 **
Position in Chapter	65.47	1	<0.0001 **
Position in in Paragraph	10.76	2	0.0046 **
Technical Word	50.08	1	<0.0001 **
Year * Subject	20.89	6	0.0019 **
Subject * Position in Paragraph	13.91	4	0.0076 **
Year * Position in Chapter	12.91	3	0.0048 **

Decrement in Chi-Square with 22 degrees of freedom = 396.59 ($p < 0.0001$)

n = 1,350 C = 0.794

**p < 0.01.

TABLE 7. Logistic regression model for perception clauses only

Variable	Chi-Square	d.f.	p
Intercept	10.32	1	0.0013 **
Year	29.24	3	<0.0001 **
Subject	120.27	2	<0.0001 **
Position in Chapter	48.38	1	<0.0001 **
Position in in Paragraph	8.07	2	0.0177 *
Technical Word	40.03	1	<0.0001 **
Year * Subject	24.39	6	0.0004 **
Subject * Position in Paragraph	11.83	4	0.0187 *
Year * Position in Chapter	13.53	3	0.0036 **

Decrement in Chi-Square with 22 degrees of freedom = 335.68 ($p < 0.0001$)

n = 1,088 C = 0.802

*p < 0.05.
**p < 0.01.

The format for Table 7 is the same as the format for Table 6. As can be seen in Table 7, the general pattern of the results are similar to those in Table 6. All of the main effect terms, with the exception of position in paragraph, are statistically significant at the 0.01 level of significance, with p-values less than 0.0001. The position in paragraph term is statistically significant at the 0.05 level of significance ($p = 0.0177$). The three interaction terms are all statistically significant. The year of publication-by-subject term and the year of publication-by-position in chapter term are both statistically significant at the 0.01 level of significance ($p = 0.0004$ and $p = 0.0036$ respectively). The third interaction term, subject-by-position in paragraph, is statistically significant at the 0.05 level of significance ($p = 0.0187$).

Analysis of the effects of the independent variables

Tables 6 and 7, while providing information about the statistical significance of the terms in the model, do not provide any insight about how the dependent variable varies with respect to the various levels of the independent variables. This can be observed by plotting the log odds of inquiry associated with the different levels of the independent variables. Plots were generated in two ways, first as log odds and then as "adjusted" log odds. The term adjusted refers to whether or not the effects of a variable were adjusted for the effects of the other variables in the model. The unadjusted log odds plots do not take into account the effect of the other variables in the model while the adjusted log odds plots do. Two sets of graphs of the log odds and adjusted log odds were generated, one set for the perception and recognition clauses combined and one set for the perception clauses only. The two sets of graphs were essentially identical, therefore only one set will be shown, those for the total data set of perception and recognition clauses combined.

The plots in Figures 1-9 are of the unadjusted log odds. These log odds were calculated as the average logit of the probability for the presence of inquiry for the various levels of the independent variables.

Figure 1 shows a plot of the log odds of inquiry versus year of publication. Log odds of inquiry are shown on the vertical axis while year of publication is shown on the horizontal axis.

Figure 2 shows a plot of log odds of inquiry versus subject area. The log odds of inquiry are plotted along the vertical axis. The three subject areas are indicated along the horizontal axis.

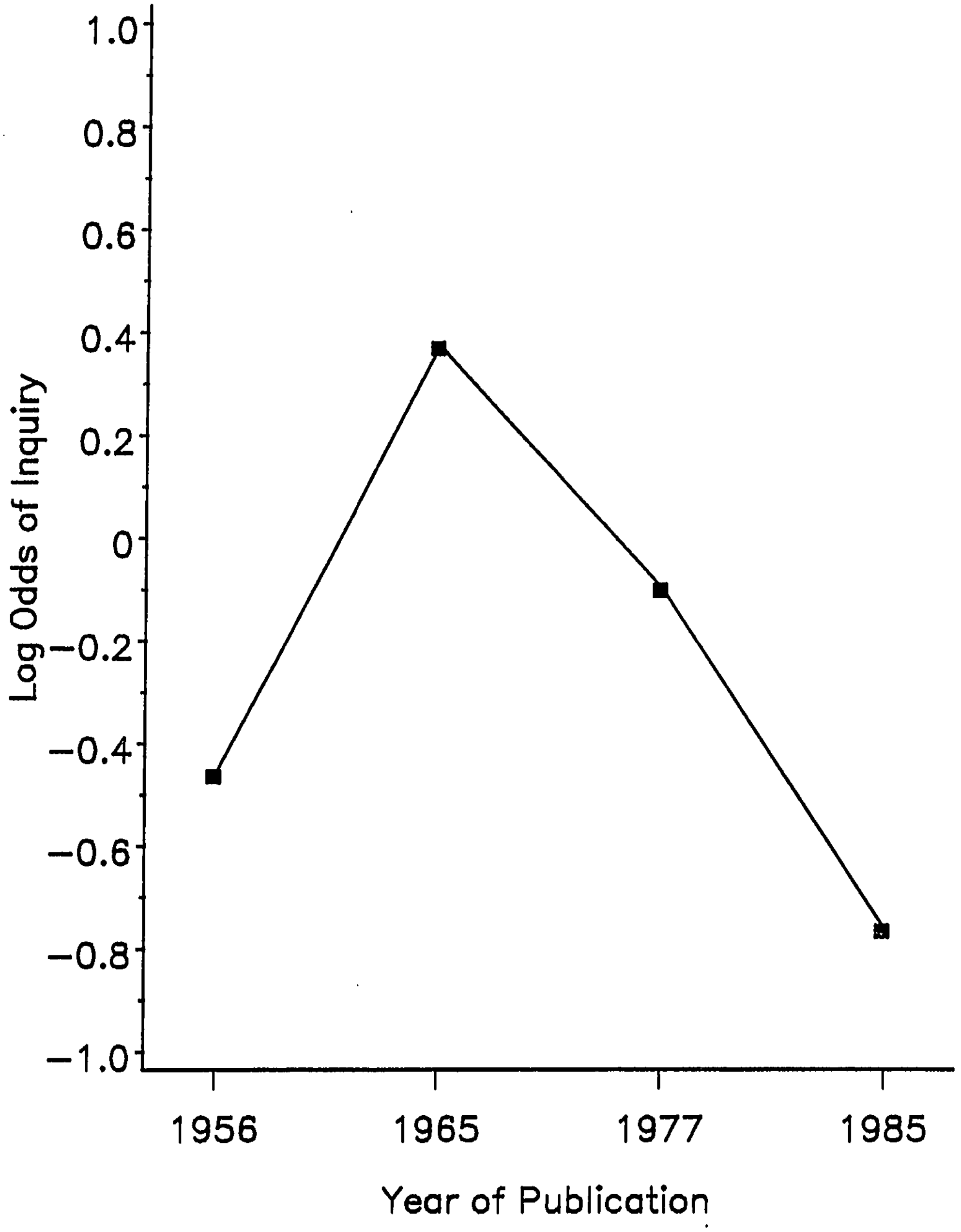

FIGURE 1. Log odds of inquiry versus year of publication

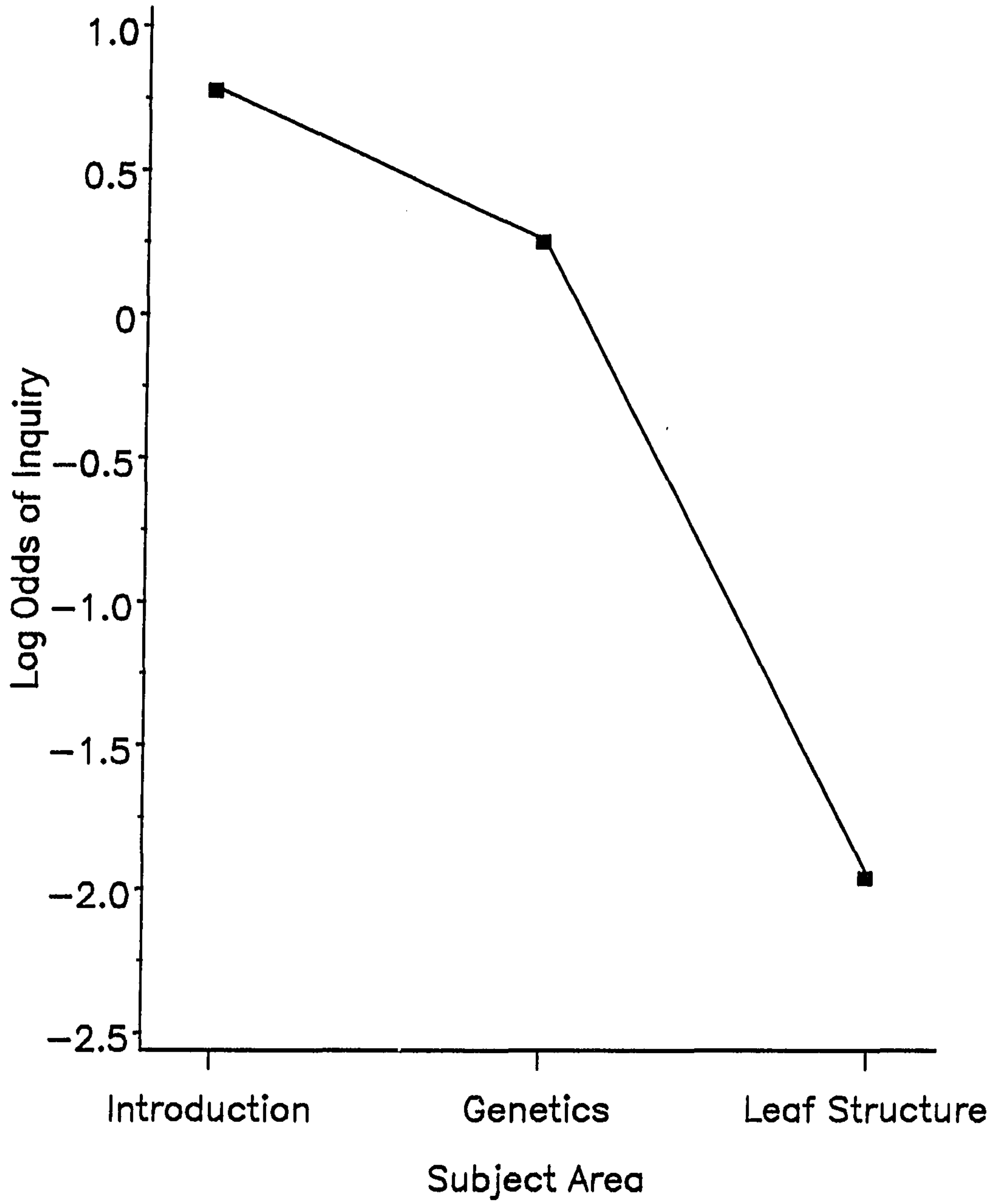

FIGURE 2. Log odds of inquiry versus subject area

The change in log odds of inquiry with respect to the position of the sentence within a chapter is shown in Figure 3. The variable of position in chapter was a continuous variable and observed log odds could not be calculated at each point along the continuum. As a consequence, the variable was altered to generate the plot. The variable was divided into four quartiles, representing the four quarters of the chapter. The log odds of inquiry in each of the four quartiles are shown on the vertical axis, while the four quartiles are indicated along the horizontal axis.

The change in log odds of inquiry with respect to the position of the sentence within a paragraph is shown in Figure 4. The log odds of inquiry are shown along the vertical axis. The three possible positions of sentences within a paragraph are shown along the horizontal axis. These positions are the first sentence in a paragraph, any of the middle sentences in a paragraph and the last sentence in a paragraph. In the event that a sentence was the only sentence in a paragraph, it was treated as being the first sentence in a paragraph.

Figure 5 shows the change in log odds of inquiry in clauses with and without technical words. Log odds of inquiry are shown along the vertical axis, while presence or absence of technical words in a clause is indicated along the horizontal axis.

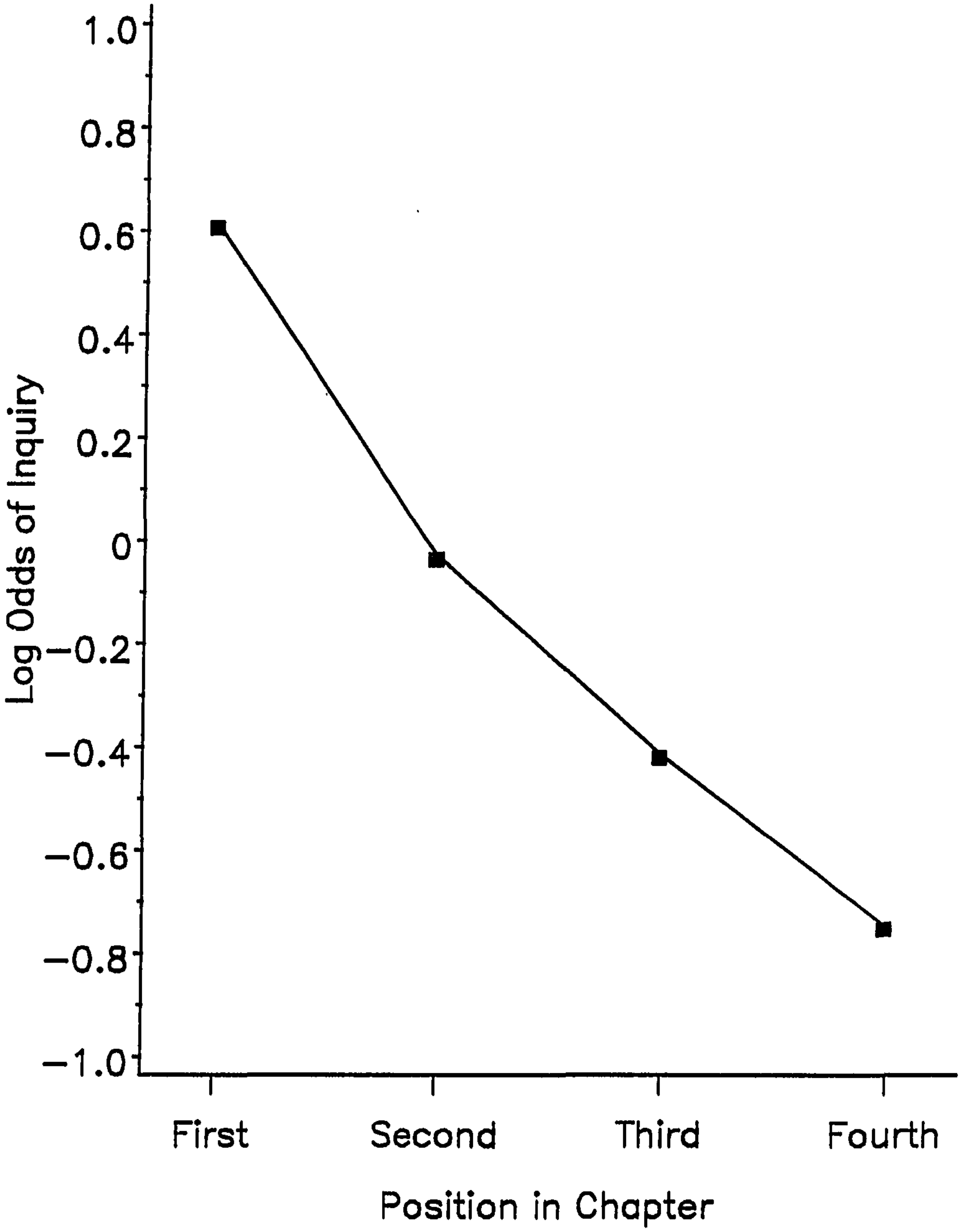

FIGURE 3. Log odds of inquiry versus position in chapter

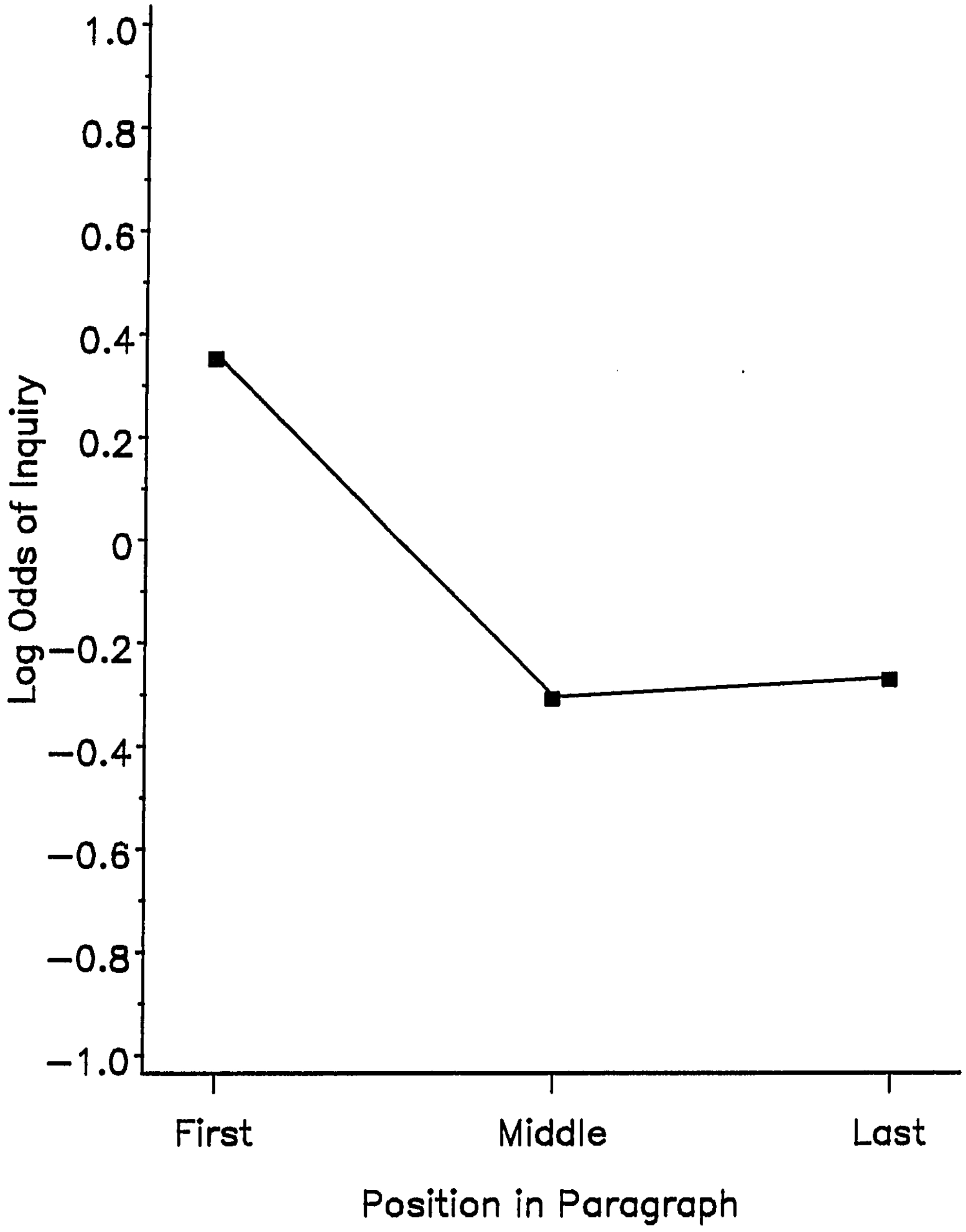

FIGURE 4. Log odds of inquiry versus position in paragraph

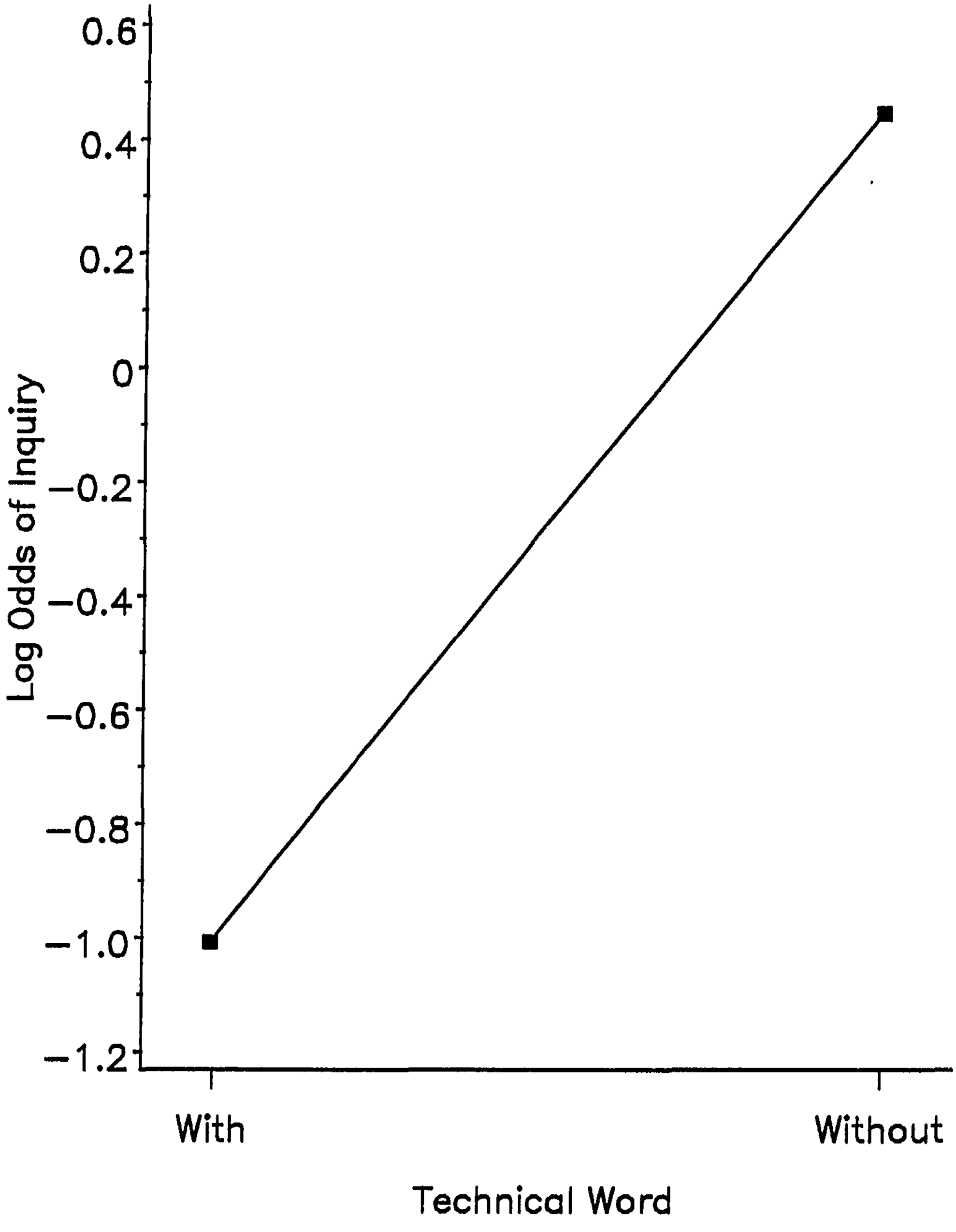

FIGURE 5. Log odds of inquiry versus use of technical words

The interaction between year of publication and subject area is depicted in Figure 6. Log odds of inquiry are shown along the vertical axis. Year of publication is indicated along the horizontal axis. The three separate lines within the body of the graph represent the three subject areas, as indicated on the figure.

The interaction between subject area and position in paragraph is shown in Figure 7. Log odds of inquiry are shown along the vertical axis. Position of the sentence within a paragraph is indicated along the horizontal axis, as in Figure 4. The three lines within the body of the graph represent the three subject areas, as indicated on the figure.

Figures 8 and 9 both show the interaction between the year of publication and the position in chapter. The interaction was plotted in two different ways because interactions involving two variables with four levels of each variable can be confusing to interpret. Viewing the interaction plotted both ways can sometimes alleviate some of the confusion. In both of the graphs log odds of inquiry are shown along the vertical axis. In Figure 8 the year of publication is indicated along the horizontal axis. The four lines within the body of the graph represent the four quartiles within the chapter, as indicated on the figure.

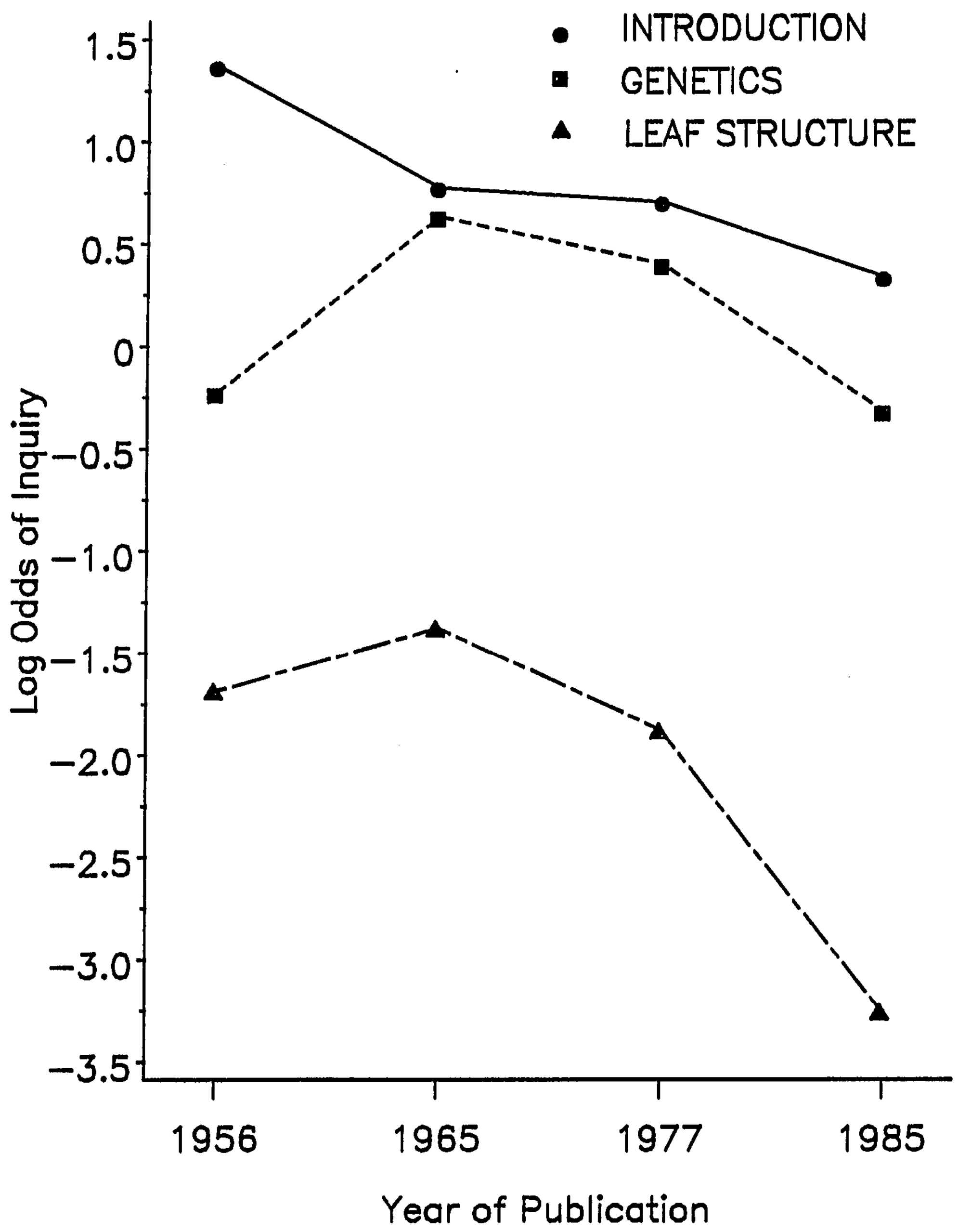

FIGURE 6. Interaction between year of publication and subject area

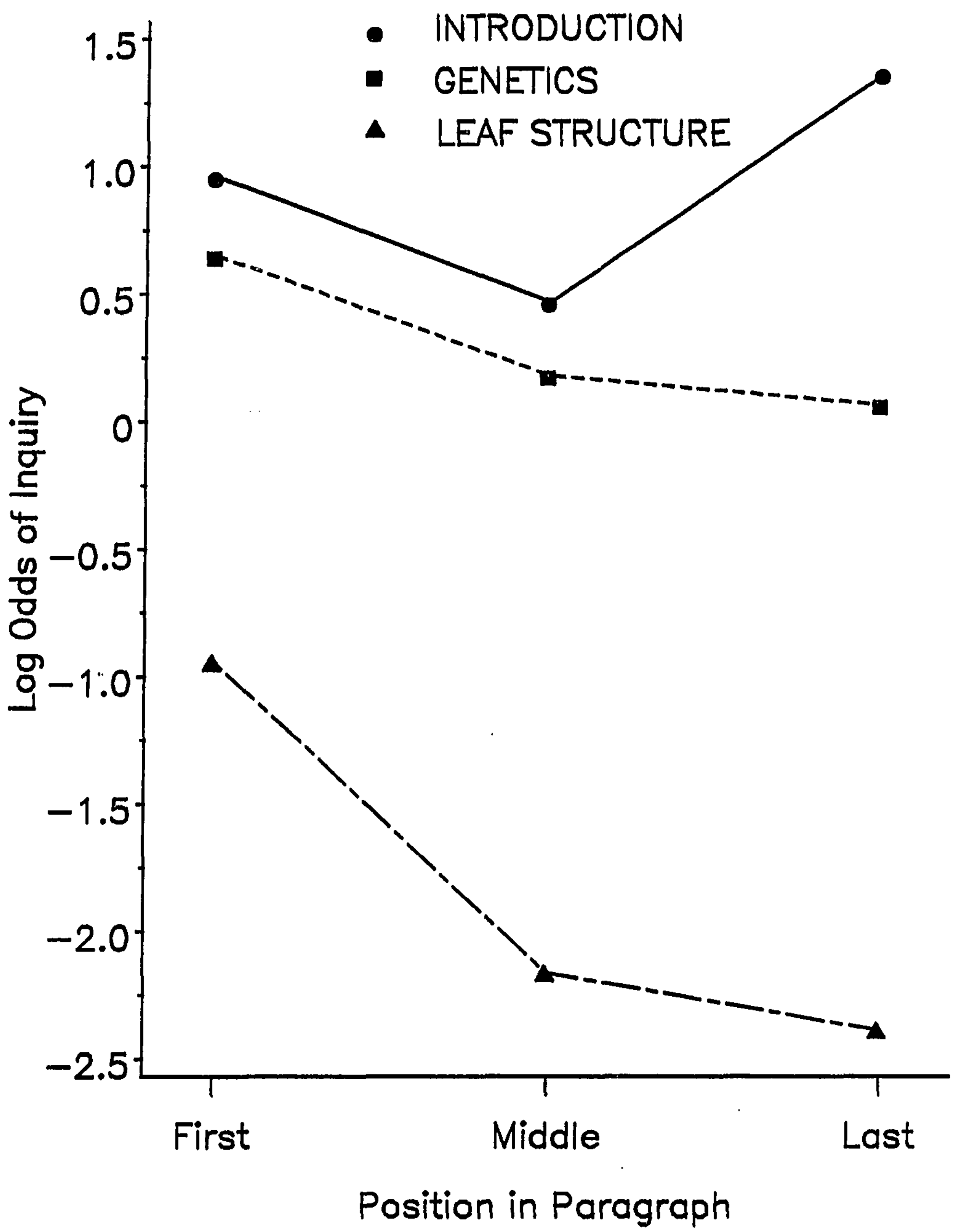

FIGURE 7. Interaction between subject area and position in paragraph

In Figure 9 the quartile, or position in chapter, is indicated along the horizontal axis while the four lines represent the four years of publication, as indicated on the figure.

Adjusted log odds were obtained by multiplying two times the appropriate additive parameters obtained from an appropriate log-linear model (see Knoke and Burke, 1980). The appropriate log-linear model for these graphs was one which modeled cell frequencies as a function of the six variables in the logistic regression model (one dependent and five independent variables), all two-way interactions of the variable "inquiry" with the five independent variables, three three-way interactions involving the variable "inquiry" and the variables in the significant interactions shown in Table 5, as well as all possible five-way and lower interactions among the five independent variables. The additive parameters used to generate the graphs shown in Figures 10-18 were the parameters of the interactions involving the variable "inquiry" and the independent variables. These parameters represent deviations from the average log odds of inquiry for all the clauses in the sample. Since all clauses fall into one and only one level of each independent variable, the unique deviations of each level from the overall mean sum to zero for each variable.

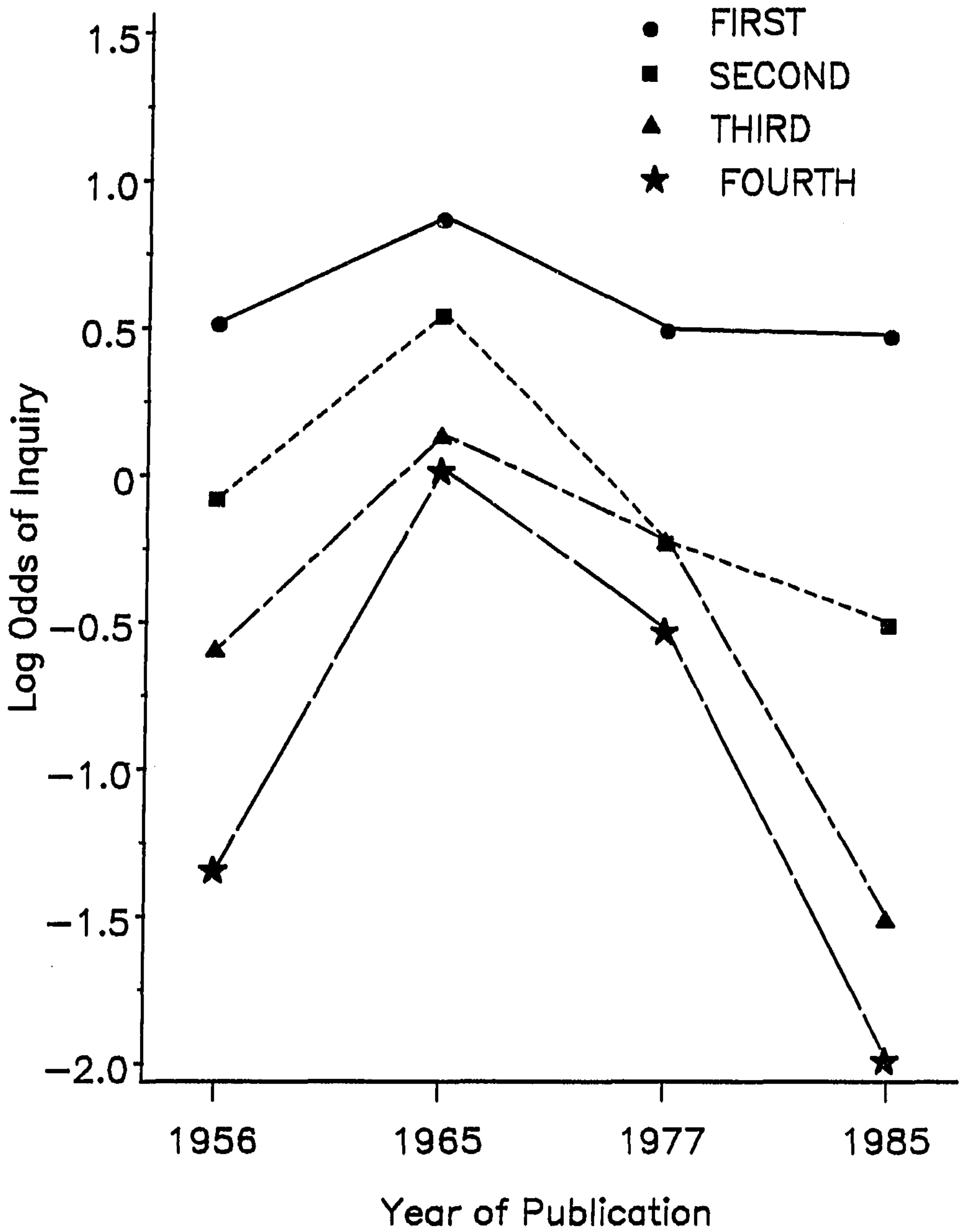

FIGURE 8. First graph of the interaction between year of publication and position in chapter

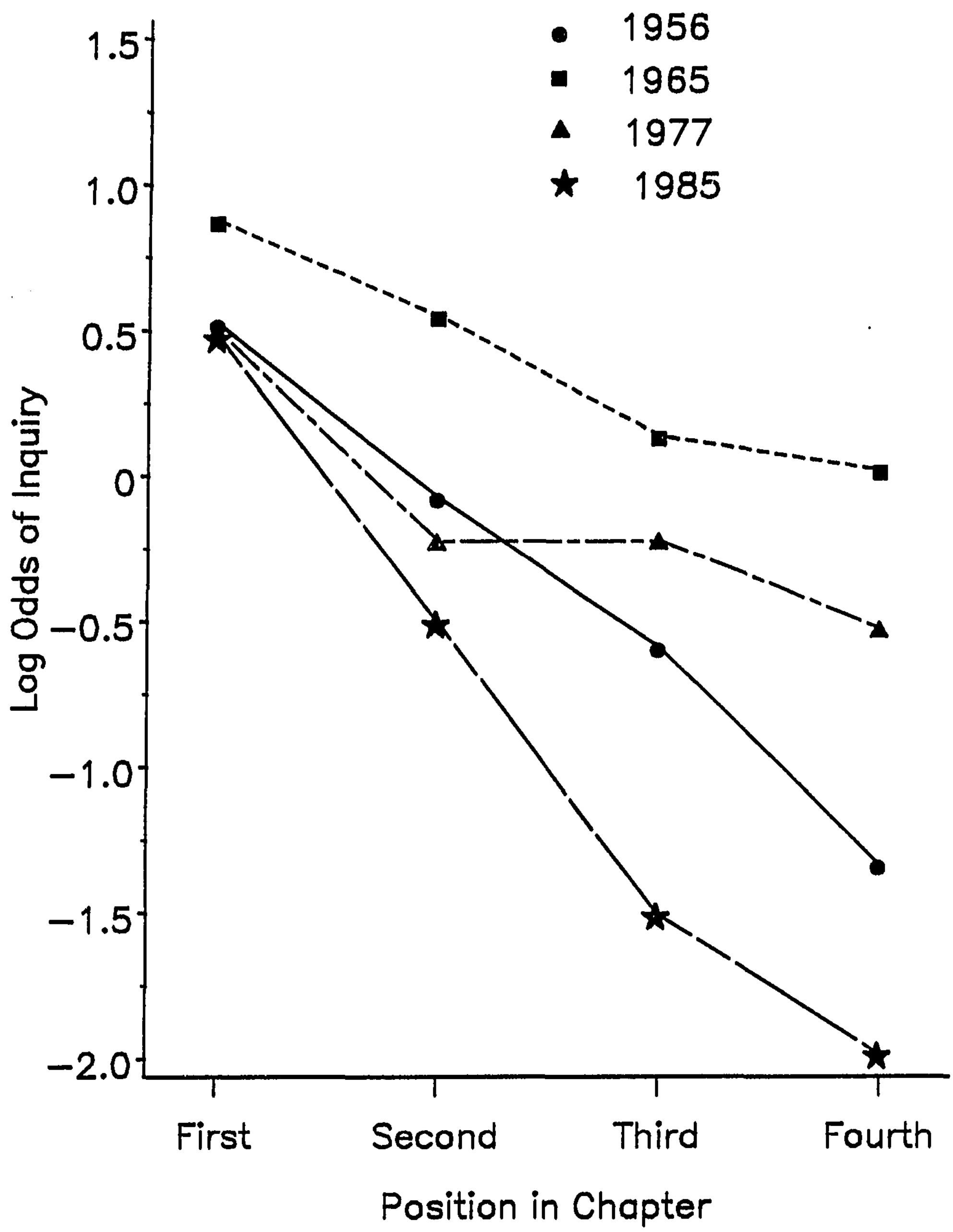

FIGURE 9. Second graph of the interaction between year of publication and position in chapter

These log odds can be interpreted as "adjusted" log odds, meaning they take into account the effects of the other variables in the model. This means the effects plotted are effects above and beyond any effects attributable to other terms in the model. This interpretation is especially important when studying the graphs depicting the interaction terms in the model.

Figure 10 shows a plot of adjusted log odds of inquiry versus year of publication. Adjusted log odds of inquiry are shown on the vertical axis while year of publication is shown on the horizontal axis.

Figure 11 shows a plot of adjusted log odds of inquiry versus subject area. The adjusted log odds of inquiry are plotted along the vertical axis. The three subject areas are indicated along the horizontal axis.

The change in adjusted log odds of inquiry with respect to the position of the sentence within a chapter is shown in Figure 12. As in Figure 3, the continuous variable indicating position in chapter was divided into a categorical variable consisting of four quartiles. The adjusted log odds of inquiry in each of the four quartiles are shown on the vertical axis, while the four quartiles are indicated along the horizontal axis.

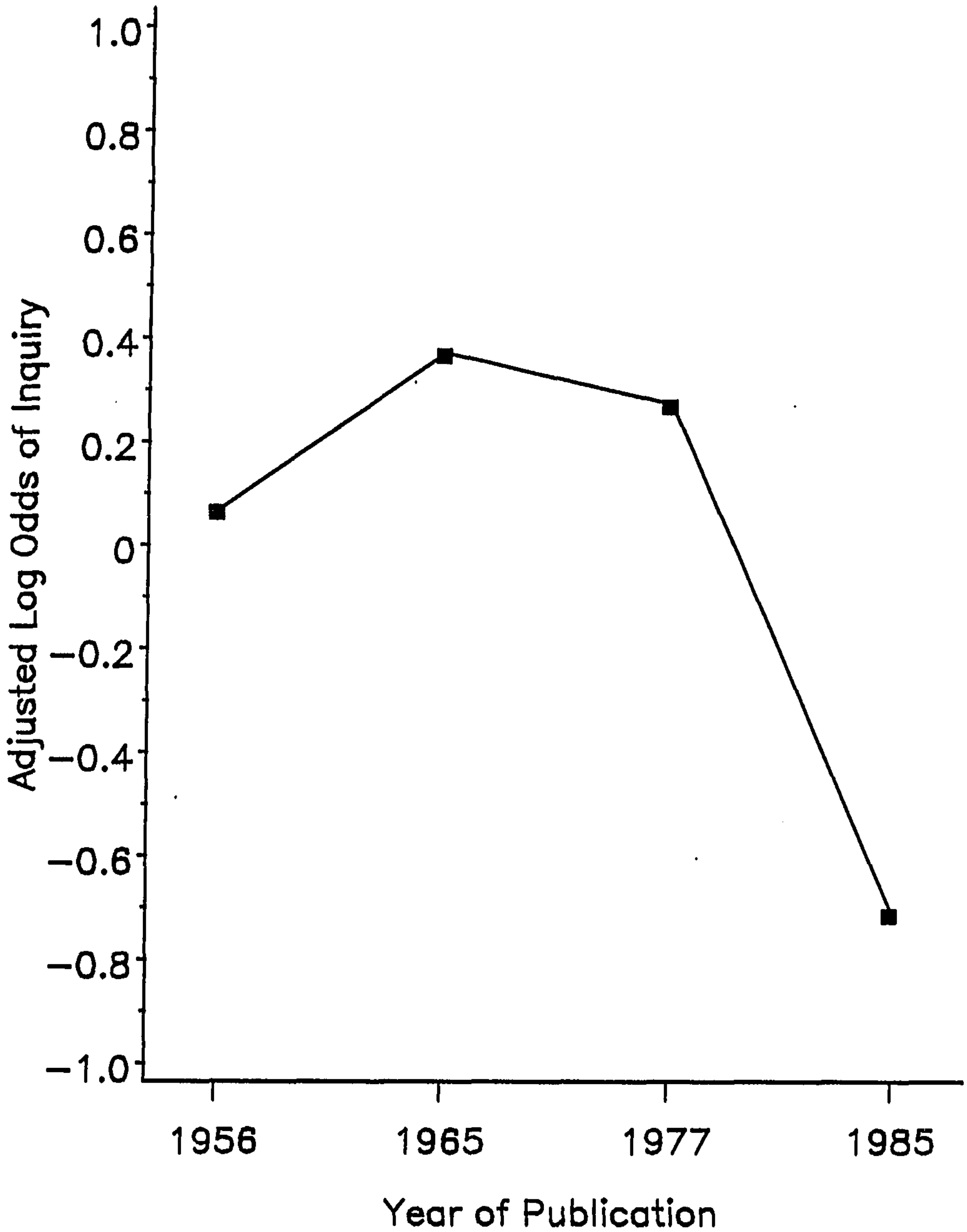

FIGURE 10. Adjusted log odds of inquiry versus year of publication

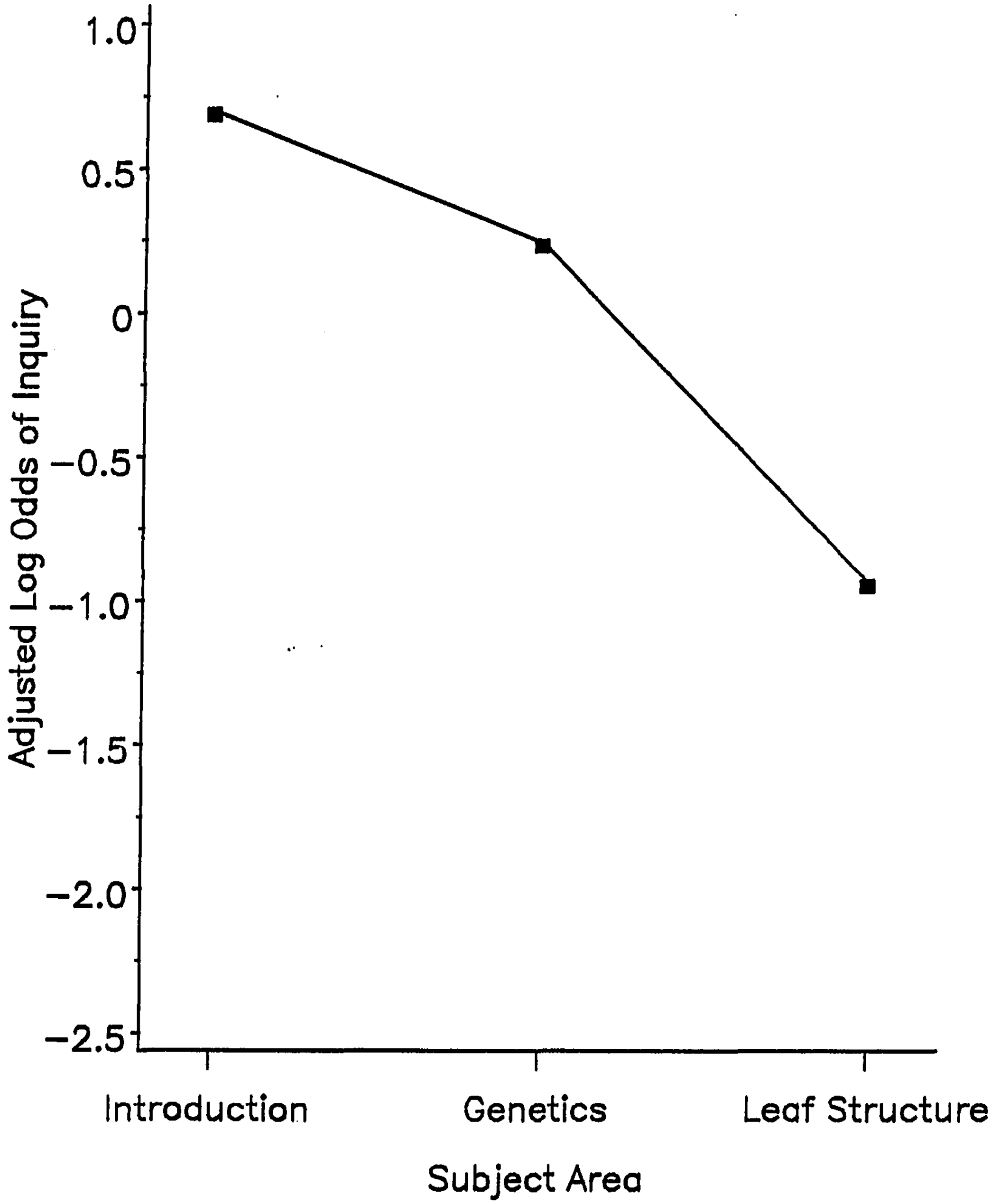

FIGURE 11. Adjusted log odds of inquiry versus subject area

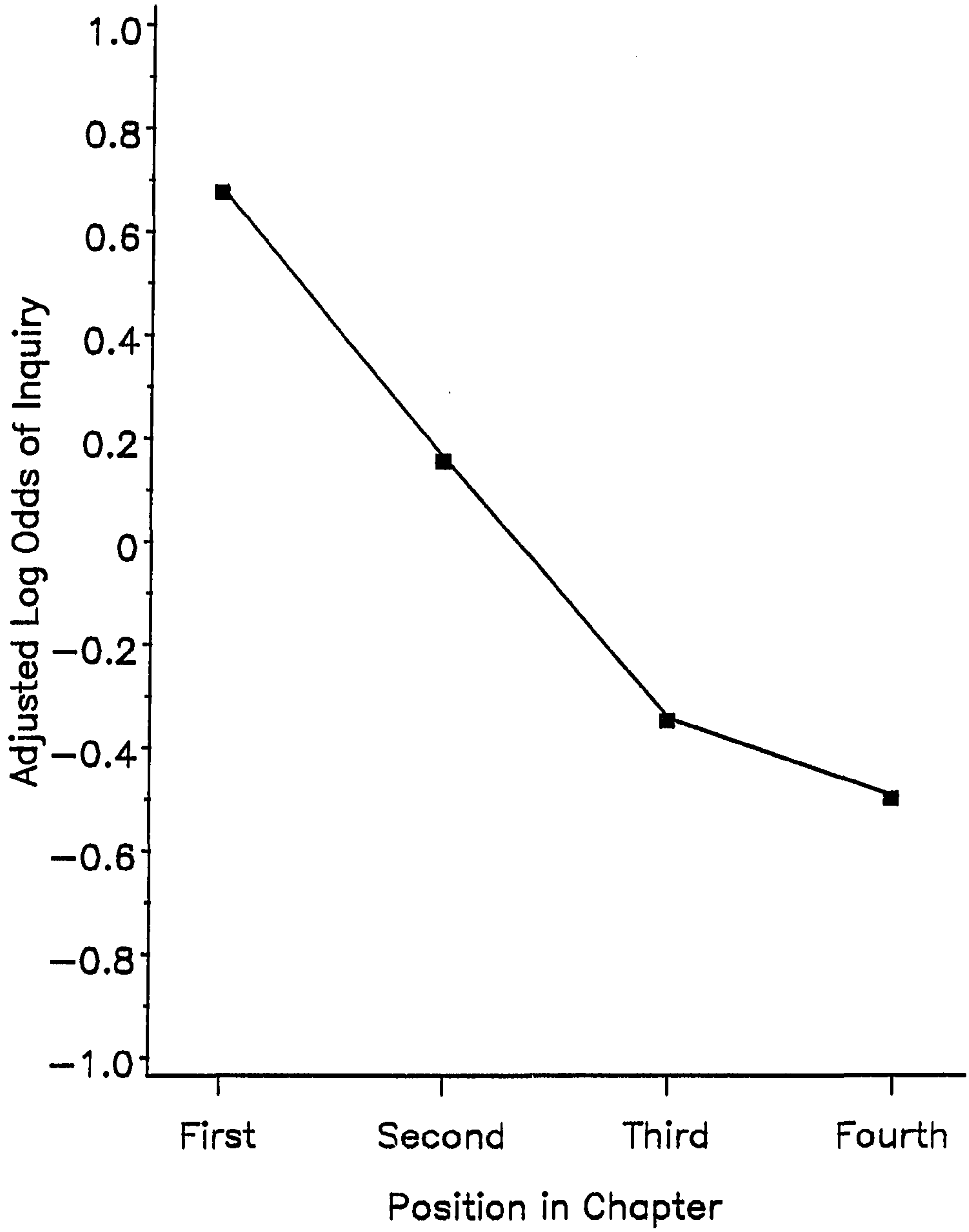

FIGURE 12. Adjusted log odds of inquiry versus position in chapter

The change in adjusted log odds of inquiry with respect to the position of the sentence within a paragraph is shown in Figure 13. The adjusted log odds of inquiry are shown along the vertical axis. The three possible positions of sentences within a paragraph are shown along the horizontal axis. These positions are the same as for Figure 4.

Figure 14 shows the change in adjusted log odds of inquiry in clauses with and without technical words. Adjusted log odds of inquiry are shown along the vertical axis, while presence or absence of technical words in a clause is indicated along the horizontal axis.

The interaction between year of publication and subject area using adjusted log odds is depicted in Figure 15. Adjusted log odds of inquiry are shown along the vertical axis. Year of publication is indicated along the horizontal axis. The three separate lines within the body of the graph represent the three subject areas, as indicated on the figure.

The interaction between subject area and position in paragraph using adjusted log odds is shown in Figure 16. Adjusted log odds of inquiry are shown along the vertical axis. Position of the sentence within a paragraph is indicated along the horizontal axis, as in Figure 4. The three lines within the body of the graph represent the three subject areas, as indicated on the figure.

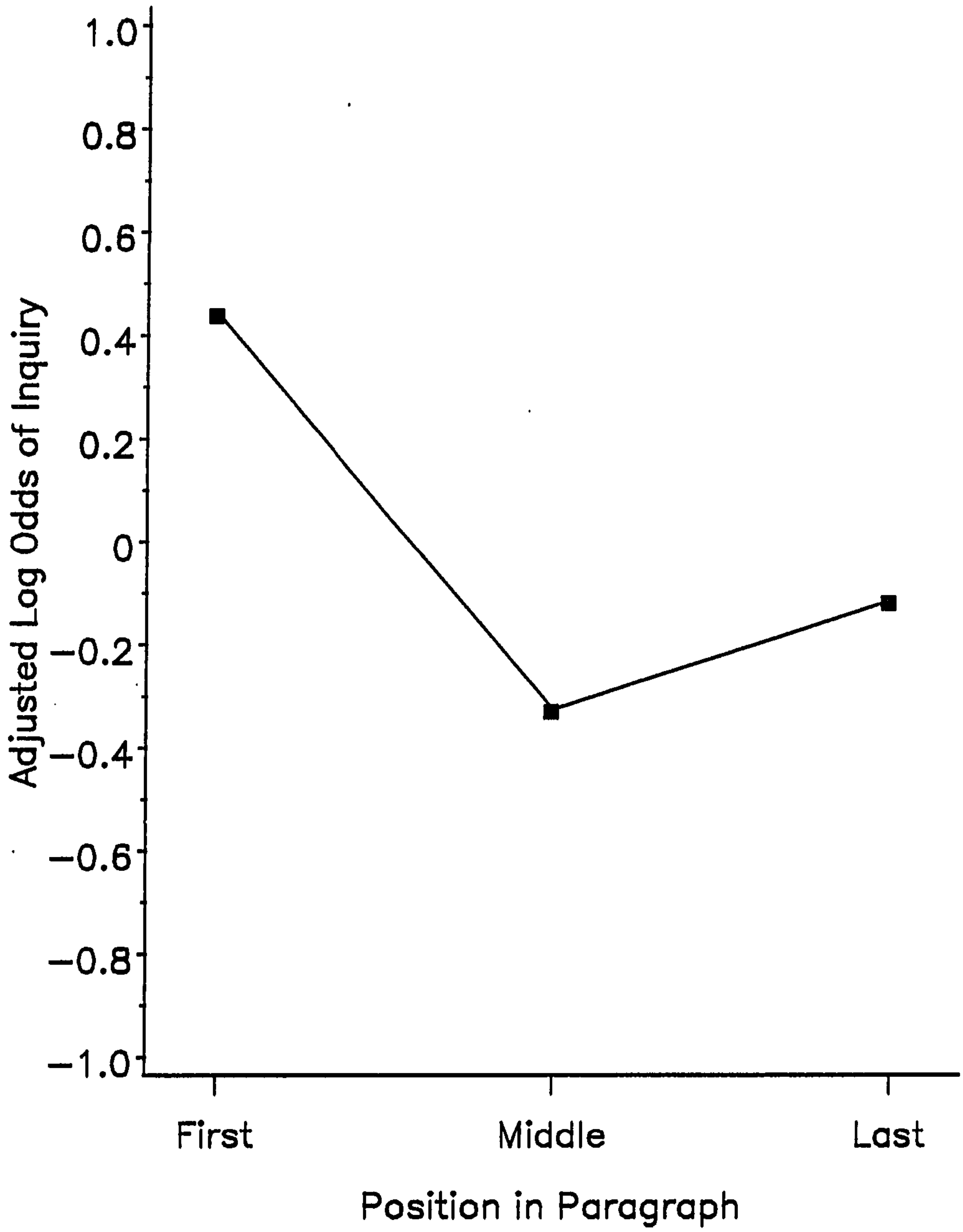

FIGURE 13. Adjusted log odds of inquiry versus position in paragraph

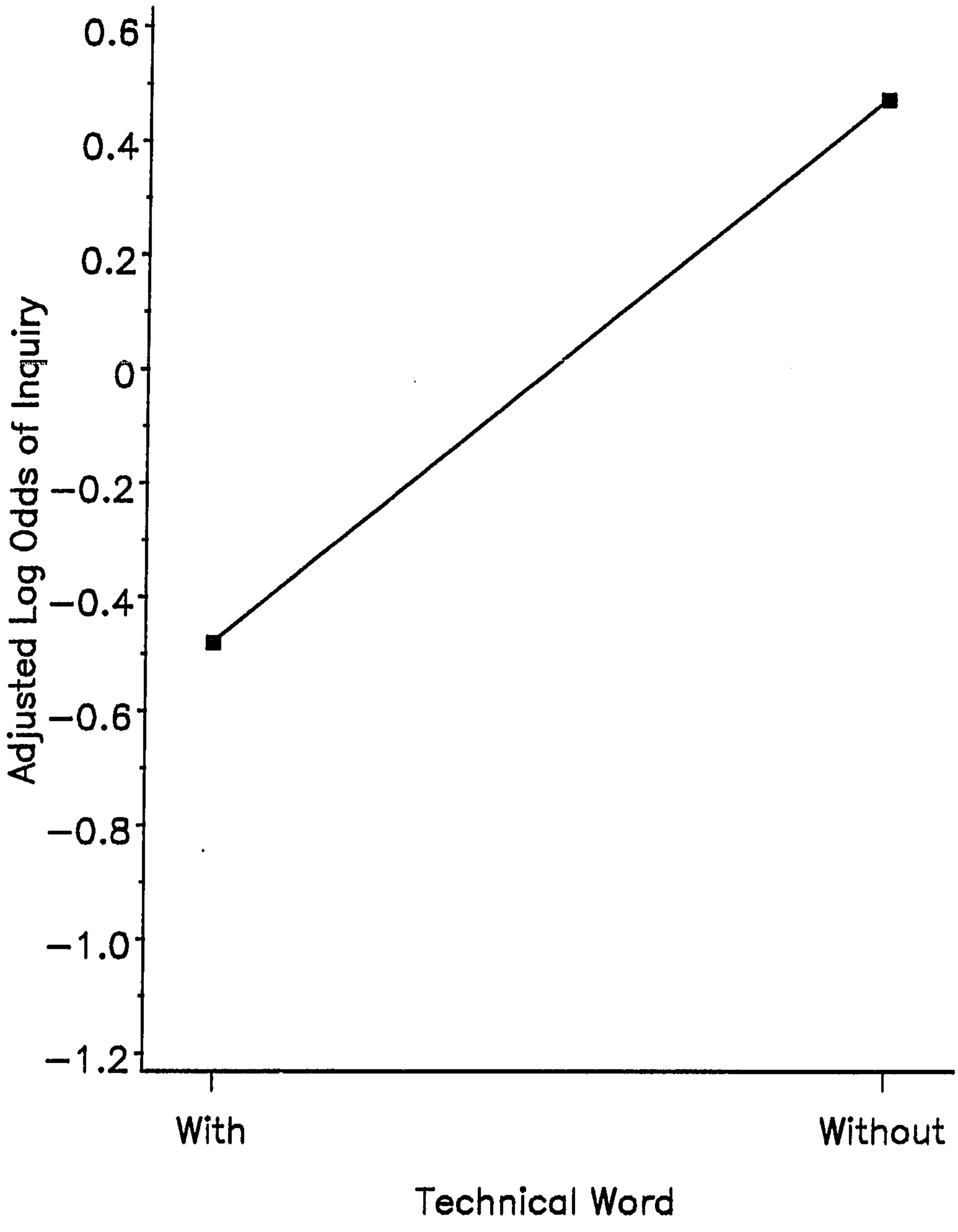

FIGURE 14. Adjusted log odds of inquiry versus use of technical word

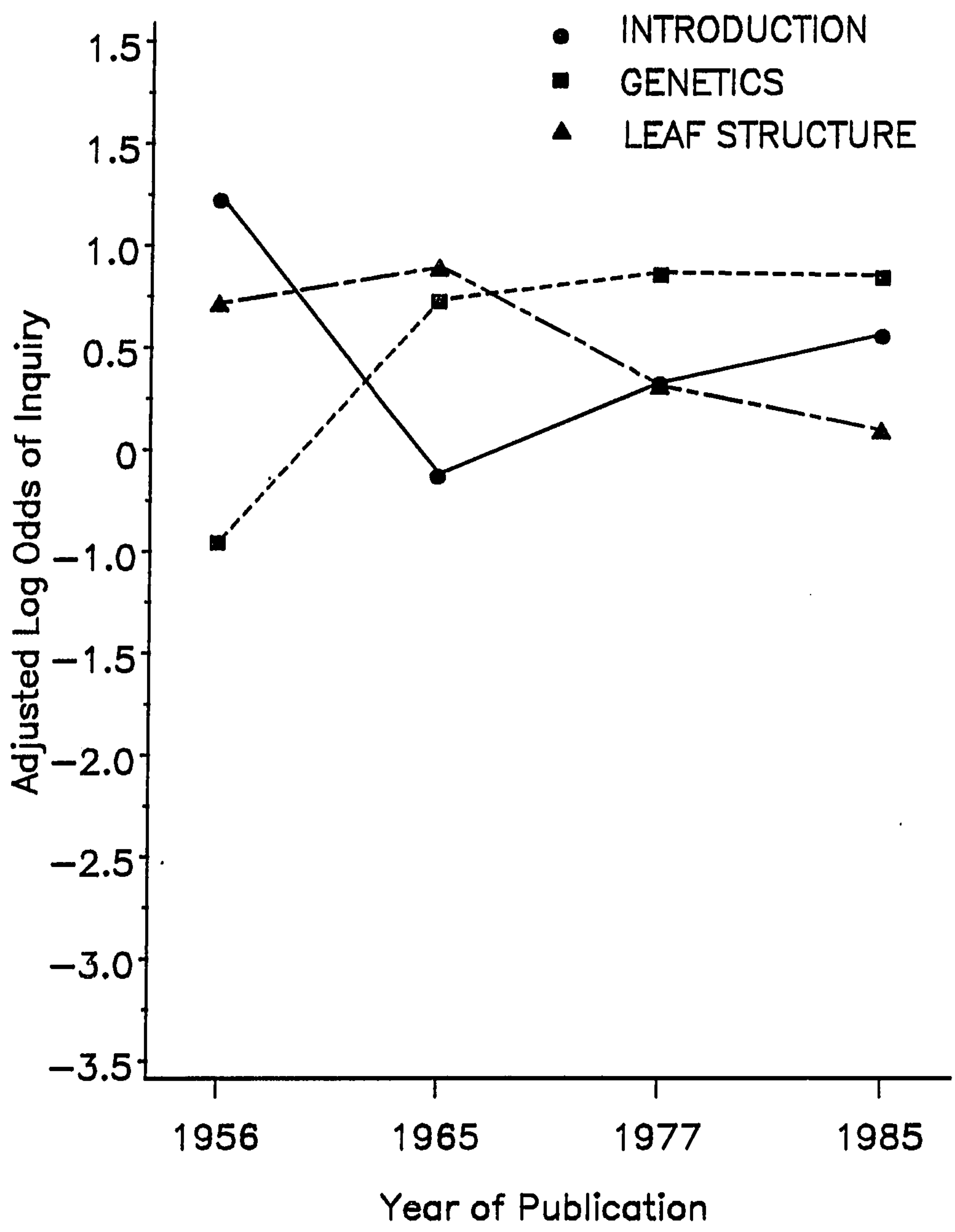

FIGURE 15. Interaction between year of publication and subject area using adjusted log odds

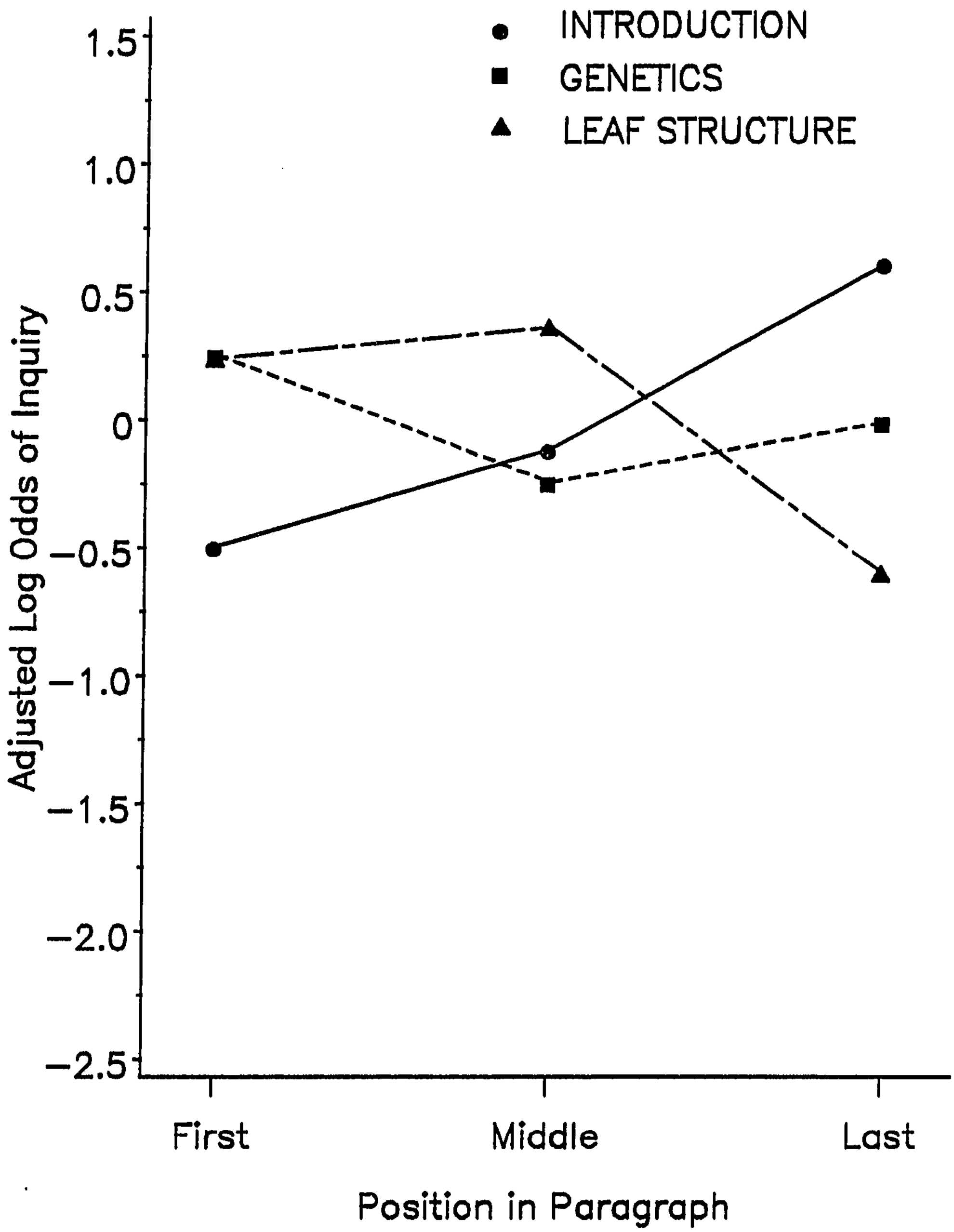

FIGURE 16. Interaction between subject area and position in paragraph using adjusted log odds

Figures 17 and 18 both show the interaction between the year of publication and the position in chapter. As with the previous treatment of this interaction, both forms of the graph were plotted to aid in the interpretation of the interaction. In both of the graphs adjusted log odds of inquiry are shown along the vertical axis. In Figure 17 the year of publication is indicated along the horizontal axis. The four lines within the body of the graph represent the four quartiles within the chapter, as indicated on the figure. In Figure 18 the quartile, or position in chapter, is indicated along the horizontal axis while the four lines represent the four years of publication, as indicated on the figure.

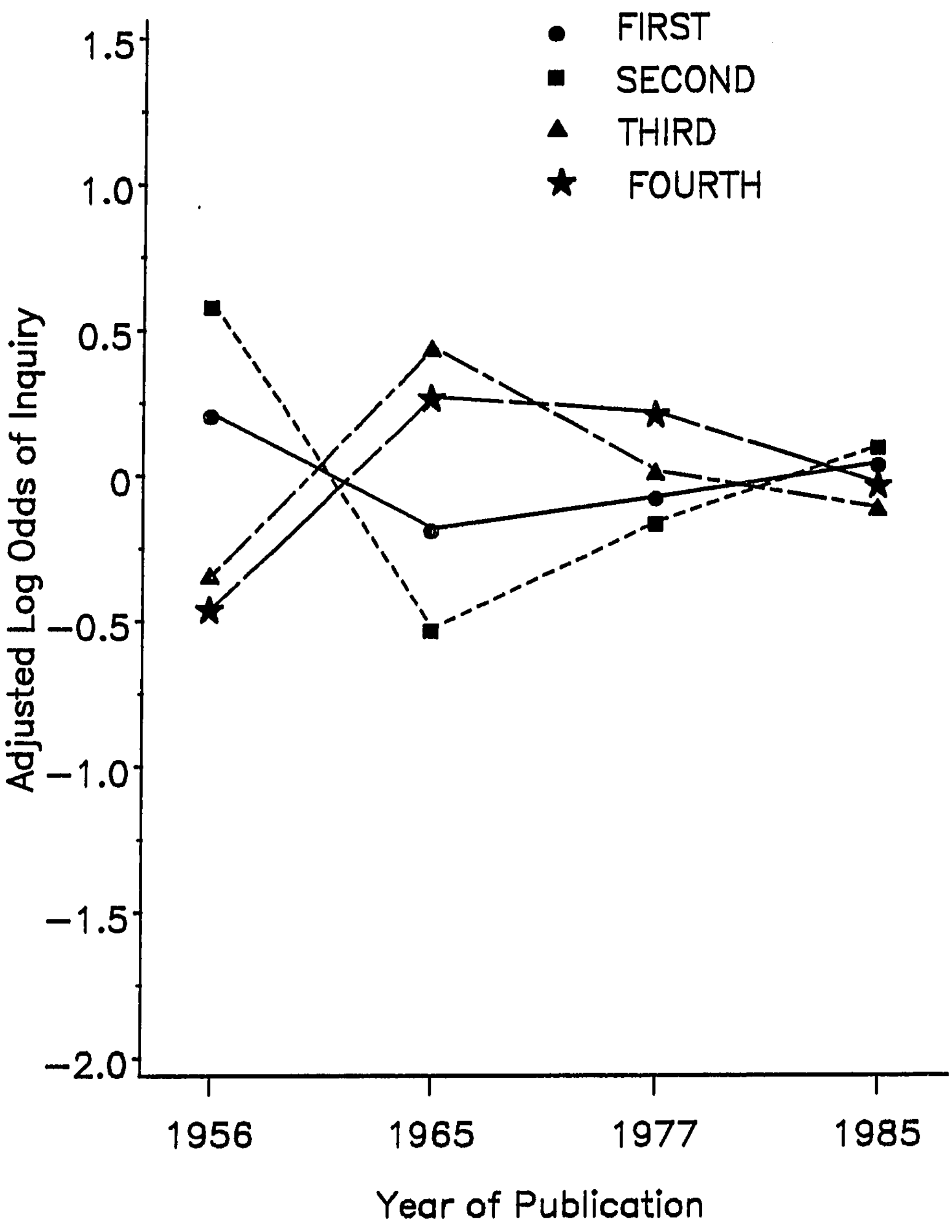

FIGURE 17. First graph of the interaction between year of publication and position in chapter using adjusted log odds

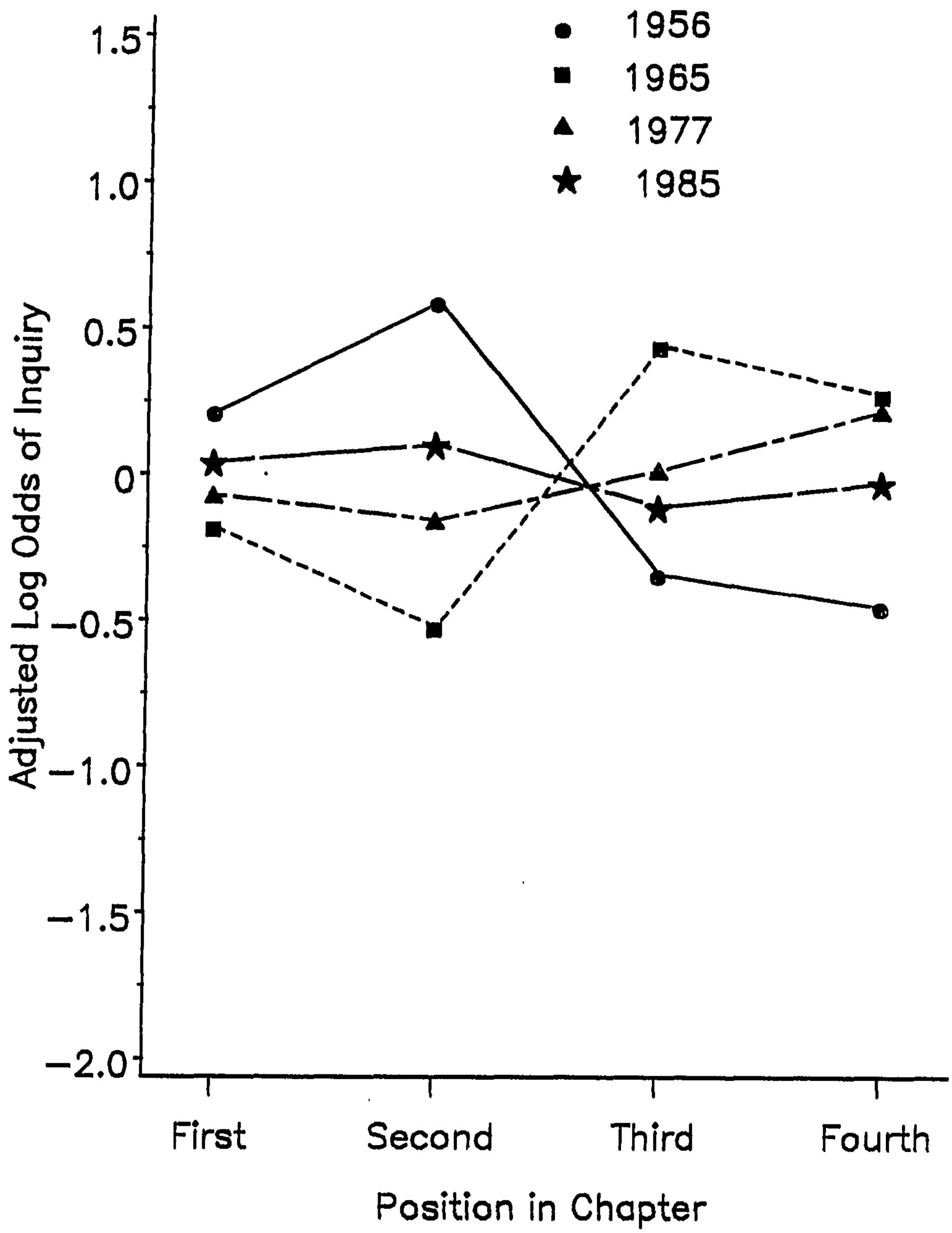

FIGURE 18. Second graph of the interaction between year of publication and position in chapter using adjusted log odds

CONCLUSIONS

The results of analyzing only the perception clauses are very similar in pattern to those obtained when analyzing the perception and recognition clauses combined. This might be expected because the majority (roughly 81%) of the clauses are perceptions. Therefore the influence of the perception clauses is dominant in the combined data set. Because the two approaches showed such similar results, the discussions about the conclusions drawn on the six proposed research questions will be combined for both of the approaches. It should be noted when comparing the plots of the log odds and adjusted log odds for a given main effect or interaction term the scales of the plots are the same. However, comparisons within Figures 1-9 or Figures 10-18 may involve a change in scale.

Answers to the Research Questions

Does level of inquiry vary across years?

Level of inquiry does vary across years. This variable was statistically significant when studying the perception and recognition clauses combined, and also when studying the perception clauses only. In both cases, the p-values associated with the variable "year" were less than 0.0001.

The pattern of the response in the level of inquiry across the four years studied was very similar for both log odds and adjusted log odds (see Figures 1 and 10). Level of inquiry was lower at the two end-points in the study (1956 and 1985) and higher at the two mid-points (1965 and 1977), with a peak in level of inquiry in 1965. This trend is of particular interest when one considers the fact of the large amount of national attention being paid to science education during the early 1960s. As mentioned in the first chapter, there was a large, federally funded curriculum effort which resulted in the publication of three Biological Sciences Curriculum Studies textbooks (BSCS green, BSCS yellow and BSCS blue versions). A major theme of the BSCS materials was the treatment of science as inquiry. This theme was seen as one of the most radical departures of the BSCS textbooks from conventional patterns (Schwab, 1963). Although it was not the original intent of the BSCS curricula, these materials did seem to serve as a model for the other biology textbooks, in that the format of other textbooks soon adopted that of the BSCS textbooks (Grobman, 1988, past director of the BSCS project).

One possible theory for explaining the change in level of inquiry across the years in the Holt biology textbook series could be a response to the influence of the BSCS textbooks. Pilot versions of the BSCS textbooks were released in 1960, and the first editions were published in 1962. Holt published editions of its Modern Biology textbook in 1960, 1963 and 1965. Allowing two years for preparation of a revised textbook, the 1963 or 1965 editions are the first editions where one could anticipate detecting possible changes due to the BSCS influence. The fact that the 1965 edition was published after such a short interval might indicate an effort on the part of the Holt publishers to keep abreast of current changes in the field, suggesting that the 1965 edition would be the first one most likely to reflect any BSCS influence.

The drop in the level of inquiry from 1965 to 1985 could be attributed to several factors. One explanation could be that while Holt eagerly embraced the ideas put forth by the BSCS project, such as science as inquiry, these ideas gradually lost in favor. Perhaps some of the ideas sounded new and appealing, but when tried in the actual market they were not widely accepted. There could be various reasons for this possible lack of acceptance. For example, teachers are not trained to present science as inquiry, and they are much more likely to present the facts of science (Duschl, 1986; Harms, Bybee, Kahle, and Yager, 1980; Herron, 1971). Teachers are not trained to be "model inquirers" for their students. Biology teachers also lecture more than 75% of the time, leaving little time for inquiry. Also, many biology teachers feel that only highly motivated, gifted students can benefit from using a great deal of inquiry (Hurd, Bybee, Kahle and Yager, 1980). Therefore, teachers may be more critical of a textbook which shows a great deal of science as inquiry. Holt, in trying to be sensitive to the requests of the classroom teacher, may have altered the textbook to show less inquiry.

Another possible reason for the decline in the treatment of science as inquiry from 1965 to 1985 could be the staggering amount of scientific knowledge which has accumulated during this time period. With such an increase in scientific knowledge, it becomes very difficult to decide what should and should not be included in a science textbook (Trump and Miller, 1979). It is possible that the ever-growing body of facts is judged as more important to include in textbooks than explanations as to the processes of acquiring these facts. This would certainly be true if textbooks were intended to be used mainly as a reference source, and not as the main curriculum guide. Also, as noted previously, facts are much easier to teach and to test for than are processes of inquiry.

A third reason for the decline in the level of science as inquiry after 1965, and particularly for the very low level of science as inquiry in 1985 could be due to the influence of the "back to basics" movement of the 1980s. Many national reports in education published during the 1980s seem to recommend an increased focus on the basic skills of learning (see for example: National Commission on Excellence in Education, 1983; Boyer, 1983; Conference on Goals for Science and Technology Education, Grades K-12, 1983; National Science Board Commission on Precollege Education in Mathematics, Science and Technology, 1983). This "back to basics" movement could also mean a movement away from teaching science as inquiry and toward a more conservative approach to science teaching.

The change in level of inquiry in textbooks across years, therefore, could reflect a change in attitudes toward teaching science as inquiry. In the 1950s the attitudes in science education were more supportive of descriptive science (Nietz, 1966; Reynolds, 1967). During the 1960s, the movement in science education was directed more at creating a collaboration between teachers, curriculum developers and scientists. It was hoped that this collaborative effort would greatly improve the quality of the science curricula by creating closer ties between what is taught in a science classroom and what occurs within the domain of science (Schwab, 1963; Shymansky, 1984; Shymansky, Kyle and Alport, 1983). The curriculum movements of the 1960s also emphasized inquiry as a goal of biology education (Grobman, 1969; Hurd, Bybee, Kahle and Yager, 1980; Schwab, 1963).

During the 1980s, while the goals of the 1960s are still in evidence, the emphasis has changed (see for example *Conference on Goals for Science and Technology Education, Grades K-12*, 1983; Hurd, Bybee, Kahle and Yager, 1980). The applications of science in society have become more prevalent in the 1980s (Bybee, 1985; Hickman and Kahle, 1982; Hofstein and Yager, 1982; McConnell, 1982). The basic goal of the importance of inquiry in science has been maintained, but less emphasis is placed on it. The following quote perhaps sums up one of the current views of the role of inquiry in biology education:

> So it seems that inquiry is not perceived to be as important as curriculum developers thought; it is not an efficient teaching method; and biology teachers do not understand inquiry either as a scientific process or as a teaching method (Hurd, Bybee, Kahle and Yager, 1980, p. 391).

Such reasoning, and the results of the changes in level of inquiry across years of publication, suggest that the Holt textbook series could be reflecting at least some of the goals put forth by science educators.

Does level of inquiry vary across subject areas studied?

Level of inquiry exhibited strong changes across the three subject areas studied. In both of the analyses this variable was associated with the largest decrement in chi-square (156.96 for perception and recognition clauses combined and 120.27 for perception clauses only, with two degrees of freedom in both cases). The p-values associated with both of these chi-square values are less than 0.0001.

The pattern in the response of the level of inquiry across subject areas was very similar when plotting either log odds or adjusted log odds (see Figures 2 and 11). The level of inquiry was highest in the introduction chapters, and lowest in the leaf structure chapters. Level of inquiry in the genetics chapters was lower than in the introduction chapters, but still much higher than in the leaf structure chapters.

This result is to be expected, as the genetics and leaf structure chapters were specifically selected to serve as contrasts, one to be high on the measure of inquiry and one to be low on the measure of inquiry. It was hypothesized that the genetics material would be high on the measure of inquiry. Due to the great amount of scientific work which has been done in the field of genetics over the past thirty years, there is a lot of opportunity to demonstrate science as inquiry in this particular area. It was hypothesized that the leaf structure material would be very low on the measure of inquiry because this material tends to be highly descriptive. Most of the chapter is devoted to describing the parts and the functions of the plant and its leaves. The results shown in Figures 2 and 11 lend support to the idea that the dependent variable is a valid measure of the concept "inquiry," because the results follow the planned contrast.

The fact that level of inquiry is highest in the introduction chapters could be due to several reasons. One possible interpretation could be that the authors of the textbook started the book with an approach high in inquiry, but did not carry out this approach across subsequent chapters. This might indicate a certain lack of commitment on the part of the authors to the treatment of science as inquiry. It could be that the introduction chapters are used to explain the processes of science, and the remaining chapters are devoted to the products of science. In the introduction chapters, a certain amount of historical background is presented, along with the "steps" of the "scientific method." These steps include the formulation and testing of hypotheses. The scientific method presented in the first chapter is usually not mentioned throughout the rest of the book. When the overall techniques of science are discussed, science is shown as a process of inquiry; but when the specific details of science are discussed, science is shown as a body of knowledge.

<u>Does level of inquiry vary with respect to the position of the sentence within a chapter?</u>

In both the analysis of combined clause types and perception clauses only, level of inquiry did vary with respect to the position of the sentence with a chapter. The chi-square values associated with this variable were both statistically significant with p-values less than 0.0001. The pattern of this response was similar when plotting either log odds or adjusted log odds, as can be seen in Figures 3 and 12. The level of inquiry showed an almost linear trend across the position in the chapter. Level of inquiry was highest at the beginning of the chapters and lowest at the end of the chapters. This pattern could be a reflection of a tendency to present historical background of scientists and their work at the beginning of the chapter as an introduction to the rest of the material in the chapter. This historical material is often followed by descriptive, factual information. Because the measure of inquiry was designed to detect historical background and mentioning of scientist names (as found in Tamir, 1985), it would be higher at the beginning of chapters.

Does level of inquiry vary with respect to the position of the sentence within a paragraph?

Level of inquiry does vary with respect to the position of a sentence within a paragraph. In the analysis combining both the perception and recognition clauses the variable was significant at the 0.01 level of significance ($p = 0.0046$). In the analysis of the perception clauses only, the chi-square value associated with position in paragraph was significant at the 0.05 level of significance ($p = 0.0177$). This difference in the statistical significance across the two analyses is rather small and could be due to random fluctuations.

The pattern of response of level of inquiry across the three paragraph positions are similar when plotting either log odds or the adjusted log odds (see Figures 4 and 13). The level of inquiry is highest at the beginning of the paragraphs. There seems to be a slight increase in level of inquiry from the middle to the last position in the paragraph, but the increase is not large enough to be meaningful. The higher level of inquiry at the beginning of the paragraphs could be in part a factor of writing style. In some cases questions are used at the beginning of paragraphs to help generate interest in the material to follow. Also, a paragraph may start with a phrase such as

"You will recall from the previous chapter that..." Both of these examples would have been counted as inquiry.

It is interesting to note that, while the variable position in paragraph is not as statistically significant as the other variables, it does seem to be a micro-example of a larger pattern throughout the textbooks. In particular, within a book, the level of inquiry was higher in the introduction chapters, which are at the beginning of the book. Within a chapter, the level of inquiry was higher in the beginning sections. Within a paragraph, the level of inquiry was higher in the beginning sentences.

Does level of inquiry vary with the use of technical words?

In both the analysis of the combined clause types and the perception clauses only, the variable indicating the presence of technical words was statistically significant ($p < 0.0001$ for both analyses). The patterns of the responses were similar in plots of both the log odds and the adjusted log odds (see Figures 5 and 14). The level of inquiry was lower in clauses which contained technical words. This result seems reasonable, because clauses with technical words tend to be descriptive, and hence not to exhibit inquiry. An example of such a clause would be "the radiation caused a mutation." Both of the terms "radiation" and "mutation" would have been recorded as being "science words" and hence "technical words." This clause is descriptive and does not contain any of the indicators of inquiry which compose the measure of inquiry.

Does level of inquiry exhibit any interactions with the five main effects variables?

With five main effect terms, there are ten possible two-way interaction terms. Of these ten possible interaction terms, five were selected as being potentially statistically significant using log-linear analysis. The statistical significance of these five interaction terms in a logistic regression model was tested by sequentially deleting them from the model (see Table 5). Based on the results shown in Table 5, three interaction terms were selected for inclusion in the model. It should be noted at this point that the testing of these interaction terms is exploratory in nature, and is not based on any previous theoretical considerations.

Year of publication-by-subject This interaction term was statistically significant at the 0.01 level of significance for both the analysis using recognition and perception clauses and the analysis using perception clauses only ($p = 0.0019$ and $p = 0.0004$ respectively). The plot of the log odds for this interaction is shown in Figure 6, and the plot of the adjusted log odds is shown in Figure 15. As can be seen, these two plots are very different. This difference is due to the fact that the values plotted in Figure 15 have been adjusted for the other terms in the model. These adjusted plots do not represent the log odds of overall levels of inquiry for combinations of categories from pairs of independent variables. They represent differences in the log odds of inquiry among combinations of categories of two independent variables, where these log odds are in addition to the log odds one would estimate based upon all other marginal and interaction effects included in the logistic regression model. Perhaps the most statistically valid way to interpret this interaction term (as well as the other two interaction terms) is to focus on the plots of adjusted log odds, rather than log odds, since these plots depict the interaction effects that were actually estimated in the model.

It will be helpful at this point to consider two possible approaches to the presentation of science as inquiry in textbooks. One approach would be to stress science as a process of inquiry in the beginning, and then to decrease the level of inquiry toward the end. The second approach would be to present moderate to low amounts of inquiry in the beginning, and to build on these levels as one proceeds further into the material. These two approaches could possibly reflect different assumptions about students.

The literature on inquiry implies that depiction of science as inquiry is a more accurate representation of what science actually is (Herron, 1971; Schwab, 1963). One could also imagine that depiction of science as inquiry, rather than of science as a collection of facts, would make the subject matter of science more interesting, as well as more realistic, for students. These two possible approaches to the placement of inquiry in science textbooks may reveal different assumptions of the student audiences of these texts. The first approach, that of high initial inquiry, could be intended for students with low motivation, who need an initial "hook" to get them interested. Once the students' interest is obtained the effort to promote inquiry decreases. The second approach, that of low initial inquiry, could be intended for highly motivated students, who may not require an initial burst of inquiry, but do require that certain levels of inquiry be maintained throughout the treatment of the material.

Evidence of shifts between these two approaches can be seen in Figure 15. The first approach, that of high initial inquiry, is seen in 1956, where the relative level of inquiry was highest in the introduction chapter, somewhat lower in the leaf structure chapter, and lowest in the genetics chapter. The approach shows a marked reversal in 1965. In the 1965 text the relative level of inquiry seems to follow the second approach, namely that of low initial inquiry. The relative level of inquiry is lowest in the introduction chapter, but higher in the two subject matter chapters. The shifts in 1977 and 1985 are progressively in the direction of the log odds of inquiry exhibited in 1956, in that the relative level of inquiry in the introduction chapters is progressively higher in these years, but the relative level of inquiry in the leaf structure chapters is progressively lower. The exception is the relative level of inquiry in the genetics chapters, which remains fairly stable from 1965 to 1985.

One other point does seem to come through from both Figures 6 and 15, and that is the point that the 1956 introduction chapter was unique in that it showed unusually high levels of inquiry. Based on informal observations, the 1956 introduction chapter did have a different format than the other introduction chapters. There were references to specific activities in which a student could apply inquiry skills, for example studying nature through interesting hobbies such as hiking. Students were encouraged to extend their study of biology from the classroom into their other activities. It is possible that the aspect of inquiry reflecting the involvement of the reader was higher in this chapter, causing the high total level of inquiry.

Subject-by-position in paragraph   This interaction term was statistically significant at the 0.05 level of significance in both analysis of the combined clause types and for the perception clauses only (p = 0.0076 and 0.0187 respectively). The plots for this interaction of log odds and adjusted log odds are shown in Figures 7 and 16. As with the plots of the year of publication-by-subject interaction, these two graphs are different. However, both graphs do show that for the introduction chapters, level of inquiry was highest at the ends of paragraphs, while for the two subject matter chapters, level of inquiry was higher at the beginning.

Again it could be helpful to consider the two proposed approaches to the treatment of inquiry, high initial inquiry or low initial inquiry. The introduction chapters seem to exhibit the low initial inquiry approach within paragraphs. The relative level of inquiry is lowest at the beginning of paragraphs and highest at the ends of paragraphs. The two subject matter chapters seem to roughly follow the approach of high initial inquiry within paragraphs. The genetics chapters have highest relative inquiry at the beginning of paragraphs, which levels off in the middle and ends of paragraphs. The leaf structure chapters have about the same initial relative level of inquiry as the genetics chapters at the beginning of paragraphs, then have a slight rise in the middle, followed by a rapid decrease in the end.

Depictions of science as inquiry are different in the introduction chapters versus in the two subject matter chapters. Not only is inquiry higher in the introduction chapters (see Figures 2 and 11), but it is also presented in a different manner. More inquiry is placed at the ends of paragraphs in the introduction chapters than in the subject matter chapters.

Year of publication-by-position in chapter This last interaction term was statistically significant at the 0.01 level of significance in the analysis of the recognition and perception clauses combined and in the analysis of the perception clauses only ($p = 0.0048$ and $p = 0.0036$ respectively). The plots for this interaction using log odds are shown in Figures 8 and 9, and the plots of this interaction using adjusted log odds are shown in Figures 17 and 18.

The plots of this interaction can be confusing. This confusion is mainly due to the fact that each of the variables involved have four levels, resulting in sixteen points on a graph. Because of the confusion in these graphs, they were plotted two ways for both log odds and adjusted log odds, first with year of publication along the horizontal axis and position in chapter in the body of the graph, and next with position in chapter along the horizontal axis and year of publication in the body of the graph.

In both Figures 8 and 17, those with year of publication plotted on the horizontal axis and position in chapter within the body of the graph, the lines representing the first and second quartiles seem to be roughly parallel, as do the lines representing the third and fourth quartiles. This could be interpreted as a difference in level of inquiry between the first and second halves of the chapter when compared across years of publication.

As with the previous interactions, the main focus for interpreting these graphs will be on the plots of adjusted log odds. One possible interpretation of Figure 17 would be to again consider the themes of high versus low initial inquiry. Again, as with the year of publication-by-subject interaction, the 1956 edition seems to be following the pattern of high initial inquiry. The relative level of inquiry is higher in the first half of the chapters, and lower in the second half of the chapters. As with the year of publication-by-subject interaction, there is a reversal of approaches between 1956 and 1965. The 1965 edition seems to follow the pattern of low initial inquiry. Relative inquiry is lower in the first half of the chapters and higher in the second half of the chapters. In the 1977 and 1985 texts depictions of science as inquiry varied little at different points within the chapters. There is a small reversal back to the pattern of high initial inquiry between 1977 and 1985, but it is not large enough to be considered significant.

In Figures 9 and 18 this same interaction is plotted with position in chapter along the horizontal axis and year of publication in the body of the graph. In Figure 18 the line depicting adjusted log odds of inquiry for the 1956 edition is almost a reflection of the line depicting adjusted log odds of inquiry for the 1965 edition. This figure further demonstrates that the 1956 edition exhibits the high initial inquiry approach while the 1965 edition exhibits the low initial inquiry approach. The lines depicting adjusted log odds of inquiry for 1977 and 1985 do not vary much across positions within the chapters. The 1977 edition does demonstrate a slight tendency toward the low initial inquiry approach, while the 1985 edition demonstrates a slight tendency toward the high initial inquiry approach, but these changes are very small.

Areas for Future Work

This study does not address the specific aspects of inquiry which are being detected. It could be of interest to perform a more detailed accounting to identify which individual indicators of inquiry are most prevalent within certain contexts. In particular, this could be useful in further clarifying the interactions, such as the year-by-subject interaction and the subject-by-position in paragraph interaction. A more detailed accounting of the measure of inquiry could suggest, for example, why the 1956 introduction chapter was unusually high on depiction of science as inquiry.

Other ares of future work arising from this study are due primarily to the limitations of the study. The sampling population, and hence the population to which inferences can be made, is the Holt *Modern Biology* textbook series. More rigidly defined, the sampling population is restricted to the chapters dealing with the introductory, genetic and leaf structure material in the 1956, 1965, 1977 and 1985 editions of the *Modern Biology* textbook series. One possible line of future work to pursue would be to expand the study into more editions of the *Modern Biology* series. Particular focus could be the treatment of inquiry in the editions published before and after the 1965 edition (1960, 1963, 1969 and

1973), since this is the time closest to the release of the BSCS materials and the BSCS materials have had such a strong impact on American biology education (Hurd, Bybee, Kahle and Yager, 1980).

Other studies might compare the Holt biology textbook series to the BSCS textbooks. One of the major stated themes in these books was the demonstration of the role of inquiry in science (Schwab, 1963). In trying to evaluate the impact of the BSCS textbooks on biology education, it is also important to carefully evaluate the BSCS textbooks themselves. It could thus be useful to use the same measure of inquiry developed for the Holt series on the BSCS textbooks, and actually determine if the BSCS textbooks demonstrate high levels of inquiry. In one comparison of high school biology textbooks, the BSCS books were rated as being generally higher in inquiry than the Holt textbooks (American Association for the Advancement of Science, 1985). However, the ratings of inquiry used in the AAAS study are more of inquiry opportunities for students, and not of science presented as a process of inquiry, so it is not possible to make a direct comparison with the current study. It is also difficult to draw comparisons across textbooks within the AAAS report, because each textbook was rated by a different group of raters, leaving any possible rater biases confounded with any reported textbook effects.

Another area of future work would be to use the same technique, linguistic content analysis, to measure different aspects of textbooks. The technique could be readily adapted to address a wide variety of topics within textbooks. Possible areas could be the treatment of evolution, the treatment of ecological issues, or the treatment of women. It could also be used to study the treatment of inquiry in other subject matter chapters to see if the patterns of responses are the same across subject matter chapters. In particular, the subject matter chapters included in this study depicted the approach of low initial inquiry within paragraphs, as compared to the approach of high initial inquiry within paragraphs depicted in the introduction chapters. It could be useful to determine if the approach of low initial inquiry is carried out in other subject matter chapters, indicating a difference in the treatment of inquiry when science in general is discussed versus when specific content areas of science are discussed.

A final area of future work could be to study the effect of various levels of the measure of inquiry on learners in the classroom. A meta-analysis performed on studies comparing students using the BSCS textbooks versus students using other, more traditional texts (such as the Holt series), indicates the BSCS texts were more effective at increasing student achievement (Shymansky, 1984; Shymansky, Kyle and Alport, 1983). If the BSCS texts are indeed higher in inquiry, this could be one piece of evidence to support the idea that higher levels of inquiry increase learning of science. In addition to studying the effect of various levels of inquiry on learners, one could also study the effect of the placement of inquiry within the text on learners. This study provides evidence of two different methods of presenting inquiry, high versus low initial inquiry. The two approaches could have different effects on learners.

A complicating factor in the measurement of the effect of a textbook on learners is, of course, teacher use of the textbook. For example, a textbook may have a very high level of inquiry, but a teacher could use it in a manner emphasizing rote learning. Alternatively, a textbook may have a low level of inquiry, but inquiry could be introduced to the students with lectures, discussions and classroom activities. Any study addressing the effect of level of inquiry in texts on student learning would have to control for the confounding variable of teacher use of textbooks.

BIBLIOGRAPHY

American Association for the Advancement of Science. (1985). Science Books and Films, 20(5).

Beard, Jean. (1986). Comment on, "Evolution in high school biology textbooks: 1963-1983" by Dorothy B. Rosenthal, Science Education, 69(5), 637-638. 1985. Science Education, 70(5), 501-502.

Bloom, Benjamin S., Engelhart, Max D., Furst, Edward J., Hill, Walker H., and Krathwohl, David R. (1956). Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain. New York: David McKay Company, Inc.

Borg, Walter R., and Gall, Meredith D. (1983). Educational Research: An Introduction. New York: Longdon.

Boyer, Ernest L. (1983). High school: A report on secondary education in America. The Carnegie Foundation for the Advancement of Teaching. New York: Harper and Row.

Burmester, Mary Alice. (1953). The construction and validation of a test to measure some of the inductive aspects of scientific thinking. Science Education, 37(2), 131-140.

Bybee, Rodger W. (1985). The Sisyphean question in science education: What should the scientifically and technologically literate person know, value and do-As a citizen? Science Technology Society. 1985 Yearbook of The National Science Teachers Association.

Champagne, Audrey B. and Klopfer, Leopold E. (1981). Problem solving as outcome and method in science teaching: Insights from 60 years of experience. School Science and Mathematics, 81(1), 3-8.

Cho, Hee-Hyung and Kahle, Jane Butler. (1984). A study of the relationships between concept emphasis in high school biology textbooks and achievement levels. J. of Research in Science Teaching, 21(7), 725-731.

Cho, Hee-Hyung, Kahle, Jane Butler and Nordland, Floyd H. (1985). An investigation of high school biology textbooks as sources of misconceptions and difficulties in genetics and some suggestions for teaching genetics. Science Education, 69(5), 707-719.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Education and Psychological Measurement, 20, 37-46.

Conference on Goals for Science and Technology Education Grades K-12; A report to the National Science Board Commission on Precollege Education in Mathematics, Science and Technology. (1983). A revised and intensified science and technology curriculum grades K-12 urgently needed for the future. Washington, D.C.: National Science Foundation.

Crowell, Victor L., Jr. (1937). The scientific method: Attitudes and skills essential to the scientific method, and their treatment in general science and elementary biology textbooks. School Science and Mathematics, 37, 525-531.

Curtis, Ruth V. and Reigeluth, Charles M. (1984). The use of analogies in written text. Instructional Science, 13(2), 99-117.

Doran, Rodney L. and Sheard, Doris M. (1974). Analyzing science textbooks. School Science and Mathematics, 74(1), 31-39.

Dukes, R. J. and Kelly, S. A. (1979). The readability of college astronomy and physics texts. The Physics Teacher, 17(7), 168-173.

Duschl, Richard A. (1986). Textbooks and the teaching of fluid inquiry. School Science and Mathematics, 86(1), 27-32.

Esler, William K. (1969). Do real differences exist between old and new curricula? Science Education, 53(1), 67-70.

Feezel, Jerry D. (1985). Toward a confluent taxonomy of cognitive, affective and psychomotor abilities in communication. Communication Education, 34, 1-11.

Finley, Fred N. (1981). A Philosophical approach to describing science content: An example from geological classification. Science Education, 65(5), 513-519.

Fiske, Edward B. (1984). Are they "dumbing down" the textbooks? Principal, 64(2), 44-46.

Fleiss, Joseph L. (1973). Statistical methods for rates and proportions. New York: Wiley.

Fleiss, Joseph L., Cohen, J. and Everitt, B. B. (1969). Large sample standard errors of kappa and weighted kappa. Psychological Bulletin, 72(5), 323-327.

Fleiss, Joseph L., Nee, John C. M. and Landis, J. Richard. (1979). Large sample variance of Kappa in the case of different sets of raters. Psychological Bulletin, 86(5), 974-977.

Fox, James Harold. (1951). Some observations. The Science Teacher, 18(2), 76-77.

Fraser, Barry J. (1978). Use of content analysis in examining changes in science education aims over time. Science Education, 62(1), 135-141.

Fuhrman, M., Lunetta, V., Novick, S., and Tamir, P. (1978). The laboratory structure and task analysis inventory (LAI): A user's handbook. (Technical Report 14). Science Education Center, University of Iowa, Iowa City, Iowa.

Fuller, Wayne A. (1987). Measurement Error Models. New York: John Wiley and Sons.

Gould, C. D. (1977). The readability of school biology textbooks. J. Biol. Education, 11(4), 248-252.

Grabiner, Judith V. and Miller, Peter D. (1974). Effects of the Scopes trial: Was it a victory for evolutionists? Science, 185(4154), 832-837.

Grobman, Arnold B. (1969). The Changing Classroom: The Role of the Biological Sciences Curriculum Study. Garden City, New York: Doubleday and Company, Inc.

Grobman, Arnold. (1988). Personal communication. Chancellor Emeritus and Research Professor of Biology. University of Missouri, St. Louis, Missouri.

Harms, N. C. and Yager, R. R. (1981). What research says to the science teacher, Vol. 3. Washington, D. C.: National Science Teachers Association.

Harrell, Frank E., Jr. (1986). The LOGIST procedure. SAS Institute Inc. SUGI Supplemental Library User's Guide, Version 5 Edition. Cary, N. C.: SAS Institute Inc., 269-293.

Hartford, Fred and Good, Ron. (1976). Assessment of cognitive requirements of instructional materials. School Science and Mathematics, 76(3), 231-237.

Helgeson, S. L., Blosser, P. E. and Howe, R. W. (1977). The status of pre-college science, mathematics, and social science education: 1955-75. The Center for Science and Mathematics Education. The Ohio State University, Columbus, Ohio. Washington, D. C.: U. S. Government Printing Office.

Herron, Marshall D. (1971). The nature of scientific enquiry (sic). School Review, 79(2), 171-212.

Hickman, Faith M. and Kahle, Jane Butler (Eds.). (1982). New Directions in Biology Teaching. A Special Publication of the National Association of Biology Teachers.

Hofstein, Avi and Yager, Robert E. (1982). Societal issues as organizers for science education in the '80s. School Science and Mathematics, 82(7), 539-547.

Hurd, Paul DeHart, Bybee, Rodger W., Kahle, Jane Butler, Yager, Robert E. (1980). Biology education in secondary schools of the United States. American Biology Teacher, 42(7), 388-404, 409-410.

Joy, Dave. (1988). Personal communication. Division Manager for Holt, Rinehart and Winston. McPherson, Kansas.

Kennedy, Keith. (1979). Determining the reading level of biology textbooks. The American Biology Teacher, 41(5), 301-303.

Knoke, David and Burke, Peter J. (1980). Log-Linear Models. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-020. Beverly Hills and London: Sage Publications.

Koval, David B. and Staver, John R. (1985). What textbooks don't teach. Science Teacher, 52(3), 49-52.

Lampkin, Richard H. (1951). Scientific inquiry for science teachers. Science Education, 35(1), 17-39.

Leonard, William H. and Lowery, Lawrence F. (1984). The effects of question types in textual reading upon retention of biology concepts. J. of Research in Science Teaching, 21(4), 377-384.

Levin, Florence S. and Lindbeck, Joy S. (1979). An analysis of selected biology textbooks for the treatment of controversial issues and biosocial problems. J. of Research in Science Teaching, 16(3), 199-203.

Lowery, Lawrence F. and Leonard, William H. (1978a). A comparison of questioning styles among four widely used high school biology textbooks. J. of Research in Science Teaching, 15(1), 1-10.

Lowery, Lawrence F. and Leonard, William H. (1978b). Development and method for use of an instrument designed to assess textbook questioning style. School Science and Mathematics, 78(5), 393-400.

McConnell, Mary. (1982). Teaching about science, technology and society at the secondary school level in the United States. An educational dilemma for the 1980s. Studies in Science Education, 9, 1-32.

Miller, W. G., Snowman, J. and O'Hara, T. (1979). Application of alternative statistical techniques to examine the hierarchical ordering in Bloom's taxonomy. American Educational Research Journal, 16, 241-248.

National Commission on Excellence in Education. (1983). A nation at risk. Washington, D. C.: U. S. Department of Education.

National Science Board Commission on Precollege Education in Mathematics, Science and Technology. (1983). Educating Americans for the 21st century. Washington, D. C.: National Science Foundation.

National Science Board Committee on Science Indicators. (1985). Science Indicators: The 1985 Report. Washington, D. C.: U. S. Government Printing Office.

Neie, Van E. (1979). A comment on, "The influence of pictorial illustrations with written text and previous achievement on the reading comprehension of fourth grade science students" by James L. Thomas. J. of Research in Science Teaching, 16(4), 371.

Nelkin, Dorothy. (1977). Science Textbook Controversies and the Politics of Equal Time. Cambridge, Mass.: The MIT Press.

Newton, D. P. (1984). A way of classifying and measuring some aspects of the illustration style of textbooks. Programmed Learning and Educational Technology, 21(1), 21-27.

Newton, David E. (1968). The dishonesty of inquiry teaching. School Science and Mathematics, 68(9), 807-810.

Nicely, Robert F., Jr. (1980-81). The analysis of mathematics instructional materials: Concepts, procedures and results. International J. of Instructional Media, 8(3), 221-233.

Nicely, Robert F., Jr. (1985a). Higher-order thinking skills in mathematics textbooks. Educational Leadership, 42(7), 26-30.

Nicely, Robert F., Jr. (1985b). Mathematics instructional materials: A decade of change. International J. of Instructional Media, 12(2), 127-136.

Nietz, J. A. (1966). The evolution of American secondary school textbooks. Rutland, Vermont: Tuttle.

Office of Educational Research and Improvement, U. S. Department of Education, Center for Statistics. (1986). Study of Excellence in High School Education: Longitudinal Study, 1980-1982. Final Report.

Pratt, Donald L. (1985). Mathematics usage in secondary science courses and textbooks. School Science and Mathematics, 85(5), 394-406.

Prosser, Michael. (1983). Relationship between the cognitive abilities of a group of tertiary physics students and the cognitive requirements of their textbooks. Science Education, 67(1), 75-83.

Reynolds, John Carl, Jr. (1967). An assessment of the cognitive content of review questions in selected secondary general biology textbooks. Unpublished dissertation. College of Education, University of Alabama. University, Alabama. University Microfilm 68-1063.

Roberts, Carl W. (1986). PLCA. Documentation materials accompanying the Program for Linguistic Content Analysis. Departments of Statistics and Sociology. Iowa State University, Ames, Iowa.

Roberts, Carl W. (1988). Other than counting words: A linguistic approach to content analysis. Social Forces (forthcoming).

Rosenthal, Dorothy B. (1984). Social issues in high school biology textbooks: 1963-1983. J. of Research in Science Teaching, 21(8), 819-831.

Rosenthal, Dorothy B. (1985). Evolution in high school biology textbooks: 1963-1983. Science Education, 69(5), 637-648.

Rumbough, William S. (1951). Use the Illustrations! The Science Teacher, 18(2), 69-70.

St. Laurence, Francis. (1951). Are heavy textbooks necessary? The Science Teacher, 18(2), 72-73.

Schwab, J. J. (1962). The teaching of science as enquiry (sic). The Teaching of Science. eds. Schwab, J. J. and Brandwein, P. F. Cambridge: Harvard University Press.

Schwab, J. J. (1963). Biology teachers handbook. New York: Wiley and Sons, Inc.

Scott, H. V. (1972). The taxonomy of educational objectives as a curriculum analysis tool: A solution to some problems encountered while coding activities. Science Education, 56(3), 411-415.

Seddon, G. M. (1978). The properties of Bloom's taxonomy of educational objectives for the cognitive domain. Review of Educational Research, 48, 303-323.

Shymansky, James A. (1984). BSCS programs: Just how effective were they? The American Biology Teacher, 46(1), 54-57.

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

Shymansky, James A. and Yore, Larry D. (1979). Assessing and using readability of elementary science texts. School Science and Mathematics, 79(8), 670-676.

Siegel, Harvey. (1978). Kuhn and Schwab on science texts and the goals of science education. Educational Theory, 28(4), 302-309.

Skoog, Gerald. (1979). Topics of evolution in secondary school biology textbooks: 1900-1977. Science Education, 63(5), 621-640.

Skoog, Gerald. (1984). The coverage of evolution in high school biology textbooks published in the 1980s. Science Education, 68(2), 117-128.

Stake, R. E. and Easley, J. (1978). Case studies in science education, volumes I and II. Center for Instructional Research and Curriculum Evaluation, University of Illinois at Urbana-Champaign. Washington, D. C.: U. S. Government Printing Office.

Stuart, John A. (1982). An identification of life science concepts in selected secondary school science textbooks. School Science and Mathematics, 82(3), 189-200.

Stuart, John A. and Burns, Richard W. (1984). The thinking process: A proposed instructional objectives classification scheme. Educational Technology, 24(7), 21-26.

Tafoya, Estelle, Sunal, Dennis W., and Knecht, Paul. (1980). Assessing inquiry potential: A tool for curriculum decision makers. School Science and Mathematics, 89(1), 43-48.

Tamir, Pinchas. (1983). Inquiry and the science teacher. Science Education, 67(5), 657-672.

Tamir, Pinchas. (1985). Content analysis focusing on inquiry. J. of Curriculum Studies, 17(1), 87-94.

Tamir, Pinchas, and Lunetta, Vincent N. (1978). An analysis of laboratory inquiries in the BSCS yellow version. The American Biology Teacher, 40(6), 353-357.

Tamir, Pinchas, and Lunetta, Vincent N. (1981). Inquiry-related tasks in high school science laboratory handbooks. Science Education, 65(5), 477-484.

Tamir, Pinchas, Nussinovitz, Rachel, and Friedler, Yael. (1982). The design and use of a practical tests assessment inventory. J. of Biological Education, 16(1), 42-50.

Thomas, James L. (1978). The influence of pictorial illustrations with written text and previous achievement on the reading comprehension of fourth grade science students. J. of Research in Science Teaching, 15(5), 401-405.

Trump, J. Lloyd and Miller, Delmas F. (1979). Secondary School Curriculum Improvement: Meeting Challenges of the Times. Boston: Allyn and Bacon, Inc.

Vachon, Myra K. and Haney, Richard E. (1983). Analysis of concepts in an eighth grade science textbook. School Science and Mathematics, 83(3), 236-245.

Vogel, Louis F. (1951). A spot-check evaluation scale for high school science textbooks. The Science Teacher, 18(2), 70-72.

Webb, Hanor A. (1951). The high school science library. The Science Teacher, 18(2), 66-68.

Weber, Robert Philip. (1986). Basic Content Analysis. Sage University series on Quantitative Applications in The Social Sciences, series no. 07-049. Beverly Hills and London: Sage Publications.

Welch, Wayne W. (1979). Twenty years of science curriculum development: A look back. In Berliner, David C. (Ed.), Review of Research in Education, Vol. 7. 282-306. American Educational Research Association.

Welch, Wayne W., Klopfer, Leopold E., Aikenhead, Glen S. and Robinson, James T. (1981). The role of inquiry in science education: Analysis and recommendations. Science Education, 65(1), 35-50.

Wright, Jill D. (1982). The effect of reduced readability text materials on comprehension and biology achievement. Science Education, 66(1), 3-13.

Weiss, I. R. (1978). Report of the 1977 national survey of science, mathematics and social studies education. Washington, D. C.: U. S. Government Printing Office.

Yager, Robert E. (1983). The importance of terminology in teaching K-12 science. J. of Research in Science Teaching, 20(6), 577-588.

Yager, Robert E. and Yager, Stuart O. (1984). The effect of schooling upon understanding of selected science terms. J. of Research in Science Teaching, 22(4), 359-364.

Yore, Larry D. (1986). Science achievement, cognitive development, and degree of structure in inquiry lessons. School Science and Mathematics, 86(4), 271-283.

Yost, Michael. (1973). Similarity of science textbooks: A content analysis. J. of Research in Science Teaching, 10(4), 317-322.

ACKNOWLEDGEMENTS

The seasons understand that balance
is the heart of change.
And understanding that, the seasons
always rearrange.
Yet still the order seems as endless
cycles round the sun.
And when we think we've learned something,
the lesson's just begun.

I believe I found the above verse in an issue of Science; and I believe it was attributed to Jackson Browne. I would like to say that I carefully recorded the volume, issue and page numbers, but alas Mrs. Bishop, this is not the case. Nevertheless, the verse accurately reflects my feelings upon completion of my dissertation.

As I finish one cycle of my professional career and prepare to embark on the next cycle, I would like to take the time to acknowledge with gratitude the help and support of a few people who have assisted me.

I would especially like to thank my major professor, Dr. Lynn Glass for his guidance, support and supervision of my doctoral program. His insight and knowledge in the field of science education have been a constant source of help to me during my professional growth at Iowa State University. I also owe a great deal of gratitude to my minor professor, Dr. Carl Roberts. I wish to thank him for the use of his linguistic content analysis program, and for the many hours he spent teaching me to use it. I also want to thank him for his assistance with the reliability studies, the statistical analyses, and for his excellent critique of the written manuscript. I am also grateful to the other members of my graduate committee, Dr. Trevor Howe, Dr. Tony Netusil, and Dr. Ann Thompson for their time and support of my efforts. I would further like to acknowledge Jerry Dunn of Ames High School, Harold Rathert of the Des Moines Public School System and Dr. Roger Volker of Iowa State University for helping me to locate the textbooks I used in the study. I also wish to acknowledge Dr. Ken Koehler of the Department of Statistics at Iowa State University for his assistance and expertise in categorical data analysis. I further wish to acknowledge the assistance of Kathy Shelly of the Numerical Analysis Section of the Statistical Laboratory at Iowa State University for helping me learn to use SASGRAPH.

In addition I wish to acknowledge the love and support of my family and friends. I am grateful to my parents, Janet and Walter Miles, to my brothers, Tom and Howard Miles, and to my sister, Jan Miller, for the nurture and love they have given to me. Finally, I am grateful to my husband, John, for his love, his support and his belief in me.

APPENDIX A: DICTIONARY CREATED FOR USE WITH PLCA

OBJECTS

Code	Noun
100	honesty/truth(s)
101	falsehood/deception(s)
102	biology
103	science
104	technology
105	science tool
106	microscope
107	plant science word
108	animal science word
109	genetic science word
110	mineral science word
111	science word
112	animal
113	plant
114	cell (egg, sperm)
115	living organism (virus)
116	nutrient/soil/water/(sugar)
117	light
118	disease
119	leaves
120	step(s)
121	part/factor(s)/(sample)
122	national boundaries
123	result(s)
124	large amount
125	laboratory
126	figure
128	neutral place
129	neutral object/thing
147	characteristic/trait(noun)
148	characteristic(adjective)
149	written materials
150	problem(s)
151	method/(process/technique)
152	research method
153	scientific method
154	hypothesis

PERSONS

Code	Noun
130	scientist/biologist(s)
132	named scientist/biologist(s)
133	technician(s)
134	neutral person(s)
135	author(s)
136	reader

MODAL VERBS

Code	Verb	Nominal Form
001	want/intend	intention
002	ought to	obligation
003	can	ability
004	must	necessity/certainty
005	try/attempt	attempt
006	refuse	refusal
007	possibly	possibility
008	would	
009	could	
010	might	

VERBS

Code	Verb	Nominal Form
*011[1]	think/consider/conceive	thought/idea/(theory)
*012	know/recognize/ (remember/learn)	knowledge
*013	observe/see	observation
*014	question/wonder	question/wonder
015	inquire/ask	question/query
*016	research/investigate/ study	research/investigation/ study
*017	describe/report/(relate)	description/report
*018	record	record/data
*019	measure	measurement/dose
*020	seek/search	search
021	progress/proceed	progress/advance
022	establish/start/ found/(begin)	groundwork
023	work/perform/(do)	work/performance
024	plan/develop	plan/development
025	reveal/show	revelation
026	agree	agreement
027	conquer	conquest
028	mean/define	definition
029	say	statement/(rule/ law/principle)
*030	prove	proof
*031	test/experiment/ (dissect)	test/experiment/ (dissection)
032	exemplify/illustrate	example/illustration
033	facilitate/encourage	cause/facilitation
034	group/categorize/classify	group category
035	inject/vaccinate	injection/vaccine
**036[2]	control/(determine)	control
037	separate/segregate	difference/separation
038	compare/contrast	comparison/contrast
039	discover/(find)	discovery
040	give/present	gift

[1]Inquiry verb.

[2]Inquiry verb only when a person is the semantic subject.

041	have/get/receive	possession
042	answer/respond/ (conclude)	answer/response/ (conclusion)
043	name/call	name
045	move/transfer	movement
046	grow/(develop)	growth
047	organize	organization/structure
048	compose/consist	composition
049	combine/join/ (fertilize/mate)	combination/ (fertilization)
051	hold/support	hold/support
052	relate/link	relationship
053	strike/hit	strike/hit
054	produce/make/form	product
055	occur	event/occurrence
056	provide/allow	provision
057	magnify	magnification
058	change/ (increase/decrease	change/ (increase/decrease)
065	exist	existence
066	succeed	success
067	fail	failure
068	assume	assumption
069	discourage/impede/(harm)	impediment/block
070	influence	influence
071	endure	duration
072	use	use
073	force	power/force/(energy)
074	expect/anticipate	expectation
098	do good	good
099	do evil/(kill)	evil
100	honest/genuine	truth/honesty
101	dishonest/deceptive	falsehood/lie

APPENDIX B: EXAMPLE OF A DATA MATRIX FOR PLCA

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
651002	050202	1	1	1	0	1	888	888	888	132	88	40	888	151
651003	070472	1	2	2	0	1	888	888	888	109	88	88	888	110
651004	100582	1	1	3	0	1	888	888	888	129	88	25	888	888
651004	110643	1	1	2	0	1	136	135	021	103	88	70	222	016
651002	040152	1	1	1	0	0	888	888	888	132	88	26	888	999
651002	040152	2	1	1	0	1	888	888	888	108	88	45	888	888
651002	040152	3	1	1	0	1	888	888	888	132	88	29	888	999
651002	040152	4	1	1	0	1	888	888	888	108	88	45	888	888
562623	020032	1	2	1	0	1	888	888	888	113	88	88	888	148
562623	020032	2	2	1	0	1	888	888	888	113	88	88	888	148
562624	050302	1	1	1	0	1	888	888	888	147	88	36	888	111
562629	210942	1	1	2	0	1	888	888	888	126	88	25	888	123
562630	221051	1	1	2	0	1	888	888	888	888	88	41	888	123
562630	221051	2	1	2	0	1	888	888	888	999	88	33	888	999
562630	221051	3	1	2	0	1	888	888	888	888	88	49	888	112
562633	281442	1	1	2	0	1	888	888	888	114	88	36	888	148
774274	010052	1	1	2	0	1	888	888	888	119	88	41	888	073
774274	030081	1	1	2	0	1	888	888	888	119	88	46	888	888
774274	030081	2	1	2	0	1	888	888	888	119	88	23	888	888
774274	040092	1	1	2	0	1	888	888	888	113	88	46	888	119
774274	040092	2	1	2	0	1	888	888	888	113	88	41	888	119
774276	120402	1	1	3	0	1	136	135	888	136	88	13	888	111
774286	602962	1	2	2	0	0	888	888	888	112	88	88	888	004

a First identifying number.
b Second identifying number.
c Clause sequence number within a sentence.
d Type of clause; 1 = perception, 2 = recognition.
e Tense of clause; 1 = past, 2 = present, 3 = future.
f Question; 0 = not a question, 1 to 9 = a question.
g Valence; 1 = positive, 0 = negative.
h Audience of clause; 136 = reader(s), 888 = not applicable.
i Speaker of clause; 135 = author(s), 888 = not applicable.
j Modifier of subject.
k Semantic subject of verb.
l Modal auxiliary of verb.
m Verb.
n Modifier of object.
o Semantic object of verb.

APPENDIX C: EXAMPLES OF PLCA TRANSLATIONS

The following examples of translated sentences correspond to the data matrix presented in Appendix B. The sentence in its original form is presented first, followed by its PLCA translation. The first five sentences were obtained from the 1956 introduction chapter, the next five from the 1965 genetics material, and the last five from the 1977 leaf structure chapter. The words in all upper case letters are derived from the dictionary. The words in upper and lower case letters are generated by the program. Words or phrases encased in double brackets are contextual cues added to the original sentence.

1. A few years later Robert Kock (kohk) of Germany gave the world a method of investigating infectious diseases and techniques for culturing bacteria in the laboratory.

   The NAMED SCIENTIST/BIOLOGIST(S) GAVE/PRESENTED a(n) METHOD.

2. Heredity is chemical.

   The GENETICS SCIENCE WORD is a(n) MINERAL SCIENCE WORD.

3. Only time will tell.

   The NEUTRAL OBJECT/THING will REVEAL/SHOW.

4. Advances in all of these fields are bound to have a dynamic influence on our study of life.

   To READER(S) AUTHOR(S) said, "The PROGRESS/ADVANCE(s) of SCIENCE INFLUENCE(s) Our RESEARCH/INVESTIGATION.

5. In the 17th century, William Harvey, an English physician, disputed the ancient belief that blood ebbed and flowed in the veins of the body like the tides of the sea and proposed that it circulated through arteries and veins.

   The NAMED SCIENTIST/BIOLOGIST(S) did not AGREE that [The ANIMAL SCIENCE WORD MOVED]. The NAMED SCIENTIST/BIOLOGIST(S) SAID that [The ANIMAL SCIENCE WORD MOVED].

6. Some plants were short and bushy, while others were tall and climbing.

   The PLANT was a(n) CHARACTERISTIC(ADJECTIVE).
   The PLANT was a(n) CHARACTERISTIC(ADJECTIVE).

7. One character dominated the other.

   The CHARACTERISTIC/TRAIT(NOUN) CONTROLLED a(n) SCIENCE WORD.

8. The diagram in Fig. 48-4 shows the result of such a cross. [[a two characteristic cross]]

   The FIGURE REVEAL/SHOWS a(n) RESULT(S).

9. When hybrid black, rough-coated guinea pigs are crossed, similar results are obtained.

   [The RESULT(S) is HAD/GOTTEN/RECEIVED], if [The ANIMAL is COMBINED/JOINED].

10. Actually, in human beings and many animals, it [[sex determination]] is determined by the sperm alone.

    The CELL CONTROL(s) a(n) CHARACTERISTIC/TRAIT(NOUN).

11. The leaves absorb energy from the sunlight, causing chemical changes in the water and minerals.

    The LEAVE(S) HAVE/GET/RECEIVE(s) a(n) POWER/FORCE.

12. A leaf usually grows and works for one season only.

    The LEAVE(S) GROW(s). The LEAVE(S) WORK/PERFORM(s).

13. Even evergreens, which keep their leaves, grow new young shoots every year.

    The PLANT HAVE/GET/RECEIVE(s) a(n) LEAVE(S). The PLANT GROW(s) a(n) LEAVE(S).

14. You will see three different kinds of tissue.

    To READER(S) AUTHOR(S) said, "You will OBSERVE/SEE a(n) SCIENCE WORD."

15. They [[insectivorous plants]] do not need the animals as food.

    The ANIMAL is not a(n) NECESSITY/CERTAINTY.

APPENDIX D: CALCULATION OF THE "C" STATISTIC

The statistic "C" is considered to be a good measure of the predictive ability of a logistic regression model (Harrell, 1986). The statistic C is somewhat analogous to a correlation coefficient between the observed values of the dependent variable and the values predicted by the model. Except for trivial cases when the sample size is less than or equal to two, the range of C is from slightly greater than zero to slightly less than one. Values of C close to one indicate greater agreement between the observed and the predicted values. The formula for calculating C is:

$$C = [c + (t/2)] / [c + d + t],$$

where "c" is the number of concordant pairs, "d" is the number of discordant pairs and "t" is the number of tied pairs in the data. Examples of two sets each of concordant, discordant and tied pairs are presented in Table 8.

A concordant pair exists when, for two units of analysis, the predicted values have the same ordered sequencing as the observed values. In the first example in Table 8, the first observed value is less than the second observed value. Because the first predicted value is also less than the second predicted value, therefore this is a concordant pair. A discordant pair exists when the

TABLE 8. Examples of concordant, discordant, and tied pairs of data points

	Observed Value	Predicted Value
Two Concordant Pairs		
first pair	0	0.2
	1	0.6
second pair	1	0.8
	0	0.5
Two Discordant Pairs		
first pair	0	0.5
	1	0.1
second pair	1	0.3
	0	0.8
Two Tied Pairs		
first pair	1	0.7
	1	0.8
second pair	0	0.3
	0	0.1

predicted values do not have the same ordered sequencing as the observed values. In the first example of a discordant pair in Table 8, the first observed value is less than the second observed value, but the first predicted value is greater than the second predicted value. A tie exists when two observed values are the same. To generate the values of "c," "d" and "t," all possible pair-wise combinations are made, and total numbers of concordant, discordant and tied pairs are counted.