THE EFFECTS OF JUNIOR GREAT BOOKS LITERATURE DISCUSSION

ON READING COMPREHENSION ACHIEVEMENT OF GIFTED FIFTH

GRADERS: APPLICATION OF GENERAL LINEAR MODEL

FOR CROSS-LEVEL INFERENCES

by

Eileen D. Sable

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Educational Research and Evaluation

APPROVED:

M. Lichtman, Co-Chairperson   J. Fortune, Co-Chairperson

Z. Weber   N. Boraks

W. Seaver

June 1987

Blacksburg, Virginia

THE EFFECTS OF JUNIOR GREAT BOOKS LITERATURE DISCUSSION

ON READING COMPREHENSION ACHIEVEMENT OF GIFTED FIFTH

GRADERS: APPLICATION OF GENERAL LINEAR MODEL

FOR CROSS-LEVEL INFERENCES

by

E.D. Sable

Committee Chairpersons: J. Fortune & M. Lichtman

College of Education

(ABSTRACT)

Research findings emphasize the need for programs
for the gifted reader, particularly the need to determine
which methods or programs best benefit the gifted student.
The purpose of this study was to assess the effectiveness
of the Junior Great Books (JGB) literature program on
reading comprehension achievement of gifted fifth grade
students due to the factors of a) discussion/non-discussion
and b) levels of students' experience in the JGB program.
This study was replicated to determine if similar results
in reading comprehension achievement occurred by using
different JGB stories. In addition, the study explored
methodological issues of cross-level inferences to
determine if different results were obtained when applying
the General Linear Model to individual-level and aggregate-
level data.

The research was experimental in design. Seventy-eight fifth grade gifted students were randomly assigned to eight groups in two schools, four treatment groups with discussion and four control groups without discussion. Students in one school had no prior JGB experience; students in the other school had a range of one to three years JGB experience. Two different JGB literature stories, randomly selected, were read by all groups. Reading comprehension was measured, pre and post, by an instrument that emerged from the JGB stories using the cloze procedure. Additional data to measure students' achievement in the form of responses to openended questions about the JGB literature stories were collected. Observations of groups featuring discussions/non-discussions were taped and rated to determine whether appropriate formats were followed by the leaders. Individual-level data and aggregate-level data were analyzed using a two-way ANOVA with nesting using the General Linear Model of the Statistical Analysis System (SAS).

It was found, and supported by study replication, that there were no significant differences between those fifth grade gifted students who participated in the JGB program with discussion and those who did not. Nor were there differences due to level of experience in the JGB

program.   Similar   results   were obtained when applying the General   Linear   Model   to   individual-level   and aggregate-level data.

Further   investigation   of   the   JGB   program and related   methods   and   procedures involved in the study were suggested.

ACKNOWLEDGEMENTS

Undertaking a study and writing a doctoral dissertation are tasks which cannot be accomplished without the guidance, encouragement and cooperation of many teachers, administrators and friends. I am deeply grateful to all those who helped me to complete this study.

I sincerely appreciate the support and wisdom extended to me by my doctoral committee, Drs. Lichtman and Fortune, my co-chairpersons, and other members, Drs. N. Boraks, W. Seaver and L. Weber. Their many hours of thoughtful and patient guidance and continuing encouragement have been invaluable.

I am particularly appreciative to the members of the Fairfax County Research Screening Committee for their approval to conduct this research project in the Fairfax County Public Schools. A very special thanks is extended to the principals, teachers and students who graciously agreed to participate in this study. Their sincere interest and generous offering of time have been the mainstay of this research.

Finally, and most importantly, I want to thank my family, and especially   , for their unwavering support and encouragement in helping me to grow.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

According to a survey on the state of gifted and talented education, a growing number of people across the United States are convinced of the need to nurture the abilities of the gifted (Mitchell, 1982). Prior to 1981 most states had an annual appropriation of between one-quarter-million and one-million dollars for gifted and talented education. Funding for gifted and talented education has seen a 112% increase over the past five years, with an additional $64.4 million being spent on these students from state revenues (Zettel, 1982). More recent figures indicate that 311 million dollars is spent annually by the states: 31 states spent over one million dollars, 21 states spent over five million dollars annually. In addition 22 states have mandated programs for the gifted and talented (O'Connell, 1986).

This surge of financial and educational support has resulted in a focus toward effective program components and teaching strategies for the gifted. As greater emphases is placed on the gifted student and his/her needs, research findings emphasize the need for programs for the gifted reader, particularly the need to determine which methods or

1

programs best benefit the gifted student (Sakiey, 1980; Stank, 1983; Witty, 1971).

Educators have suggested that gifted readers need to develop higher level cognitive skills to effectively interpret, manage, and grow in the present and future technological era. Savage (1983) specifies:

> Reading is essentially a thinking process; thus reading for gifted children must allow the use of the higher cognitive abilities they possess. They should be challenged to analyze various readings, synthesize pieces of information garnered from several sources, and critically evaluate various presentations (p.9).

However, the application of such an approach is unclear and extremely complex, particularly in the area of reading education. School program managers, reading specialists, and classroom teachers of the gifted are directed to develop inferential reading comprehension skills of the gifted reader; yet few empirical studies offer conclusive evidence of the effectiveness of suggested programs which are presently employed across the nation nor do studies reveal effective ways to measure inferential reading comprehension. Bates (1984) adds:

> Recent years have seen many attempts to meet the reading needs of gifted students. Unfortunately, the various programs have not been studied rigorously, and the results have not been consistent. In fact, most programming suggestions are based upon intuition rather than upon research (p.590).

Inconsistent results and lack of rigorous research of programming for the gifted reader reflect the fact that studying instruction in the classroom presents serious and unique problems for the researcher (Berliner, 1980; Burstein, 1980; Hopkins, 1982). A chief function of educational research is to determine factors that influence educational performance and to ascertain their effects. A complication, however, that may interfere with this purpose is that educational data are inherently multilevel; that is, "...education involves students taught by teachers in classrooms in school districts and so on" (Burstein, p. 129). As a consequence, researchers are advised to examine the multilevel data to avoid making interpretational errors due to cross-level discrepancies (Glick, 1980). In addition, related problems of defining the unit of analysis for the observance of classrooms may arise. Reseachers must choose the appropriate level of aggregation in data analysis. At issue is the choice of proper characteristics of the phenomenon of interest in relation to the questions being asked. This may be a difficult decision to make (Berliner, 1980).

Background

To provide enrichment for the gifted reader, for the past five years this researcher has implemented the Junior Great Books (JGB) Program with students in an elementary public school in Virginia. Over 250 gifted students have been placed in that school in a Gifted Center Program, in self-contained classrooms, grades three through six. Prior to admittance into the Program, each of the students in the Gifted Center Program has been identified as obtaining at least a 140 I.Q. score on the Otis Lennon Test, a minimum Stanford Binet score of 140, and a score of 90% or higher on the reading section of the Wide Range Achievement Test.

The JGB Program is designed for the average reader and pulls together classical literature units which provide a methodology of "shared inquiry" or structured small group discussions. Reading materials are available for students in grades two through twelve in the form of paperbound texts. Twelve different literature units are provided at each grade level. The types of literature selections within each level represent genres from folk literature to realistic fiction and fantasy. The Program's aims, as stated in the Leaders' Manual, are to "improve the students' ability to comprehend what they read, to think for themselves about the meaning of the author's words, and to develop the habit of reading literature for enjoyment"

(Dennis & Moldof, 1978,p.1). A recent revision of the JGB program states as its purpose "to improve reading comprehension by helping students to interpret what they read and to think independently and reflectively about what they read" (Handbook on Interpretive Reading and Discussion, 1984, p.3). The discussion method, within small groups, is used to promote these goals and is led by one or two leaders who are trained by the Great Books Foundation in two-day workshops.

The discussions are structured so that questions discussed are interpretive in nature; that is, questions have more than one possible answer. Responses are to be substantiated by references to the text. Emphasis is placed on a search for the author's meaning or "what he wants us to think about or feel through his words" (Dennis & Moldof, 1978, p.2). Leaders of the groups may ask questions about which they have real doubt, hence "shared inquiry".

The use of discussion as a pedogogical technique has been accepted as an important method of enhancing the goals of the reading instructional system (Gutherie, 1981). In addition, research has indicated that gifted students prefer discussion rather than lecture as a method of instruction (Stewart, 1981). Nevertheless, the structured discussion approach of the JGB program has not been

assessed as an effective method of reading comprehension achievement for gifted students nor have previous researchers of the JGB program made adjustments in their analyses for the effects of cross-level inferences on multilevel data to explain the effectiveness of the program.

Statement of the Problem

Increased state funding and mandates for gifted education programs continue to indicate a need for effective programs for the gifted reader. The Junior Great Books Program is a literature discussion program specifically designed to improve reading comprehension.

This researcher participated in an ethnographic study of the JGB program involving gifted and non-gifted fifth-grade students. The study examined "the impact of JGB leader behavior and student ability on a) student response during discussion and b) student acquisition of desired JGB goals" (Boraks, Early, Sable, 1986, p.307). This study was considered exploratory because intact, non-representative groups were involved, and the test instrument was in the developmental stage. In addition, fifth grade JGB program materials had changed somewhat prior to the study. Results indicated that leader behavior seemed to promote greater changes in the non-gifted. That is, non-gifted JGB children made more progress in dealing

with interpretive questions than did gifted JGB children. Tentative conclusions were drawn based on observations, interviews of leaders and students, and student behavior on the post test. Recommendations were made regarding the reevaluation of the discussion format and the use of JGB with gifted children to determine whether discussions truly provided an opportunity for the gifted to increase their reading ability or display their reading ability.

The use of JGB discussion strategy and its effect on reading comprehension achievement of gifted students has not been studied rigorously nor have researchers examined the differences in the effects of the JGB discussion program by analyzing the data at the group level and the individual level.

The effects of Junior Great Books literature discussion on the reading comprehension achievement of gifted fifth graders is investigated. The General Linear Model is used to adjust for the effects of cross-level inferences on multilevel data, a methodological correction of prior JGB research.

Purpose of the Study

The major purposes of this study are: 1) to compare the effect of JGB literature discussion and non-discussion on reading comprehension achievement of gifted fifth grade students and to determine whether or not the level of

experience in JGB program affected these variables, and 2) to implement a research analysis which corrects for methodological shortcoming, due to cross-level inference using multilevel data. The first purpose is accomplished by examining the differences between reading comprehension achievement of experienced JGB students and non-experienced JGB students. To accomplish the second purpose, data are examined at the group level and the individual level using the General Linear Model.

Research Questions

For the purpose of this study the following lines of inquiry are addressed. The first two represent questions that were investigated through hypotheses testing; the third and fourth questions were investigated in an exploratory manner. The following questions were investigated:

1. Does discussion/non-discussion make a difference in reading comprehension achievement?

2. Does the level of JGB experience make a difference in reading comprehension achievement?

3. Do similar results in reading comprehension achievement occur across one study replication using different JGB stories?

4. Does the application of the General Linear Model to the individual-level data provide a

different solution than the solution obtained

when applying the group-level data?

Hypotheses

The following hypotheses are tested:

H_1 "There is no significant difference between

discussion and non-discussion groups on reading

comprehension achievement."

H_2 "There is no significant difference between

experienced and inexperienced students in the JGB

program on reading comprehension achievement."

Operational Definitions

For the purposes of this investigation, the following

definitions are used:

1. Above average reader: a student who is capable of

reading and understanding text above his/her

present grade level as determined by a standardized

reading test.

2. Average reader: a student who is capable of reading

and understanding text at his/her present grade

level as determined by a standardized reading

test.

3. Cross-level inferences: data interpretations

developed at one level and translated to another

level.

4. Discussion: the process of a group of individuals

presenting different points of view and trying

to resolve an issue.

5. Enrichment: providing opportunities for children to

    engage in activities which will provide increased

    depth and breadth in their quest for knowledge of

    their world and universe.

6. General Linear Model: the matrix solution of N equa-

    tions in N unknowns across N conditions when variables

    are linear and normally distributed.

7. Group: A small number of students led by a teacher.

8. JGB literature: classical and contemporary fiction-

    al writing selected by the Great Books Foundation

    as works that lend themselves to different inter-

    pretations.

9. Level of JGB experience: the number of years a

    student has been in the JGB Program, ranging from no

    experience to three years experience.

10. Reading comprehension achievement: the extraction of

    implied meaning from JGB literature text as

    determined by the cloze procedure and open-ended

    questions.

Limitations

Subjects are from two schools in a suburban upper-middle class school district.

Although all discussion leaders were trained by the Great Books Foundation and used the same materials, the researcher had no control over how training was employed by leaders in the discussion groups.

Significance of the Study

The interchange of ideas in discussion has been accepted by educators as an appropriate teaching strategy for all students. Webb (1985) states:

> Often challenged by group response an investigation of the reciprocal effects of individual readers and groups could clarify the educational value of discussing literature with others...Research informed by literature studies and an understanding of schools can contribute substantially to our knowledge of the ways readers read and the way schools help or hinder readers (p.285).

In reporting research on teaching reading comprehension, Tierney and Cunningham (1984) suggest, "Further, the effects of discussion as a postreading activity in and of itself have yet to be fully addressed. Researchers should be encouraged to examine discussion's influence upon reading to learn, strategy acquisition, the nature of a reader's interpretation, and the reader's self-initiated pursuits " (p.629).

Dillon (1984) adds that a body of research exists on the relationship of questioning strategies and discussion participants but not on the effects of discussion. In other words, the questioning process has been refined to stimulate responses, but "whether any of this means that students <u>learn</u> more has not been asked" (p.55).

Millions of dollars are spent annually by the states on programs for the gifted and yet research indicates that most programming for gifted students is based upon intuition rather than upon research.

The JGB program is widely used as an enrichment program for gifted readers. Promotional literature of the Great Books Foundation attests that over one million children have been involved in its program and contends its use is appropriate for gifted students (Will, 1986). Previous studies of the JGB program have not examined the impact of discussion on reading comprehension achievement of gifted students nor have grouping effects been empirically tested to determine the amount of variance they might explain in reading comprehension achievement. This issue of analysis of multilevel data has been cited as a complication of educational research. The study methodology illustrates a way to address the complication and offers an alternative to this problem.

CHAPTER II

REVIEW OF THE LITERATURE

Due to the dual purpose of this study the review of the literature has been divided into two parts. Part I reflects research related to the conceptual framework of this investigation, the effect of literature discussion on reading comprehension, and an amplification of reading comprehension measurement procedures. Part II describes studies on the methodological issues of cross-level inference and interpretation of multilevel data investigated by educational researchers and social scientists.

Part I: Effect of Literature Discussion on Reading Compreension

Research has indicated the use of literature as an important reading component for gifted readers to develop higher level cognitive skills (Lukasevich, 1983; Monson, 1983). Dole and Adams (1983) surveyed national and state leaders in the field of both education for the gifted and reading education to determine the essential distinctions which differentiate reading curriculum for gifted readers. More than half of the educators cited specific and/or general studies of literature enrichment for the gifted

reader. These findings are corroborated by McCormick and Swassing (1982) in their nationwide survey of educators to determine how school systems provide for reading instruction for the gifted.

An extensive amount of research has been done in the general area of reading comprehension and response to literature; however, literature on inferential reading comprehension of the gifted reader has appeared in a scant amount with no specific reference to the use of group discussion as a method of enhancement of reading comprehension. This section includes related studies in the following areas: 1) literature discussion and reading comprehension, 2) JGB format, application, and experience, and 3) measures of reading comprehension.

Literature discussion and reading comprehension

While searching for effective ways to teach literature to children and adolescents, researchers have developed theories about the phenomenon of the reading process. These theories focus on the relationship of the reader to the writer and the text. Investigators have verified that as readers of all ages engage in active cognitive transactions with literary texts, different developmental patterns have emerged. A discussion of theoretical studies related to the reading process and literature response precedes research pertaining to

discussion grouping, and reading comprehension and discussion.

Theoretical aspects. The complexity of the reading process and its relationship to response to literature can be traced back to the studies of Rosenblatt (1938) and Richards (1929). Rosenblatt presented a transactional theory of literature response in which the reading process is a personal and aesthetic experience, reflecting on the nature of fictional works and the individual's response to those works. Richards focused on the text for responding to literature. He posited that the reader must understand the literal meaning of the author, recognize the writer's tone and attitude toward the reader, as well as the intent of the writer in order to fully comprehend. Hence, Rosenblatt viewed the ideas of the reader as a critical focus of literary experience whereas Richards considered the reader's experience as a hindrance to interpretation of literature.

Renewed interest in response to literature surfaced in the United States and England in the 1960's and 1970's in the studies of Squires, Purves, Petrosky, Applebee, et.al. (Cooper, 1985). This interest continued in the 1980's accompanied by a redirection of perspective in the study of reading psychology from behavioristic models to cognitive models. Prior to 1970 reading comprehension was

viewed as some degree of approximation to the text read. This view of comprehension has changed considerably. The text is now viewed as a blueprint for meaning or a set of clues that the reader uses to build a model of what the text means. This implies a much more active and constructive role in comprehension rather than a passive-receptive role (Collins, Brown & Larkin, 1980; Pearson, 1985; Spiro, 1980).

This redirection is supported by Duke (1984), Chabot (1985) and Fish (1980) who recommended that literature response focus not on individual response but on commonalities of response. Duke advocated the process of shared inquiry on the basis of the assumptions that reflections and problem-solving are primary objectives in the teaching of literature and shared discussion plays a major role in the process. Chabot argued, "Reading or understanding is necessarily a totalizing activity. At every step of the way, our understanding of the text we read depends upon our projection or anticipation of the totality it constitutes" (p.30). Fish contended that meaning is not in the individual reader but in the interpretive community of the reader and agreements and disagreements on an interpretation will occur based on shared or contrasting points of view of the community of readers.

A phenomenological theory on the reading process is espoused by Iser (1978, 1980) in which meaning lies at the intersection between text and reader. The emphasis in this theory of reading is on the processes involved in text construction as well as the effects of literature on the reader. That is, although different interpretations of a text are possible, there are limits to the range of interpretations because some constraints are imposed in the text.

A review of selected cognitive development research that emerged in the 1970's and 1980's which directly relates to literature response and reading comprehension immediately follows.

Clinical studies of Petrosky (1976, 1977) indicated that the development of response to literature is consistent with cognitive and affective development in general and that response to literature is learned and subject to the abilities of any individual 's stage of development. He associated Piaget's (1952) cognitive development theory of concrete and formal operations to children's response to literature and asserted that the transition from concrete to formal operations occurred at various ages depending on the individual. His case studies investigating the effect of reality perception and fantasy on response to literature of two ninth-grade girls

indicated that individual personal responses were at times quite divergent. However, oftentimes they shared broad concerns and themes and engaged in discourse that led to consensus. Petrosky theorized that through sharing of themes and concerns of pieces of literature by utilizing free response, consensus over a work of literature would evolve. This would occur if not initially than continually, through activities of interpretation, analysis, synthesis and evaluation.

Applebee (1978) also linked Piaget's cognitive development theory to a child's concept of story in a series of investigations designed to explore the changes in ideas about and responses to literature of students whose ages ranged from six through seventeen. Broad and discernible patterns emerged from the study and they paralleled findings in other areas of psychological study, in particular Piaget's investigations of the development of various scientific concepts. Developmental stages related to the formulation of responses found by Applebee were: a) preoperational (two-six years of age) - narration which lacked integration, b) concrete operational (ages seven to eleven) - summarization and categorization attributed to the work, c) formal operational Stage I (ages 12 to 15) - analysis of the structure of the work or the motives of the characters, and d) formal operational Stage

II (ages 16 to adult) - generalization about the work; understanding gained through the work. However, Applebee noted that these patterns emerged from responses of children who had a choice about what they said and how they said it, and reflected a preferred way of responding. Different possibilities for response might emerge in interactions with a teacher or peer.

In a naturalistic study to learn more about students' response patterns to literature Matsuhashi (1980) observed small group discussions of five tenth-graders which centered on a short story all the children had read. Patterns that emerged from the unguided small group discussion were characterized by the students' ability to put oneself in the character's position and to treat the character and incident as if they were real. On the other hand the children were not aware of the story as expression of an author's intentions, nor could they discuss the story as existing in its own right.

A later ethnographic study (Hickman, 1981) investigated response patterns of elementary school children and a linkage emerged between the quality of responses and social-instructional contexts of various classrooms. Observations indicated that the teacher had considerable power to influence expressions of response through ability to manipulate the classroom context either

by the choice of reading materials available or the amount of discussion that took place.

Galda (1982a, 1983) related the ability of a child to approach a fictional text with a certain attitude, comparable to a spectator stance, to the development of formal operational thought. That is, as a spectator one can analyze and evaluate texts as a whole and consider the text events as valid possibilities or alternate interpretations of reality. She examined oral literary responses of three fifth-grade girls ages 10.7, 10.11 and 11.3, reading above grade level, and found differences in their ability to assume and maintain a spectator stance toward selected pieces of literature. Two different works of realistic fiction were read by the girls and Galda observed and recorded how these three girls differed in their ability to respond to the stories. Three distinct evaluative behaviors were identified. From the responses analyzed, only the oldest reader was able to assume the spectator stance. Analysis of the data suggested that the abilities necessary for assuming a spectator stance seemed to be characteristic of formal operational thought.

Purves (1984) intensive research on literature response has shown that school children tend to change their responses based on a developmental point of view. Third, fourth, and fifth graders seem to dwell on the

literal aspects of a story. By the sixth grade there is introduction of interpretation which increases in quantity and quality through high school where formal evaluations relate mainly to meaning rather than engagement. That is, "Younger children are more literal and egocentric; they make snap judgements about characters and events. The older children are more abstract and formal; they tend to talk more about the work as something apart from themselves and they tend to be a bit more tentative in their judgements " (p.143). Purves hastened to add, however, that there are some studies that suggest that a simple developmental perspective is not adequate. In fact he suggested that Piaget's stage of concrete operations might occur later with respect to literature.

In summary, cognitive development research of the 1970's and 1980's has associated the development of response to literature with Piaget's cognitive development theory suggesting that response to literature is learned and subject to the abilities of any individual's stage of development. This has been reflected in the recent redirection of reading psychology from behavoristic models to cognitive models which focus on active and constructive roles in reading comprehension.

Discussion Grouping. The use of discussion as a teaching strategy has been employed to initiate and culminate educational projects that have necessitated interaction to some degree. It has been implied, but rarely empirically tested that discussion improves learning (Gall & Gillet, 1980; Hill, 1962; Nelson & Abraham, 1976; Rudduck, 1978; Singer & Dolan, 1980). However, researchers have examined the effects of discussion grouping on the achievement of university, high school, and elementary school students. The JGB technique of shared inquiry through structured discussion is applied in small groups to promote the goal of improved reading comprehension. Therefore, research related to the JGB grouping format is reviewed.

In a study conducted to evaluate whether learning would be facilitated more in a structured method of discussion than in an unstructured method, Maloney (1956) examined the achievement level of university students after engaging in ten sessions of either a strong leader group or a facilitator-type leader group. Data analysis of two achievement measures revealed insignificant differences between the groups. However, more verbal participation occurred in the non-structured group.

Beane and Lemke (1971) examined the effect of heterogeneous and homogeneous grouping of high ability (GMA 122 and above) and low ability (GMA 112 and below) undergraduate students on the transfer of concepts. Statistically significant ($p < .05$) findings showed that high ability subjects benefit from being grouped heterogeneously, but that training high ability subjects as homogeneous quads inhibited individual transfer performance. It was concluded that in the training there was a high probability that each subject developed an individual strategy and when placed in a group setting the group interaction forced the subject to adopt a different individual strategy causing a great deal of interference.

Considerable evidence has indicated that students working cooperatively perform better than do students working individualisticly (Johnson, Maruyama, Johnson, Nelson & Skon, 1981; Yager, Johnson & Johnson, 1985). In a review of 60 years of studies comparing individual and group performance Hill (1982) concluded that group performance was generally qualitatively and quantitatively superior to the performance of the average individual. Hill qualified this by adding, "However, group performance was often inferior to that of the best individual in a statistical aggregate and often inferior to the potential suggested in statistical pooling model" (p.535). This

research would then confirm the belief that the individual performance can be superior to that of a group if the group contains a number of low ability students.

Discussion and reading comprehension. Small group and whole class discussion has been used by educators in some form or another as a means of facilitating reading comprehension. Researchers have linked reading comprehension improvement with discussion (Barrett, 1967; Barrett & Smith, 1974; Caskey, 1970; Hardy, 1980; Hunkins, 1968; LePere, 1975; Reder, 1978; Tierney & Spiro, 1979) but only one study (Hahn & Avery, 1985) could be located in which the hypothesis was specifically tested. Following the review of Hahn and Avery's research, other studies related to discussion and reading comprehension are considered.

Hahn and Avery (1985) investigated the effects of value analysis on students' political attitude and reading comprehension. Using fifteen intact tenth and twelfth grade classes, they examined these constructs in three situations: a) structured value analysis discussions of controversial issues read in a newspaper, b) reading only of controversial isses from newspaper, and c) no reading or discussing of issues - control situation. A pre-test/post-test design was used in the study and reading comprehension was measured with cloze tests, constructed by

the researcher from the reading material used in the study. The results obtained in the experiment demonstrated no significant difference in reading comprehension between the groups. The use of the value analysis technique did not significantly improve reading comprehension. Factors that researchers considered in viewing the results were the effects of using intact groups (SES between groups was different) and the effect of teachers who experienced difficulty teaching the value analysis technique.

Smith (1972) investigated the difference in content learning of seventh grade literature classes in two different classroom situations, one utilizing teacher directed communication and the other student-centered communication of small group discussions. Ten literature classes in four schools in Denver studied a short unit in mythology; five classes were teacher directed, five were student centered. Since seven teachers volunteered for the study each teacher did not use both instructional methods which resulted in a loss of statistical power of the data analyses. Measurement of achievement was assessed by an experimenter-made test which was previously piloted. No reliability or validity of the measurement was reported. Correlational analyses of data indicated that no appreciable difference in achievement was produced by the different instructional methods.

In order to examine the impact on reading comprehension on primary grade children (K,1,2) Galda (1982b) compared the results of children's exposure to drawing, discussion and play activity as a follow-up activity for reading aloud. All the children were evaluated by a criterion-referenced test of cognitive taxonomy as well as a retelling of the folktale. Play, discussion and retelling were video-taped and audiotaped. Results indicated that the two conditions which demanded verbal response, play and discussion, aided in comprehension for kindergarteners and first graders. Second graders showed no difference in comprehension on play and discussion. Statistical data were not reported.

A questioning/discussion technique that has been applied by reading and literature teachers to enhance reading comprehension is Kohlberg's Cognitive Developmental Approach to Moral Education (1975). It is a questioning method based on Kohlberg's levels of moral reasoning in which discussion participants reflect upon moral dilemmas in a passage or story to arrive at a course of action. Dilemmas have no one "right answer", however, by discussing the dilemmas, and by exposure to other points of view, participants arrive at their own personal answer.

Kennon (1980) investigated the effectiveness of this technique on the development and literal and inferential comprehension in sixth grade students. Eighty-six sixth graders in three intact classes were randomly assigned to one of three treatments: 1) moral dilemma questioning and discussion, 2) directed reading activity, 3) routine classwork. A pre/post test design using ANCOVA was employed with two forms of the Stanford Diagnostic Reading tests used as measures. No significant (p<.05) differences of reading comprehension achievement was exhibited by the three different treatments.

In a similar study designed to determine whether second graders' inferential comprehension could be improved, Hansen (1981) developed and tested two instructional strategies. She either gave students several opportunities to discuss inferential guided-reading questions and follow-up questions or employed a prereading strategy that encouraged students to rely on prior knowledge to predict story outcomes. She found that both strategies produced reliable increases in the children's ability to answer inferential comprehension questions.

In a follow-up study, Hansen and Pearson (1983) combined the prereading strategy training and inferential postreading questions and compared it with a control group using the traditional reading method for both good and poor

fourth grade readers. Experienced teachers were trained to administer the treatments instead of having the experimenters do so as in the previous study. The subjects were 40 fourth graders, randomly assigned to experimental or control treatments. Twenty subjects were labeled good readers (mean SAT reading comprehension grade level, 6.3) and 20 poor readers (mean SAT 3.2). Ten weeks of the study involved pre/postreading discussion strategies. At the end of the ten weeks all subjects were evaluated on three separate measures of literal and inferential questions. Results showed that poor readers benefited significantly from the instruction, but good readers did not. Differences in performance were observed on both literal and inferential measures but were striking on the inferential measure. Researchers concluded that the lack of consistant reliable differences among good readers might be attributed to to the fact that many good readers learn similar strategies on their own whereas poor readers seem to need more careful guidance from a teacher.

In summary, although it has been implied that discussion improves learning a scant amount of research exists that tests this hypothesis. Discussion grouping has been extensively studied with evidence that high ability students tend to benefit from being grouped heterogeneously, and students working cooperatively

generally perform qualitatively and quantitatively superior to those of the average individual. No significant differences in learning have been found between structured and unstructured discussions.

Junior Great Books - format, application and experience.

The JGB Program is a literature discussion program for students in grades two through twelve designed to teach students to interpret what they read. The program was established in 1962 by the Great Books Foundation and was developed from the Adult Great Books program founded in 1947. Classic and contemporary literature is read and discussed in small groups through a method of shared inquiry. In shared inquiry leaders center discussion on questions of interpretation to which they themselves are not sure of the answers. This method is used to help participants think harder about what they read and to explore the author's meaning. An explanation of the program's structured format, recently revised in 1984, precedes a review of the studies related to the JGB program, and the factor of JGB experience.

JGB format. Leaders are given handbooks and trained by representatives of the Great Books Foundation in two day-long workshops "to explain, demonstrate, and practice the elements of shared inquiry discussion" (Handbook, 1984, p.2). Major aspects of the shared inquiry approach are:

1) interpretive reading, 2) questions for discussion, and 3) rules of shared inquiry.

Interpretive reading is accomplished by the participants' reading of the literature selection at least twice with a pencil in hand. On the first reading the reader is to note responses that occur while reading by jotting ideas in the margin or underlining words and passages. On the second reading the reader is to look for new passages of significance. This process is designed to promote the development of interpretative questions.

Three kinds of questions are explained and analyzed in JGB: 1) questions of fact, 2) questions of interpretation and 3) questions of evaluation. "Questions of fact concern what the author says, questions of interpretation consider what he means by what he says, and questions of evaluation ask whether what the author says and means is true or consistent with your experience and values" (Handbook, 1984, p.69). The major focus of the JGB Program is the development and application of interpretive questions. Leaders are encouraged to develop and apply their own questions, however, interpretive questions for each selection are provided in each leader's manual. A good question has the following characteristics: 1) focuses on a comprehensive problem of interpretation, 2) requires examination of many lines and passages in the

selection, and 3) generates a number of other interpretive questions" (Handbook, 1984, p.24). Follow-up questions relate to the basic question and are used to clarify responses, substantiate opinions, to solicit more opinions, to keep discussion on track and to overcome a lack of response.

The rules of shared inquiry are:

1. No one may take part in the discussion without first reading the selection.
2. Discuss only the selection that everyone has read.
3. Do not introduce any outside opinions unless you can back it up with evidence from the selection.
4. Leaders may only ask questions - they may not answer them (Handbook, 1984, p.79).

Leaders are to use a seating chart during the discussion to write down participants' comments for follow-up questions and to keep a record of individual participation in each discussion. During the discussion participants are to support their responses with evidence from the text and to agree or disagree with other participants' opinions. After the group has heard and discussed a number of responses to a basic question, the leader is to seek resolution to the basic question by summarizing those responses liked best and supported with evidence from the selection.

JGB studies. Previous Junior Great Books studies indirectly refer to the use of group discussion in the program and examine the program's effectivenes in areas other than inferential reading comprehension.

Casper (1964) investigated whether a type of the JGB program enhanced higher level thinking skills. This program resembled the JGB program in method however it was designed by, and run in, parochial schools. His subjects were gifted fifth graders who volunteered for the program. The instrument used to assess higher level thinking skills was in the developmental stage and inconclusive results were found.

Cashman (1977) examined the effectiveness of the JGB program by measuring verbal meaning and reasoning skills of non-gifted fourth, fifth and sixth graders. A pre/post test design was employed using the Primary Mental Abilities Test to measure skills. Intact groups of fourth and fifth graders were randomly assigned to a control group (no JGB experience) and a treatment group (five months of weekly discussions of JGB). Discussions were led by untrained volunteers. The results of the study indicated a significant difference favoring the experimental group at all grade levels in both verbal meaning and reasoning. A correlation was found between intelligence and performance in the program.

In a study designed to compare systematically the
effects that specific questioning strategies have on
children, Biskin, Hoskisson and Modlin (1976) employed JGB
literature in their study but modified the format in the
control group and experimental group by reading the stories
to the children. Two experimental and one control group
were used in the study. Subjects were from a summer school
program and were a year or more below grade level in
reading achievement. They were randomly chosen from first
and third grades and randomly assigned into one of two
groups, treatment or control group. Experimental groups
discussed a story using two different questioning
stategies, predictive and reflective. All children were
tested for reading comprehension by retelling the story
after the experience and then again two weeks later.
Responses were rated using a revised procedure of the
story-retelling format of the Goodman-Burke Reading Miscue
Inventory (Goodman & Burke, 1971). Retelling was recorded
and qualitatively scored blind by raters. The results of
the study indicated that the JGB strategy of reflective
questioning was superior to other treatments in recall of
factual information, immmediately and on the delayed test.

Using a quasi-experimental nonequivalent control
group design, Bird (1984) examined whether or not the JGB
approach enhanced critical reading skills, critical

thinking skills, attitude toward toward reading, whether sex or previous experience with JGB affected these variables, and whether there was a relationship among the variables. These variables were investigated using "top level" fifth grade readers from four different school districts and assigned into three treatment groups: groups using the traditional reading approach of basal readers, groups using JGB approach to reading, and one group using a mixture of traditional and JGB approaches. All intact groups took a pretest of attitude and critical thinking. Following the completion of the JGB program, each group was retested with these tests as well as a test of critical reading. Data were analyzed using ANCOVA and ANOVA statistical tests. No consideration of multilevel data was made in interpretation of data. Results of the study indicated that students in the JGB program, both on a mixed and full time basis, did better in critical thinking and critical reading than students in basal programs alone. There was no significant $(p<.05)$ difference between the mixed and full time groups. There were no significant $(p<.05)$ differences in attitude toward reading between groups nor was there a differential effect due to sex or previous experience with the JGB program.

JGB Experience. The JGB program is a structured discussion process in which students are trained to examine literature for depth of meaning; that is, learning behaviors and interpretation skills are systematically taught at each level. Program materials are available for grades two through twelve. The JGB program is generally offered as an enrichment program or in addition to the traditional reading program. Since it can be an optional program the length of time in the program will vary from student to student and differences in students' skills may be present. Research has verified that experience is one dimension of the reader that influences individual responses to literature (Wilson, 1966). Galda (1983) explains, "For example, if teachers consistently ask students to interpret fiction, these students may, after enough experience, automatically interpret when asked to explicitly respond to fiction" (p.2).

Of the four related JGB studies that were revealed through a computer search of the literature only one, Bird (1984), considered experience in the JGB program as a possible source of variance. Analyses of covariance and variance were performed to determine whether the JGB program had a different effect on attitude, critical thinking and critical reasoning according to experience in the program. Results of the analyses indicated no

substantiation of the difference and suggested that these results might indicate that there was little cumulative effect to the program. Bird also suggested that since all groups were tested before and after twelve weeks of the program, time might have added up to the erasure of difference due to experience; that is, twelve weeks in the JGB program might approximate saturation point of experience and eliminate any initial differences. She added, "If groups were to be formed and studied on the basis of uniform experience, perhaps results would be different and it could be determined more clearly whether previous experience influences achievement." (p.87). As a result of this conclusion and recommendation, experience in the JGB program is analyzed in this study to determine the effect of this factor on reading comprehension achievement.

Measurement of reading comprehension

Comprehension is not easily defined or measured. In some cases it may refer to literal recall, in other cases it may mean extracting the main theme of a text and in other cases it may mean drawing inferences from what is read. Researchers and educators agree that reading comprehension is not directly observable nor directly measurable (Page & Vacca, 1979; McDonald, 1970) and that no single test should be used as a sole indicator of an individual's reading comprehension (Hutson & Niles, 1983;

Kendall, Mason & Hunter, 1980; Readence & Moore, 1983; Wulz, 1979).

It has been shown that the results of reading comprehension tests can be influenced by types of questions asked (Yost, Avila & Vexler, 1977), passage content, passage organization and testing procedure (Kendall, Mason & Hunter, 1980; Scherick & Hanna, 1977), and prior knowledge (Kingston, 1970; Royer & Cunningham, 1981).

Reading specialists have expressed discontent with most existing tests of reading comprehension as measures of assessing reading comprehension and educational gain (Curtis & Glaser, 1983; Purves, 1984; Royer & Cunningham, 1981). Arguments for change have been based on recent theoretical and empirical activity surrounding the reading process and human performance. Schwartz (1984) explains:

> Criticisms of reading comprehension tests have appeared more or less constantly since the tests were first introduced. The extent to which comprehension tests merely measure a general factor of intelligence, how much the tests measure logical inference as opposed to restating the facts in a paragraph, and whether it is possible to answer comprehension questions solely on the basis of prior knowledge without actually reading the selections are all questions debated by test proponents and critics (see the Buros series for the flavor of this debate) (p.62).

Nevertheless, reading comprehension continues to be measured by standardized normed tests, criterion-referenced tests and other measures designed specifically to assess reading comprehension subskills. Studies related to these measures are reviewed in brief.

Standardized tests. The standardized reading comprehension test generally contains passages, followed by a set of multiple-choice questions. These objective tests provide information about a student relative to a norm group, or an individual with his/her class, or group within classes and are usually timed, easily administered and rated. An accumulation of studies have shown the inadequacy of such tests. Researchers have cited problems of interpretation (Bauman & Stevenson, 1982a, 1982b; Rogers, 1976) and invalidity of tests due to passage independence (Johns, 1978; Pyrczak, 1981; Scherick & Hanna, 1977).

Criterion-referenced tests. Reading comprehension may also be measured by criterion-referenced tests which assess whether or not an individual has reached a stated level of competence in comprehension skills. Criterion-referenced tests have gained in popularity in the minimum competency movement's quest for measurement of specific skills. However, Lyons (1984) correlational studies of Massachusetts State Board Education's criterion tests

evidenced the inability of domain subskills to discriminate between constructs thereby negating the validity of the tests. Although criterion-referenced reading comprehension tests are designed to be skill specific for local application and planning of instruction, Lyons states, "No research supports the contention that criterion reading tests differentiate comprehension subskill abilities better than do norm referenced measures" (p. 296). Additional criticism came from Popham (1978) who questioned the uniqueness and proper construction of criterion-referenced tests.

Cloze tests. The cloze procedure has also been employed as another technique for measuring reading comprehension. Initially developed in 1953 by Wilson Taylor as a readability measure, it has since been employed as a measure of intelligence, a measure of reading comprehension, a method of teaching reading comprehension and a method of monitoring comprehension in metacognition studies (Bailey & Harrison, 1984; Culhane, 1970; DiVesta, Hayward & Orlando, 1979; Rankin, 1970). Recent research has linked the cloze procedure with an understanding of the writing process (Ashby-Davis, 1985) and as a measure for yielding both quantitative and qualitative information of reading ability (DeSanti, Casbergue & Sullivan, 1986).

The conventional cloze procedure is the systematic method of deleting words from a passage, such as leaving out every fifth or tenth word, and substituting underlined blank spaces. The reader of the passage, or cloze test, is to determine the exact words which are deleted in the passage. Rankin (1970) adds, "It is an objective measure of language correspondence between reader and writer. It may be inferred, however, that the ability to make correct word predictions of the precise words deleted in a cloze test is indicative of the respondent's grasp of 'meaning' contained in the message" (p.239). Standardized cloze tests have been published and used to assess reading comprehension and include the Schonell Silent Reading Test B, GAP Reading Comprehension Test, Wide-Span Reading Test, and the London Reading Test, (Rye, 1982). The Degrees of Reading Power Test, a test of reading comprehension, grades 3-12, published by the College Board in 1979, 1980, employs a form of the cloze procedure and has been accepted as well constructed with a wide range of potential uses (Kibby, 1981).

According to Russell (1978) the advantages of the cloze procedure in measuring reading comprehension include: 1) absence of examiner intrusion - the reader deals with the text directly, 2) questions are not delayed until after the fact, 3) subject can be tested on materials used in the

classroom, and 4) it is easily constructed and scored. Disadvantages listed are 1) inadequate for diagnosis of reading comprehension deficiencies, 2) produces high anxiety for the disabled reader, and 3) inappropriate for content subjects since it needs content redundancy to permit educated guessing. Despite these disadvantages Russell suggests more use of the cloze procedure to assess reading comprehension.

Adaptations of the conventional cloze format have been constructed and tested by researchers (Bormuth, 1968; Carver, 1977; Helfeldt & Hank, 1985; Rye, 1982). Inconsistencies in cloze criterion studies suggest that researchers still are in disharmony.

Literary style is an aspect of comprehension that has been related to the cloze procedure. Bormuth and MacDonald (1965) investigated the correlation of cloze test scores with scores on tests to measure the ability to detect an author's literary style. The investigators carefully developed tests to detect the literary style of two authors. In addition, they constructed cloze tests, using every-fifth deletions. A pre-cloze test was given to 150 female college students the beginning of a literature course. After reading and studying the authors' works, the students took the post-cloze test and the tests to detect literary style. Because the two tests correlated highly,

the investigators suggested that a person's sensitivity to literary style was one of the variables which effected performance on cloze tests. However, the extended length of the cloze tests, 1000 words in length, may have allowed a student a much better chance for getting a taste of the author's style.

Dupis (1976) tested the cloze procedure as a predictor of reading comprehension success with literature for 392 tenth graders. Students' comprehension was tested with the conventional cloze as a pretest and with a multiple choice test as a post test. Results of the study revealed that pre and post test scores were significantly correlated at the .01 level.

In a later study Cunningham and Tierney (1979) employed three types of cloze tests from fiction and non-fiction to see if such tests showed promise for use in a pre/post testing paradigm for investigating the acquisition of information from texts by reading. The types of cloze tests varied in deletions and format and were subjected to criteria of reliability, sensitivity to reading, and sensivity to individual differences in reading. Subjects consisted of 457 seventh and eighth graders from schools in North Carolina and Arizona. Results suggest that cloze tests of various formats are usually reliable, but may not be sensitive to reading and individual subject

differences. Both fiction regular cloze and limited cloze were the only tests that passed the three criteria tested. It was speculated that more may be memorable from reading a fiction text between cloze testings than a non-fiction text. "Indeed, it is possible that the ability either to predict or to retrieve the author's choice of words may be easier for a narrative than an informational selection" (p.291).

The relationship between the performance on conventional cloze test (before reading) and the performance on a post oral reading cloze test has been studied and researchers have demonstrated that scores from both formats share significant variance (Garnier, 1976; Page & Vacca, 1979). This would suggest the post oral-reading test as well as the convention test can be a useful indicator of comprehension.

Open-ended tests. Researchers have also utilized the open-ended question as a comprehension measure to provide for possible individual responses and to tap students comprehension abilities in specific reading comprehension subskills (Readence & Moore, 1983). This less precise measure is based on the theory that meanings discerned from a written text may be different for each individual and allows for the possiblility of more than one correct response. Response to literature studies have

examined a number of written responses to literature and instruments have been developed to describe modes, or ways of responding. Categories are used to classify written responses and are the basis of instruments designed to measure or describe response. Instruments employed consist of three basic forms: statement analysis, essay analysis, and response preference measure. The essay analysis has been proven to have greater construct validity than statement analysis and the response preference measure (Galda, 1983).

Purves (1984) used the open-ended questioning format to assess achievement in the 1979-1980 National Assessment of Educational Progress which examined the broad spectrum of reading and literature achievement of students. Using a stratified multistage probability sample design, the reading assessment surveyed approximately 29,000 9-year-olds, 41,000 13-year-olds, and 36,000 17-year-olds who were in school. Reading comprehension was assessed by students' response to multiple choice questions as well as open-ended questions based on three types of expository and three types of literary passages. Open-ended questions were scored by independent raters and interrater reliability was estimated at .94. The most significant finding from this assessment was that students develop few skills for examining the ideas that they glean from their reading,

and, although they could express their own judgements of a work, students did not substantiate their opinions by referring to the text to explain their interpretations.

Additional studies of response to literature are too numerous to include in this review, but research in this area that have related to discussion and reading comprehension have been cited in the subdivision of reading and comprehension.

Summary

The complexities of the reading process and its relationship to literature has been studied since the early 1900's. Theories of literature response have evolved which relate comprehension to the personal experience of the individual reader, the intention of the author, an intersection of the reader experience and the text, and the interpretive community of the reader. Studies have linked Piaget's cognitive development theory to a child's concept of story and response patterns. Recently, reading psychology has redirected its perspective from behavoristic models to cognitive models which focus on an active and constructive role in reading comprehension.

It has been implied that discussion improves learning yet rarely has this hypothesis been empirically tested. Discussion grouping, however, has been extensively studied. Research has indicated that high ability students

tend to benefit from being grouped heterogeneously, and students working cooperatively generally perform qualitatively and quantitatively superior to those of the average individual. No significant differences in learning have been found between application of structured and unstructured discussions.

A limited amount of research investigating the effects of discussion on reading comprehension has resulted in statistically insignificant evidence for gifted readers. It has been suggested, however, that poor readers benefit significantly from pre and post reading strategies related to discussion.

Since its formulation in 1962, the JGB program, a structured literature discussion program, has been implemented in the classroom to improve gifted and non-gifted student's reading comprehension. Researchers have examined the JGB program for its ability to enhance higher level thinking and reasoning skills, improve attitude toward reading, and increase literal and recall comprehension. JGB experience has been considered a possible source of variance on program effects. In some cases, selective reading skills of non-gifted readers improved significantly. In others, results have been insignificant or inconclusive due to study design or use of undeveloped measurement instruments. In most cases,

data was obtained from intact classes and analyzed at one level.

The use of objective and subjective tests as effective measurements of reading comprehension has been examined by researchers. Although standardized objective tests are most frequently employed to assess reading comprehension achievement, it is generally agreed that since reading comprehension is not directly observable, no single measure should be used a a sole indicator of reading comprehension.

Part II: Methodological Issue of Cross-level Inference

To determine individual success in reading comprehension as a result of JGB literature discussion the issues of cross-level inferences need to be addressed to assure consistency in data collection and explanation. Data obtained from this study is multilevel, since it reflects reading comprehension scores obtained by students taught by teachers to groups within two schools. To specify the educational effects of the JGB program as posed by study research questions, selection of the appropriate unit of analysis had to be considered as well as the specification of appropriate models in the estimation of relationships from multilevel data obtained.

This interdependence among units has led to difficulties in making inferences across levels of analyses and has been extensively examined by social scientists (Burstein, 1980; Glick, 1980). Researchers have investigated change in units of analysis, ecological inference, aggregation bias, correlations based on grouped data and grouping observations. Interpretational errors due to inconsistency across levels of theory, design, measurement and analysis may yield cross-level discrepancies, or the difference between the observed relation and the theoretical or inferred relation.

Research has indicated that using group-level data to determine educational effects on the student level will result in biased estimates unless proper judgements are made about the nature of causal relations between grouping criteria and variables in the model employed in the research (Firebaugh, 1978; Hannan & Burstein, 1974). This is reemphasized by researchers in the field of reading comprehension who are similarly concerned with the domain or population to which the results of their studies will reasonably generalize. Pearson (1979) explains:

> In the past, many of us have used inappropriate if not illegimate design and statistical procedures to draw the inferences we have. In the future we need either to adjust our statistical models and procedures used to draw inferences, or we need to hedge

our conclusions so that they do not outweigh
our methodology (p.157).

Part II of the literature review is subdivided to
include first, studies that relate to the theoretical
issues of aggregation or problems of cross-level inferences
and second, studies of practical application that have
attempted to examine, estimate, or contrast parameters of
individual level data from grouped level data.

Problems of cross-level inferences

The issue of cross-level inferences was brought to
the attention of social scientists more than three decades
ago when Robinson (1950) showed that correlations between
variables at the aggregate level differed from correlations
between the same variables at the individual level. Using
the 1930 U.S. census data, Robinson computed the
correlation between race (black/nonblack) at the individual
level and at the regional level. The correlations were .20
and .95 respectively confirming his contention that
aggregate data misestimates individual-level correlations.

As a result of Robinson's findings the difficulties
involved in making inferences across levels has been widely
studied. Research stemmed from concern about making
inferences about relationships at one level from
relationships found in data at a different level. For the
purpose of this study literature pertaining to the
following problems of cross-level inference is reviewed:

1) selecting appropriate units of analysis, and 2) drawing inferences across different levels of analysis.

Selecting units of analysis. Haney (1980) defines the unit of analysis as "the primary entity of which data are analyzed in an evaluation or other study" (p.1). For example, the student, the study group or the school might be selected as the unit of analysis and its selection can sharply affect results. In order to determine the proper unit of analysis, arguments continue to exist over the units of treatment, independence of units and treatment effects, and the appropriate size of units.

Cox (1958) suggested that the unit of analysis should be "the smallest division of the experimental material such that any two units may receive different treatments in the actual experiment" (p.2). This differentiation may not be clear in educational research where, for example, although all students in a classroom may receive the same instructional treatment it has been argued that different pupils in any single classroom undergoes different experiences and thus receives different treatments (Haney, 1980).

Another rationale for choosing a unit of analysis deals with the independence of units and their response to treatment effects. Cox (1958) asserts that different experimental units should respond independently of one

another so that treatment applied to one unit does not affect the observation obtained on another unit. In addition, the occurrence of an extremely high or low observation on one unit should not effect another unit. "The first requirement is necessary to allow the effects of the different treatments to be sorted out from one another, the second ensures that a proper estimate of error is obtained from the comparison of observation on units receiving the same treatment" (p.196).

Contrasts between analyses using the individual or the aggregate as the unit of analysis can be found in the literature. Pedhazur (1982) notes the distinction between the R^2 obtained when individual and aggregates are compared. He states, "When individuals are used as the unit of analysis, R^2 indicates the proportion of the <u>total</u> variance accounted for by the independent variables. When, on the other hand, aggregates (e.g., classes, schools) are used as the unit of analysis, R^2 indicates the proportion of variance of the <u>between</u> aggregates that is accounted for by the independent variables " (p.539). He warns that when analyzing group data, a small variance between groups will produce an inflated R^2 yet it will explain only a minute proportion of the total variance.

Hopkins (1982) has shown that the common recommendation to use group means where there may be nonindependence among the observational units is unnecessary, unduly restrictive, and impoverishes the analysis as well as limits the questions that can be addressed in a study. He suggests using the proper statistical model as an answer to the question of choosing the proper experimental unit or unit of analysis for treatment effects. He states, "When random factors are properly identified and included in the analysis, the results for all common effects (F's and critical F's) are identical in balanced ANOVA designs, regardless of the observational unit employed" (p.17).

Drawing inferences across different levels of analysis. The previously cited study by Robinson (1950) alerted researchers to hazards of attempting to generalize statistical results from one level of analysis to another. This fallacy can be committed in either direction working downwards, by projecting from groups or categories to individuals, or upwards, by projecting from individuals to higher units. It is generally acknowledged that aggregation may lead to more change in correlation analysis across units and levels of analysis than in regression relations (Blalock, 1964; Burstein 1978, 1980; Knapp, 1977).

To alleviate this problem it has been recommended that researchers conduct analyses at different levels and with different units and conduct cross-level analyses (Burstein, 1980). Burstein advises, "Clearly, models for analyzing educational effects at the elementary level need to emphasize accurate assessment of the role of the classroom and the teacher and to be sensitive to the dependence associated with instruction in intact groups. Classroom level and within classroom analyses are likely required" (p.177). When the purpose is to determine factors affecting pupil performance, analyses of between group (class, school, etc.) means can hide important differences in the within group distribution of pupil outcomes and educational inputs. Different groups can have the same mean performance yet vary on the other moments of the groups' distribution.

In a review of the major issues about the role of level of analysis in the specification of educational effects, Burstein (1980) summarizes the main findings about the differences between regression models at two levels. His findings indicate that: 1) estimates of regression coefficients from different levels of analysis are inconsistent unless groups are formed randomly, 2) the magnitude of the differences between coefficients from individual and group level data is a function of the

relationship of the grouping variables to the regressor, to the dependent variable net of the regressors, and to the ratios of the variances of the regressors at the two levels of analysis, 3) differences in coefficents across levels of inference are a clear indication of specification bias except in cases when groups are formed directly on the basis of values of the dependent variable, 4) estimates of coefficients from one level of analysis are inefficient for estimating coefficents from another level unless observations are groups according to values of the regressors, 5) collinearity among regressors seriously affects the consistency and efficiency of estimation across levels, 6) aggregation gain is possible when grouping minimizes grouped variation in confounding varables and when regressors at the lower level of aggregation are measured with error, and 7) in order to assess group effects, knowledge of the process that groups observations and the nature of the problem and research design is crucial. Burstein adds, "Research on problems of cross-level inference has shown that analyses of educational effects at different levels reveal substantial differences across levels for specific models. Aggregation typically inflates the estimated effects of background on outcomes and decreases the likelihood of identifying effective

teacher - classroom - school characteristics and practices" (p.181).

Practical application studies.

An example of the controversy that exists over the appropriate level of analysis in educational research has been illustrated in a study which examined the determinants of organization effectiveness using group-level data from 104 school districts in Colorodo (Bidwell & Kasarda, 1975, 1976). The researchers hypothesized that environmental conditions that confront a school district (size, fiscal resources, percent non-white population, education and income levels of parental risk population) would affect levels of student achievement primarily through their effects on the structure and staff composition of these districts. Measures of district structure were pupil-teacher ratio, administrative intensity, and the ratio of supporting professional staff to teachers. These conditions were linked in a causal model to the median reading and mathematics achievement test scores of the districts' high school students. Results indicated that pupil-teacher ratio and administrative intensity depressed median levels of achievement; whereas, staff qualifications fostered student achievement. Of the environmental

conditions, only percent non-white had consistently significant direct effects on median achievement levels.

The authors of this research stated that all measured relationships were at the level of the school district and that their intent was to measure overall effectiveness of a school district as measured by the aggregate achievement level for all its students at a given grade not analyzing antecedants of the academic achievement of individual students. However, the chosen unit of analysis was challenged (Alexander & Griffin, 1976; Hannan, Freeman & Meyer, 1976) since the theoretical argument of the study concerned the experience of individual students and not aggregates of students. It was contested that the analysis with data aggregated above the theoretically appropriate level led to grouping bias since the social processes which allocated students to school districts would tend to select them on individual variables which affect achievement. Comparable data were reanalyzed by both commentators at the individual level and included variables that described characteristics which pupils bring with them into the organization. It was then concluded that while organization effects may be found in a properly specified model, these effects were smaller than reported by Bidwell and Kasarda (1975).

The originators of the study counterreplied that they were justified in using the aggregate unit of anlysis since their research question was about the relative effectiveness of school districts and not the academic achievement of individual students (Bidwell & Kasarda, 1976).

The most frequently cited educational effects' studies conducted at many levels and with data from multiple levels has been the Coleman Report (Coleman et al., 1966). Results are based on between school analyses and a mixed model with individual level measures of school and teacher characteristics. Langbein (1977) reanalyzed a portion of Coleman's Equal Educational Opportunity Survey, completed in 1964, to investigate the difference between identical model of student achievement estimated for both student and school. Her study indicated that standardized aggregate measures of the influence of race and status on achievement were inflated versions of their individual level counterparts. Reliance on unstandardized estimates was shown to be equally hazardous. Individual level analysis disclosed socioeconomic status and ability of peers had a direct effect on the achievement of students. When using aggregate data it was not possible to estimate uniquely the effects of these types of contextual properties. In addition, Langbein advocated the

simultaneous analysis of individual and grouped data to assess whether aggregate estimates were more biased then those from individual level data. She concluded that divergence of individual from group level estimates was an indicator of bias in group level estimates and convergence suggested either that both models were properly specified or that the omitted variable had no relation to the grouping variable. The exclusive reliance on aggregate data, in this study, precluded an understanding of whether manipulable variables affected average as well as individual achievement levels.

Researchers examining the relationship between grades and students rating of instruction (Stumpf & Freedman, 1979) investigated whether university students could provide unbiased appraisals. Individual and class effects were analyzed to see if there was a distinguishing variance between them. During the 1976-1978 semesters, over 27,000 undergraduate and graduate students from NYU anonymously completed a Course-Faculty Instrument during the last week of classes before exams. Expected grade rather than actual grade was used in the research. Correlations between CFI dimensions and expected grade were computed at both the student and class level of analysis. In order to obtain an unconfounded individual effect, it was necessary to partial out the effects of classes for the

individual level of analysis. Results of the study indicated that 1) at both levels of analysis expected grades positively correlated with instructor ratings, 2) expected grades accounted for more instructor rating variance at the class level than at the student level, and 3) the grades variable was considered two constructs, one for each level of anlysis. These results of the study suggested a need to offer alternate explanations consistent with the differences found at the two levels of analysis.

Greene's (1980) ATI (aptitude x treatment interaction) study supported the importance of including class analysis in ATI research. Greene investigated effects on performance and motivation of four motivational appitudes and two cognitive aptitudes with the manipulation of a choice versus no-choice over learning procedure treatments. Both treatments were implemented in each of nine fourth or fifth grade classrooms (N=165) over a four-week period. Students were randomly assigned into one of two choice groups. Three sets of simple regression results (by treatment, within class and between class) evidenced differential effects of treatment on individuals versus classes. In the results of the two selected-outcome pairs (general ability and confidence in one's ability) the pooled regression lines indicated that students in high ability or high confidence classes learned better when

given some choice, while students in low ability or low confidence classes learned better under more structured conditions. "That is, the significant interaction was not with individual student aptitudes, but rather with the social effects of being a member of a class that was relatively high or low in ability or academic confidence" (p.300). Although limited sample size did not present conclusive evidence for a given learning phenomenum, Greene argued for classes rather than individuals as the proper unit of analysis for statistical inference.

Summary

Extensive research has indicated that utilizing aggregate level data to determine educational effects on the student level may result in biased results. Of particular concern has been the selection of appropriate units of analysis and drawing inferences across different levels of analysis. Controversy continues to exist over selection of units of treatment as well as the independence and size of units. Contrasts and differences have been found between regression models analyzing individual and aggregate level data. Since analyses of educational effects at different levels have revealed substantial differences across levels for specific regression models, it has been recommended that researchers conduct analysis at different levels and with different units.

A review of selected studies, in which researchers challenged the unit of analysis or examined the effect of inferences drawn across different levels of analyses, suggested that the prime choice of unit of analysis was based upon the nature of the research question. However, exclusive reliance on aggregate data tended to produce inflated results of individual level data. Therefore, a simultaneous analysis of individual and grouped data was preferred to assess whether aggregate estimates were more biased than those from individual level data and to offer the possibility of alternate explanations if differences were found at two levels of analyses.

CHAPTER III

METHODOLOGY

This investigation utilized the Junior Great Books literature program to examine the effects of discussion on reading comprehension of gifted fifth graders. The following sections describe sampling, measures, procedures, design, and data analysis employed in the study.

Sampling

Samples were drawn from two elementary schools in a suburban school district in Northern Virginia. One school sample represented two classes of gifted fifth graders with no previous JGB experience. The second school sample represented two classes of gifted fifth graders with length of experience in the JGB program ranging from from one year to three years. Samples within classes were randomly assigned to experimental and control groups with a proportional ratio of male and female represented in each group. The initial total sample size consisted of 94 students. They were fifth graders who had been placed in two Gifted Centers, or subschools within community schools, in self-contained classrooms. Due to student absenteeism and attrition, actual sample sizes varied between stories; 84 participated in story one and 83 participated in story two.

Four JGB trained teachers were employed in the study as leaders of one experimental and one control group in each of the four classes. Three of the teachers were newly trained JGB leaders; the fourth teacher was trained six years ago, and had actively led JGB groups for the past four years. All were certified teachers with a range of teaching experience from ten to nineteen years.

<div align="center">Measures</div>

To capture the immediate effect of group JGB literature discussion on students' reading comprehension with minimum disruption in the regular classroom daily schedule, the researcher developed measures that emerged from a JGB literature unit using the cloze procedure and the open-ended question format (see Appendix). These measures were employed in the study to obtain students' reading comprehension scores. Scores obtained from cloze tests (pre and post) and open-ended questions (post test) represent dependent variables in the study.

Cloze tests.

The Cloze procedure was initially designed as a tool to measure readability by Taylor in 1953 and has since been effectively employed to measure reading ability and develop reading ability (Rye, 1982). Rye argues that performance on a cloze test involves the child in the following:

i. recognising words;
ii. using semantic, syntactic and at times stylistic information to infer and predict;
iii. drawing meaning from outside the context of the immediate sentence;
iv. skimming, to recap on what has been read;
v. scanning, in search of unspecified information that may help the prediction (p.32).

Apart from the ease of test construction there is evidence for the validity of cloze reading tests for the purpose of measuring reading comprehension and its subskills (Bormuth, 1969; Rankin, 1970; Rankin & Culhane, 1969; Smith & Zinc, 1977). Jenkinson (1957) correlated cloze scores with results of standardized reading tests and obtained correlations of .78 and .73 in vocabulary and comprehension respectively. The highest correlation in Jenkinson's study was .82 between cloze test results and objective questions based upon the same material as were the cloze tests. This evidence suggests that " the cloze procedure produces tests which tend to measure specific comprehension of an article better than general comprehension." (Rankin, 1970, p.247).

The highest validity correlations for small-scale cloze tests have come from Bormuth (1969). He obtained correlations ranging from .73 to .84 on cloze and multiple-choice tests he constructed to measure the comprehension of vocabulary, explicitly stated facts, sequences of events, inferences, causal relationships, main ideas and the

author's intent in each of nine passages. The tests were administered to 150 students enrolled in grades four, five and six. Results of principal components analysis found only one factor that accounted for variance in the matrix and this was interpreted as providing evidence that the tests measured reading comprehension skills.

There is also evidence of reliability of cloze tests. Studies of Taylor (1953) employed the stringent Test/Retest method for calculating reliability. In a series of experiments using 24 and 18 juniors and seniors at the University of Illinois journalism courses, Taylor obtained coefficients ranging from .80 to .88. In a large study undertaken by Landsheere (1972), twelve cloze tests of between 250 and 300 words were used with a population of nearly 4,000 French children aged between nine and eighteen. A total of 72 reliability coefficients, ranging from .79 to .97, were obtained using internal consistency methods. Rye (1982) reports internal consistency (Kuder-Richardson20) coefficients of .76 and .97 on cloze tests taken by 54 and 36 twelve-year-olds in public and private schools respectively.

The cloze procedure was used as one measure of reading comprehension in this study. This researcher randomly selected two, 300-word, passages of equal readability, determined by the Fry Readability Formula

(Fry, 1968), from two stories in <u>JGB Series Five</u> (1984). Stories selected for the study were "The Nightingale" (R.L.,6.7) and "Spit Nolan" (R.L.,6.6). Story passages served as pre and post cloze tests to assess students' reading comprehension achievement. Students were asked to complete the blanks with words that might be used by the author. Although minor misspellings were allowed, only the author's original word was counted as correct. This method was employed in previously cited validity studies.

Cloze tests, derived from "Spit Nolan", were piloted by fifth grade gifted students to determine reliability of measures and assess the reaction of the students to this form of test. Six groups of gifted fifth graders, three classes, participated in the pilot test at one school in Nothern Virginia. The groups were all intact groups that had just completed the final story. One cloze test of 31 deletions was administered to two groups and two cloze tests of 25 deletions (alternate forms) were administered to four groups of students. A total of six groups piloted three forms of the cloze test, a total of 72 students were in the six groups.

Although 50 deletions have been recommended for measures of reading comprehension (Rye, 1982), fewer deletions were employed to avoid obtaining meaningless data as a result of possible student fatigue due to repeated

tests of one pretest and two posttests at one sitting. This also allowed for a minimum interruption in the regular school schedule. Initial results of the Kuder-Richardson20 index of internal consistency of test items indicated reliability coefficients of .60 and .72 for the 31 deletions and .47, .59, .72, and .74 for the 25 deletions. A coefficient of equivalence of .71 was obtained on the alternate forms of the 25-deletion tests.

From this evidence it appeared that use of the 31 deletions would provide more adequate data as a measure to assess reading comprehension achievement for the purposes of this study. Students reaction to the cloze test format was for the most part positive; the pilot cloze tests were completed by most students within fifteen minutes.

Open-ended test

The open-ended questioning format was used as an additional measure of reading comprehension to assess the individual student's ability to extend the understanding or meaning of a written work through interpretation, a major aim of the JGB program, and to allow for more than one correct response.

Using three groups of JGB fifth grade gifted students, the following question was piloted, along with the cloze test pilot: "What was the main idea of (name of Story) and how does the author use the characters to

arrive at the main idea in this story?". Observations made by the researcher during piloting of this question and assessment of students' written responses revealed that approximately two percent of the students were able to appropriately respond to the question. A similar result was obtained in a related ethnographic JGB study, cited previously (Boraks, Early & Sable, 1986), when students were asked to make an analysis of the author's intent in a literature selection. Research by Applebee (1978) and Carter (1985) has suggested that fifth grade students are not able to respond to higher level questions that require analysis of the structure of the work since they have not yet reached the formal level of cognitive development. According to Carter's study, gifted students at age ten are in the transitional level of cognitive development and begin to approach the formal level at age thirteen.

Therefore, this question was replaced in the study with an open-ended question similar to one employed in the 1979-1980 National Assessment of Educational Progress Assessment of Educational Progress (Reading/Literature Released Exercise Set, 1981). The following questions assessed the student's understanding of a written work through interpretation of the text:

1. Think about the story again. What kind of person was ____(name of major character in story)____?

Describe ___ (name of major character in story)
in a few words below.

2. What was it about the story that led you to
describe ___ (name of major character in story) the way
you did in question 1? Write your answers on the lines
below.

Responses were evaluated using similar criteria
employed in the National Assessment of Educational Progress
1979-1980. A seven-point, cumulative, unidimensional
scale, constructed by the researcher, was employed to
assess students' responses to the open-ended questions.
Two reading experts, not involved in the study, read both
JGB stories and established four critical character traits
for each major story character. With this information,
thirty-four responses were rated blind by two trained JGB
leaders, not employed in the study, to establish interrater
reliability. Raters were instructed to take into
consideration the entire student response, question one and
question two, before rating the response. Criteria for
scoring open-ended tests can be found in the Appendix.

Initially, interrater reliability was assessed at
.71, or, of the 34 papers scored, 24 received identical
scores. Following alterations of criteria and amplification
of directions, raters rescored the same thiry-four papers
and rater reliability of .82 was achieved; 28 received

identical scores. A Kappa Coefficient (Cohen, 1960) was used to correct for chance that the two raters would assign the same score to a paper. The resulting Kappa Coefficient, or chance corrected value, of .76 indicated that the interrater reliability represented "substantial" agreement beyond chance (Landis & Koch, 1977). Five trained JGB leaders, who were not part of the study, rated the remaining responses following a training session by the researcher.

Procedures

Following random assignment of students into experimental (discussion) and control (non-discussion) groups, each group met twice, during the school day, for approximately ninety minutes, to complete the study. Two JGB literature stories, selected at random by the researcher, were read twice by students in the experimental and control groups. Stories were assigned by teachers one day in advance of the treatment, and reread immediately before the experiment." Each teacher lead one experimental
"
and one control group; four teachers were included in the study. Prior to the experiment, the researcher explained the study to the students and reviewed elements of the cloze procedure with each class.

For approximately thirty minutes discussion groups followed the prescribed format of the JGB program.

Leaders of each experimental group selected questions that were discussed in the groups. A summary of the questions can be found in the Appendix. Students in non-discussion groups worked independently on JGB written activity developed by the researcher (See Appendix). Pre and post cloze tests, constructed from two randomly selected JGB literature units, were administered immediately before and after discussion/non-discussion treatment to assess students' reading comprehension achievement. Upon completion of the post cloze test, all students also responded to open-ended questions related to JGB literature selection to assess interpretive reading comprehension achievement and to allow for the possibility of more than one correct response.

To determine whether teachers followed the prescribed format of the study, discussion and non-discussion sessions were tape recorded and assessed by the researcher and one JGB-trained teacher not involved in the study. Tape recordings of all discussion groups were evaluated according to an adaptation of a system developed by previous JGB researcher, Bird (1984). Tapes were listened to for five minutes on, five minutes off, three times, providing three observations totaling fifteen minutes for each discussion. Ratings were made on student behavior, teacher behavior, and overall focus of the

discussion (see Appendix). Total coding scores of each rater were compared to establish interrater reliability. Tape recordings of all control groups working independently were listened to by same raters for evidence of JGB story discussion.

Design

The study is experimental in nature. Factors are crossed and nested. When both factorial and nested factors appear in the same experiment, it is considered a nested-factorial experiment (Hicks, 1964). The crossed independent factors are schools (experience/inexperience) and methods (discussion/non-discussion). The nested independent factor is teachers nested within schools. In light of this design the effect of schools and experience/inexperience were confounded and thus inseparable.

Data Analyses

Subjects who had no previous experience in the JGB program and participated in the study in the experienced school were eliminated. Subjects were then randomly discarded to achieve a balanced design across comparisons. In story one, six subjects were randomly deleted; in story two, five subjects were randomly deleted. This created thirty-nine subjects in each comparsion, schools and method, in each story. A balanced design was needed to

implement the proper models to analyze multilevel data and to provide information on additional sources of variation (Hopkins, 1982).

Pretest cloze scores were partialed out from posttest cloze scores through regression techniques to create a dependent variable based on residual gain scores. Residual gain scores were found by calculating the difference between the posttest score and the predicted posttest score. The following formula was used:

$$Y_{ijkm} = Y - [bX + C]$$

where

Y_{ijkm} = residual gain score

Y = posttest score

X = pretest score

b = $\dfrac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}$

C = $\bar{Y} - b\bar{X}$

$\bar{Y}$ = posttest mean

$\bar{X}$ = pretest mean

The residual gain score was employed rather than the raw gain score to prevent random error of measurement and to provide a "corrected" measure of gain (Cronbach & Furby, 1970). Tracy and Rankin (1970) add, "This technique permits the measurement of differences in improvement with subjects who have been equated statistically on the basis

of the pre-training measurement....It removes the influence of regression effects upon the measurement of improvement" (p. 366).

A parametric procedure, crossed and nested analysis of variance (ANOVA), was used to analyze the data. This procedure was chosen since it would facilitate the analysis and interpretation of the data. In order to estimate stability across replications, the experiment was replicated once. Procedure GLM on SAS was used to run separate analyses of each story on individual-level data and group-level data.

Two different models were used to analyze individual level data and group level data. Individual-level data was analyzed using the following model:

$$Y_{ijkm} = \mu + E_i + T_{j(i)} + G_k + GE_{ki} + GT(E)_{kj(i)} + \epsilon_{m(ijk)}$$

Aggregate-level data was analyzed using the following model:

$$Y_{ijkm} = \mu + E_i + T_{j(i)} + G_k + GE_{ki} + \epsilon_{m(ijk)}$$

where;

Y_{ijkm} represents residual gain score,

E_i represents experience/inexperience,

$T_{j(i)}$ represents teachers nested within experience,

G_k represents discussion/non-discussion,

GE_{ki} represents interaction between discussion

               factor and experience factor,

$GT(E)_{kj(i)}$ represents interaction between discussion

factor and teacher within schools factor,

and

$\in_{m(ijk)}$ represents random error.

The expected mean square values of sources of variance were calculated and applied as the test statistic in testing the null hypothesis and are shown in Table 1. In this study factors of Experience and Teachers (Exp.) were considered fixed since they were systematically selected. The Group factor was designated as a random effect because the desired inference is to JGB groups "like these" and were randomly selected from a population of groups along a continuum extending from no discussion through extensive discussion. To test for significance of main effect, Experience, the mean square value of Teacher nested within experience was used as the error term. To test for main effect, Groups (discussion/non-discussion) and interaction effect, Experience*Groups, the mean square value of Teacher(Exp.)*Group was used as an error term.

Assumptions underlying the ANOVA model were examined for possible violations. Minimum degrees of non-normality of error terms in story one were investigated. Data were ranked and transformed on SAS and reanalyzed using the General Linear Model. Analysis of data using both test procedures produced similar p vaules. This indicated that

Table 1

EMS for Individual Level Data and Aggregate Level Data

Model: Individual Level Data	2 F i	2 F j	2 R k	10 R m	EMS
Experience $_i$	0	2	2	10	$\sigma_e^2 + 20\sigma_T^2 + 40\sigma_E^2$
Teacher(Exp.) $_{j(i)}$	1	1	2	10	$\sigma_e^2 + 20\sigma_T^2$
Group $_k$	2	2	0	10	$\sigma_e^2 + 10\sigma_{TG}^2 + 40\sigma_G^2$
Group*Exp. $_{ki}$	0	2	0	10	$\sigma_e^2 + 10\sigma_{TG}^2 + 20\sigma_{GE}^2$
Group*Tea.(Exp) $_{kj(i)}$	1	1	1	10	$\sigma_e^2 + 10\sigma_{TG}^2$
$\epsilon_{m(ijk)}$	1	1	1	1	σ_e^2

Model: Aggregate Level Data	2 F i	2 F j	2 R k	1 R m	EMS
Experience $_i$	0	2	2	1	$\sigma_e^2 + 2\sigma_T^2 + 4\sigma_E^2$
Teacher(Exp.) $_{j(i)}$	1	1	2	1	$\sigma_e^2 + 2\sigma_T^2$
Group $_k$	2	2	0	1	$\sigma_e^2 + 4\sigma_G^2$
Group*Exp. $_{ki}$	0	2	0	1	$\sigma_e^2 + 2\sigma_{GE}^2$
$\epsilon_{m(ijk)}$	1	1	1	1	σ_e^2

the minor violation of non-normality did not affect the conclusions (Pavur & Nath, 1986).

A correlation between both dependent variables, cloze test gain scores and open-ended test scores, was calculated to determine whether one or two analyses would be performed. Results of the Pearson and Spearman correlation procedures (story one, $r=.02$, $p<.84$, $r=.05$, $p< .64$; story two, $r=.27$, $p<.02$, $r=.26$, $p<.02$) indicated that the dependent variables were not significantly correlated, $p<.01$ or lower. Therefore, the dependent variables were analyzed separately.

Results of analyses were examined to determine if differences in reading comprehension scores existed between story one and story two. Models were also examined for significant effects and to determine if they adjusted out the effects of nesting. All hypotheses were tested using the F tests of significance. The alpha level chosen for this study is $p<.05$. The results of all analyses are reported in Chapter Four.

CHAPTER IV

RESULTS

The results of each analytic procedure is summarized in this chapter. First the results of the observational analysis of the study format application and description of the data are presented. Then the results of the investigation are reported. The questions are:

1. Does discussion/non-discussion make a difference in reading comprehension achievement?

2. Does the level of JGB experience make a difference in reading comprehension achievement?

3. Do similar results in reading comprehension achievement occur across study replication using different JGB stories?

4. Does the application of the General Linear Model to the individual-level data provide a different solution than the solution obtained when applying the group-level data?

Application of Study Format

An essential element of the experiment was the compliance of discussion and non-disucussion groups to the study format by students and teachers. During the experiment students in non-discussion groups remained in the classroom and independently worked on a written JGB

78

activity (see Appendix). Students in discussion groups met in a separate room and discussed JGB story with the teacher. Results of the discussion format and non-discussion format follow.

Discussion format

All teachers received JGB training and used the same materials in the study. Because implementation of the JGB discussion format could vary between groups, discussions for each story were audiotaped and rated by the researcher and a JGB leader not involved in the study. Tapes were rated according to the directions on the form for the experimental treatment in the Appendix. Scores were translated into a four-point scale with four representing the category "always", three "frequently", two "sometimes" and one "never". Total raw scores obtained by raters gave an interrater reliability of .94 for story one and interrater reliability of .99 for story two.

The results of the raters' observation of discussions are found in Table 2. The overall means of 24.8 in story one, and 23.2 in story two, indicate that teachers fell between "always" (28 points) and "frequently" (21 points) in both stories. Leaders were rated consistently high on avoiding personal interpretation with a mean rating of 11.8 in story one and story two which falls between "frequently" (9 points) and "always" (12 points). In both stories lower

Table 2

Mean Ratings of Discussion Observation in Experimental
Groups Story One and Two

Criterion	Story One Rating	Story Two Rating
Interpretation is made by students not leader*	11.8	11.8
Answers are substantiated from text*	9.2	7.4
Discussion is appropriate to the format**	3.9	4.0
Total	24.8	23.2

*Possible range 12 for "always" to 3 for "never."

**Possible range 4 for "always" to 1 for "never."

ratings of 9.2 (story one) and 7.4 (story two) were achieved on substantiation of answers by students. Story one ratings fell between "always" (12 points) and "frequently" (9 points), whereas, story two ratings fell between "frequently" (9 points) and "sometimes" (6 points). The discussions were rated consistently high on their appropriateness to the format with mean ratings in the "always " (12 points) range. All major questions asked by the leaders were taken directly from the JGB manual and were interpretive in nature.

Non-discussion format

Control groups, non-discussion, in stories one and two were tape recorded as they completed the written independent activity. Although the researcher and, on occasion, instructional aides monitored all non-discussion groups, the actual taping of independent work fostered complete student and teacher compliance to the format and insured similarity of taping treatment for both experimental and control groups. No obvious sounds of story discussion by subjects in the control group could be discerned on the tapes. Therefore, JGB stories were not discussed by subjects in control groups and study format for non-discussion groups was followed.

Description of Data

Initially 167 subjects completed both stories, 84 from story one and 83 from story two. However eleven subjects were randomly discarded, six from story one and five from story two, to create a balanced design across comparisons. The final analyses of data consisted of 39 subjects in discussion/nondiscussion groups and 39 subjects in experienced/inexperienced schools; a total of 78 subjects for each story. Since the JGB is an optional program offered to students, the amount of experience in the program will vary from student to student. Final data from subjects in experienced schools represented a range of one to three years of subjects' experience in the JGB program. Data from subjects in inexperienced schools represented no student experience in the JGB program. There were 25 females and 53 males that participated in experiment one, story one; 24 females and 54 males participated in study replication, story two.

In the analyses which follow, scores termed "cloze" refer to residual gain scores obtained by partialing out pre-test cloze scores from posttest cloze scores through regression procedures (possible range -31 to +31). The residualized score is a corrected measure of gain since it removes from the posttest score the portion that could have been predicted linearly from the pretest status. It is the

part of the true gain not predictable from true pretest status (Cronbach & Furby, 1970).

Those scores termed "openended" refer to raw scores obtained from openended post tests (possible range 0 to 7) and represent scores rated on a cumulative, unidimensional scale (see Appendix). Cloze and openended scores reflect subjects' reading comprehension achievement in this study.

JGB stories, chosen at random by the researcher, represent two types of literature genre. Story one, "The Nightingale", is fantasy fiction and story two, "Spit Nolan", is realistic fiction.

Results of Investigation of Research Questions

Questions 1 and 2 were investigated by hypothesis testing. The following null hypotheses were tested:

H_1 There is no significant difference between discussion and non-discussion groups on reading comprehension achievement.

H_2 There is no significant difference between experienced and inexperienced groups in the JGB program on reading comprehension achievement.

In presenting results related to the research questions of this investigation, Questions 1 and 2 are combined, and separately reported for each story, because Analysis of Variance accounts for effects of treatment and experience on individual level data. The results of the

ANOVA's on both reading comprehension measures, cloze and openended, are included within each story. Two analyses were necessary since cloze and openended scores were not significantly correlated.

Results in relation to question 1 and 2: Story one

In Table 3 is shown the results of analysis of variance of story one scores, cloze and openended, on individual level data. Results indicate no significant (p<.05) effects on reading comprehension achievement for groups, discussion or nondiscussion, or levels of experience in the JGB program for both sets of scores. Therefore, the null hypotheses are accepted for both conditions.

In Table 4 is summarized the means and standard deviations for groups, levels of experience and teachers within schools. There are no significant (p<.05) differences for the main effect means of levels of experience and discussion/non-discussion groups on cloze gain scores and openended scores for story one.

Results in relation to question 1 and 2: Story two

In Table 5 is shown the results of analysis of variance of story two scores, cloze and openended, on individual level data. Results indicate no significant (p<.05) effects on reading achievement scores for groups, discussion and non-discussion, or levels of experience in

Table 3

Results of Analysis of Variance of Cloze and Openended
Scores, Individual Level Data: Story One

Source	SS	df	MS	F	PR <F
Cloze					
Experience[1]	14.84	1	14.84	13.28	.07
Teacher(Experience)	2.23	2	1.12	.18	.83
Group[2]	13.13	1	13.13	.93	.45
Group*Experience[2]	21.32	1	21.32	1.51	.34
Group*Teacher(Exp.)	28.26	2	14.13	2.33	.11
Error	425.28	70	6.08		
Total	505.55	77			
Openended					
Experience[1]	7.33	1	7.33	2.25	.27
Teacher(Experience)	6.51	2	3.26	1.38	.26
Group[2]	4.19	1	4.19	13.12	.07
Group*Experience[2]	.94	1	.94	2.94	.23
Group*Teacher(Exp.)	.64	2	.32	.14	.87
Error	165.19	70	2.36		
Total	185.18	77			

[1]Teacher (Experience) MS used as error term

[2]Group*Teacher(Exp.) MS used as error term

Table 4

Summary of Means and Standard Deviations for Pre and Posttest Cloze Scores and Openended Scores, Story One: Individual-level Data

Condition	n	Pre Cloze*	SD	Post Cloze*	SD	Open-ended**	SD
Group							
Discussion	39	18.87	3.59	19.62	3.45	2.33	1.53
Nondiscussion	39	19.92	3.62	19.38	3.09	2.79	1.56
Total	78						
Schools							
Experienced	39	19.05	4.01	19.77	3.14	2.91	1.54
Discussion	20	19.05	3.85	19.65	3.10	2.75	1.68
Nondiscussion	19	19.05	4.27	19.89	3.26	3.00	1.41
Inexperienced	39	19.74	3.21	19.23	3.39	2.56	1.51
Discussion	19	18.68	3.40	19.58	3.88	1.89	1.24
Nondiscussion	20	20.75	2.73	18.90	2.92	2.60	1.70
Total	78						
Teachers within Schools							
Teacher 1(Experienced)							
Discussion	10	19.00	4.19	19.50	3.60	2.90	1.45
Nondiscussion	9	19.89	4.48	20.00	3.39	2.88	1.62
Teacher 2(Experienced)							
Discussion	10	19.10	3.70	19.80	2.70	2.60	1.96
Nondiscussion	10	18.30	4.16	19.80	3.33	3.10	1.29
Teacher 3(Inexperienced)							
Discussion	10	18.80	2.39	18.90	3.35	1.50	.97
Nondiscussion	10	20.20	3.16	19.50	3.14	2.20	1.81
Teacher 4(Inexperienced)							
Discussion	9	18.56	4.42	20.33	4.47	2.33	1.41
Nondiscussion	10	21.30	2.26	18.30	2.71	3.00	1.56
Total	78						

*Possible range (0 to 31)
**Possible range (0 to 7)

Table 5

Results of Analysis of Variance of Cloze and Openended
Scores, Individual Level Data: Story Two

Source	SS	df	MS	F	PR < F
Cloze					
Experience[1]	22.26	1	22.26	1.50	.35
Teacher(Experience)	29.77	2	14.89	1.89	.16
Group[2]	5.73	1	5.73	1.13	.40
Group*Experience[2]	3.53	1	3.53	.69	.49
Group*Teacher(Exp.)	10.18	2	5.09	.65	.53
Error	550.10	70	7.86		
Total	622.59	77			
Openended					
Experience[1]	5.25	1	5.25	.63	.51
Teacher(Experience)	16.77	2	8.39	2.67	.08
Group[2]	12.19	1	12.19	6.73	.12
Group*Experience[2]	.25	1	.25	.14	.75
Group*Teacher(Exp.)	3.62	2	1.81	.58	.56
Error	219.72	70	3.14		
Total	257.45	77			

[1] Teacher (Experience) MS used as error term

[2] Group*Teacher(Exp.) MS used as error term

the JGB program for both sets of scores. The null hypotheses are accepted for both conditions.

In Table 6 is summarized the means and standard deviations for groups, levels of experience and teachers within schools for story two. As in story one, there are no significant (p<.05) differences for the main effect means of levels of experience and discussion/non-discussion groups on cloze and openended scores for story two.

Results in relation to question 3

To determine whether or not similar results in reading comprehension achievement occur across study replication using different JGB stories, the results of each analytic procedure are explored and reported.

1. With respect to study hypotheses, there are no significant differences between discussion and non-discussion groups or experienced and inexperienced groups in the JGB program on reading comprehension achievement at the .05 level of significance for both stories.

2. Results of correlation procedures reported in the data analyses section of Chapter III show different correlations between reading comprehension scores, cloze gain scores and openended scores, for story one and story two. Story two reading comprehension scores are more correlated with each other than story one scores. Pearson and Spearman correlations obtained are: Story one, r=.02,

Table 6

Summary of Means and Standard Deviations for Pre and Posttest Cloze Scores
and Openended Scores, Story Two: Individual-level Data

Condition	n	Pre Cloze*	SD	Post Cloze*	SD	Open-ended**	SD
Group							
Discussion	39	15.23	3.50	15.03	3.38	3.22	1.93
Nondiscussion	39	16.36	3.48	15.05	3.05	2.45	1.65
Total	78						
Schools							
Experienced	39	16.13	3.77	14.62	3.18	3.13	1.94
Discussion	20	15.50	3.75	14.40	3.63	3.55	2.06
Nondiscussion	19	16.79	3.78	14.84	2.71	2.65	1.73
Inexperienced	39	15.46	3.25	15.46	3.20	2.57	1.70
Discussion	19	14.95	3.29	15.68	3.06	2.90	1.78
Nondiscussion	20	15.95	3.22	15.25	3.40	2.25	1.58
Total	78						
Teachers within Schools							
Teacher 1(Experienced)							
Discussion	10	16.10	3.48	14.80	2.10	3.90	1.85
Nondiscussion	10	16.10	3.14	14.90	3.11	3.40	1.84
Teacher 2(Experienced)							
Discussion	10	14.90	4.09	14.00	4.81	3.20	2.30
Nondiscussion	9	17.56	4.45	14.78	2.39	1.90	1.27
Teacher 3(Inexperienced)							
Discussion	9	15.22	3.46	15.44	3.05	2.30	1.32
Nondiscussion	8	14.89	2.76	13.33	2.55	2.10	1.69
Teacher 4(Inexperienced)							
Discussion	10	14.70	3.30	15.90	3.21	3.50	2.01
Nondiscussion	11	16.82	3.43	16.82	3.28	2.40	1.57
Total	78						

*Possible range (0 to 31)
**Possible range (0 to 7)

p<.84, r=.05, p<.64, Story two, r=.27, p<.02, r=.26, p<.02.

3.   Mean pretest and posttest cloze scores are generally higher for story one than for story two. Subjects from both experienced and inexperienced schools were randomly and informally interviewed to explore possible reasons for this difference in cloze scores. When asked to compare story one with story two, general reactions from students include "more familiarity with content of story one than story two," "too much description in story two," "more of a plot in story one" and "no flashbacks in story one".

4. Mean openended scores generally lie in the lower range of possible scores 0 to 7 for both stories. The highest score of 7 (substantiation of characteristic trait with reference to author's intent), the major goal of interpretive reading in the JGB program, was not obtained by students in either story. All students responded legibly, therefore, no one received a score of 0.

In summary, results of information obtained from analytical data support null hypotheses for both stories at significant level of .05.

Results to question 4

To determine whether or not the application of the General Linear Model to the individual-level data provides a different solution than the solution obtained when

applying the aggregate level data, results of the ANOVA's and means for both type of data are compared. In addition, the R 's obtained by application of GLM to aggregate level data and individual-level data are examined to determine if significantly different R 's result.

In Table 7 and Table 9 are shown the results of analyses of variance of story one and story two scores, cloze and openended, on aggregate-level data. Results indicate no significant effects on reading comprehension achievement for groups, or levels of experience in the JGB program at the .05 level for both stories. These results are similar to those obtained when analyzing individual level data reported in Table 5 and Table 7.

A summary of the means and standard deviations for cloze gain scores and openended scores for story one and story two, aggregate-level data, is shown in Table 8 and Table 10. Similar non-significant F tests for main effects of experience and discussion/non-discussion groups are reported for indvidual-level data and aggregate level data.

Table 11 shows the proportions of variance accounted for by independent variables that are obtained by application of the General Linear Model on individual-level and aggregrate-level data. In story one the greatest proportion of variance is explained by individual-level

Table 7

Results of Analysis of Variance of Cloze and Openended
Scores, Aggregate Level Data: Story One

Source	SS	df	MS	F	PR < F
Cloze					
Experience[1]	1.53	1	1.53	13.21	.07
Teacher(Experience)	.23	2	.12	.08	.93
Group	1.34	1	1.34	.92	.44
Group*Experience	2.18	1	2.18	1.50	.35
Error	2.92	2	1.46		
Total	8.21	7			
Openended					
Experience[1]	.78	1	.78	2.43	.26
Teacher(Experience)	.65	2	.33	10.28	.09
Group	.45	1	.45	14.44	.06
Group*Experience	.10	1	.10	3.24	.21
Error	.06	2	.03		
Total	2.04	7			

[1]Teacher (Experience) MS used as error term

Table 8

Summary of Means and Standard Deviations for Pre and Posttest Cloze Scores and Openended Scores, Story One: Aggregate-level Data

Condition	n	Pre Cloze*	SD	Post Cloze*	SD	Open-ended**	SD
Group							
Discussion	4	18.87	.24	19.63	.60	2.33	.60
Nondiscussion	4	19.92	1.24	19.40	.76	2.80	.41
Total	8						
Schools							
Experienced	4	19.07	.65	19.78	.21	2.88	.21
Discussion	2	19.05	.07	19.65	.21	2.75	.21
Nondiscussion	2	19.10	1.12	19.90	.14	3.00	.14
Inexperienced	4	19.72	1.28	19.26	.87	2.25	.61
Discussion	2	18.68	.17	19.62	1.01	1.90	.57
Nondiscussion	2	20.75	.78	18.90	.85	2.60	.57
Total	8						
Teachers within Schools							
Teacher 1(Experienced)							
Discussion	1	19.00		19.50		2.90	
Nondiscussion	1	19.89		20.00		2.90	
Teacher 2(Experienced)							
Discussion	1	19.10		19.80		2.60	
Nondiscussion	1	18.30		19.80		3.10	
Teacher 3(Inexperienced)							
Discussion	1	18.80		18.90		1.50	
Nondiscussion	1	20.20		19.50		2.20	
Teacher 4(Inexperienced)							
Discussion	1	18.56		20.33		2.30	
Nondiscussion	1	21.30		18.30		3.00	
Total	8						

*Possible range (0 to 31)
**Possible range (0 to 7)

Table 9

Results of Analysis of Variance of Cloze and Openended
Scores, Aggregate Level Data: Story Two

Source	SS	df	MS	F	PR $<$ F
Cloze					
Experience[1]	2.32	1	2.32	1.50	.35
Teacher(Experience)	3.10	2	1.55	2.98	.25
Group	.61	1	.61	1.18	.39
Group*Experience	.37	1	.37	.72	.49
Error	1.04	2	.52		
Total	7.44	7			
Openended					
Experience[1]	.55	1	.55	.62	.51
Teacher(Experience)	1.77	2	.99	4.89	.17
Group	1.20	1	1.20	6.63	.12
Group*Experience	.03	1	.03	.17	.72
Error	.36	2	.18		
Total	3.92	7			

[1]Teacher (Experience) MS used as error term

Table 10

Summary of Means and Standard Deviations for Pre and Posttest Cloze Scores and Openended Scores, Story Two: Aggregate-level Data

Condition	n	Pre Cloze*	SD	Post Cloze*	SD	Open-ended**	SD
Group							
Discussion	4	15.23	.62	15.04	.82	3.23	.68
Nondiscussion	4	16.34	1.14	14.96	1.43	2.45	.67
Total	8						
Schools							
Experienced	4	16.17	1.09	14.62	.42	3.10	.85
Discussion	2	15.50	.85	14.40	.57	3.55	.49
Nondiscussion	2	16.83	1.03	14.84	.08	2.65	1.06
Inexperienced	4	15.41	.97	15.37	1.48	2.58	.63
Discussion	2	14.96	.37	15.67	.33	2.90	1.85
Nondiscussion	2	15.86	1.36	15.08	2.47	2.25	.21
Total	8						
Teachers within Schools							
Teacher 1(Experienced)							
Discussion	1	16.10		14.80		3.90	
Nondiscussion	1	16.10		14.90		3.40	
Teacher 2(Experienced)							
Discussion	1	14.90		14.00		3.20	
Nondiscussion	1	17.56		14.78		1.90	
Teacher 3(Inexperienced)							
Discussion	1	15.22		15.44		2.30	
Nondiscussion	1	14.89		13.33		2.10	
Teacher 4(Inexperienced)							
Discussion	1	14.70		15.90		3.50	
Nondiscussion	1	16.82		16.82		2.40	
Total	8						

*Possible range (0 to 31)

**Possible range (0 to 7)

Table 11

Proportions of Variance Accounted for by Independent Variables:
Application of General Linear Models on Individual-level and
Aggregate-level Data

	Individual-level Data			Aggregate-level Data		
	R^2	F Value	P<F	R^2	F Value	P<F
Story One						
Cloze	.16	1.89	.08	.64	.73	.66
Openended	.11	1.21	.31	.97	12.65	.07
Story Two						
Cloze	.12	1.32	.25	.86	2.47	.31
Openended	.15	1.72	.12	.91	3.92	.23

model applied on cloze scores, R^2 =.16, F(7,70)=1.89, p<.08. This is contrasted by story one the greatest proportion of variance is explained by aggregate-level model applied on openended scores, R^2 =.97, F(5,2)=12.65, p<.08. In story two all individual-level models and aggregate-level models applied on cloze and openended scores indicate less significant proportions of variance, p<.12.

The percent of variation explained by aggregate level models consistently exceeds the percent of variation explained by the individual-level model. In fact, less than one-fourth of the differences in cloze and openended scores in individual-level data are attributable to all the variables combined.

In summary, results obtained from analytical data do not show significantly (p<.05) different solutions when the General Linear Model is applied to the individual-level data and aggregate-level data.

CHAPTER V

SUMMARY, CONCLUSIONS AND DISCUSSION

This study had two major purposes of investigation: 1) to compare the effect of Junior Great Books literature discussion and non-discussion on reading comprehension achievement of gifted fifth grade students and to determine whether or not the level of experience with JGB program affected these variables and 2) to implement a research analysis which corrects for methodological shortcoming due to cross-level inference using multilevel data.

The following section summarizes the findings and draws conclusions from the results of the investigation. Suggestions are made for further research.

Major Findings

Application of study format

In story one and story two observations of taped discussions in JGB experienced and inexperienced schools show that most of the time interpretations are made by the students and not the leaders. Observations also reveal that leaders do not regularly request students to substantiate their answers from the text. These lower ratings may reflect the complexity and fluidity of the discussion process as it occurs in small groups. In

addition, these ratings may reflect the impact of students' levels of cognitive ability since discussions focus on different literature genre, fantasy fiction in story one and realistic fiction in story two. Previously cited research (Petrosky, 1976) indicates that response to literature is learned and subject to the abilities of the individual's stage of cognitive development. In his clinical studies of adolescents' response to literature, Petrosky found divergent personal individual responses to realistic and fantasy literature.

Leaders consistently adhered to the appropriateness of the discussion format in both stories. Taping of non-discussion groups (independent activity), experienced students and inexperienced students, for both stories, served to promote students' compliance to the study format. Taping of students in control groups may have shown them that their role in the experiment was equally valued as the discussion groups' role.

Questions 1 and 2: Story one

The hypothesis that there is no significant difference between discussion and non-discussion groups on reading comprehension achievement could not be rejected by the results of this investigation in story one. The data revealed that the differences between the means of the gain

cloze and opended scores of the discussion and non-discussion groups were not statistically different.

The hypothesis that there is no significant difference between levels of experience in the JGB program on reading comprehension achievement could not be rejected by the results of this investigation in story one. Similarly, the data revealed that the differences between the means of the cloze and openended scores of the experienced and inexperienced students were not statistically different.

Questions 1 and 2: Story two

The hypothesis that there is no significant difference between discussion and non-discussion groups on reading comprehension achievement is supported by the results of this investigation on story two. The data also revealed that the differences between the means of the cloze and openended scores of the discussion and non-discussion groups were not statistically different.

The hypothesis that there is no significant difference between the level of experience in the JGB program on reading comprehension achievement is also supported by the results of this investigation on story two. As in story one, the data revealed that the differences between the means of the cloze and openended

scores of experienced and inexperienced groups were not statistically different.

Question 3

It was found that there were no significant differences between discussion and non-discussion groups or experience and inexperienced groups in the JGB program on reading comprehension achievement of gifted fifth graders at the .05 level of significance for two different JGB stories.

Reading comprehension scores, cloze gain scores and openended scores, for story two were more correlated with each other than for story one. This may reflect inabilitly of measures to reliably assess construct of reading comprehension, particularly the skill of reading interpretation as presented in the JGB program.

It was found that students achieved higher mean pretest and posttest cloze scores in story one than in story two. These differences may reflect individual student's reaction to textual language and story structure in passages of both stories. Research has indicated that passage structure, redundancy of content and prior knowledge of story content may affect the measurement of reading comprehension (Bailey & Harrison, 1984; Kendall, Mason & Hunter, 1980). Familiarity with the content of story one was reported by students as a possibile reason

for obtaining higher cloze scores on story one than in story two.

When asked to respond to openended questions, none of the students achieved the major goal of interpretation in the JGB program, to express the author's intention. To some degree, this supports previous NAEP research findings (Purves, 1984) that students could express their own judgements of a work but did not substantiate their opinions by referring to the text. It also may reflect cognitive development research in which it is suggested that response to literature is subject to individual stages of cognitive development (Purves, 1984). Subjects in this study were gifted fifth graders in the transitional stage of cognitive development, as theorized by Piaget (1952). It was found that these subjects were unable to assume a spector's stance, or express the author's intention, when they approached a fictional text. Students may not have been able achieve this JGB goal because it is characteristic of formal operational thought, and fifth graders have not yet fully reached this stage of cognitive development.

Question 4

When data were analyzed at the individual level and aggregate level using the General Linear Model, no statistically significant (p<.05) differences in solutions

were found. However, different proportions of variances
accounted for by the independent variables were obtained
by the application of the analytical models to group-level
data and individual-level data in story one and story two.
R^2's were inflated for the aggregate-level data and
reflected the impact of variance between groups that is
accounted for by the independent variables, whereas the
lower R^2's account for the proportion of the total variance
accounted for by the independent variables. This supported
similar findings of researchers (Pedhazur et al., 1982) in
which analyses using the individual as the unit of analysis
were contrasted with ones in which aggregates are used as
the unit of analysis. It was found that small variances
between groups in aggregate-level data produced inflated
R^2's yet explained only a minute proportion of the total
variance.

Conclusions

The conclusions of this study are that literature
discussion using the Junior Great Books program does not
provide greater improvement in reading comprehension
achievement of gifted fifth grade students. Similarly, the
level of experience in the JGB program does not provide
greater improvement in reading comprehension achievement of
gifted fifth grade students. Since there might be some
debate about whether there is a continuum of possibilities

between discussion and non-discussion, as is assumed in stories one and two, data was reexamined as if all factors were designated fixed. All factors were tested using the mean square error. The p values for factors of experience and groups for cloze and openended scores, individual-level and aggregate-level data, stories one and two, changed but there were still no statistically significant differences at the $p < .05$ level. The only exception was for the experience factor, aggregate-level data, openended scores, which was significant at the $p < .05$ level. There was some confirmation of this exception from the openended scores, individual-level data, with a $p < .08$ obtained for the experience factor.

It may be inferred from this study that gifted fifth graders are unable to fully interpret JGB literature fiction due to the students' level of cognitive development. After working independently and in discussion groups, both experimental and control groups were unable to achieve the goal of the JGB literature program, determining the author's intention. This occurred despite the fact that the experimental groups discussed only interpretive questions and the control groups wrote responses to a majority of content questions.

It is also concluded that application of the General Linear Model to data at the group level and individual

level does not offer different solutions to the effectiveness of the JGB literature program on reading comprehension achievement of gifted fifth graders. A greater proportion of variance was accounted for by aggregate-level data, however, this was inflated and explained only a small proportion of the total variance.

Although the cloze tests were chosen on the basis of previously established reliability and validity and their practicality in the experiment, the effects of the JGB discussion program might not be adequately assessed by such a specific measure. There may be the possibility of a conflict between the general nature of the program and the discriminative isolation of the individual factors.

Suggestions for Further Research

From the preceding major findings of this study, and conclusions, the following suggestions are offered both for further studies on the Junior Great Books program itself and for related methods and procedures involved:

1. Investigate the amount of unexplained variance in reading comprehension achieved by individual students. Some potential variables which were not included in this study could include levels of student intelligence, levels of reading ability and levels of cognitive reasoning.

2.   Develop   an   instrument   to   measure   reading
comprehension achievement with a pretest and post test that
would   reliably   assess   gifted   readers'   ability   to   interpret
literature.   Variations   of   the   standard   cloze   procedure
might   be   examined   based   on   passage   structure   and   text
redundancy   to   determine their impact on the measurement of
reading comprehension.

3.   Investigate   the   effects   of   discussion   and   non-
discussion   of   different   literature   genre   on   reading
comprehension achievement of gifted students.

4.   Design   a   study   to   investigate   the   effects   of
adaptations   of   the   JGB   format   on   gifted   students'   reading
comprehension   achievement.   Design   could   reflect   recent
cognitive   development   research   which   suggests   that
student's   prior   knowledge and experience influence reading
comprehension achievement.

5.   Explore further the application of the discussion
format   by different teachers using more detailed discussion
observation   techniques   to   examine   the   extent   to   which
differences   in   discussion   techniques   effect   reading
comprehension achievement outcomes.

6.   Study   the   effect   in   the   program   on   reading
comprehension achievement   of   gifted children at different
grade   levels.   The   JGB   program   may   be   employed   with
students   in   grades   two   through   twelve.   Research in this

area   may provide insight into current theories of cognitive
development and reading comprehension.

REFERENCES

Alexander, K.L. & Griffin, L.J. (1976). School district
effects on academic achievement: A reconsideration.
American Sociological Review, 41, 144-152.

Applebee, A. (1978). The child's concept of story.
Chicago: The University of Chicago Press.

Ashby-Davis, C. (1985). Cloze and comprehension: A quali-
tative analysis and critique. Journal of Reading, 28
(7), 585-589.

Bailey, M. & Harrison, C. (1984). Cloze procedure as a
measure of reading comprehension: The effects of varia-
tions in subject matter on textual redundancy. Human
Learning, 3(3), 185-201.

Barrett, T.C. (1967). Goals of the reading program: The
basis for evaluation. In T.C. Barrett (Ed.), Reading
the evaluation of children's reading achievement.
Newark, Delaware: International Reading Association.

Barrett, T.C. & Smith, R.J. (1974). Teaching reading in
the middle grades. Reading, Maine: Addison-Wesley.

Bates, G.W. (1984). Developing reading strategies for the
gifted: A research-based approach. Journal of Reading,
27(7), 590-593.

Baumann, J.F. & Stevenson, J.A. (1982a). Using scores
from staandardized reading achievement tests. Reading
Teacher, 35(5), 528-533.

Baumann, J.F. & Stevenson, J.A. (1982b). Understanding
reading achievement test scores. Reading Teacher, 35
(6), 648-655.

Beane, W.E. & Lemke, E.A. (1971). Group variables
influencing the transfer of conceptual behavior.
Journal of Educational Psychology, 62(3), 215-218.

Berliner, D.C. (1980). Studying instruction in the ele-
mentary classroom. In R. Dreeben & J. Thomas (Eds.),
The analysis of educational productivity vol 1: Issues
in microanalysis. Cambridge, Mass.: Ballinger Publish-
ing Co.

Bidwell, C.E. & Kasarda, J.D. (1975). School district organization and student achievement. American Socio-logical Review, 40, 55-70.

Bidwell, C.E. & Kasarda, J.D. (1976). Reply to Hannon, Freeman & Meyer, and Alexander & Griffin. American Sociological Review, 41, 152-160.

Bird, J.J. (1984). Effects of fifth graders' attidudes and critical thinking/reading skills resulting from a Junior Great Books Program. (Doctoral dissertation, Rutgers University). Dissertation Abstracts International. 45, 02A.

Biskin,D.S., Hoskisson, K & Modlin, M. (1976). Prediction, reflection and comprehension. Elementary School Journal, 77(2), 131-139.

Blalock, H.M. (1964). Causal inferences in nonexperi-mental research. New York: W.W. Norton.

Boraks, N., Early, J. & Sable, E. (1986). Ethnographic research on Junior Great Books: Discussion leader behavior. In D. Lumpkin, M. Harshberger, & P. Ransom (Eds.), Evaluation in reading: Learning teaching administering, 6 (pp. 306-316). Muncie, IA: Ball State University.

Bormuth, J.R. (1968). The cloze readability procedure. Elementary English, 45, 429-36.

Bormuth, J.R. (1969). Factor validity of cloze tests as measures of reading reading comprehension ability. Reading Research Quarterly, 4(3), 358-365.

Bormuth, J. & MacDonald, O.L. (1965). Cloze tests as a measure of ability to detect literary style. Procedings of the International Reading Association, 10, 287-290.

Burstein, L. (1978). Assessing differences between grouped and individual-level regression coefficents: Alternate approaches. Sociological Methods and Research, 7(1), 5-28.

Burstein, L. (1980). The role of levels of analysis in the specification of education effects. In R. Dreeben & J. Thomas (Eds.), The analysis of educational productivity Vol. 1: Issues in microanalysis. Cambridge, MA: Ballinger Publishing Co.

Carter,K. (1985). Cognitive development of intellectually gifted: A piagetian perspective. Roeper Review, 7(3), 180-184.

Carver, R.P. (1977). Measuring reading comprehension using the paraphrase test and the reading storage test. Journal of Reading, 9(4), 381-389.

Cashman, R. (1977). The effects of the Junior Great Books Program at the intermediate grade level (4,5,6) on two intellectual operations - verbal meaning and reasoning ability. (Doctoral dissertation, Boston College). Dissertation Abstracts International, 38 03A. (University Microfilms No. 77-19933)

Caskey, H.J. (1970). Guidelines for teaching comprehension. Reading Teacher, 24, 649-654;669.

Casper, L. (1964). Effects of the Junior Great Books program at the fifth grade level on four intellectual operations and certain of their component factors as defined by J.P.Guilford (Doctoral dissertation, St. Louis University). Dissertation Abstracts International 25, 4570.(University Microfilms No. 64-13,453)

Chabot, C.B. (1985). Understanding interpretive situations. In C. Cooper, (Ed.), Researching response to literature and the teaching of literature:Points of departure. New York: Ablex Publishing Corp.

Cohen, J. (1960). A Coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20 (1), 37-46.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, S., Weinfeld, F.D. & York, R.L. (1966). Equality of educational opportunity. Washington, D.C.: Government Printing Office.

Collins, A., Brown, J.S. & Larkin, K.M. (1980). Inference in text understanding. In R. Spiro, B. Bruce, & W. Brewer (Eds.), Theoretical issues in reading compre- hension: Perspectives and cognitive psychology, linguistics, artifical intelligence and education. Hillsdale, N.J.: Erlbaum.

Cooper, C. (Ed.), (1985). Researching response to literature and the teaching of literature: Points of departure. New York: Ablex Publishing Corp.

Cox, D.R. (1958). Planning of experiments. New York:
    John Wiley & Sons, Inc.

Cronbach, L.J. & Furby, L. (1970). How we should measure
    "change"-Or should we?. Psychological Bulletin, 74(1),
    68-80.

Culhane, J.W. (1970). Cloze procedures and comprehension.
    Reading Teacher, 23, 410-413.

Cunningham, J.W. & Tierney, R.J. (1979). Evaluating cloze
    as a mmeasure of learning from reading. Journal of
    Reading Behavior, 11(3), 287-292.

Curtis, M.E. & Glaser, R. (1983). Reading theory and
    assessment of reading achievement. Journal of Educa-
    tional Measurement, 20(2), 133-147.

Dennis, R.P. & Moldof, E.P. (1978). A handbook of inter-
    pretive reading and discussion. Chicago: The Great
    Books Foundation.

DeSanti, R.J., Casbergue, R.M. & Sullivan, V.G. (1986).
    The cloze reading inventory as a qualitative and quan-
    titative measure of reading ability. In D. Lumpkin,
    M.Harshberger & P.Ransom (Eds.), Evaluation In
    Reading: Learning Teaching Administering (pp. 194-203)
    Muncie,IA: Ball State University.

Dillon, J.T. (1984). Research on questioning and dis-
    cussion. Educational Leadership, 42(3), 50-56.

DiVesta, F.J., Hayward, K.G. & Orlando, V.P. (1979).
    Developmental trends in monitoring text for compre-
    hension. Child Development, 50(1), 97-105.

Dole, J.A. & Adams, P.J. (1983). Reading curriculum for
gifted readers. A survey. Gifted Child Quarterly, 27(2),
64-72.

Duke, C. (1984). The role of reflections, problem-solving
and discussion in the teaching of literature. A paper
presented at the annual meeting of the National Council of
Teachers of English, Spring Conference, April, 1984.
Columbus, Ohio.

Dupis, M.M. (1976). The cloze procedure: Can it be used
    with literature? Reading Improvement, 13(4), 199-203.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. American Socio-logical Review, 43, 557-572.

Fish, S. (1980). Is there a text in this class? The authority of interpretive communities. Cambridge, MA.: Harvard University Press.

Fry, E. (1968). A readability formula that saves time. Journal of Reading, 11(7), 513-578.

Galda, L. (1982a). Assuming the spectator stance: An examination of the responses of three young readers. Research in the Teaching of English, 16(1), 1-20.

Galda, L. (1982b). Playing about a story: Its impact on comprehension. Reading Teacher, 36(1), 52-55.

Galda, L. (1983). Research in response to literature. Journal of Research and Development in Education, 16 (3), 1-7.

Gall, M.D. & Gillet, M. (1980). The discussion method in classroom teaching. Theory Into Practice, 19(2), 98-103.

Ganier, A.S. (1976). The post oral reading cloze test: Does it really work? Reading World, 16(1), 21-27.

Glick, W. (1980). Problems in cross-level inferences. In K. Roberts, L. Burstein (Eds.), New directions for methodology of social and behavioral science: Issues in aggregation. San Francisco, CA: Jossey-Bass Inc.

Goodman, Y. & Burke, C. (1971). Reading miscue inventory manual: Procedure for diagnosis and evaluation. New York: Macmillan Publishing Co.

Greene, J.C. (1980). Individual and teacher/class effects in aptitude treatment studies. American Educational Research Journal, 17(3), 291-302.

Gutherie, J.T. (1981). Comprehension and teaching: Research reviews. Newark, DE (ERIC Document Reproduction Service No. ED 203299).

Hahn, C. & Avery, P. (1985). Effect of value analysis
    discussion on students' political attitudes and
    reading comprehension. Theory and Research in Social
    Education, 13(2), 47-60.

Handbook on Interpretive Reading and Discussion. (1984)
    Chicago, Illinois: The Great Books Foundation.

Haney, W. (1980). Units and levels of analysis in large-
    scale evaluation. In K.Roberts & L.Burstein (Eds.),
    Issues in aggregation. San Francisco,CA: Jossey-Bass
    Inc.

Hannan, M.T. & Burstein, L. (1974). Estimation from
    grouped observations. American Sociological Review,
    39, 374-392.

Hannan, M.T., Freeman, J.H. & Meyer, J.W. (1976). Speci-
    fication of models for organizational effectiveness.
    American Sociological Review, 41, 136-143.

Hansen, J. (1981). The effects of inference training and
    practice on young children's reading comprehension.
    Reading Research Quarterly, 16(3), 391-417.

Hansen, J. & Pearson, P.D. (1983). An instructional study:
    Improving the inferential comprehension of good and poor
    fourth-grade readers. Journal of Educational Psycho-
    logy, 75(6), 821-829.

Hardy, M. (1980). Talking in school. Forum for the dis-
    cussion of new trends in education, 22(2), 45-46.

Helfeld, J.P. & Henk, W.A. (1985). Usefulness of conven-
    tional vs. total random cued cloze tests as measures
    of reading comprehension. Journal of Reading, 28(8),
    719-725.

Hickman, J. (1981). A new perspective on response to
    literature: Research in an elementary school setting.
    Research in the Teaching of English, 15, 343-354.

Hicks, C.R. (1964). Fundamental concepts in the design of
    experiments. New York: Holt, Rinehart and Winston.

Hill, G. (1982). Group versus individual performance:
    Are N+1 heads better than one? Psychological Bulletin,
    91(3), 515-539.

Hill, W. (1962). Learning thru discussion. Beverly Hills, CA: Sage Publications.

Hopkins, K.D. (1982). The unit of analysis:Group means vs. individual observations. American Educational Research Journal, 19(1), 5-18.

Hunkins, F.P. (1968). The influence of analysis and evaluative questions in achievement of sixth grade social studies. Educational Leadership Journal Research Supplement, 326-332.

Hutson, B. & Niles, J. (1983). Reconciling differences in test results: Comprehension. Reading Horizons, 23 (4), 235-242.

Iser, W. (1978). The act of reading: A theory of aesthetic response. Baltimeore, MD: Johns Hopkins University Press.

Iser, W. (1980). Texts and readers. Discourse Processes, 3, 327-343.

Jenkinson,M.E. (1957). Selected processes and difficulties in reading comprehension. Unpublished doctoral dissertation. University of Chicago.

Johns, J.L. (1978). Do comprehension items really test reading? Sometimes! Journal of Reading, 21(7), 615-619.

Johnson, D.W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competetive and individualistic goal structures on achievement: A meta-analysis. Psychological Bulletin, 89(1), 47-62.

Junior Great Books Series Five. (1984). Chicago, IL: The Great Books Foundation.

Kendall, J., Mason, J. & Hunter, W. (1980). Which comprehension? Artifacts in the measurement of reading comprehension. Journal of Educational Research, 73(4), 233-236.

Kennon, C. (1980). Utilizing moral dilemmas to enhance comprehension. The importance of context variables in research on teaching skills. Journal of Teacher Education, 28, 43-48.

Kibby, M. (1981). Test review, the degrees of reading power. Journal of Reading, 34, 416-427.

Kingston. A.J. (1970). The measurement of reading comprehension. In R. Farr (Ed.), Measurement and evaluation of reading. New York: Harcourt, Brace & World, Inc.

Knapp, T. (1977). The unit-of-analysis. Problem in applications of simple correlation analysis to education research. Journal of Educational Statistics, 2(3), 171-186.

Kohlberg, L. (1975). The cognitive-developmental approach to moral education. Phi Delta Kappan, 56, 670-677.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrica, 33, 159-174.

Landsheere, G. de (1972). Application du cloze test de W.L. Taylor a la langue francaise. Scientia Paedagogica Experimentalis, 9, 207-56.

Langbein, L.I. (1977). Schools or students: Aggregation problems in the study of student achievement. In M. Guttentag (Ed.) Evaluation studies: Review Annual 2. Beverly Hills, CA: Sage Publications.

Le Pere, J.M. (1975). Beyond the literal level. Elementary English, 52, 476-480.

Lukasevich, A. (1983). Three dozen useful information sources on reading for the gifted. Reading Teacher, 36(6), 542-548.

Lyons, K. (1984). Criterion referenced reading comprehension tests: New forms with old ghosts. Journal of Reading, 29(4), 293-298.

Maloney, R.M. (1956). Group learning through group discussion: A group discussion implementation analysis. Journal of Social Psychology, 43(1), 3-9.

Matsuhashi, A. (1980). Small group response. English Education,. 12(1), 26-31.

McCormick, S. & Swassing, R. (1982). Reading instruction for the gifted: A survey of programs. Journal for the Education of the Gifted, 5(1), 34-43.

McDonald,A.S. (1970). Measuring reading performance. In R.
    Farr (Ed.), Measurement and evaluation of reading. New
    York: Harcourt, Brace & World, Inc.

Mitchell, B. (1982). An update on the stage of gifted/
    talented educaion in the U. S. Phi Delta Kappan,
    63, 357-358.

Monson, D. (1983). Literature in programs for gifted and
    talented children. In N. Roser & M. Frith (Eds.),
    Children's Choices (pp.48-65). Newark, DE: Interna-
    tional Reading Association.

National Assessment of Educational Progress. (1981).
    Reading/literature released exercise set 1979-1980
    assessment supplement. Denver: Education Commission
    of the States. (ERIC Document Reproduction Services
    No. ED 205589).

Nelson, M.A. & Abraham, E.C. (1976). Discussion strate-
    gies and student cognitive skills. Science Education,
    60(1), 13-27.

O'Connell, P. (Ed.), (1986). State of the State's Gifted
    and Talented Education. The Council of State Directors
    of Programs for the Gifted. Augusta, ME: Maine
    Department of Educational and Cultural Services.

Page, W.D. & Vacca, R.T. (1979). Overt indications of
    reading comprehension: Product and process considera-
    tion. In J. Harste & R. Carey (Eds.), Monograph
    in language and reading studies: New perspectives
    on comprehension, 3 (pp.49-63). Indiana University.

Pavur, R. & Nath, R. (1986). Parametric versus rank
    transform procedures in the two-way factorial experi-
    ment: A comparative study. Journal of Statistical
    Computation and Simulation, 23, 231-240.

Pearson, P.D. (1979). Methodological concerns in compre-
    hension research. In J.C. Harste & R.F. Carey (Eds.),
    Monograph in language and reading studies. New perspec-
    tives on comprehension. 3 (pp.153-170). Indiana
    University.

Pearson, P.D. (1985). Changing the face of reading compre-
    hension instruction. Reading Teacher, 38(8), 724-738.

Pedhazur, E.J. (1982). Multiple regression in behavioral
    research. New York: Holt, Rinehart and Winston.

Petrosky, A. (1976). The effects of reality perception and fantasy on response to literature: Two case studies. Research in the Teaching of English, 3(10), 239-257.

Petrosky, A. (1977). Genetic epistemology and psychoanalytic ego psychology: Clinical support for the study of response to literature. Research in the Teaching of English, 11(1), 28-38.

Piaget, J. (1952). The origins of intelligence in children. Translated by Margaret Cook. New York: International Universities Press.

Popham, W.J. (1978). Well-crafted criterion-referenced tests. Educational Leadership, 36, 91-95.

Purves, A.C. (1984). The potential and real achievement of U.S. students in school reading. American Journal of Education, 93(1), 82-105.

Pyrczak, F. (1981). A closer look at the quality of multiple choice reading comprehension items. Reading Improvement,. 18(2), 108-109.

Rankin, E.F. (1970). The cloze procedure - Its validity and utility. In R. Farr (Ed.), Measurement and evaluation of reading. New York: Harcourt, Brace & World, Inc.

Rankin, E.F. & Culhane, W.J. (1969). Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 13, 193-198.

Readence, J.E. & Moore, D.W. (1983). Why questions? A historical perspective on standardized reading comprehension tests. Journal of Reading, 26(4), 306-313.

Reder, L.M. (1978). Comprehension and retention of prose: A literature review. Technical Report No. 108. Illinois University, Urbana. Center for the Study of Reading. (ERIC Document Reproduction Service No. ED 165114).

Richards, I.A. (1929). Pratical criticism. New York: Harcourt, Brace.

Robinson, W.S. (1950). Ecological correlation and the behavior of individuals. Sociological Review, 15, 351-357.

Rogers, C.D. (1976). What standardized reading tests don't tell us and we forget to ask. Reading Improvement, 13(3), 135-136.

Rosenblatt, L. (1938). Literature as exploration. New York: Noble and Noble.

Royer, J.M. & Cunningham, D.J. (1981). On the theory of and measurement of reading comprehension. Contemporary Educational Psychology, 6(3), 187-216.

Rudduck, J. (1978). Learning through small group discussion. Surrey, England: Society for Research into Higher Education Ltd. at the University of Surrey.

Russell, S.N. (1978). How effective is the cloze technique in the measurement of reading comprehension? Viewpoints in Teaching and Learning, 54(3), 90-96.

Rye, J. (1982). Cloze procedure and the teaching of reading. London: Heinemann Educational Books.

Sakiey, L. (1980). Reading for the gifted: Instructional strategies based on research. A paper presented at the 3rd Eastern Regional Conference of the International Reading Association, Niagra Falls, N.Y.

Savage, J.F. (1983) Reading guides: Effective tools for teaching the gifted. Roeper Review, 5(3), 9-11.

Scherick, H.H. & Hanna, G.S. (1977). Passage-dependency data on the selection of reading comprehension test items. Educational and Psychological Measurement, 37(4), 991-997.

Schwartz, S. (1984). Measuring reading competence. New York: Plenum Press.

Singer, H & Dolan D. (1980). Reading and learning from text. Boston: Little, Brown and Company.

Smith, P.G. (1972). An empirical study of learning taking place in two different classroom communication situation. (Doctoral Dissertation, University of Colorado). Dissertation Abstracts International. 73 18,595.

Smith, P.G. & Zinc, A. (1977). A cloze-based investigation of reaading comprehension as a composite of subskills. Journal of Reading, 9(4), 395-398.

Spiro, R.J. (1980). Constructive processes in prose comprehension and recall. In R. Spiro, B. Bruce & W. Brewer. (Eds.), Theoretical issues in reading comprehension: Perspectives and cognitive psychology, linguistics, artifical intelligence and education. Hillsdale, N.J.: Erlbaum.

Stank, L. (1983). Reading programs for the gifted: A necessity!. G/C/T, 27, 39-41.

Stewart, E. D. (1981). Learning styles among gifted/talented students: Instructional techniques preferences. Exceptional Children, 48(2), 134-138.

Stumpf, S.A. & Freedman, R.D. (1979). Expected grade covariation with student ratings of instruction: Individual versus class effects. Journal of Educational Review, 71(3), 293-302.

Taylor, W.L. (1953). Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 30, 415-433.

Tierney, R.J. & Cunningham, J.W. (1984). Research on teaching reading comprehension. In P. David Pearson (Ed.), Handbook of reading research. New York: Longman Inc.

Tierney, R.J. & Spiro, R.J. (1979). Some basic notions about reading comprehension: Implications for teachers. In J. Harste & R. Carey (Eds.), Monograph in language and reading studies. New perspectives on comprehension. 3. (pp.132-137). Indiana University.

Tracy, R. & Rankin, E. (1970). Methods of computing and evaluating residual gain scores in the reading program. In R. Farr (Ed.), Measurement and Evaluation of Reading. New York: Harcourt, Brace & World, Inc.

Webb, A.J. (1985). Studying the effects of literary instruction in classrooms: Collaborative research in schools. In C.R. Cooper (Ed.), Researching response to literature and the teaching of literature: Points of departure (pp.77-87). New York: Ablex Publishing Corp.

Will, H. (1986). Toward a broader definition of the more able learner: Junior great books. G/C/T, 9(1), 6-7.

Wilson, J.R. (1966). Response of college freshmen to three novels. Urbana, Illinois. National Council of Teachers of English.

Witty, P.A. (1971). Reading for the gifted and creative student. Newark, Delaware: International Reading Association.

Wulz, S. V. (1979). Comprehension testing: Functions and procedures. Reading Teacher, 33(3), 295-299.

Yager, S., Johnson, D. & Johnson, R.T. (1985). Oral discussion, group-to-individual transfer, and achievement in cooperative learning groups. Journal of Educational Psychology, 77(1), 60-66.

Yost, M., Avila, L. & Vexler, E.B. (1977). Effect on learning of postinstructional responses to questions of differing degrees of complexity. Journal of Educational Psychology, 69(4), 399-408.

Zettel, J.J. (1982). The education of gifted and talented children from a federal perspective. Special Education in America: Its Legal and Governmental Foundation. Reston, Va.: Council for Exceptional Children.

APPENDIX

SPECIMENS OF MEASURES, INDEPENDENT ACTIVITIES, DISCUSSION
QUESTIONS AND EVALUATIVE CRITERIA

The following passage from "The Nightingale" has certain words omitted. Each gap in the passage stands for just ONE word that is missing. Read the passage carefully and write the missing word in the blank alongside the passage.

If you can think of more than one word that would fit into the gap choose the one word that you think the author would have used.

From all over the world travelers came to the emperor's city to admire his palace and gardens; but when the heard the nightingale sing, they all declared that it was the loveliest of all. When they returned to their

own _____ they would write long and learned   1._____

books _____ the city, the palace, and the   2._____

garden; _____ they didn't forget the nightingale.   3._____

No, that _____ always mentioned in the very   4._____

first chapter. _____ who could write poetry wrote   5._____

long odes _____ the nightingale who lived in the   6._____

forest, _____ the shores of the deep blue sea.   7._____

_____ books were read the whole world   8._____

over; _____ finally one was also sent to the   9._____

_____. He sat down in his golden chair _____   10._____

started to read it. Every once in _____ while   11._____

he would nod his head because _____ pleased him   12._____

to read how his own _____ and his own palace and   13._____

gardens were _____; but then he came to the   14._____

sentence "_____ the song of the nightingale is the   15._____

_____ of all."   16._____

"What!" said the emperor. "The _____?   17._____

I don't know it, I have never _____ of it; and yet   18._____

it lives not _____ in my empire but in my   19._____

     20._____

very _____. That is the sort of thing one
     <u>21</u>

_____ only find out by reading books."
  <u>22</u>

        He _____ his chief courtier, who was
          <u>23</u>

so very _____ that if anyone of a rank lower
       <u>24</u>

_____ his own either talked to him or _____
  <u>25</u>   <u>26</u>

ask him a question, he only answered "_____".
                               <u>27</u>

And that didn't mean anything at all.

        "_____ is a strange and famous bird
         <u>28</u>

called _____ nightingale," began the emperor.
      <u>29</u>

"It is thought _____ be the most marvelous thing
            <u>30</u>

in my _____. Why have I never heard of it?"
     <u>31</u>

        "I have never heard of it," answered the

courtier. "It has never been presented at court."

21._____

22._____

23._____

24._____

25._____

26._____

27._____

28._____

29._____

30._____

31._____

NAME_____ _____TEACHER_____Post Cloze 1

The following passage from "The Nightingale" has certain words omitted. Each gap in the passage stands for just ONE word that is missing. Read the passage carefully and write the missing word in the blank alongside the passage.

If you can think of more than one word that would fit into the gap choose the one word that you think the author would have used.

_____

One day a package arrived for the emperor; on it was written "Nightingale".

"It is probably another book about our famous

bird," said the emperor. But he was _____; it   1._____
                                            1

was a mechanical nightingale. It lay _____ a   2._____
                                         2

little box and was supposed to _____ like the   3._____
                                    3

real one, though it was _____ of silver and gold   4._____
                            4

and studded with _____, diamonds, and rubies.   5._____
                      5

When you wound it _____, it could sing one of the   6._____
                       6

songs _____ real nightingale sang; and while it   7._____
           7

performed, _____ little silver tail would go up   8._____
                8

and _____. Around its neck hung a ribbon on   9._____
        9

_____ was written: "The Emperor of Japan's   10._____
    10

nightingale _____ inferior to the Emperor of   11._____
               11

China's."

"It _____ beautiful!" exclaimed the   12._____
        12

whole court. And the _____ who brought it had the   13._____
                          13

title of _____ Imperial Nightingale Deliverer   14._____
             14

bestowed upon him at _____.   15._____
                          15

"They ought to sing together, it will _____   16._____
                                            16

a duet," said everyone. And they did. _____ that   17._____
                                            17

didn't work out well at all, _____ the real bird   18._____
                                  18

sang in his own _____ and the mechanical one had a   19._____
                     19

cylinder _____ its chest instead of a heart.　　　　20._____
　　　　　　　20

"It _____ not its fault," said the imperial　　　　21._____
　　　21

music _____. "It keeps perfect time. It　　　　22._____
　　　　22

belongs to _____ school of music." Then the　　　23._____
　　　　　　　23

mechanical nightingale _____ to sing solo.　　　24._____
　　　　　　　　　　24

Everyone agreed that its _____ was just as　　　25._____
　　　　　　　　　　　25

beautiful as the real _____; and besides,　　　26._____
　　　　　　　　　　26

the artificial bird was much _____ to　　　　27._____
　　　　　　　　　　　27

look at, with its sapphires, rubies, _____　　　28._____
　　　　　　　　　　　　　　　28

diamonds that glittered like bracelets and

brooches.

　　　　　_____ mechanical nightingale sang its　　　29._____
　　　　　　　29

song thirty-three times _____ did not grow　　　30._____
　　　　　　　　　　　　30

tired. The court would _____ liked to hear　　　31._____
　　　　　　　　　　　31

it the thirty-fourth time, but the emperor thought

that the real nightingale ought to sing now. But

where was it? Nobody had noticed that he had flown out

through an open window to his beloved green forest.

NAME_____TEACHER_____

OPEN-ENDED QUESTION 1

1.   Think about the story, "The Nightingale", again. What
kind of person is the emperor? Describe the emperor in a
few words on the line below.

_____

2.   What was it about the story that led you to describe
the emperor the way you did in question 1? Write your
answers on the lines below.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

NAME _____TEACHER_____

INDEPENDENT ACTIVITY 1

Answer the following questions about "The Nightingale" on the attached paper. You may look back at the story to complete your response. You may not discuss your responses with anyone.

1. (a) The author states that this story happened a long time ago. List details from the story that indicate it does not take place today. (b) Could this story take place today? Explain your answer.

2. The author compares a mechanical nightingale to a real nightingale in the story. How were these birds alike? How were they different? In your answers compare (a) the birds' song, (b) the birds' usefulness to the emperor and the townspeople, and (c) the birds' appearance.

3. Many unfamiliar words are simple words with affixes added to them. For example, joyously = joy + ous + ly. Sometimes spelling changes occur when suffixes are added. For example, in " desparation" the final e of " desparate" has been dropped. Copy the following words on your paper. After each, write the root word.

1. perfectly
2. loveliest
3. shiny
4. motionless
5. artistic

6. embarrassment
7. ungrateful
8. illness
9. terrifyingly
10. misfortune

4. Write 5 questions for this story that could be used for a group quiz.

NAME_____TEACHER_____PRECLOZE 2

The following passage from "Spit Nolan" has certain words omitted. Each gap in the passage stands for just ONE word that is missing. Read the passage carefully and write the missing word in the blank alongside the passage.
If you can think of more than one word that would fit into the gap choose the one word that you think the author would have used.

It was the custom for lads to gather at

the street corner on summer evenings and, trolleys

parked at hand, discuss trolleying, road surfaces,

and also show off any new gadgets. Then when Spit

gave the sign, we _____ to set off for Cemetery   1._____
                    1
Brew. There _____ scarcely any evening traffic   2._____
               2
on the roads _____ those days, so that we could   3._____
                3
have _____ good practice before our evening race.   4._____
         4
Spit, _____ unbeaten champion, would inspect   5._____
          5
every trolley and _____, and allow a start   6._____
                      6
which was reckoned _____ the size of the wheels   7._____
                       7
and the _____ of the rider. He was always the   8._____
            8
_____ in the line of starters, though no   9._____
    9
_____ how long a start he gave it _____   10._____
   10   11
impossible to beat him. He knew that _____   11._____
                                         12
like the palm of his hand, every _____ lump   12._____
                                     13
or pothole, and he never came _____ cropper.   13._____
                                  14
Among us he took things easy, _____   14._____
                                  15
when occasion asked for it he would _____ all   15._____
                                        16
out. Once he had to meet _____ challenge from   16._____
                             17
Ducker Smith, the champion of _____ Engine Row   17._____
                                 18
gang. On that occasion Spit _____ a wheel from   18._____
                                19
                                                          19._____

the baby's pram, removing _____ nearest the wall,   20._____
                              20

so it wouldn't be _____, and confident he could   21._____
                    21

replace it before _____ mother took baby out. And   22._____
                    22

after fixing _____ to his trolley he made that   23._____
               23

ride _____ what was called the "belly-down"   24._____
      24

style - that _____, he lay full stretch on his   25._____
          25

stomach, _____ as to avoid wind resistance.   26._____
        26

Although Ducker _____ away with a flying start he   27._____
           27

had _____that sensitive touch of Spit, and his   28._____
  28

_____ bumps and swerves lost him valuable inches,   29._____
  29

_____ that he lost the race with a _____   30._____
  30   31

three lengths. Spit arrived home just in time   31._____

to catch his mother as she was wheeling young Georgie

off the doorstep, and if he had not made a dash for it

the child would have fallen out as the pram overturned.

NAME_____TEACHER_____Post Cloze 2

      The following passage from "Spit Nolan" has certain words omitted.
Each gap in the passage stands for just ONE word that is missing. Read the
passage carefully and write the missing word in the blank alongside the
passage.
      If you can think of more than one word that would fit into the gap
choose the one word that you think the author would have used.

      Those beautiful ball-bearing wheels, engineer-

made, encased in oil, were holding the road and bringing

Leslie along faster than spirit and skill could carry

Spit. Dead level they sped into the final stretch.

Spit's slight figure was poised fearlessly on _____   1. _____

trolley, drawing the extremes of speed from _____.   2._____
                                    2

Thundering beside him, anxious but determined, came

_____. He was actually drawing ahead - and   3._____
       3

forcing _____ way to the top of the camber.   4._____
            4

_____ they came like two charioteers - Spit   5._____
     5

delicately _____ to the side, to gain inches by   6._____
          6

_____ extra downward momentum. I kept my eyes   7._____
    7

_____ clean across the road as they came _____   8._____
    8   9

past the winningpost.   9._____

      First past was the _____ The British Queen.   10._____
                10

I saw that first. _____ I saw the heavy rear wheel   11._____
           11

jog _____ a pothole and strike Spit's front wheel -   12._____
    12

_____ him in a swerve across the road. _____   13._____
    13   14

then, from nowhere, a charabanc* came speeding   14._____

_____ the wide bend.   15._____
    15

      Spit was straight in _____ path. Nothing   16._____
              16

could avoid the collision. I _____ a cry of fear as   17._____
          17

I saw _____ heavy solid tire of the front wheel   18._____
   18

     *charabanc. A large bus.

_____ the trolley. Spit was flung up and   19._____

_____ back hit the radiator. Then the   20._____
  20

driver _____ dead.   21._____
    21

      I got there first. Spit was _____   22._____
               22

on the macadam road on his side. _____ face was 23._____
          23

white and dusty, and coming _____ between his   24._____
      24

lips and trickling down his _____ was a   25._____
      25

rivulet of fresh red blood. _____ all about   26._____
      26

him were yellow rose petals.

     "_____ my fault," I heard the driver   27._____
    27

shouting. "_____ didn't have a chance. He came 28._____
     28

straight _____ me."   29._____
    29

    The next thing we were surrounded _____   30._____
               30

women who had got out of the _____. And then   31._____
      31

Leslie and all the lads came up.

    "Somebody send for an ambulance!" called

a woman.

"I'll run an' tell the gatekeeper to

telephone," said Ernie Haddock.

NAME_____TEACHER_____

OPEN-ENDED QUESTION 2

1. Think about the story, "Spit Nolan", again. What kind
of person is Spit?   Describe Spit in a few words on the
line below.

_____

2. What was it about the story that led you to describe
Spit the way you did in question 1? Write your answers on
the lines below.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

NAME_____TEACHER_____

INDEPENDENT ACTIVITY 2

Answer the following questions about "Spit Nolan" on the attached paper. You <u>may</u> look back at the story to complete your response. You <u>may not</u> discuss your response with anyone.

1.   The person telling the story, the narrator, is a character in the story. What is his name? What important role does he play in the final race?

2.   What do you suppose might be a possible ending of the story if the narrator was Leslie?

3.   Who is Ducker Smith and Chick Dale? How do these characters show respect for Spit?

4.   Describe the two trolleys "Egdam" and "The British Queen". How are they the same? How are they different?

5.   The author considers the importance of names and mottoes for trollies. What are other possible names and mottoes for trolleys? Design and sketch your own trolley.

6.   Use the context of the paragraph to explain the meaning of the following words:

        imperturbability (p.60, line 17)

        slanced (p.55, line 18)

        reckoned (p.56, line 18)

        intently (p.58, line 19)

        optimistic (p.60, line 11)

SUMMARY OF QUESTIONS DISCUSSED IN EXPERIMENTAL GROUPS:

STORIES ONE AND TWO

Story One

1. Which is more dependable - a real nightingale or a
mechanical one?

2. Is the author criticizing the very idea of a mechanical
nightingale?

3. Why does the nightingale feel there is "a strange
power" in an emperor's tears?

4. If the mechanical nightingale keeps perfect time, why
isn't his song better than the real nightingale's?

5. Why does the nightingale's song sound best in the green
woods?

6. Why are the fisherman and the kitchenmaid better able
than the people at court to appreciate the nightingale's
song?

7. Why does the nightingale's song drive away the evil
phantoms but appeal to Death?

8. Why will the emperor be a better ruler at the end of
the story?

Story Two

1. Why does the author have Spit die as a result of losing
the race?

2. Why do Spit's friends respect and admire him?

3. Why is Spit given an incurable disease?

4. Why does the author have the superior driver lose to
the superior machine? Why doesn't he have Spit's spirit
and skill defeat Leslie's professionally made trolley?

5. Why are we told that Spit is right when he says that
Leslie's new trolley will never really belong to him?

6. Why are the boys' loyalties equally divided between
Spit and Leslie?

7. Why does Spit want the boys to know he admits defeat?

8. Are we meant to think Spit is right when he says that
Leslie's new trolley will never really belong to him?

CRITERIA FOR SCORING OPEN-ENDED TEST

0   No response, illegible response,

1   <u>Unable to identify critical</u> character traits.

        *Opinion about the action of the character is
        offered.
        *Some material is quoted from text with no clear
        identification of character.

2   <u>One critical</u> character trait identified without
        substantiation.

        *Respondents name something but cannot go on.
        *They identify a character trait but do not substan-
        iate the choice with evidence from the text.
        *Responses tend to provide 1)vague reasons, 2)circu-
        lar evidence, or 3)a subjective reaction as sub-
        stantiation.

3   <u>Two or more critical</u> character traits identified with-
        out substantiation. (See examples in criteria 2.)

4   <u>One critical</u> character trait identified and substanti-
        ated with <u>literal</u> parallel substantiation.

        *"He said in the story....."
        *"It said in the story....."

5   <u>One critical</u> character trait identified and substanti-
        ated with <u>inference</u> substantiation.

        *Interpretation of an event in story.

6   <u>Two or more critical</u> character traits identified and
        substantiatied with <u>one or more reasons</u> or pieces of
        evidence related to text (literal or inferential).

7.   <u>More than one critical</u> character trait identified and
        substantiated with <u>reference to author's purpose</u> or
        intent.

JGB Audiotape Rating of Discussion Observation

Directions:
1. Listen 5 minutes and stop tape.
2. Complete observation #1.
3. Play tape 5 minutes with sound off.
4. Repeat steps 1-3 twice to complete observations #2-3.
5. Complete last observation.

TAPE #_____   OBSERVER_____

Leader does not make interpretations, participants do.

                    always   frequently   sometimes   never

observation #1   _____

observation #2   _____

observation #3   _____

Answers are substantiated from text material.

observation #1   _____

observation #2   _____

observation #3   _____

Discussion is centered on appropriate interpretive
questions as defined in Junior Great Books Program format.

                    always   frequently   sometimes   never

                    _____

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2