1977

A cohort analysis of elementary reading achievement scores in two school districts which implemented the Individually Guided Education program

Michael Szymczuk
Iowa State University

77-26,014

SZYMCZUK, Michael, 1950-
A COHORT ANALYSIS OF ELEMENTARY READING ACHIEVEMENT SCORES IN TWO SCHOOL DISTRICTS WHICH IMPLEMENTED THE INDIVIDUALLY GUIDED EDUCATION PROGRAM.

Iowa State University, Ph.D., 1977
Education, tests and measurements

Xerox University Microfilms, Ann Arbor, Michigan 48106

A cohort analysis of elementary reading achievement scores in two school districts which implemented the Individually Guided Education program

by

Michael Szymczuk

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Professional Studies
Major: Education (Research and Evaluation)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1977

TABLE OF CONTENTS

I. INTRODUCTION

A program's philosophy affects the processes of education in many ways. It influences not only the mode of each classroom teacher but also the learning environment and who is responsible for each student's education. In short, its aim is based in its philosophy.

> "... the primary aim of all education effort should be to help boys and girls achieve the highest degree of individual development of which they are capable..." (66, p. 121).

The above quote embodies the spirit of the Individually Guided Education (IGE) program. IGE promotes the notion that each student has various forms of ability, and it is up to the school to find out what that ability is and help develop it. However, IGE also recognizes that the student too must play a role. Instruction is a joint venture.

The focus of the IGE program is the individual student and the creation of an environment which is conducive to learning. The ultimate goal of this program is student self-directed learning. Charles Cols best sums up this notion in these words:

> "... education is not something that is done for a student nor to a student. It is what the student does for himself in developing his own powers. Teachers can help, so can a curriculum and an atmosphere of devotion to things of the mind. But ultimately the problem is entirely the student's..." (18, p. 253).

Respectively, the principal element of schooling is knowledge and skill building. Averch, Carroll, Donaldson, Kiesling and Pincus (2) point out that public schools have been viewed as carrying out four other important functions: socialization; sorting; custodial care; and enhancement of individual creativity and self-reliance. Traditionally, however, a school program's effectiveness is measured in terms of student scores on standardized achievement tests. Student achievement has become the yardstick by which educators and laymen judge the success or failure of an educational program.

The National Assessment of Educational Progress enterprise and the recent back to basics movement (64) serve as evidence for the claim that student achievement is of principal concern. In light of this, how IGE affects student achievement is a vital issue.

The development of the IGE program has been well-documented by Halvorsen (36), Doyle (22) and Stow (84). In light of these documents this investigation intends to provide only a brief overview.

In 1965, the Institute for the Development of Education Activities, I/D/E/A, was founded by the C. F. Kettering Foundation. I/D/E/A along with the Wisconsin Research and Development Center developed an approach to schooling that provided a framework for individualized instruction - IGE. In promoting its program, I/D/E/A encouraged participating schools to link together in cooperative Leagues, whereby experiences and ideas could be exchanged.

In 1972, the Central Iowa Individually Guided Education League was founded by the College of Education at Iowa State University through Dr. George Hohl and by the Department of Public Instruction through Dr. A. John Martin, who became the League's facilitators. Several years after the intervention, research concerned with the effects of IGE was initiated, resulting in selected doctoral studies. Topics spanned from principal leadership perception to how well the IGE program had been implemented. Thus far, however, no doctoral study has been written relative to student achievement.

A. Statement of the Problem

The purpose of this study is to investigate the long term effects of the IGE program and to investigate differences among schools relative to the amount of IGE concept implementation on student reading comprehension test scores. Other studies related to the IGE model have not considered the long term effects of the model on student achievement. Further, studies which have measured achievement and other facets have tended to simply dichotomize the experimental unit into IGE and non-IGE. Charters and Jones (17) have pointed out the fallacy of this approach. They have contended that an investigator needs to know the level of program implementation. This knowledge must extend beyond an overall district level of program implementation. It should include building and within building information. Charters and Jones define four levels of a school's program upon which data should be gathered: Institutional Commitment; Structural Context; Role or Teacher Performance; and Learning Activity.

Institutional Commitment and Structural Context are both easily attainable and measurable. The former generally comes from an authoritative source and is designed to set directions and goals. The latter refers to those changes in formal arrangements and physical conditions in which staff members carry out the educational program.

Evaluation of the level of program implementation based on Role or Teacher Performance and Learning Activities are generally better founded tests of a school's program. How thoroughly teachers embrace the new program's philosophy and how well they utilize the necessary resources of the program are the underlying basis of the notion of Role Performance. Finally, it is the student's own activities and experiences while in the program's environment which influence his measurable outcomes, achievement.

In light of the argument presented by Charters and Jones, this investigation is designed to consider participating schools relative to their level of IGE concept implementation rather than by the simple dichotomized category of IGE and non-IGE.

B. Purpose of the Study

The purpose of this investigation is twofold. First, by using an experimental design which is both cross-sectional and longitudinal one can study the effects of the program implementation on student comprehensive reading achievement test scores within and among schools for each district. In addition, by using a newly developed monitoring instrument to measure the extent to which IGE concepts were implemented one can identify the IGEness of a school or levels within a school and use this measure as an index to identify schools rather than simple program labels such as IGE schools versus non-IGE schools. The subsequent combination of design and measure allows one to investigate the effects of IGE relative to program duration and concept implementation on student outcomes, which in this investigation is reading achievement as measured by standardized tests.

The second purpose of this investigation is to consider the differences among instrument raters. Raters are defined as expert raters and non-expert or amateur raters. Results from this second issue may affect the future cost of IGE program monitoring.

C. Objectives of the Study

The objectives of the investigation may be summed up as follows:

1. To investigate elementary reading test scores among several cohort groups from selected schools of two local school districts,
   A. relative to the level of IGE concept implementation of each building as measured by A Survey of Effective School Processes
   B. relative to the implementation date of IGE.

2. To investigate male versus female elementary reading test scores among several cohort groups from selected schools of two local school districts,
   A. relative to the level of IGE concept implementation of each building as measured by A Survey of Effective School Processes.

3. To investigate the differences among expert and amateur raters on several scales of A Survey of Effective School Processes for the primary units or Learning Community 1's in the Ames and Indianola school districts.

These objectives can be easily translated into the following null hypotheses.

Hypothesis 1: There is no significant difference for the school by cohort interaction in the Ames school district on elementary reading test scores with students as observational units.

Hypothesis 2: There is no significant difference for the school by cohort by sex interaction in the Ames school district on elementary reading test scores with students as observational units.

Hypothesis 3: There is no significant difference for the school by cohort by grade interaction in the Ames school district on elementary reading test scores with students as observational units.

Hypothesis 4: There is no significant difference for the school by cohort by sex by grade interaction in the Ames school district on elementary reading test scores with students as observational units.

Hypothesis 5: There is no significant difference for the school by cohort interaction in the Ames school district on elementary reading test scores with means as observational units.

Hypothesis 6: There is no significant difference for the school by cohort by sex interaction in the Ames school district on elementary reading test scores with means as observational units.

Hypothesis 7: There is no significant difference for the school by cohort by grade interaction in the Ames school district on elementary reading test scores with means as observational units.

Hypothesis 8: There is no significant difference for the school by cohort by sex by grade interaction in the Ames school district on elementary reading test scores with means as observational units.

Hypothesis 9: There is no significant difference for the school by cohort interaction in the Indianola school district on elementary reading test scores with students as observational units.

Hypothesis 10: There is no significant difference for the school by cohort by sex interaction in the Indianola school district on elementary reading test scores with students as observational units.

Hypothesis 11: There is no significant difference for the school by cohort by grade interaction in the Indianola school district on elementary reading test scores with students as observational units.

Hypothesis 12: There is no significant difference for the school by cohort by sex by grade interaction in the Indianola school district on elementary reading test scores with students as observational units.

Hypothesis 13: There is no significant difference for the school by cohort interaction in the Indianola school district on elementary reading test scores with means as observational units.

Hypothesis 14: There is no significant difference for the school by cohort by sex interaction in the Indianola school district on elementary reading test scores with means as observational units.

Hypothesis 15: There is no significant difference for the school by cohort by grade interaction in the Indianola school district on elementary reading test scores with means as observational units.

Hypothesis 16: There is no significant difference for the school by cohort by sex by grade interaction in the Indianola school district on elementary reading test scores with means as observational units.

Hypothesis 17: There is no significant difference among expert and amateur raters' scores on the Overall scale of A Survey of Effective School Processes for Learning Community 1 in the Ames school district.

Hypothesis 18: There is no significant difference among expert and amateur raters' scores on the Commitment scale of A Survey of Effective School Processes for Learning Community 1 in the Ames school district.

Hypothesis 19: There is no significant difference among expert and amateur raters' scores on the Structure scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Ames school district.

Hypothesis 20: There is no significant difference among expert and amateur raters' scores on the Teacher role scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Ames school district.

Hypothesis 21: There is no significant difference among expert and amateur raters' scores on the Student Role scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Ames school district.

Hypothesis 22: There is no significant difference among expert and amateur raters' scores on the Overall scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Indianola school district.

Hypothesis 23: There is no significant difference among expert and amateur raters' scores on the Commitment scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Indianola school district.

Hypothesis 24: There is no significant difference among expert and amateur raters' scores on the Structure scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Indianola school district.

Hypothesis 25: There is no significant difference among expert and amateur raters' scores on the Teacher role scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Indianola school district.

Hypothesis 26: There is no significant difference among expert and amateur raters' scores on the Student Role scale of _A Survey of Effective School Processes_ for Learning Community 1 in the Indianola school district.

D. Definitions

Cohort	a group of students who experienced schooling during a certain number of school years within the same school district and within the same respective school building.
Cohort Analysis	a statistical design which embodies the properties of both a longitudinal and a cross-sectional design.
Comprehensive Reading Score	grade-equivalent scores obtained from the Word Meaning Subscale of the Stanford Achievement Test for the Ames school district and grade-equivalent scores obtained from the Comprehensive Reading subscale of the Gates-McGinitie Reading Test for the Indianola school district.
Expert Rater	an individual who is employed by I/D/E/A or aided in the development of the monitoring instrument.
Grade	a major division of the instructional program of an elementary school, representing the work of one school year.
Incomplete Block Design	a specific statistical design whereby blocks do not include all treatment levels or combinations of individuals contained in an experiment.
IGE	an abbreviation for Individually Guided Education. A program which incorporates and encourages such concepts as: multi-age grouping, continuous progress learning, individualized learning, differentiated staffing, team teaching, tutoring and other innovations.

Learning Community	the instructional unit of a school which includes a unit leader, teacher, associates, and a multi-age group of students.
Reading	recognition of printed or written symbols which serves as stimuli for the recall of meaning built-up through the reading process involving both the acquisition of the meanings intended by the writer and the reader's own contribution in the form of interpretation, evaluation and reflection about their meaning.

E. Delimitations

The scope of this study is limited to those school districts which share similar characteristics and experiences with the two local school districts in this investigation. Careful examination of the findings and conclusions should enable local school officials to determine whether or not implementing IGE concepts enhances student comprehensive reading test scores.

It should be said that this investigation does not attempt to provide data which would support or refute the relative worth of IGE concepts in other subject areas, such as vocabulary, mathematics, etc.

II. REVIEW OF LITERATURE

In light of the topic of this study the literature review was limited to four major areas. These areas were as follows: A.) Achievement studies in IGE related areas; B.) Previous IGE achievement studies; C.) IGE research conducted in the non-achievement domain within the local IGE League; and D.) Developmental designs.

A. Achievement Studies in IGE Related Areas

IGE is a change program which incorporates and encourages the following innovations: (1) team teaching; (2) differentiated staffing; (3) multi-age grouping; (4) continuous progress learning; (5) tutoring; (6) inquiry directed learning; (7) open classrooms; (8) individualized instruction; and (9) flexible scheduling (36). One may consider the IGE program as an aggregate of ideas which tend to foster a non-traditional or progressive philosophy. In general, however, these notions have not significantly influenced American education. In fact, a brochure entitled I/D/E/A's Guide to an Improvement Program for Schools exclaims this to be the case:

> "... all these inherently sound concepts have failed to exert a significant impact on American education because in most instances, they have been tried in isolation from each other..." (43, p. 6).

A review of literature of each of the above topics tended to support I/D/E/A's contention concerning isolated experimentation. However, some of the innovations tended to be more effective in producing a noticeable, statistical, difference. Multi-aged grouping, individualized instruction, continuous progress learning and tutoring seemed to effect or to result in more studies which claimed significant differences than team teaching, differentiated staffing, etc.

Many, if not most, of these innovative concepts are not new to American education and tend to be confounded with broader classifications. For example, the notion of non-gradedness can be traced back to Western Springs, Illinois, 1937 (89). As the non-graded concept developed it also incorporated the notions of multi-age grouping, continuous progress, individualized instruction and tutoring. The underlying theme of non-graded organization was best expressed by Goodlad and Anderson:

> "The nongraded school is designed to implement a theory of continuous pupil progress. Since the differences among children are great and since these differences cannot be substantially modified, school structures must facilitate the continuous educational progress of each pupil. Some pupils, therefore, will require a longer period of time than others for achieving certain learning and attaining certain developmental levels" (32, p.52).

Investigations into the effects of the non-graded concept in the elementary school level tended to result in mixed but favorable results toward the non-graded concept. Paven (72), Brody (10), Buffie (13), Brown and Theimer (11) and Hillson, Jones, Moore and Van Devender (39) studied the effects of non-gradedness by using standardized achievement tests. They found significant differences favoring the non-graded concept. Mitchell and Zoffness (60) and Muck (63) found no significant differences between graded and non-graded schools on various standardized tests. Finally, Vogel and Bowers (88), Carbone (16) and Ingram (42) found differences favoring the graded system of education.

Bowman (9), Halliwell (35), Killough (47) and Skapski (82) studied the effects of non-gradedness upon various dimensions, such as sex, grade level and levels of ability. Halliwell compared the graded and the non-graded concept relative to grade level and found highly significant differences, at the .01 level, among the pupils in the first grade in favor of non-gradedness, significant differences, at the .05 level, among pupils in the second grade, again in favor of the non-graded concept, but no differences among pupils in the third grade. Bowman's results, however, were directly opposite those of Halliwell. Bowman found no differences between primary treatment groups but found significant differences, at the .05 level, between the intermediate treatment groups.

Skapski considered student ability level when she compared graded and non-graded schools. Her study revealed that students who were categorized as very superior tended to benefit more from non-gradedness than students categorized as either superior or average. She inferred this by examining the mean difference score of each of the categories by treatment group. However, it should be pointed out that Skapski errored in her comparison between different ability levels. She neglected the scaling problem that is, students of different abilities start at different levels and proceed at different rates regardless of treatment.

Killough considered the effects of non-gradedness relative to sex differences. He concluded that girls in a non-graded program will achieve at a faster rate in arithmetic computation than will boys, and boys in a non-graded program will achieve at a faster rate in reading comprehension than will girls. These findings are opposite those which are generally found. However, sex by school program did not result in a significant interaction.

The follow-up or time lapse aspect of the effects of a non-graded program were studied by Jones, Moore and Van Devender (46), Morris, Prager and Morrell (61) and Brown (12). Here, too, mixed results are reported. In 1967, Jones, Moore and Van Devender conducted a follow-up study and found no significant differences between the non-graded and graded treatment groups. They accounted for the initial superiority of the non-graded program by exclaiming possible novelty effects and by reasoning that the non-graded program is more appropriate for the very young or beginning students.

Morris, Prager and Morrell found that significant differences between the different programs were not only maintained but enhanced. Furthermore, Morris, Prager and Morrell found that both males and females in the non-graded program tended to score higher than males and females within the graded program. However, within each program females scored equally as well as males.

Brown confirmed the findings of Morris, Prager and Morrell. He found that significant differences between the non-graded and graded programs were maintained and again favored the non-graded program.

In 1960 the philosophical tenets of continuous progress learning were expressed by John Gardner, on behalf of the President's Commission on National Goals, in these words:

"... there must be diverse programs within the educational system to take care of the diversity of individuals. No child should be required to fit a space and pattern of education designed for children of other capacities..." (74, p. 209).

Findings of studies conducted in this concept generally reflected those reported by investigators of the non-graded concept. Evans (23) reported that high ability students tended to benefit more from continuous progress learning than low ability students. Resnich and Wang (75) and Williams (90) indicated that students in a continuous progress organization score significantly higher than students in a traditional classroom organization. Thomas (87) and Amendola (1), however, found no significant differences between their treatment groups but they did report that students in a continuous progress environment tended to score higher than pupils in a traditional classroom environment.

The notion of individualized instruction plays a major role in the IGE program. Halvorsen (36) felt that it was the central concept of the entire program. Sentiments concerned with the need for individualized instruction were expressed by Gronland in these words:

"The wide range of individual differences among students makes it unlikely that group instruction alone, with or without ability grouping can meet the varied needs of the students. Some type of adaptation in the instructional program is needed so that more individualized learning expression can be provided..." (34, pp. 1-2).

Among the various subject areas within which one can individualize, reading is probably most apt. Bond and Tinker bear this notion out in these words:

> "... any one child must be given material that is as nearly suitable to his level of reading growth as possible. He must be taught by methods compatible with his characteristics and capacities. For him, in addition, those phases of reading instruction must be emphasized that demand immediate attention. Reading instruction, to be effective, must be provided on an individual basis..." (8, p. 48).

Research relative to individualized instruction and reading supports Bond and Tinker's contention. The investigations of Spencer (83), Davis and Lucas (21) and Tausig (86) tend to show significant differences in favor of individualized reading. Parker (71) took into consideration ability level and found that students in the high and middle ability levels benefited from individualized instruction, whereas low ability students did not benefit.

In other subject areas the results of various investigations are mixed. Using the Iowa Test of Basic Skills and the Iowa Test of Educational Development, Hatfield (37) found that students who experienced individualized instruction scored significantly better on each subscale than students within the control group. However, when he dichotomized each group into high and low ability groups, Hatfield found no differences between the low ability groups but found the high ability individualized learning group scoring significantly greater than the high ability control group.

Somewhat counter to Hatfield's results, Given (31) and Huser (41) found no overall significant differences between individualized learning and group learning. However, Hauser had found significant differences favoring individualized learning when students were divided into grades. In his investigation, the highest individualized learning grade (sixth grade) scored significantly better than the highest group learning grade.

Finally, in recent years educational researchers have been investigating student to student tutoring with regard to its effects upon both the tutee and the tutor. Investigators tend to report that both benefit from the encounter. Significant differences favoring tutoring have been found in all subject areas. Unlike the previous investigations, however, individuals who are classified as low ability students do benefit from tutoring. These findings are born out by Frager (26), Frager and Stern (27), Bailey (3), Rust (77), Brown (12), and Schaver and Nuhn (79). Rogers (76), however, found no significant differences in his investigation. Student to student tutoring may be the most beneficial element within the IGE program.

By virtue of IGE's school organization the notion of team teaching is implicitly and explicitly promoted as a key concept. The preponderance of research, however, tends to indicate that team teaching does not affect student achievement. Studies by Beasley (6), Burchyett (14), Cooper and Stan (19), Fraenkel and Gross (28), Heathers (38), and Lutenbacher (55) affirmed this notion. Floyd (25) and Miller (59), however, found that team teaching had affected student achievement. Both studies indicated statistically significant gains at lower as well as higher academic levels.

McCallum and Roark (57) explained the failure of team teaching as a result of a basic misconception with regard to the capability of members within the teaching team. It is not the academic or leadership ability that is in question, rather the notion that at least one member of the team is sure to understand a child who has a problem. Their research did not support the assertion that teachers in teams can identify children's problems in some superior way. Hence, a student with a special academic problem may not be identified.

Several other concepts were previously mentioned, such as differentiated staffing, flexible scheduling, open classrooms and inquiry directed learning. Investigations concerned with these innovations tend to show either no significant differences or differences favoring a traditional, one classroom to one teacher, organization.

1. Summary

A review of achievement studies in IGE areas reveals a hodgepodge of results. Innovations such as continuous progress learning, individualized learning, multi-age grouping or non-gradedness and tutoring do seem to enhance student achievement, whereas other innovative concepts which are equally encouraged and incorporated into the IGE program do not tend to benefit students. How the aggregate of these ideas affect student achievement is not yet known.

B. Previous IGE Achievement Studies

Through a review of literature and personal communications five IGE achievement studies were reviewed. The first of these studies was reported by Morrow and Quilling (62). Using a pre-test post-test design, they investigated the effects of the IGE program on first, third and fifth graders. The investigators attempted to match students relative to their IQ and previous achievement. However, no account was given of the level of IGE program implementation.

First graders were tested on their arithmetic concepts and skills, third graders were tested on various arithmetic and reading attributes and fifth graders were tested on various arithmetic, reading and scientific attributes. In several instances, the investigators used different tests to measure cognitive change. For example, first graders were pretested on the Metropolitan Achievement Test, Arithmetic Concepts and Skills and were post-tested on the Scientific Research Associates (SRA) Achievement Test, Arithmetic Concepts. It was not indicated whether or not these tests were equated. Hence, spurious gain scores may have been reported. However, in most cases the same tests were used.

The results of this first investigation of the effects of IGE upon student achievement reported no significant differences between the innovative program (IGE) and the traditional self-contained classroom program.

The second IGE achievement study was reported by Klausmeier, Quilling and Sorenson (49). These investigators selected two "smoothly" operating IGE primary units and used a criterion referenced (CR) measure and a standardized reading test as their achievement measures.

The CR test was only administered to three groups in the IGE schools in 1969 and in 1970. The groups were comprised of students who were starting their second through fourth years in school respectively. Gains made by the three groups were reported. The median number of objectives attained by the three groups were 8, 19 and 11 respectively. These gains were supposed greater than gains experienced by students in a traditional classroom program. However, the investigators failed to report statistics relative to these differences.

Using the Doren Diagnostic Reading Test and a pre-test post-test design the investigators compared the gain scores of IGE students with those in a traditional classroom program. The investigators found no significant differences at the .05 level. They explained the lack of significance by citing unusually high gains by the control group which may have resulted from communication among students, IGE and non-IGE, or failure of teachers to adhere to the specified control treatments. Furthermore, the investigators felt that the large gains would not likely be maintained and eventually the differences between the two treatment groups would become apparent.

The third study was a dissertation by Schneiderhan (80). Schneiderhan compared achievement scores of students in an IGE program, an Individually Guided Instruction (IGI) program and a traditional self-contained classroom at the intermediate level. The study failed to indicate the number of years the IGE program was in operation and the possible level of program implementation. It did note that the IGI program was a recent development.

The investigator used the Iowa Test of Basic Skills (ITBS) as the achievement measure. The experimental design incorporated both a pre-test and a post-test. Socio-economic status, sex and treatment were considered categorical variables. IQ and the pre-test were used as covariants.

The results obtained by this investigator generally indicated no significant differences among the treatment groups. Mean scores on the Vocabulary subscale of the ITBS revealed that students in the self-contained classroom treatment group scored highest, whereas students in the IGE program scored lowest. However, the differences among the groups were not statistically significant.

The fourth investigation was performed by Paden (70), an I/D/E/A staff member. He was aware of the likelihood of no initial differences between IGE and non-IGE scoools.

> "Though I/D/E/A makes no claim that beginning IGE schools will necessarily have achievement gains, it is important to understand and to communicate the early effects to schools entering the program..." (70, p. 1).

In Paden's investigation, standardized achievement test results were obtained from 19 schools in their second year of IGE implementation. The ITBS was administered to third, fourth, fifth and sixth grade students in participating schools. The mean grade equivalent scores for each test group were calculated for the year prior to participating in IGE and for two years following.

Paden monitored the implementation effects by looking at the number of average grade equivalent scores that were above and below the national average on the ITBS. Forming a 2 X 3 contingency table for each ITBS subscale, using the Chisquare statistic with 2 degress of freedom, Paden found none of the subscales statistically significant.

Paden concluded that the initial two years of using IGE did not have a measurable effect upon student achievement. Paden also noted that only when the processes of the entire program are brought to bear on a specific problem does IGE make claims. The schools used in Paden's investigation were estimated to be at the 40% to 60% implementation level.

In 1975, Wilson (91) evaluated the Des Moines IGE program. Wilson, like Paden, recognized that it was inappropriate to attempt to measure the impact of IGE upon student achievement before determining the level of IGE program implementation.

Through the combined effort of the IGE Facilitators, other IGE groups and the Department of Evaluation a three year investigation was planned. The implementation phase was broken into six phases. The first two phases were exploratory and preparatory. The following phases were stepping stones toward the sixth phase which was full implementation.

To investigate the impact of IGE upon student achievement 13 IGE and 13 non-IGE schools were matched as closely as possible on student enrollment and socio-economic levels.

A random sample of 302 IGE students and 307 non-IGE students from the fourth and sixth grades was selected. A subsample from the IGE schools was selected based on whether or not students were housed in an open space environment. A complementary subsample was selected from the non-IGE schools.

Using the ITBS and a post-test design with treatment type, sex and grades as categorical variables, Wilson computed t-tests on various combinations of categorical variables for both the total sample and the subsample.

The results of the t-tests on the entire sample were mixed. Mean difference scores tended to favor the non-IGE schools. Fourth grade males seemed to vary most. Significant differences favoring non-IGE were reported for this category in three subscales: Language Skills, Work Study and Composite.

The results of the t-tests on the special sample, students in an open space environment and their counterparts, revealed differences favoring IGE. These differences were prominent among females, especially sixth grade females. In fact, the largest difference favoring IGE was reported on the Work Study subscale when IGE sixth grade female student scores were compared with non-IGE sixth grade female student scores. Although none of the differences were found to be statistically significant, a positive trend possibly related to age was observed.

1. Summary

The previous IGE achievement studies tended to report either non-significant differences between IGE and non-IGE programs or differences favoring the non-IGE programs. Recent investigators, Paden and Wilson, recognized the need to determine the level of program implementation. Their estimates, however, were only aimed at the IGE schools. None of the achievement studies thus far have investigated whether or not the control schools were implementing the IGE concepts. If a control school implements many of the same concepts and practices as that of an IGE school, or if an IGE school does not implement the concepts, then which of the two schools is the IGE school? In other words, a label does not make the program.

C. IGE Research Conducted in the Non-achievement Domain within the Local IGE League

Student outcomes, such as student achievement, are generally affected not only by educational hardware, teaching machines or student carrel, but also by the attitudinal and organizational changes in the school staff. In order to better understand these variables and their possible effects upon student achievement, this section reviews the research conducted in areas of school organization, student attitude and teacher role performance.

Before measuring the possible effects of any program one must first determine whether or not the program has been implemented. Halvorsen's (36) doctoral dissertation provided future researchers of IGE with an instrument which measures the amount of IGE program implementation. About the same time Doyle (22), Lindaman (51) and Olney (68) began research into facets of IGE. Unfortunately Doyle, Lindaman and Olney began their research before Halvorsen completed his dissertation, hence measurement of schools relative to amount of IGE implementation was not part of their experimental designs. Later, however, Stow (84) utilized a refined version of Halvorsen's instrument.

Doyle investigated the effects of IGE in conjunction with measurements from Indicators of Quality. Scores from participating schools within the Central Individually Guided Education League (CIEL) and selected controls were collected over a two year period, 1972 and 1973. Doyle also examined how IGE principals and non-IGE principals compared to each other in their behavior, as measured by the Leadership Behavior Descriptive Questionnaire (LBDQ). The results of Doyle's study indicated no significant differences between IGE and non-IGE. However, trends favorable toward IGE were evident.

Lindaman researched the aspect of student self-esteem and the ability of elementary teachers to infer learner self-concept. He administered the Self-Esteem Inventory (SEI) and Florida Key (FK) to a selected group of students and teachers within the CIEL. He found that non-IGE students reported a slightly more positive self-esteem and that non-IGE teachers inferred their students' learning self-concept to be higher.

Olney surveyed the relationship of organizational patterns of IGE schools to the opinions and goals of teachers. To gather his data he used three instruments: (1) an opinionnaire; (2) Continuous Progress (CP) -- A Test of Current Instructional Principles and Practices; and (3) Perception of Educational Trends (PET). The results of the opinionnaire indicated that the opinions of teachers differed significantly in favor of the IGE schools with regard to interaction patterns, division of labor and decision making. However, data gathered regarding teacher objectives indicated no significant differences between IGE and non-IGE schools.

The results from the CP indicated that IGE teachers scored highly upon the concepts of continuous progress learning, small groups, brainstorming, individualized assessment and learning stations. Also, IGE scored highly upon PET with respect to individualized curriculum, team teaching and the use of para-professionals.

Olney's results indicated that teachers within IGE labeled schools do indeed incorporate the concepts promoted by the IGE program. Hence, institutional commitment among IGE teachers would be relatively high.

Stow (84) took a third measure of Indicators of Quality and analyzed the data relative to program label (IGE or non-IGE) and three levels (high, medium and low) of IGE concept implementation. In cooperation with I/D/E/A Stow utilized a modified version of Halvorsen's instrument. Measurements of Indicators of Quality and the level of IGE concept implementation were observed and gathered from selected grade groupings within buildings. These groups were defined as learning communities (LC). In addition, Stow investigated the differences between IGE and non-IGE schools on four scales defined by the I/D/E/A staff.

Stow's results indicated that IGE schools tend to have significantly higher scores across the observation years than non-IGE schools on several Indicators of Quality subscales: Interpersonal Regard; Group Activity; and Pupil Signs. Interpersonal Regard was significant at all levels of measurement. Group Activity and Pupil Signs were significant at the LC1 level (primary or grades k-2). When IGE concept implementation levels were compared high level implementation was statistically significant from medium and low level on Pupil Signs in LC1. Levels did not differ on any other subscale.

Further, no significant differences were found between IGE and non-IGE schools on those subscales defined by the I/D/E/A staff.

1. Summary

In general it appears that the IGE program does affect the school community over time. This notion is most dramatically illustrated in Stow's study. The lack of significance in Doyle's and Lindaman's studies may be attributed to the fact that they collected their data only a year or two years after IGE program implementation.

D. Developmental Designs

The purpose of this section is to discuss, in general terms, various research designs for investigating human development or change. In turn, four strategies are discussed: change scores; longitudinal designs; cross-sectional designs; and cohort or mixed designs. Each subsection examines the advantages and disadvantages of the particular strategy.

1. Change scores

Since developmental designs are concerned with the notion of change over time, it is tempting to assume that the basic datum unit should be the change score, that is post-measurement minus pre-measurement. This seemingly logical assumption is the most appealing attribute of the change score. Educators often use this technique to secure data for purposes of evaluating teaching effectiveness and learning. However, a survey of the literature reveals that the change score is not a viable datum unit and is ridden with many psychometric and conceptual faults.

Nunnally (67) listed several major problems with the change score: 1.) regression effect; 2.) errors of measurement; and 3.) misinterpretation. Cronbach and Furby (20), Furby (29) and Lord (54) have expounded on these major problems and tend to agree that the change score is not a desirable datum unit.

The notion of regression effect or regression toward the mean was best expressed by Nunnally in these words:

> "If one computes change scores, he finds that people who scored above the mean on the first occasion tend to have negative change scores, and that the people who scored below the mean on the first occasion tend to have positive change scores..." (67, p. 88).

Lord further stated that

> "The reasons for the so-called regression are to be found in real life and not in statistical theory" (53, p. 447).

Finally, regression toward the mean occurs when the correlation between the pre-test and post-test measures is less than unity.

The notion of errors of measurement further complicates the accurate measurement of growth. For example, if an investigator employs two test forms which are not perfectly reliable, then the apparent gains or losses may be due to errors in measurement and the true score of a student may in reality remain the same, Lord (52). Furby explains this notion in these words:

> "... subjects with large positive error contributing to their scores on x (pre-test) are likely to have higher x scores on the average than those subjects with negligible or large negative error contributing to their scores. However, it is highly unlikely that these subjects with large positive error in their x scores will also have large positive error in their y (post-test) scores (since error in x is uncorrelated with error in y). Therefore, their y scores tend to be lower (closer to the mean) than their x scores" (29, p. 175).

The above notion assumes an unbiased test. Errors of measurement contribute to correlations less than unity between the pre-test and post-test measures.

The misinterpretation of change scores and their usage is still far from becoming a rare phenomenon. Furby cites the investigations of Eysench and Jensen. They both misinterpreted regression effect as supporting evidence for genetic sources of individual differences in the variables under consideration. Lord (52) further elaborates that equal change scores do not necessarily mean equal ability, that is the test-score scale does not ordinarily provide units of measurements that are clearly equal in different parts of the scale.

In light of the preceding evidence the usage of the change score as a measure of human development should be discarded. Further, tests of treatment effects via student cognitive change scores should be suspect. If an investigation is designed with regard to randomized treatment assignment, then post-test scores alone are a suitable dependent variable.

2. Longitudinal designs

Another measure of human development is the longitudinal design. This design is characterized by its repeated measures of one sample group over time. It, unlike the change score, permits the investigator to directly analyze subject change. Further, the longitudinal design suffers less severe problems than the cross-sectional design. Finally, it allows the investigator to employ powerful statistical methods. Goulet (33, p. 506) points out that the longitudinal design is amenable to both the between-subject and within-subject (i.e. repeated measurement) testing procedures.

Unfortunately, the disadvantages of the longitudinal design tend to outweigh its advantages. Bell (7, p. 145) pointed out that the sampling selection involved generally restricts a longitudinal study to cooperative groups which, in turn, limits the generalization of findings. Another problem related to sampling, attrition, was pointed out by Jones (45). Baltes (4) contended that the survival rate may often be correlated with the measurable variable.

Other disadvantages of the longitudinal design are possible practice effects, methodological lag, researcher commitment and, probably most important, funding of longitudinal research. Schaie (78), for example, suggested that practice effects may influence successive findings. This problem, however, is easily handled by either testing for possible practice effects or, as recommended by Nunnally (67), developing several reliable alternative instrument forms and counterbalance these over subgroups.

The very practical problems of commitment and funding are the major stumbling blocks of this developmental design. Very few investigators are motivated to undertake research that might not be completed in their lifetime. Finally, funds for longitudinal research are indeed scarce. In essence, the theoretical and statistical problems of the longitudinal design may be overcome but the very human problem of commitment and financing may never be overcome.

3. Cross-sectional designs

The cross-sectional design is distinguished from the longitudinal design by the time of observation. Unlike the long and involved process of longitudinal study, the cross-sectional study investigates the variables of interest at a point of time with several strata of a particular variable, usually age. The social scientist tends to use this sampling strategy more than the longitudinal design. Several reasons for its preference are as follows: 1.) Cross-sectional data contain large variations in some variables whose variations over time are only moderate and often subject to trend; 2.) The size of a cross-sectional sample of data can usually be increased enough to make sampling variance relatively negligible; 3.) Multicolinearity among variables in a cross-section is less acute than among corresponding variables in a longitudinal design; 4.) The problem of interdependent disturbances usually does not arise in the analyses of cross-section data; 5.) In some cases the data are more reliable (44, pp. 523 - 524).

An example of a cross-sectional design is the Survey of Equality of Education Opportunity of the Coleman Study. This investigation sampled several grade strata: first, third, sixth, ninth and twelfth.

This very appealing design, however, has major problems. The principal problem is a by-product of time itself. In sampling several different strata the data is liable to generation effects. With regard to educational research, Hilton and Patrick (40, p. 24) point out that if strata differences are large, perhaps from migration patterns in the school population, or if the dropout rate is high, cross-sectional data may provide highly inaccurate estimates of the students' growth in a particular time. In addition, Goulet (33, p. 507) notes that the amount of schooling and other components of age related behavior change tends to be confounded within the cross-sectional design. Finally, the cross-sectional design assumes that the effects of schooling for children in comparable grades are the same irrespective of the year in which the children are enrolled.

For the educator, the necessary assumptions involved may preclude the use of the cross-sectional design. In other words, if the educational researcher cannot compensate for the various nuisance variables such as migration patterns, dropout rate, etc., then should he proceed with the investigation using a cross-sectional design?

4. Cohort or mixed designs

> "A cohort is defined as an aggregate of individual elements each of which experienced a significant event in its life history during the same chronological interval" (44, p. 546).

The notion of cohort analysis originated within demography, particularly in the study of time series of fertility. In 1959, the notion of cohort analysis was suggested for use in Public Opinion Research (24). However, cohort analysis as we now know it was not developed until 1965.

The forerunner of the modern cohort analysis, the convergence approach, was introduced by Bell (7). It, like the cohort analysis, involved the combination of the cross-sectional and longitudinal techniques in such a way that developmental changes could be computed in a shorter period. The technique involved selecting a sample of subjects, say fourth graders, sixth graders and eighth graders and taking repeated measures from each stratum until the strata overlapped, that is fourth graders became sixth graders. Bell contended that this technique would generate a continuum of observations like that of longitudinal designs without the expenditure of research time and foundation monies.

Bell's convergence technique never became a popular technique. Unfortunately, it inherited the ills which plagued the cross-sectional design plus some of those which troubled the longitudinal design, such as attrition.

In 1965, Schaie (78) proposed three models which took into account three sources of developmental change: cohort difference; time difference; and age difference. The first of these models was dubbed the cohort-sequential method since longitudinal sequences for two or more cohorts were examined simultaneously. This method allowed the investigator to make inferences relative to age changes at all points of the age range covered and to also make inferences about cohort differences. However, utilization of this design tended to confound time differences. Schaie pointed out that the cohort-sequential method would yield unambiguous results only when the assumption is made that changes in the variable under study are unrelated to cultural change.

The second design suggested by Schaie is entitled the time-sequential method. This model allows the investigator to examine age differences at all points of the age range, as well as time differences at all ages. In this case, the method is valid only when the change in the variable under study is not environmentally liable, that is unrelated to genetic or cohort-environmental change.

Finally, Schaie's third model is the cross-sequential method. This procedure permits inferences as to time-lag differences at all points of the time range, as well as to cohort differences at all times of measurement. In this design, age is confounded. One must assume that the variable under study is unrelated to age change. For educational research involving achievement this model would probably not be a viable technique, especially for research in the primary and intermediate levels where students are most subject to maturation.

Baltes (4) criticized Schaie's models with regard to the formal definitions of the three components and their measurability. Baltes asserted that only two of the three variables were necessary since they are interdependent. Baltes (4, p. 157) formulated these relations:

$$\text{Age} = \text{Time} - \text{Cohort}$$
$$\text{Time} = \text{Age} + \text{Cohort}$$
$$\text{Cohort} = \text{Time} - \text{Age}$$

Using these, Baltes suggested that Schaie's basic trifactorial formula,

$$R = f(A, C, T)$$

could be reduced to a formula which involved only two components plus a linear composite. For example, the following formula illustrates Baltes' contention:

$$R = f(A, C, A + C)$$

Baltes also suggested that direct measurement of each component within the model is virtually impossible. The variation of age, cohort and time consists only of a classification of individuals into different time segments of the time continuum. A functional interpretation of the components as suggested by Schaie would require direct measurements and variation of neuro-physiological and environmental conditions (4, p. 157).

Afterward, Baltes proposed his own developmental model which was composed of an age and a cohort component. He conceptualized his bifactorial model as a bifactorial analysis of variance design with age and cohort as fixed effects. The bifactorial design was extended to accommodate both single independent observations and the repeated measure design on the age factor.

Schaie responded to Baltes' allegation by acknowledging the non-independence of the three components. However, with regard to the lack of measurability, Buss (15, p. 467) reported that Schaie contended that the matrix sampling design determines the different information which, in turn, accommodates the various tripartite designs.

Buss in turn agreed with the adaption of Baltes' bifactorial model but also maintained that Schaie's model was equally acceptable given clarification of various concepts. However, he suggested that the time-sequential and cross-sequential methods would be of little value because of their underlying assumptions. He further asserted that, even though age and cohort related behavioral changes may be independent, it is difficult to separate out the meaning of time of measurement from either age or cohort effects.

Counter to the claims of Baltes and Buss, Mason, Mason, Winsborough and Poole (56) noted that performing cohort analysis by ignoring one of the three independent variables may in some situations be justified. However, it may not prove satificatory in others. They contended that if age, cohort and time of measurement have distinct causal interpretations, then an analysis which omits one of these variables is subject to spurious findings.

5. Summary

The usage of cohort or mixed designs developed by either Schaie or Baltes in educational research has not become in vogue. Applications of these designs requires careful planning. Goulet (33) cites several instances in which these developmental designs could be implemented. For example, the amount of time spent in study may vary with the season of the year or the proximity to important holidays. Goulet also points out that general uncontrollable variables, such as social interaction, environmental context and parent and peer demands do not vitiate the use of these designs in educational research. These designs are viable tools.

III. METHODS AND PROCEDURES

This chapter details the methodology and procedures employed during the course of this investigation. It consists of four sections which are as follows: A.) Subjects; B.) Instruments and Measures; C.) Variables; and D.) Design and Statistical Procedure.

A. Subjects

Two local school districts, Ames and Indianola (Iowa), were selected to participate in this investigation. All (four) of the elementary schools in the Indianola school district and five elementary schools in the Ames school district cooperated. These schools were selected primarily because of their involvement in the field test of A Survey of Effective School processes (hereafter reported as A Survey), Stow (84). Data generated by the field test on these schools were used in this investigation.

For this investigation 875 students from the Ames school district and 711 students from the Indianola school district were selected to be subjects. Student scores from grades two, three and four in the Ames schools and student scores from grades one and two in the Indianola schools were employed. These grades were selected for two reasons. First, the IGE program was implemented initially at the primary and lower elementary grade levels. Hence, if the IGE program or its concepts do indeed affect student behavior, then its effect would most likely be observed among these grade levels. Second, these grades were the earliest in which a standardized reading test was administered.

Selection of subjects from the two districts was based on a student's enrollment. All students who were enrolled in a participating school for at least two consecutive years were potential candidates. In other words, if a student entered the second grade of a participating school during the 1972-1973 school year and was enrolled within the same school for the 1973-1974 school year, then he was a subject.

By virtue of the investigation's design student scores were collected beginning with the 1970-1971 school year in each district (see Figures 1 and 2). In addition, there were five cohort groups within each school building. A cohort group was defined as a group of students who experienced schooling during a certain number of school years and within the same school district and within the same respective school building. For example, a cohort group in building A of the Indianola school district consisted of those students who were in the primary unit, first and second grade, for two consecutive years. Within each grade a classroom

	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77
Cohort1	2	3	4				
Cohort2		2	3	4			
Cohort3			2	3	4		
Cohort4				2	3	4	
Cohort5					2	3	4

Figure 1. Sampling design of grades in the Ames school district

	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76
Cohort1	1	2				
Cohort2		1	2			
Cohort3			1	2		
Cohort4				1	2	
Cohort5					1	2

Figure 2. Sampling design of grades in the Indianola school district

substructure was administratively defined. However, for this investigation the classroom unit was not a viable observational unit for in many cases the classroom teacher was not necessarily the reading teacher. Also, the classroom unit was never kept intact from year to year. Students were evaluated on affective and cognitive dimensions and were regrouped among classrooms.

B. Instruments and Measures

This section discusses the instruments and measures used in this investigation. It is divided into two sections: 1.) A Survey and the opinionnaire and 2.) Gates-MacGinitie and Stanford. The instruments consisted of A Survey and an opinionnaire. The former instrument was developed by I/D/E/A and Halvorsen and first used by Stow in conjunction with its field test. The opinionnaire was developed by this investigator and used primarily as a guide during interviews with selected individuals within the Ames and Indianola school districts. The measures were student comprehensive reading scores on the Gates-MacGinitie Reading Test from the Indianola school district and word meaning scores on the Stanford Achievement Test from the Ames school district.

1. A Survey and the opinionnaire

In 1974, Halvorsen (36) developed An Objective Measure of Education Practices as part of his dissertation at Iowa State University. The instrument was designed to measure the amount of IGE concept implementation in both IGE and non-IGE schools. Its format was a paper-and-pencil questionnaire with a three point response scale.

The following year Halvorsen worked with the I/D/E/A staff to improve his instrument. The resulting instrument was couched in an interview/observation format. In addition, information was obtained from the central office, teachers, students, parents and principals. Instrument items were classified into one of 35 pre-determined categories, i.e. outcomes. The outcomes were further categorized into four scales: Institutional Commitment; Organizational Structure; Teacher Role; and Learning Activities (Student's Role) (see Appendix A). The categorization of items into outcomes and of outcomes into one of four scales was principally a matter of expert judgment on the part of I/D/E/A staff members.

Upon consulting with Dr. Jon Paden, an I/D/E/A staff member, it was agreed that scores from the Learning Activities scale were most appropriate. Measures of IGE implementation amount were available for each building and for clusters of grades within each building which were referred to as learning communities (LCs). This investigation employed measures from LC1 (grades one and two) within the Indianola school district and a computed LC score for the Ames school district which requried a two to one weighting. Ames LC measures spanned two LCs namely those communities which contained grades K through 3 and 4 through 6. Hence, a weighted expert Student Role scale score was used. The implication of using the learning activity scale was that there was a relationship between the scale score and student reading achievement scores. In other words, if students were behaving in a manner such that the building or LC scored highly on IGE concept implementation, then it followed that student achievement was greater among these students than among students of a school with a lower IGE concept implementation score.

Data for this investigation were gathered in conjunction with the field testing of A Survey. During the field test expert and non-expert raters were assigned to schools in pairs (expert and amateur) so that each rater visited at least two schools and was teamed with a different rater in each school (see Table 1 and Table 2).

Table 1. Pairings of expert and amateur raters for the field test of A Survey of Effective School Processes in the Ames school district

	Schools				
	A	B	C	D	E
Raters[1]					
E1	*			*	
A1	*				*
E2		*			*
A2		*		*	
A3			*	*	
E3			*		*

[1]E = Expert; A = Amateur

Table 2. Pairings of expert and amateur raters for the field test of A Survey of Effective School Processes in the Indianola school district

	Schools			
	A	B	C	D
Raters[1]				
E1	*			*
A1	*	*		
E2		*	*	
A2			*	*

[1]E = Expert; A = Amateur

The scores of expert raters were used as reflecting IGE concept implementation. It was felt that experts would be more able to recognize all the facets of IGE than individuals classifed as non-experts (see Table 3 and Table 4).

Table 3. Expert scores on the Student Role scale for each Learning Community 1 or Primary Unit in the Ames school district

Building	Scores
A	23.3
B	18.7
C	26.7
D	11.5
E	19.3

Table 4. Expert scores on the Student Role scale for each Learning Community 1 or Primary Unit in the Indianola school district

Building	Scores
A	24
B	28
C	20
D	6

By virtue of the posthoc nature of this investigation, information concerning non-achievement variables was not available. In light of this, conversations with selected expert members of the participating school districts provided some information concerning unmeasured variables, such as socio-economic status of areas served by the participating elementary schools in this investigation. In order to maintain a congruent nature to the interviews, a short list of specific questions was developed (see Appendix B). It should be noted that in order to avoid possible bias, conscious or unconscious, experts were not affiliated with any particular elementary school.

2. Gates-MacGinitie and Stanford

The Indianola school district tested its first and second grade students with the Gates-MacGinitie Reading Test, 1965 version. The test was administered every Spring by the classroom teachers and was hand scored. As a result no classroom level, building level or district level reliability estimates were calculated. Scores were reported as grade-equivalent scores and percentile scores. The Gates-MacGinitie Technique manual (30) reported a split-half reliability estimate of .94 for the first grade Reading Comprehension test scale and .93 for the second grade Reading Comprehension test scale.

Powell (73) reviewed the Gates-MacGinitie Reading Test and summarized that the test would provide usable data on achievement in comprehension, vocabulary and speed. Similar notions were expressed by Mehrens and Lehman (58).

The Ames school district tested its second, third and fourth grade students with the Stanford Achievement Test. Word Meaning is one among several academic domains tested by the Stanford Achievement Test. The test was administered every October and was computer scored. Grade-equivalent and percentile scores were reported. Again, no reliability estimates were available at the building or district level. The Stanford Achievement Test Norms Booklet for Primary Level 1 (85) reported a split-half reliability estimate for the word meaning scale of .95. Similarly, the manual for primary level 2 reported a split-half reliability estimate of .96, and the manual for the primary level 3 reported a split-half reliability estimate of .94.

C. Variables

The principal variables in this investigation were as follows: student grade-equivalent scores on the comprehensive dimension of reading (dependent variable); IGE concept implementation rate, which is the treatment per building, as measured by A Survey; cohort differences within and among buildings across time on reading achievement scores (classificatory variables); male and female achievement score differences relative to treatment (classificatory variable). In addition, variables which were not directly measured but were considered as being influential on student achievement such as the socio-economic status of the geographic area served by the participating elementary schools was subjectively surveyed through personal interviews with individuals who were judged as experts on the social fabric of the district relative to each elementary school.

Reading was selected as the achievement domain for three reasons. First, reading and mathematics were judged most compatible with the precepts of IGE. Secondly, IGE concepts were implemented into the reading curriculum of several buildings across districts. Finally, the testing programs relative to IGE program implementation were most compatible in the primary and lower elementary grades.

Support for the first reason is suggested by Bond and Tinker in these words,

> "... any one child must be given material that is as nearly suitable to his level of reading growth as possible. He must be taught by methods compatible with his characteristics and capacities. For him, in addition, those phases of reading instruction must be emphasized that demand immediate attention. Reading instruction, to be effective, must be provided on an individual basis" (8, p. 48).

Charters and Jones (17) have suggested that testing for program effect via student outcomes, e.g. achievement test scores, necessitates knowledge of whether or not a program has been implemented. Since IGE notions were implemented into the reading curriculum, it followed that the existence and the effects of IGE precepts would most likely be evident in the reading curriculum.

Finally, by virtue of the standardized test program of the Ames and Indianola school districts and the amount of IGE implementation it was judged that reading at the primary and lower elementary grades would best reflect IGE concept effect upon achievement.

IGE concept implementation rates per building and LCs within building were calculated from scores given by the expert member of an observation team during the field test of _A Survey_. This variable was selected to test the notion that schools which highly utilize IGE concepts such as individualization, team teaching, peer tutoring, etc. tend to affect student achievement differently than those schools which lowly espouse IGE concepts. It should be noted that buildings were randomly assigned letters (A, B, C, D and E).

In the Ames school district school B and school C implemented the IGE program in 1972. In the Indianola school district all schools implemented the IGE program but in different years: school A in 1973; school B in 1972; school C in 1975; and school D in 1974.

Cohort differences within and among buildings per district were intended to monitor the effects of the IGE program across time. If the program was effective, then the rate of cohort achievement within a building would increase over the duration of the IGE program. Testing the cohort variable across time among buildings within districts would reveal possible program effects and alert the investigator to across district phenomena such as sudden decreases or increases in test scores for all participating schools within a district.

Generally, female students tend to score higher than males on reading achievement tests. If the IGE concepts of non-gradedness and individualization enhance student motivation, then one may expect that male and female achievement scores in high IGE implementing schools would be greater than those of lower IGE implementing schools.

One of the principal advantages to either longitudinal or cohort designs is the ability to observe the effect of a program over time. Hence, the comparison of student outcomes before and after the implementation of any program should be of great interest. However, it must be realized that this variable is only valid if generation effects are minimal and is only one of many variables to be taken under consideration in a program's evaluation.

Finally, it was considered that other variables such as socio-economic status could influence student achievement as much as a school program. Hence, buildings within districts were ranked by experts relative to the socio-economic status of their local geographic area (see Table 5 and Table 6). A building which was highest in socio-economic status was ranked one. Second highest school was ranked two, etc. Building ranks would be particularly informative when contrasted with student achievement scores within district as effected by treatment.

Table 5. District expert ranking of schools relative to neighborhood socio-economic status for the Ames school district

Building	Rank
A	1
B	4
C	2.5
D	5
E	2.5

Table 6. District expert ranking of schools relative to neighborhood socio-economic status for the Indianola school distict

Building	Rank
A	2.5
B	4
C	1
D	2.5

D. Design and Statistical Procedure

A cohort design was used as the sampling design. The format of the design required groups (cohorts) to be defined for each year interval and at least two measures to be taken of each group. This particular design utilized a repeated subject mode whereby students were tracked across grades.

One consequence of an investigation with a wide range of year intervals, for example 1910 to 1970, is generation effect. This effect may be a debit in most cohort and longitudinal investigations. However, generation effect was considered negligible in this study because of its narrow scope, that is first grade students of Indianola in 1971 were relatively equal to first grade students of Indianola in 1975 with regard to their socio-economic environment relative to building. However, it was recognized that there were socio-economic differences among the areas served by the various elementary schools within a district.

Multiple regression (MR) was used as the statistical technique to analyze the data. For each school district two different MR equations were fitted. One MR assumed that students were independent observational units, that is treatment was unique to each student. This assumption was predicated on the notion of individualized instruction. However, individualized instruction was not the only style of teaching.

Classroom instruction and small group activity were also prevalent. In these cases, the notion of a student as an independent measure was considered tenuous. Hence, the second MR utilized the cohort by grade by sex within building average as the observational unit, hereafter referred to as the mean observational unit. It should be noted that neither MR was absolutely correct.

The independence of an observation cannot be judged solely on individual beings. The effects of environment must also be considered. In other words, a subject's response to a stimulus may be affected not only by treatment, say teaching mode, but also by interaction with his fellow students. Hence, students in a classroom environment would have higher correlated scores than students in an isolated instructional environment.

The statistical model for the case where students were the observational units was a non-orthogonal repeated measures (split-plot) design with the between plot composed of school (treatment), sex, cohort and their subsequent interactions. The split plot or within portion of the design was composed of grade and interactions with the whole plot variables. In order to compute the appropriate partial sums of squares and calculate the error terms for the whole and split plots two regression models were fitted. For the whole plot the dependent variable (reading score) was summed across the repeated measure (grade) and was regressed onto an unbalanced X-matrix (Model 1.1). The statistics for the design were computed by regressing the independent variable onto a completely crossed factorial composed of all the independent variables (Model 1.2). Using these designs on non-orthogonal data with no missing repeated observations in the within plot a split-plot design was constructed.

The error term for the split plot or within subjects portion of the design was computed as follows:

SS Error (Within) = SS Error (Four way Factorial) -
(SS Error (Between)/ # of Repeated Measures)

Since school was considered as a random variable, Expected Mean Squares for each variable were formulated (48, pp. 208-212). Subsequent F-tests were made under the assumption of an orthogonal design. According to Overall, Spiegel and Cohen (69) the F-tests were reasonably approximate when the number of observations was large.

When the sex within grade within cohort within building means were used as the observational units the split-plot design was both balanced and orthogonal. In this case, however, the interaction among school, sex and cohort was the whole plot error term, and the interaction among school, sex, cohort and grade was the split plot error term (Model 2).

Model 1.1:

$$Y_{ijkm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \xi_{ijkm}$$

α = school; β = sex; and γ = cohort

Model 1.2:

$$Y_{ijkln} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \delta_l + \alpha\delta_{il} + \beta\delta_{jl} + \alpha\beta\delta_{ijl} + \gamma\delta_{kl} + \alpha\gamma\delta_{ikl} + \beta\gamma\delta_{jkl} + \alpha\beta\gamma\delta_{ijkl} + \xi_{ijkln}$$

α = school; β = sex; γ = cohort; and δ = grade

Model 2:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \delta_l + \alpha\delta_{il} + \beta\delta_{jl} + \alpha\beta\delta_{ijl} + \gamma\delta_{kl} + \alpha\gamma\delta_{jkl} + \beta\gamma\delta_{jkl} + \xi_{ijkl}$$

α = school; β = sex; γ = cohort; and δ = grade

This study was also concerned with the differences between expert and non-expert or amateur raters' scores on _A Survey_ in the Indianola and the Ames school districts. To investigate these differnces a balanced incomplete block design was used for the analysis of Indianola's data, and a partially balanced incomplete block design was used for the analysis of Ames' data. The basic model of the incomplete block design was as follows;

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ijk}$$

α = rater; and β = school

In order to analyze the data via regression, an X-matrix for each district's statistical design was composed (Appendix C). Through the Statistical Analysis System 76 (SAS76) (5), a computer package, statistics were computed for each design. The procedures employed to compute the statistics differed for each design. For the balanced design only one regression was required. However, due to the confounding of rater and building in the partially balanced design several regressions were necessary.

In separating out the various sources of variation for the partially balanced incomplete block design four regressions of the X-matrix were necessary. The first regression fitted the entire X-matrix which included a column differentiating between experts and amateurs, several columns accounting for differences among expert raters and differences among amateur raters, and several columns accounting for school differences. The second fit excluded the column which differentiated between expert and amateur raters. The third fit ignored those columns which corresponded to differences among raters of the same type. Finally, the fourth regression excluded those columns which related to school differences. The sums of squares for the sources of variation, Expert versus Amateur (EA), Raters within Expert and Amateur (R/EA) and Schools (S), were computed by the subtractions of the Multiple R Squared (R^2) for the various models (Table 7).

Table 7. Computation of the Sums of Squares for the Partially Balanced Incomplete Block design through the subtraction of appropriate Multiple R^2s

Sources of Variation	Sums of Squares
EA	R^2(First Fit) - R^2(Second Fit)
R/EA	R^2(First Fit) - R^2(Third Fit)
S	R^2(First Fit) - R^2(Fourth Fit)

Finally, all statistical tests were made at either the .05 level or the .01 level and were respectively referred to as significant or highly significant.

IV. FINDINGS

Three different statistical techniques were used to analyze the data from each school district in this investigation. The first analysis employed a split-plot design with students as observational units. Likewise, the second analysis utilized a split-plot design, but cell means, each reflecting the average reading score of a particular cohort within a school for a certain grade and sex, were used as the observational units. Finally, an analysis between expert and amateur raters was computed through an incomplete block design.

A. Split-Plot Analyses of the Ames School District's Data

The major concerns of this investigation were to analyze elementary reading as a function of the duration of the IGE program and of the school which reflected an amount of IGE concept implementation as measured by A Survey. In addition, a concurrent concern was to analyze how a school's IGEness influenced elementary reading scores as a function of a child's sex.

Employing the computational capabilities of the Statistical Package for the Social Sciences (SPSS) (65) and SAS76 the split-plot statistics were computed for the analysis with students as observational units (Table 8). Through the split-plot ANOVA the following hypotheses were tested to determine the effect of IGE and IGEness.

> Hypothesis 1: There is no significant difference for the school by cohort interaction in the Ames school district on elementary reading test scores with students as observational units.
>
> Hypothesis 2: There is no significant difference for the school by cohort by sex interaction in the Ames school district on elementary reading test scores with students as observational units.
>
> Hypothesis 3: There is no significant difference for the school by cohort by grade interaction in the Ames school district on elementary reading test scores with students as observational units.
>
> Hypothesis 4: There is no significant difference for the school by cohort by sex by grade interaction in the Ames school district on elementary reading test scores with students as observational units.

The analysis revealed the main effects were highly significant. Means of the main effects were reported in Table 9. However, results of the tests for hypothesis 1 through hypothesis 4 failed to reject the null hypotheses, that is no significant differences were found relative to the interactions of school by cohort, of school by cohort by sex, of school by cohort by grade and of school by cohort by sex by grade.

Table 8. ANOVA for the Ames school district with students as observational units

Sources of Variation	Df	Mean Squares	F-Test	F-Value
School(S)	4	8481.26	MS(S)/MS(BE)	8.77**
Sex(X)	1	17878.14	MS(X)/MS(SX)	37.74**
SX	4	473.71	MS(SX)/MS(BE)	0.49
Cohort(C)	4	13041.55	MS(C)/MS(SC)	11.42**
SC	16	1142.30	MS(SC)/MS(BE)	1.18
XC	4	523.99	MS(XC)/MS(BE)	0.54
SXC	16	814.04	MS(SXC)/MS(BE)	0.84
Between Error(BE)	825	967.56		
Grade(G)	2	3875.31	MS(G)/MS(SG)	60.20**
SG	8	64.37	MS(SG)/MS(WE)	1.30
XG	2	13.05	MS(XG)/MS(SXG)	0.32
SXG	8	41.07	MS(SXG)/MS(WE)	0.83
CG	16	72.71	MS(CG)/MS(SCG)	1.29
SCG	32	56.02	MS(SCG)/MS(WE)	1.13
XCG	8	22.91	MS(XCG)/MS(SXCG)	0.33
SXCG	32	70.03	MS(SXCG)/MS(WE)	1.42
Within Error(WE)	1650	49.47		

** Denotes significance beyond the .01 level

Table 9. Means of statistically significant variables from Ames' ANOVA where students were the observational units

School	A	B	C	D	E
Amount of IGEness	23%	19%	27%	11%	19%
	3.85	3.35	3.54	3.25	3.44
Cohort	1	2	3	4	5
	3.40	3.24	3.38	3.62	3.98
Sex		Male		Female	
		3.37		3.69	
Grade	Second		Third		Fourth
	2.34		3.53		4.71

The second analysis used cell means as observational units and through it the following hypotheses were tested.

> Hypothesis 5: There is no significant difference for the school by cohort interaction in the Ames school district on elementary reading test scores with means as observational units.

> Hypothesis 6: There is no significant difference for the school by cohort by sex interaction in the Ames school district on elementary reading test scores with means as observational units.

> Hypothesis 7: There is no significant difference for the school by cohort by grade interaction in the Ames school district on elementary reading test scores with means as observational units.

> Hypothesis 8: There is no significant difference for the school by cohort by sex by grade interaction in the Ames school district on elementary reading test scores with means as observational units.

The second analysis found that the main effects were highly significant (Table 10). In addition, the school by grade, sex by grade and cohort by grade interactions were significant. Again, means for both the main effects and interactions are reported in a subsequent table (Table 11). However, evidence was not sufficient to reject null hypotheses 5 through 8. Hence, implementation of the IGE program had not changed student reading scores as a function of duration.

Table 10. ANOVA for the Ames school district with means as observational units

Sources of Variation	Df	Mean Squares	F-Test	F-Value
School(S)	4	140.31	MS(S)/MS(BE)	6.91**
Sex(X)	1	309.36	MS(X)/MS(SX)	52.52**
SX	4	5.89	MS(SX)/MS(BE)	0.29
Cohort(C)	4	265.87	MS(C)/MS(SC)	11.56**
SC	16	23.00	MS(SC)/MS(BE)	1.13
XC	4	9.91	MS(XC)/MS(BE)	0.49
Between Error(BE)	16	20.32		
Grade(G)	2	6824.34	MS(G)/MS(SG)	460.17**
SG	8	14.83	MS(SG)/MS(WE)	3.01*
XG	2	2.62	MS(XG)/MS(SXG)	3.64*
SXG	8	0.72	MS(SXG)/MS(WE)	0.15
CG	16	25.45	MS(CG)/MS(SCG)	10.22**
SCG	32	2.49	MS(SCG)/MS(WE)	0.51
XCG	8	1.31	MS(XCG)/MS(WE)	0.27
Within Error(WE)	32	4.93		

* Denotes significance beyond the .05 level
** Denotes significance beyond the .01 level

Table 11. Means of statistically significant variables from Ames ANOVA where means were the observational units

School	A	B	C	D	E
Amount of IGEness	23%	19%	27%	11%	19%
	3.84	3.33	3.53	3.30	3.44

Cohort	1	2	3	4	5
	3.34	3.23	3.31	3.59	3.96

Grade	Second	Third	Fourth
	2.32	3.49	4.65

Sex	Male	Female
	3.34	3.63

		Grade Second	Third	Fourth
School	School A	2.51	3.82	5.19
	School B	2.16	3.35	4.48
	School C	2.31	3.59	4.69
	School D	2.29	3.24	4.38
	School E	2.33	3.45	4.52

		Grade Second	Third	Fourth
Cohort	Cohort 1	2.27	3.18	4.57
	Cohort 2	2.17	3.19	4.33
	Cohort 3	2.20	3.17	4.57
	Cohort 4	2.23	3.74	4.80
	Cohort 5	2.72	4.16	5.06

		Grade Second	Third	Fourth
Sex	Male	2.20	3.34	4.49
	Female	2.44	3.64	4.82

B. Split-Plot Analyses of the Indianola School District's Data

Paralleling the analyses of the Ames data, two split-plot analyses were computed. The first analysis used students as observational units and was used to test hypotheses 9 through 12.

Hypothesis 9: There is no significant difference for the school by cohort interaction in the Indianola school district on elementary reading test scores with students as observational units.

Hypothesis 10: There is no significant difference for the school by cohort by sex interaction in the Indianola school district on elementary reading test scores with students as observational units.

Hypothesis 11: There is no significant difference for the school by cohort by grade interaction in the Indianola school district on elementary reading test scores with students as observational units.

Hypothesis 12: There is no significant difference for the school by cohort by sex by grade interaction in the Indianola school district on elementary reading test scores with students as observational units.

The analysis found the main effects, school, sex and grade highly significant. Also, school by cohort and sex by cohort interactions were found significant (Table 12). Means for the significant variables were reported in Table 13.

Table 12. ANOVA for the Indianola school district with students as observational units

Sources of Variation	Df	Mean Squares	F-Test	F-Value
School(S)	3	1920.80	MS(S)/MS(BE)	7.19**
Sex(X)	1	5727.76	MS(X)/MS(SX)	25.46**
SX	3	225.00	MS(SX)/MS(BE)	0.84
Cohort(C)	4	490.31	MS(C)/MS(SC)	0.61
SC	12	806.36	MS(SC)/MS(BE)	3.02**
XC	4	785.80	MS(XC)/MS(BE)	2.94*
SXC	12	451.01	MS(SXC)/MS(BE)	1.69
Between Error(BE)	711	266.99		
Grade(G)	1	1183.19	MS(G)/MS(SG)	65.95**
SG	3	17.94	MS(SG)/MS(WE)	0.56
XG	1	58.10	MS(XG)/MS(SXG)	1.79
SXG	3	32.29	MS(SXG)/MS(WE)	1.01
CG	4	61.86	MS(CG)/MS(SCG)	1.19
SCG	12	25.00	MS(SCG)/MS(WE)	1.63
XCG	4	22.72	MS(XCG)/MS(SXCG)	0.62
SXCG	12	36.24	MS(SXCG)/MS(WE)	1.13
Within Error(WE)	711	31.97		

* Denotes significance beyond the .05 level
** Denotes significance beyond the .01 level

Tests constructed of hypotheses 10, 11 and 12 were not significant. However, result of the test for hypothesis 9 was highly significant, that is schools and cohorts interacted. Figure 3 best illustrates this complex interaction.

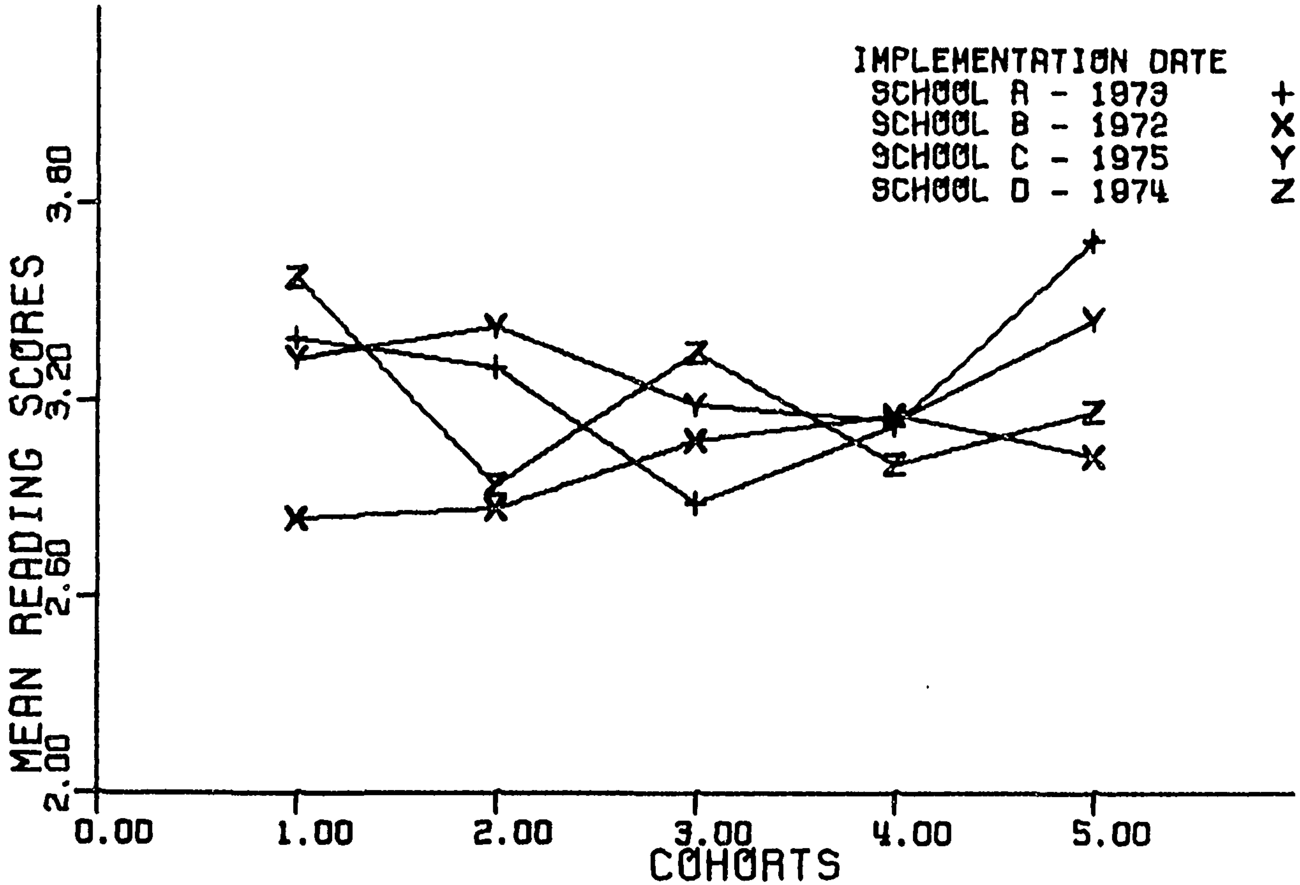

Figure 3. School by cohort interaction from Indianola's ANOVA with students as observational units

Table 13. Means of statistically significant variables from Indianola's ANOVA where students were the observational units

School	A	B	C	D
Amount of IGEness	24%	28%	20%	6%
	3.26	2.98	3.30	3.26

Sex	Male	Female
	3.08	3.33

Grade	First	Second
	2.57	3.84

	Cohort				
	1	2	3	4	5
School A	3.39	3.33	2.89	3.13	3.69
School B	2.84	2.87	3.08	3.16	3.03
School C	3.33	3.43	3.19	3.14	3.45
School D	3.58	2.94	3.35	3.01	3.17

	Cohort				
	1	2	3	4	5
Male	3.17	3.03	2.90	3.04	3.30
Female	3.35	3.28	3.48	3.18	3.35

As with the Ames data a split-plot analysis was computed for the Indianola data with cell means as observational units. It was used to test hypotheses 13 through 16.

Hypothesis 13: There is no significant difference for the school by cohort interaction in the Indianola school district on elementary reading test scores with means as observational units.

Hypothesis 14: There is no significant difference for the school by cohort by sex interaction in the Indianola school district on elementary reading test scores with means as observational units.

Hypothesis 15: There is no significant difference for the school by cohort by grade interaction in the Indianola school district on elementary reading test scores with means as observational units.

Hypothesis 16: There is no significant difference for the school by cohort by sex by grade interaction in the Indianola school district on elementary reading test scores with means as observational units.

The second analysis found the main effects, school, sex and grade, highly significant (Table 14). Means for these variables were reported in Table 15. However, tests for hypothesis 13 through hypothesis 16 failed to reject the null hypotheses.

Besides finding the main effects significant it was observed that the sex by grade interaction was highly significant. Means for this interaction are reported on Table 15. However, it should be noted that the denominator for the sex by grade interaction F-test was very small which artifically inflated the F-value and resulted in a highly significant interaction. The significance of this interaction was not surprising nor was it of great import.

Table 14. ANOVA for the Indianola school district with means as observational units

Sources of Variation	Df	Mean Squares	F-Test	F-Value
School(S)	3	48.24	MS(S)/MS(BE)	3.51*
Sex(X)	1	155.71	MS(X)/MS(SX)	26.75**
SX	3	5.82	MS(SX)/MS(BE)	0.42
Cohort(C)	4	14.43	MS(C)/MS(SC)	0.68
SC	12	21.11	MS(SC)/MS(BE)	1.54
XC	4	23.01	MS(XC)/MS(BE)	1.67
Between Error(BE)	12	13.74		
Grade(G)	1	3282.85	MS(G)/MS(SG)	713.10**
SG	3	1.21	MS(SG)/MS(WE)	0.57
XG	1	7.28	MS(XG)/MS(SXG)	91.00**
SXG	3	.08	MS(SXG)/MS(WE)	.04
CG	4	1.98	MS(CG)/MS(SCG)	0.70
SCG	12	2.82	MS(SCG)/MS(WE)	1.31
XCG	4	0.26	MS(XCG)/MS(WE)	0.12
Within Error(WE)	12	2.15		

* Denotes significance beyond the .05 level
** Denotes significance beyond the .01 level

Table 15. Means of statistically significant variables from Indianola's ANOVA where means were the observational units

School	A	B	C	D
Amount of IGEness	24%	28%	20%	6%
	3.27	2.99	3.35	3.22

Sex	Male	Female
	3.07	3.35

Grade	First	Second
	2.57	3.85

		Grade	
		First	Second
Sex	Male	2.46	3.68
	Female	2.68	4.02

C. Analyses between Expert and Amateur Ratings on Several Scales of _A Survey_ in the Ames School District

The second purpose of the investigation was to compare expert and amateur raters on several scales of _A Survey_ which measured the amount of IGE implementation. A partially balanced incomplete block design was selected as the appropriate analysis of variance model to test hypothesis 17 through hypothesis 21.

Hypothesis 17: There is no significant difference among expert and amateur raters' scores on the Overall scale of A Survey of Effective School Processes for Learning Community 1 in the Ames school district.

Hypothesis 18: There is no significant difference among expert and amateur raters' scores on the Commitment scale of A Survey of Effective School Processes for Learning Community 1 in the Ames school district.

Hypothesis 19: There is no significant difference among expert and amateur raters' scores on the Structure scale of A Survey of Effective School Processes for Learning Community 1 in the Ames school district.

Hypothesis 20: There is no significant difference among expert and amateur raters' scores on the Teacher role scale of A Survey of Effective School Processes for Learning Community 1 in the Ames school district.

Hypothesis 21: There is no significant difference among expert and amateur raters' scores on the Student Role scale of A Survey of Effective School Processes for Learning Community 1 in the Ames school district.

Evidence was not sufficient to reject null hypotheses 17, 18, 20 and 21. However, the test for hypothesis 19 was highly significant (Table 16). Expert raters differed significantly from amateur raters on the Structure scale of A Survey. Means for the raters were reported in Table 17.

Table 16. ANOVA of expert versus amateur ratings in the Ames school district on the Structure scale of A Survey

Sources of Variation	Df	Mean Squares	F-Test	F-Value
Rater(R)	5	0.012	MS(R)/MS(E)	0.63
Expert vs Amateur(EA)	1	0.040	MS(EA)/MS(TR)	53.33**
Type/Rater(TR)	4	0.0007		
School(S)	4	0.230	MS(S)/MS(E)	12.11
Error	2	0.019		

** Denotes significance beyond the .01 level

Table 17. Means of expert versus amateur raters from Ames' ANOVA on the Structure scale of A Survey

	Expert	Amateur
Raters	8.5	11.3

D. Analyses between Expert and Amateur Ratings on Several Scales of A Survey in the Indianola School District

Using a balanced incomplete block design, expert and amateur raters were compared across several scales of A Survey. Hypothesis 22 through hypothesis 26 were formulated to test for differences between rater types on each scale.

> Hypothesis 22: There is no significant difference among expert and amateur raters' scores on the Overall scale of A Survey of Effective School Processes for Learning Community 1 in the Indianola school district.

> Hypothesis 23: There is no significant difference among expert and amateur raters' scores on the Commitment scale of A Survey of Effective School Processes for Learning Community 1 in the Indianola school district.

> Hypothesis 24: There is no significant difference among expert and amateur raters' scores on the Structure scale of A Survey of Effective School Processes for Learning Community 1 in the Indianola school district.

> Hypothesis 25: There is no significant difference among expert and amateur raters' scores on the Teacher role scale of A Survey of Effective School Processes for Learning Community 1 in the Indianola school district.

> Hypothesis 26: There is no significant difference among expert and amateur raters' scores on the Student Role scale of A Survey of Effective School Processes for Learning Community 1 in the Indianola school district.

All of the null hypotheses were not rejected. However, a comparison among raters ignoring type, that is, expert and amateur, found a significant difference at the .05 level (Table 18). Means for each rater were reported in Table 19.

Table 18. ANOVA of expert versus amateur ratings in the Indianola school district on the Student Role scale of A Survey

Sources of Variation	Df	Mean Squares	F-Test	F-Value
Rater (R)	3	448.17	MS(R)/MS(E)	35.85*
Expert vs Amateur (EA)	1	364.50	MS(EA)/MS(TR)	0.74
Type/Rater (TR)	2	490.00		
School (S)	3	810.83	MS(S)/MS(E)	64.87*
Error	1	12.50		

* Denotes significance beyond the .05 level

Table 19. Means of statistically significant variables from Indianola's ANOVA among raters on the Student Role scale of A Survey

Raters	1	2	3	4
	15.0	60.5	24.0	5.5
School	A	B	C	D
	46.0	40.5	14.0	4.5

V. DISCUSSION

A. Cohort Analysis of Elementary Reading Scores in the Ames School District

Using a cohort analysis in conjunction with measurements of IGE concept implementation at the elementary level the principal objectives of this investigation were satisfied. The statistical analyses of the Ames data by two split-plot designs provided the investigator statistical estimates for concluding remarks regarding IGE's effect on reading over time and IGEness.

Both analyses revealed that the main effects, school, sex, cohort and grade, were highly significant. However, while school effect was found statistically significant, a comparison of school means and amounts of IGEness revealed that the school having the highest IGEness score did not have the highest reading score (see Table 9 and Table 11), whereas a comparison between school means and school ranks showed a closer relationship (see Table 5, Table 9 and Table 11).

Females scored higher than males, and cohort group differences tended to reveal a positive upward trend. Neither of these results was surprising. Females generally score higher than males on standardized reading tests. A positive cohort trend may have resulted from local factors and test changes.

The second analysis which used means as observational units found the school by grade interaction and the cohort by grade interaction significant. An examination of the school by grade means in Table 11 and the plot of the respective means on Figure 4 (see Appendix D) revealed that the pattern among grades in school D resulted in the significant interaction. The first grade of school D appeared equally competitive to school C and school E. However, in succeeding grades both school C and school E scored higher than school D.

An examination of the cohort by grade interaction means on Table 11 and a subsequent plot of means on Figure 5 (see Appendix D) revealed that all grades in cohort four and cohort five experienced an increase in mean reading scores. However, the school by cohort by grade interaction was not statistically significant, which meant that increases in reading scores were multilateral across schools.

In summary, the Ames analyses revealed that while the main effects were significant they were probably influenced more by local conditions than either the implementation of IGE or the amount of IGE concept implementation. This notion was supported by the lack of significance for the school by cohort interaction and other interactions of school by cohort with sex and grade main effects. However, it should be noted that IGEness was not high among the cooperating schools and was only measured during the last period of this investigation, 1976. Relative equality of IGEness among the Ames schools could be attributed to the organization of the school district above and beyond that of IGE. For example, the Ames district has both vertical and horizontal curriculum committees.

B. Cohort Analysis of Elementary Reading Scores in the Indianola School District

A parallel study was conducted in the Indianola school district. Again, a cohort analysis in conjunction with a measure of IGE concept implementation was employed to satisfy the investigation's principal concerns. Also, two split-plot designs were used to compute the statistics.

The analyses found the main effects, school, sex and grade, highly significant. Like the Ames analyses school means did not correspond with the amount of IGEness in a linear relationship. In fact, a comparison between the school means and the amount of IGEness was considered erratic (Table 13 and Table 15). For example, school B scored the highest IGEness rating but scored lowest on the standardized reading test (Table 6).

Unlike the Ames analyses cohort groups were not significant. The lack of significance in both analyses implied that there was an overall stability across cohorts. Cohort group one which experienced no IGE scored equally as well as cohort group five which was completely immersed into the IGE program. However, the analysis with students as observational units also found the school by cohort interaction significant. This was interpreted as meaning that while the overall cohorts were about equal, schools fluctuated across cohorts. An examination of the means for the school by cohort interaction on Table 13 and a plot of the means on Figure 3 revealed that after a school implemented IGE the immediate succeeding cohort group experienced an increase in the mean reading score. This was observed for schools B, A and D. No mean reading score was available for the immediate successor cohort group in school C. It should be noted, however, that while the immediate succeeding cohorts' mean reading scores increased, maintenance and enhancement of reading scores was inconsistent among schools. For example, school B which was longest in the IGE program and had the highest IGEness score experienced a decrease in the mean reading score for cohort five, whereas school A experienced a substantive increase in its mean reading score. Possible reasons for this inconsistency may be staff turnover or a change in the local neighborhood.

The sex by cohort interaction was found statistically significant in the first Indianola analysis. Means of the interacting variables (Table 13) and a corresponding plot of means (Figure 6 in Appendix D) revealed that females tended to fluctuate in a cyclic pattern about the female overall mean. Males, however, decreased among cohorts one, two, and three but steadily increased between succeeding cohorts. It should be noted that by cohort group five all of Indianola's schools had implemented the IGE program. Hence, IGE may have had a positive effect upon male reading scores. Further, lack of significance for the school by cohort by sex interaction supports the notion that the rise in male reading scores was multilateral, that is one particular school did not bias the sex by cohort mean. The second analysis found the sex by grade interaction significant. However, it was felt that the divisor Mean Square of the F-test artificially inflated the F-value. Hence, a significant interaction was found.

Finally, Indianola's analyses, like that of Ames' analyses, found the main effects, school, sex and cohort, significant. In addition, the analysis with students' scores as observational units found the school by cohort interaction significant. An examination of the plot of interaction means revealed that when schools A, B and D implemented the IGE program their respective immediate succeeding cohort group's reading scores rose. However, as noted previously maintenance and increase of mean reading scores was inconsistent among schools. Male reading scores for cohorts four and five steadily increased as schools C and D implemented the IGE program. However, it appears that the amount of IGEness was not influential since schools C and D scored lowest in IGE-ness. Again, local conditions appeared to play a predominate role. In summary, null hypotheses 10 through 16 were not rejected. Null hypothesis 9 was rejected and its tests was highly significant.

C. Expert and Amateur Rater Differences on A Survey for LC1's in the Ames and Indianola School District

Another major concern of this investigation was the comparison of expert and amateur raters on several scales of A Survey. Raters were paired expert-amateur in each district. Pairings were arranged so that a partially balanced incomplete block design could be used to analyze the Ames data and a balanced incomplete block design could be used to analyze Indianola's data.

Of the Ames analyses experts differed from amateurs only on the Structure scale of A Survey. According to Table 17, amateurs tended to score Ames schools higher than experts. Hence, amateurs felt that the Ames school tended to conform with the IGE organizational model. It should be noted, however, that a 12% rating is not very high. In addition, the F-value was probably artificially inflated by the small Mean

Square of the F-test between expert and amateur.

No significant differences were found on any of the scales for the Indianola analyses. However, a comparison among all raters on the Student Role scale was highly significant. Rater mean scores from Table 19 indicate that rater 2 radically differed from rater 4. Whereas, rater 1 and 3 were about the same. It should be noted that rater 2 and rater 4 were amateurs. These means illustrate that while amateur raters and expert raters may average about the same individually, amateur raters may differ quite extremely from the expert raters.

In conclusion, in both analyses expert and amateur raters on the average differed little. This may be due in part by the way amateur raters were defined. Ability, experience and knowledge of educational practices and constructs were not taken into account. Hence, no real differences between amateurs and experts may have existed other than by whom they were employed. However, it should be noted that experts tended to be more homogeneous in their scoring. In summary, except for hypothesis 19 all other hypotheses, 17 through 26, were not rejected.

D. Recommendations for Further Study

One of the principal concerns of this investigation was how a school's IGEness affected student reading scores. Further, it was assumed that a high IGE concept implementing school would have a higher reading score than a low IGE concept implementing school. However, schools in both districts were not high implementing schools. Thus, school differences were probably affected more by local conditions than the IGE model or IGEness as measured by A Survey. The present study, while approaching the problem correctly by considering IGEness, could not conclude that students in high implementing schools would not score higher reading scores than students in low implementing schools. In light of this a study which consists of low concept implementing schools and high concept implementing schools is recommended.

For this study IGEness was measured only during the last year of the investigation period, 1976. This was considered a shortcoming of this study. No previous history of the schools' IGEness was available. An average measure of IGEness would have better reflected the amount of IGE concept implementation than any single measure. Hence, it is recommended that future long term studies measure participating schools at regular intervals, say once a year.

How IGE or IGEness affects different academic achievement strata was not investigatied in this study. It is recommended that succeeding studies consider the aspect of different academic levels. However, it should be noted that academic strata would be relative to the investigator's sample. In other words, a classification of low academic stratum in Ames may not be appropriate for New York. Ames' low academic stratum may be New York's middle stratum.

Finally, the present study and most studies of alternative programs are confined to school systems which allocate students to schools solely upon geographic location. A school, therefore, is composed of a collection of children with varied learning styles and of a heterogeneous body of parents with different notions concerning education. Such a conglomeration of styles and notions may, in part, explain why certain programs fail. Parents, the silent partners of education, may not support a particular program. Lack of support by the community usually results in a change or elimination of a program. In light of this, it is recommended that IGE's effect upon student outcomes, such as achievement scores, be studied within a non-traditional student allocation environment.

One way to improve the allocation practices might be through implementing a limited voucher system based upon the latent trait structure analysis (50) of questionnaires relegated to parents. In other words, based upon the response patterns of the parents several schools which espouse compatible philosophies may be suggested to parents as potential schools for their children. In this way parents through their own acclamation should support the school of their choice. Parent involvement through home-school communication is a vital cog in the IGE program (81).

VI. SUMMARY

A. Purpose

The purpose of this investigation was to study the effects of IGE implementation and the amount of IGE concept implementation on reading scores in two local school districts. In addition, expert and amateur raters were compared as to how they scored several scales of A Survey.

B. Methodology

An experimental design referred to as a cohort design was used to study the effect of IGE implementation over time. A different experimental design was employed for each district. In addition, districts used different standardized achievement tests. Hence, no comparisons between districts were made.

For both districts, two split-plot statistical designs were used to analyze the reading scores. The analyses within each district differed with regard to the observational units. In one split-plot analysis student reading scores were used as observational units, whereas in the other split-plot analysis, mean reading scores were employed as observational units.

In the selection of students for observation across grade levels, a criterion of no missing reading achievement scores was enforced, that is, only students with no missing reading scores were used as observational units. In this way, students exposed to only one learning environment were selected. In addition, the orthogonality between the whole plot and the split plot of the analyses was maintained.

Finally, the amount of IGE concept implementation was obtained from the field test of A Survey. The scores of expert raters were employed as reflecting the amount of IGE concept implementation of each cooperating school. Further, the field test of a A Survey supplied data for comparisons between expert and amateur raters.

C. Findings

1. Analyses of Ames' data

The results of the split-plot analyses found no evidence to reject null hypotheses 1 through 8. The implementation of IGE had not affected student reading as measured by the Stanford Achievement Test. Local and test scale effects tended to play a dominant role. This was evidenced by comparisons between school effect and expert ranking, the school by grade interaction and the cohort by grade interaction.

IGEness was relatively small and varied little among the Ames schools. Finally, with regard to reading achievement the IGE program or its concept implementation was judged ineffective as a treatment in the Ames cooperating schools. However, as noted previously, the small amount of IGEness and its single measurement during the 1975-76 school year makes the results less conclusive with regard to IGEness effect upon reading achievement.

2. Analyses of Indianola's data

Unlike the Ames analyses, Indianola's analysis with students as observational units found the school by cohort interaction significant. A careful study of the interaction revealed that when a school implemented IGE the reading score of the immediate cohort rose. This finding was consistent for each school with a succeeding cohort. Unfortunately, the school by cohort interaction was not significant when means were used as observational units. Lack of significance diminished the finding of the former analysis. However, the plot of the school by cohort means from the former analysis did reveal a consistent and a substantive pattern of increased reading scores. Unfortunately, maintenance or enhancement of the reading scores was inconsistent among schools. In addition, the sex by cohort interaction was significant when students were observational units. A study of the plot for the sex by cohort interaction means revealed that as all of Indianola's schools adopted the IGE program the male mean reading scores increased, almost equaling the female mean reading scores.

3. Expert and amateur raters

On the average there was no difference between expert and amateur raters in either district. Differences appeared to occur on the Structure scale of A Survey in the Ames school district, but a close study revealed evidence of an inflated F-value. Also, the means were substantively too close for the differences to be meaningful.

It was observed that while on the average expert and amateur raters differed little. Amateur raters tended to be more divergent than their expert counterparts. This was best illustrated by the Student Role individual rater scores from the Indianola school district. Expert raters were more homogeneous in their scoring.

D. Conclusions

No conclusive evidence regarding the effects of IGE implementation or of IGEness was found by this investigation. In one case, Ames, all the null hypotheses were not rejected. For Ames, the IGE program had not affected the reading scores of students. However, in Indianola IGE implementation was beneficial. Student reading scores rose for at least the succeeding cohort, and as all of the Indianola schools implemented the IGE program male reading scores rose.

The amount of IGEness was small in both districts and its potential effects were never really tested.

In both districts the amount of IGE concept implementation was about the same. Yet, although in the Ames district no significant differences were found, in the Indianola district a school by cohort interaction was found significant. This leads one to suspect that the program itself may be ineffective with regard to student reading scores. Recall that Jones, Moore and Van Devender found a similar phenomenon in their study of non-gradedness. Finally, the enthusiasm a new program spawns may itself be that latent intervening variable which causes significant differences in educational research in student achievement.

VII. BIBLIOGRAPHY

1. Amendola, A. A. "Changes in Attitudes and Achievement Effected by a Continuous Progress Education Program at the Elementary School Level." Dissertation Abstracts International 33A (9) (1973): 4702.

2. Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J., and Pincus, J. How Effective Is Schooling: A Critical Review of Research. Englewood Cliffs, New Jersey: Educational Technology, 1974.

3. Bailey, D. J. "The Effects of Tutoring First, Second and Third Grade Tutees on the Academic Achievement Potential and Self-concept of the Seventh Grade Tutors." Dissertation Abstracts International 33A (7) (1972): 3248.

4. Baltes, P. B. "Longitudinal and Cross-Sectional Sequencies in the study of age and generation effects." Human Development 11 (3) (1968): 145-171.

5. Barr, A. J., Goodnight, J. H., Sall, J. P. and Helwig, J. T. A User's Guide to SAS76. Raleigh, North Carolina: SAS Institute, Inc. 1976.

6. Beasley, K. L. "An Investigation of the Effect of Team Teaching Upon Achievement and Attitude in U.S. History Classes." Dissertation Abstracts International 35 (5) (1974): 2574.

7. Bell, R. Q. "Convergence: An Accelerated Longitudinal Approach." Child Development 24 (2) (June, 1953): 145-152.

8. Bond, G. L. and Tinker, M. A. Reading Difficulties: their diagnosis and correction. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

9. Bowman, B. L. "A Comparison of Pupil Achievement and Attitude in a Graded School with Pupil Achievement and Attitude in a Non-graded School, 1968-69, 1969-70 School Years." Dissertation Abstracts International 32A (2) (1971): 660.

10. Brody, E. B. "Achievement of First and Second Year Pupils in Graded and Non-graded Classrooms." Elementary School Journal 70 (7) (April, 1970): 391-394.

11. Brown, E. K. and Theimer, W. C. "An Evaluation of the Ungraded Program at Powell Elementary School." ERIC ED 029 368, 1968.

12. Brown, J. C. "Effects of Token Reinforcement Administered by Peer Tutors on Pupil Reading Achievement and Tutor Collateral Behavior." Dissertation Abstracts International 32A (7) (1971): 3775.

13. Buffie, E. G. "Comparison of Mental Health and Academic Achievement: the Nongraded School vs. the Graded School." Dissertation Abstracts 23A (1) (1962): 4225.

14. Burchyett, J. A. "A Comparison of the Effect on Nongraded, Multiage Team Teaching versus the Modified Self-contained Classroom at the Elementary School Level." Dissertation Abstracts International 33 (11) (1972): 5998.

15. Buss, A. R. "An Extension of Developmental Models that Separate Ontogenetic Changes and Cohort Differences." _Psychological Bulletin_ 80 (6) (December, 1973): 466-479.

16. Carbone, R. F. "A Comparison of Graded and Non-graded Elementary Schools." _Elementary School Journal_ 62 (2) (November, 1961): 82-88.

17. Charters, W. W. and Jones, J. E. "On Neglect of the Independent Variable in Program Evaluation." _Project MITT Occasional Paper_, February, 1974.

18. Cols, C. W. _Quotable Quotes on Education._ Edited by Kerger, A. Detroit: Wayne State University Press, 1968.

19. Cooper, D. H. and Stan, H. N. "Team Teaching: Student Adjustment and Achievement." _Journal of Educational Research_ 66 (7) (March, 1973): 323-327.

20. Cronbach, L. J. and Furby, L. "How We Should Measure "Change" of Should We?" _Psychological Bulletin_ 74 (1) (January, 1970): 68-80.

21. Davis, F. W. and Lucas, J. S. "An Experiment in Individualized Reading." _The Reading Teacher_ 24 (8) (May, 1971): 737-743.

22. Doyle, R. L. Effectiveness of Individually Guided Education Schools as Measured by Indicators of Quality. Unpublished Ph.D. Dissertation, Iowa State University, February 1976.

23. Evans, C. L. "A Comparison of Continuous Progress and Traditional Classes in 3 Middle Schools." ERIC ED 096 334, 1973.

24. Evans, W. M. "Cohort Analysis of Survey Data: A Procedure for Studying Long-term Opinion Change." Milbank Memorial Fund Quarterly 23 (1) (Spring, 1959): 63-72.

25. Floyd, D. B. "A Comparison of Pupil Achievement in a Team Teaching Program and a Traditional Program in the Lawrence Elementary School Program in Indianola, Missippi." Dissertation Abstracts International 30A (11) (1969): 4728.

26. Frager, S. R. "Achievement of Elementary School Children in a Tutor-tutee Relationship." Dissertation Abstracts International 31A (6) (1969): 675.

27. Frager, S. and Stern, C. "Learning by Teaching." The Reading Teacher 23 (5) (February, 1970): 403-405.

28. Fraenkel, J. R. and Gross, R. E. "Team Teaching: A Note of Caution is in Order." NEA Journal 56 (4) (April, 1967): 16-17.

29. Furby, L. "Interpreting Regression Toward the Mean in Developmental Research." Developmental Psychology 8 (2) (March, 1973): 172-179.

30. Gates-MacGinitie Reading Test: Technical Manual. New York, New York: Teachers College Press, 1965.

31. Given, H. "A Comparison Study of Achievement and Attitudinal Characteristics of Black and White Intermediate Pupils in Individualized, Multi-grade and Self-contained Instructional Programs." Dissertation Abstracts International 33A (3) (1972): 893.

32. Goodlad, J. I. and Anderson, R. H. The Nongraded Elementary School. New York: Harcourt, Brace and World, 1963.

33. Goulet, L. R. "Longitudinal and Time-Lab Designs in Educational Research: An Alternative Sampling Model." Review of Educational Research 45 (4) (Fall, 1975): 505-523.

34. Gronland, N. E. Individualized-Classroom Instruction New York, New York: Macmillan Publishing Company Inc. 1974: 1-2.

35. Halliwell, J. W. "Comparison of Pupil Achievement in Graded and Non-graded Primary Classrooms." Journal of Experimental Education 32 (1) (Fall, 1963): 59-64.

36. Halvorsen, J. R. Development and Testing of an Instrument to Measure the Degree of Implementation of Individually Guided Education Process. Unpublished Ph.D. Dissertation, Iowa State University, August 1974.

37. Hatfield, D. J. "A Three Tear Study of the Effects of an Individualized Learning Center on Students' Attitudes and Achievement." Dissertation Abstracts International 31A (7) (1970): 3166.

38. Heathers, G. "Research on Implementing and Evaluating Cooperative Teaching." National Elementary Principal 44 (3) (January, 1965): 27-33.

39. Hillson, M., Jones., J. C., Moore, J. W. and Van Devender, F. "A Controlled Experiment Evaluating the Effects of a Non-graded Organization on Pupil Achievement." Journal of Educational Research 61 (2) (October, 1967): 75-77.

40. Hilton, T. L. and Patrick, C. "Cross-Sectional Versus Longitudinal Data: an Empirical Comparison of Mean Differences in Academic Growth." Journal of Educational Measurement 7 (1) (Spring, 1970): 15-24.

41. Huser, M. K. "The Effects of Individualized Reading on Achievement and Attitude." Dissertation Abstracts 26A (5) (1965): 2547.

42. Ingram, V. "Flint Evaluates Its Primary Cycle." Elementary School Journal 61 (2) (November, 1960): 76-80.

43. Institute for the Development of Educational Activities, Inc. I/D/E/A's Guide to an Improvement Program for Schools. Dayton, Ohio: I/D/E/A.

44. International Encyclopedia of the Social Sciences. Edited by Sills, D. L. New York: Macmillan Company and the Free Press, 1968.

45. Jones, H. "Problems of Method in Longitudinal Research." Vita Humana 1 (2) (1958): 93-98.

46. Jones, J. C., Moore, J. W. and Van Devender, F. "A Comparison of Pupil Achievement After One and One-half and Three Years in a Non-graded Program." Journal of Educational Research 61 (2) (October, 1967): 75-77.

47. Killough, C. K. "An Analysis of the Longitudinal Effects that a Nongraded Elementary Program, Conducted in an Open-space School, Had on the Cognitive Achievement of Pupils." Dissertation Abstracts International 32A (6) (1971): 3614.

48. Kirk, R. E. Experimental Design: Proceudues for the behavioral sciences. Belmont, California: Wadsworth Publishing Company, Inc. 1968.

49. Klausmeier, H. J., Quilling, M. R. and Sorenson, J. S. "The Development and Evaluation of the Multiunit Elementary School, 1966-1970. Report From the Project on Variables and Processes in Cognitive Learning." ERIC ED 051 589, 1971.

50. Lazarsfeld, P. E. "Latent Structure Analysis." Psychology: A Study of a Science. Edited by Koch, S. New York, New York: McGraw-Hill Book Company, Inc. 1959.

51. Lindaman, A. IGE versus Conventional Schools: Pupil Self-concept. Unpublished Ph.D. Dissertation, Iowa State University, 1976.

52. Lord, F. M. "The Measurement of Growth." Educational and Psychological Measurement 16 (3) (1956): 421-437.

53. Lord, F. M. "Further Problems in the Measurement of Growth." Educational and Psychological Measurement 18 (3) (1958): 437-451.

54. Lord, F. M. "Elementary Models for Measuring Change." Problems in Measuring Change. Edited by Harris, C. W. Madison: University of Wisconsin Press, 1973.

55. Lutenbacher, D. A. "An Investigation of the Effect of Team Teaching upon Achievement, Motivation and Attitude." Dissertation Abstracts 23A (9) (1962): 3256.

56. Mason, K. O., Mason, W. M., Winsborough, H. H. and Poole, W. K. "Some Methodological Issues in Cohort Analysis of Archival Data." American Sociological Review 38 (2) (April, 1973): 242-258.

57. McCallum, C. J. and Roark, A. E. "Children's problems in Team Teaching Schools." Elementary School Journal 74 (7) (April, 1974): 436-439.

58. Mehrens, W. A. and Lehman, I. J. Standardized Tests in Education. New York, New York: Holt, Rinehart and Winston, Inc. 1969.

59. Miller, I. "To Coordinate Existing Resources to Merit the Needs of Students at the Primary Level." ERIC ED 115 586, 1975.

60. Mitchell, J. and Zoffness, R. "Elementary Pupils Favor a Multiage Class." Education 91 (3) (February-March, 1971): 270-273.

61. Morris, V, R., Prager, B. B. and Morrell, J. E. "Pupil Achievement in a Non-graded Primary Plan After Three and Five Years of Instruction." Educational Leadership 28 (6) (March, 1971): 621-625.

62. Morrow, R. G. and Quilling, M. R. "Student Achievement and Attitudes in Instruction and Research Units in Two Elementary School in Janesville, Wisconsin, 1967-68." ERIC ED 064 657, 1973.

63. Muck, R. E. S. "The Effects of Classroom Organization on Academic Achievement in Graded and Non-graded Classes." Dissertation Abstracts 27A (9) (1966): 2923.

64. Newsweek. "Back to basics in the schools." Newsweek (October 21, 1974): pp. 87-95.

65. Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. and Bent, D. H. Statistical Package for the Social Sciences. New York, New York: McGraw-Hill Book Company, 1975.

66. Nunn, T. P. "Education and the Endangered Individual." Social Scientist at Odds. Edited by Hill, B. V. New York, New York: Teachers College Press, 1973.

67. Nunnally, J. C. "Research Strategies and Measurement Methods for Investigating Human Development." Life Span Deveopmental Psychology: Methodological Issues. Edited by Nesselroade, J. R. and Haynes, W. R. New York, New York: Academic Press, 1973.

68. Olney, G. Opinion and Goals of IGE and Non-IGE Teachers. Unpublished Ph.D. Dissertation, Iowa State University, 1976.

69. Overall, J. E., Spiegel, D. K. and Cohen, J. "Equivalence of the Orthogonial and Non-orthogonial Analysis of Variance." Psychological Bulletin 82 (2) (March, 1975): 182-186.

70. Paden, J. S. "An Interim Report on Student Achievement in 19 IGE Schools." I/D/E/A Staff Paper Affiliate of the Charles F. Kettering Foundation, Dayton, Ohio, (October, 1973).

71. Parker, C. E. "Effects of an Individualized Reading Program on Achievement in Reading." Dissertation Abstracts 26A (8) (1965): 4394.

72. Paven, B. N. "Good News: Research on the Nongraded Elementary School." Elementary School Journal 73 (6) (March, 1973): 333-342.

73. Powell, W. R. "Gates-MacGinitie Reading Test." Journal of Educational Measurement 6 (2) (Summer, 1969): 114-116.

74. Read, E. A. "Exemplory Centers for Continuous Progress Education." Curriculum Development in Nongraded Schools. Edited by Buffie, E. G. and Jenkins, J. M. Bloomington: Indiana University Press, 1971.

75. Resnich, L. B. and Wang, M. C. "Improvement of Academic Performance of Poor Progress Children Through the Use of Individualized Instruction Program." ERIC ED 101 860, 1972.

76. Rogers, M. S. "A Study of an Experimental Tutoring Reading Program in which Sixth-grade Underachievers Tutored Third-grade Children Who Were Experiencing Difficulty in Reading." Dissertation Abstracts International 30A (4) (1969): 4695.

77. Rust, S. P. "The Effect of Tutoring on the Tutor's Behavior, Academic Achievement and Social Status." Dissertation Abstracts International 31A (11) (1969): 4862.

78. Schaie, K. W. "A General Model for the Study of Developmental Problems." Psychological Bulletin 64 (2) (August, 1965): 92-107.

79. Schaver, J. P. and Nuhn, D. "The Effectiveness of Tutoring Underachievers in Reading and Writing." Journal of Educational Research 65 (3) (November, 1971): 107-112.

80. Schneiderhan, R. M. "A Comparison of an Individually Guided Education(IGE) Program, Individually Guided Instruction(IGI) and a Traditional Elementary Education Program at the Intermediate Level." Dissertation Abstracts International 34A (7) (1973): 3848.

81. Schultz, K. M. Communicating with parents. Dayton, Ohio: I/D/E/A, 1971.

82. Skapski, M. K. "Ungraded Primary Reading Program: An Objective Evaluation." Elementary School Journal 61 (1) (October, 1960): 41-45.

83. Spencer, D. U. "Individualized Versus a Basal Reader Program in Rural Communities - Grades One and Two." The Reading Teacher 21 (1) (October, 1967): 11-17.

84. Stow, S. B. A Criticial Appraisal of Indicators of Quality and A Survey of Effective School Processes. Unpublished Ph.D. Dissertation, Iowa State University, August 1976.

85. Stanford Achievement Test: Administrator Handbook. New York, New York: Harcourt Brace Jovanvich, Inc. 1973.

86. Tausig, C. M. "The Effects of Directive Individualized Instruction Upon Reading Achievement and Locus of Control." Dissertation Abstracts International 31A (12) (1970): 6480.

87. Thomas, B. G. "Continuous Progress Advanced Algebra in Lincoln Public School - a study of achievement and attitude toward mathematics." Dissertation Abstracts International 32A (11) (1971): 6113.

88. Vogel, F. X. and Bowers, N. D. "Pupil Behavior in a Multi-age Non-graded Primary Classrooms." Journal of Experimental Education 41 (2) (1972): 78-86.

89. Wheat, L. B. "Flexible Progress Group System." Elementary School Journal 38 (3) (November, 1937): 175-183.

90. Williams, B. G. "An Evaluation of a Continuous Progress Plan in Reading and Mathematics on the Achievement and Attitudes of 4, 5 and 6 Grade Pupils." Dissertation Abstracts International 34A (11) (1973): 7115.

91. Wilson, M. Part 1: Student Impact, Individually Guided Education (IGE) 1974-75. Des Moines Independent Community School District, Iowa: 1975.

VIII. ACKNOWLEDGMENTS

This dissertation cannot be complete without acknowledging those individuals who had helped and encouraged me during its composition. Dr. Trevor Howe, Dr. Ray Bryan, Dr. George Hohl, Dr. Rex Thomas, Dr. Richard Warren and Dr. Leroy Wolins were judicious and ardent committee members, fellow men with whom one could find solace.

I also want to thank Mr. James Seim of Ames, Mrs. Freda Caviness, Mr. Robert Himes and Mr. Merle M. Reisner of Indianola and others who assisted me during my data collection from their respective communities.

To Dr. Jon Paden of I/D/E/A for allowing me to utilize the field test results of _A Survey of Effective School Processes_ and for his consultations, I do thank.

Finally, last but most, I do especially thank my wife Marsha who not only proofread but also spent long nights with me at the Iowa State Computation Center.

IX. APPENDIX A: IGE OUTCOMES AS GROUPED BY A SURVEY OF EFFECTIVE SCHOOL PROCESSES SCALES

Thirty-five basic notions espoused by the IGE program were grouped by the I/D/E/A staff in their development of A Survey of Effective School Processes. The decision for placement of each outcome was based on the experience of the I/D/E/A staff.

The outcomes are listed relative to their original listing in various I/D/E/A publications.

IGE Outcomes

Institutional Commitment:

2. The school district has approved the schools staff's decision to implement the I/D/E/A Change Program for Individually Guided Education.

1. All staff members have had an opportunity to examine their own goals and the IGE outcomes before a decision is made to participate in the program.

Organizational Structure:

3. The entire school is organized into Learning Communities (L.C.) with each L.C. composed of students, teachers, aides, and a leader.

33. The Program Improvement Council assures continuity of educational goals and learning objectives throughout the school and assures that they are consistent with broad goals of the school system.

10. Sufficient time is provided for Learning Community staff members to meet.

4. Each Learning Community contains a cross section of staff.

6. Each Learning Community is composed of approximately equal numbers of two or more student age groups. (Ages 5-12)

22. The Program Improvement Council formulates school-wide policies and operational procedures and resolves problems brought to it involving two or more Learning Communities.

28. The Program Improvement Council analyzes and improves its operations as a functioning group.

23. The Program Improvement Council coordinates school-wide in-service programs for the total staff.

34. Students are involved in making decisions regarding school-wide activities and policies.

25. The school as a member of a League of IGE schools stimulates an interchange of solutions to existing educational problems and serves as a source of ideas for new development.

24. The school is a member of a League of schools implementing IGE processes and participating in an interchange of personnel to identify and alleviate problems within League schools.

Teacher's Role:

13. Each student's learning program is based on specified learning objectives.

5. Learning Community members have an effective working relationship as evidenced by responding to one another's needs, trusting one another's motives and abilities, and using techniques of open communication.

16. Students pursue their learning programs within their own Learning Communities except on those occasions when their unique learning needs can only be met in another setting using special human or physical resources.

15. Both student and teacher consider the following when a student's learning activities are selected:
Peer relationships
Achievement
Learning styles
Interest in subject area
Self-concept

21. Teachers and students have a systematic method of gathering and using information about each student which affects his learning.

17. Learning Community members make decisions regarding the arrangements of time, facilities, materials, staff, and students within the Learning Community.

27. Learning program plans for the Learning Community and for individual students are constructively critiqued by members of the Learning Community.

11. Learning Community members select broad educational goals to be emphasized by the Learning Community.

19. A variety of data sources is used when learning is assessed by teachers and students becoming increasingly more responsible for self-assessment.

7. Each student has an advisor whom he views as a warm, supportive person concerned with enhancing the student's self-concept; the advisor shares accountability with the student for the student's learning program.

26. The Learning Community analyzes and improves its operations as a functioning group.

12. Role specialization and a division of labor among teachers are characteristics of the Learning Community activities of planning, implementing and assessing.

14. A variety of learning activities using different media and modes is used when building learning programs.

8. Personalized in-service programs are developed and implemented by each Learning Community staff as a whole as well as by individual teachers.

35. Teacher performance in the learning environment is observed and constructively critiqued by members of the Learning Community using both formal and informal methods.

9. The Learning Community maintains open communication with parents and the community at large.

Learning Activities (Student's Role):

20. Each student (individually, with other students, with staff members, with his parents) plans and evaluates his own progress toward educational goals.

18. The staff and students use special resources from the local community in learning programs.

32. Each student demonstrates increasing responsibility for pursuing his learning programs.

29. Each student can state learning objectives for the learning activities in which he is engaged.

30. Each student accepts increasing responsibility for selecting his learning objectives.

31. Each student accepts increasing responsibility for selecting or developing activities for specific learning objectives.

X. APPENDIX B: LIST OF QUESTIONS ASKED OF LOCAL EXPERTS REGARDING THE SOCIO-ECONOMIC STATUS OF COOPERATING SCHOOLS

1.) Have you been a resident of the Ames (or Indianola) community since 1970?

2.) Are you familiar with the schools in this study and the local area they serve?

3.) Define the social fabric of those areas served by the schools in this study for your community.

4.) Based upon your best judgment please rank, from high to low, the schools relative to their expected academic performance based solely upon their socio-economic environment.

XI. APPENDIX C: X-MATRICES

Table 20. Partially balanced incomplete block design used in the Ames analyses of expert versus amateur raters[1]

Y(Rater, School)	Overall Mean	E vs A	E1 vs E3	E1+E3 vs E5	A2 vs A4	A2+A4 vs A6	S1 vs S5	S2 vs S5	S3 vs S5	S4 vs S5
Y(1, 1)	1	1	1	1	-1	-1	3	0	0	0
Y(2, 1)	1	-1	-1	-1	1	1	3	0	0	0
Y(3, 2)	1	1	-1	1	1	-1	0	3	0	0
Y(4, 2)	1	-1	1	-1	-1	1	0	3	0	0
Y(5, 3)	1	1	0	-2	1	2	0	0	3	0
Y(6, 3)	1	-1	0	2	-1	-2	0	0	3	0
Y(1, 4)	1	2	1	1	1	1	0	0	0	1
Y(4, 4)	1	-1	-1	1	-2	1	0	0	0	1
Y(6, 4)	1	-1	0	-2	1	-2	0	0	0	1
Y(2, 5)	1	-2	1	1	2	1	-2	-2	-2	-1
Y(3, 5)	1	1	-1	1	-1	1	-2	-2	-2	-1
Y(5, 5)	1	1	0	-2	-2	-2	-2	-2	-2	-1

[1]E = Expert; A = Amateur; S = School; and odd numbered raters were experts

Table 21. Balanced incomplete block design used in the Indianola analyses of expert versus amateur raters[1]

Y(Rater, School)	Overall Mean	E vs A	E1 vs E2	A1 vs A2	S1+S2 vs S3+S4	S1 vs S2	S3 vs S4
Y(1, 1)	1	1	-1	-1	1	1	0
Y(2, 1)	1	-1	1	1	1	1	0
Y(2, 2)	1	-1	-1	1	1	-1	0
Y(3, 2)	1	1	1	-1	1	-1	0
Y(3, 3)	1	1	1	1	-1	0	1
Y(4, 3)	1	-1	-1	-1	-1	0	1
Y(4, 4)	1	-1	1	-1	-1	0	-1
Y(1, 4)	1	1	-1	1	-1	0	-1

[1]E = Expert; A = Amateur; S = School; and odd numbered raters were experts

XII. APPENDIX D: PLOTS OF SIGNIFICANT INTERACTIONS

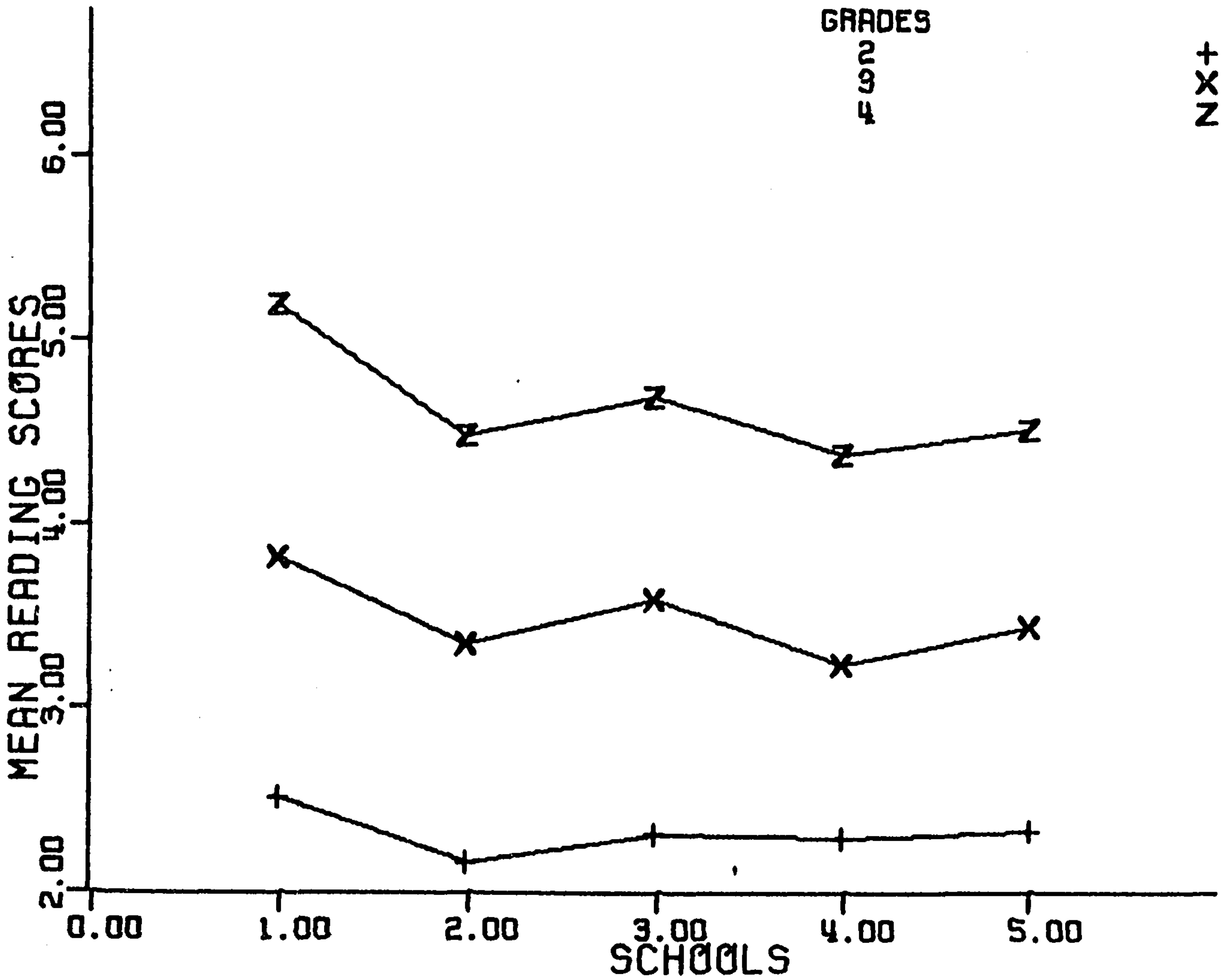

Figure 4. School by grade interaction from Ames' ANOVA with means as observational units

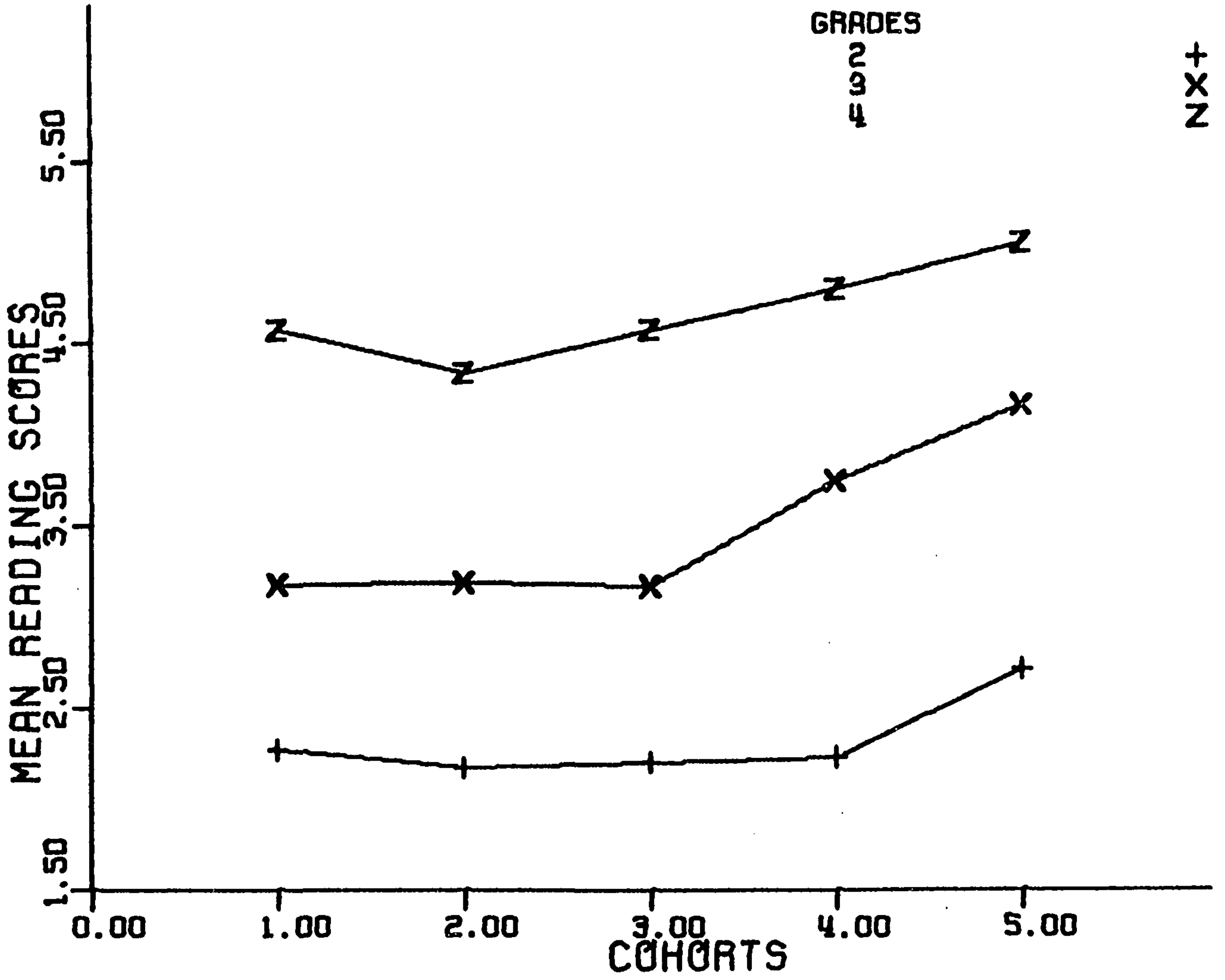

Figure 5. Cohort by grade interaction form Ames' ANOVA with means as observational units

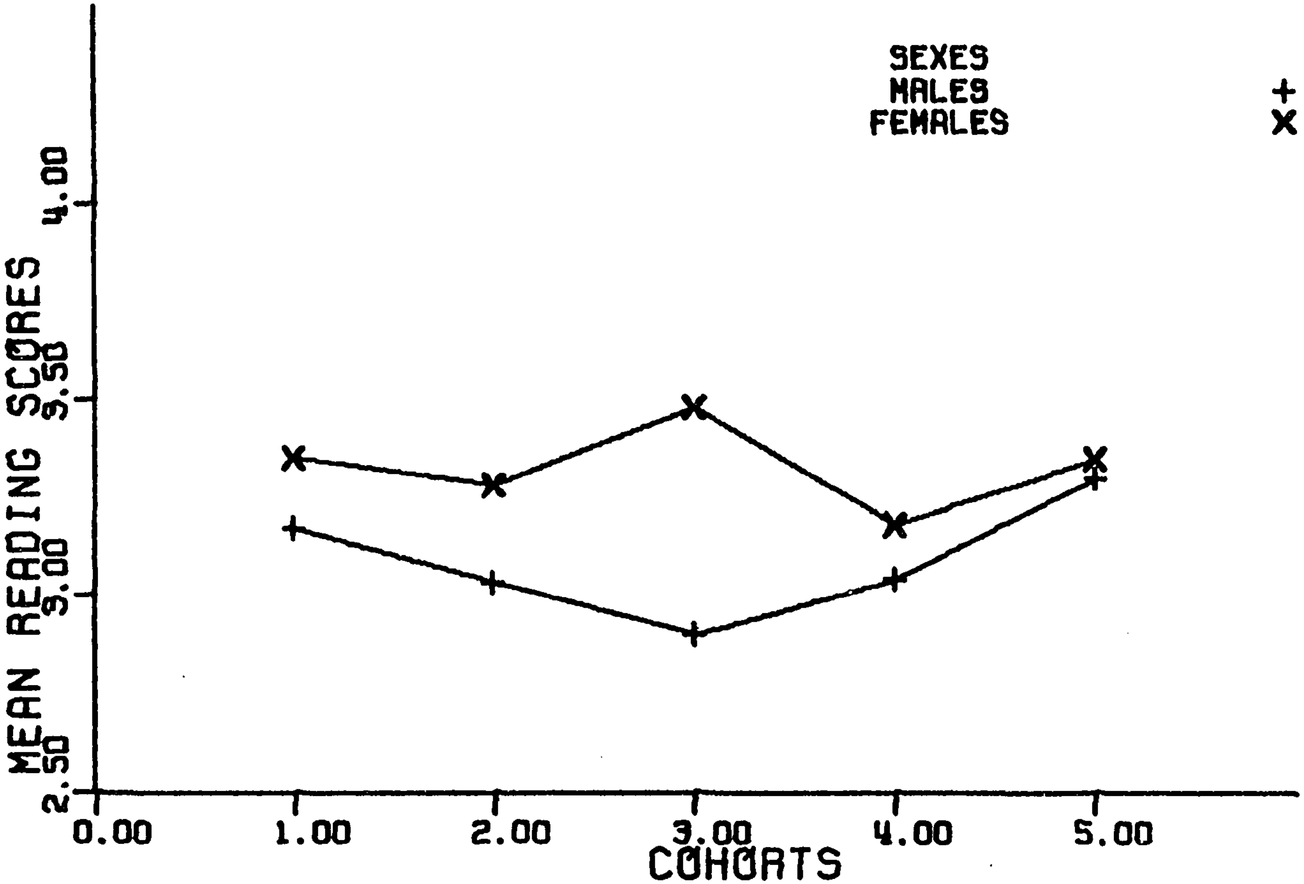

Figure 6. Sex by cohort interaction from Indianola's ANOVA with students as observational units