1985

# An analysis of the Iowa Tests of Educational Development performance of participants and non-participants in a ninth grade contemporary science course

Ronald D. Millsap
*University of Northern Iowa*

# An analysis of the Iowa Tests of Educational Development performance of participants and non-participants in a ninth grade contemporary science course


## Abstract

The need to document the results of instruction in a school system has been recorded as early as 1912 (Thorndike, 1912). Today, there is little argument about the necessity of evaluating pupil progress. Recent national attention has focused on public education and student performance on standardized achievement tests. Reports such as "Nation at Risk: The Imperative for Education Reform" (1983) agree that public schools need to be more accountable for student achievement and that standardized tests may be one of the best tools for assessing and/or evaluating the performance of students and schools.

# AN ANALYSIS OF THE IOWA TESTS OF EDUCATIONAL DEVELOPMENT PERFORMANCES OF PARTICIPANTS AND NON-PARTICIPANTS IN A NINTH GRADE CONTEMPORARY SCIENCE COURSE

Department Paper
Submitted
In Partial Fulfillment
of the Requirements for the Degree
Master of Arts in Education
Educational Psychology: Teaching Major

UNIVERSITY OF NORTHERN IOWA

by

Ronald D. Millsap

July, 1985

This Research Paper by: Ronald D. Millsap

Entitled: An Analysis of the Iowa Tests of Educational Development Performances of Participants and Non-Participants in a Ninth Grade Contemporary Science Course

has been approved as meeting the research paper requirement for the

Degree of Master of Arts in Education

Len Froyen

Co-Director of Research Paper
Len Froyen

Charles Dedrick

Co-Director of Research Paper
Charles Dedrick

Len Froyen

Graduate Faculty Adviser

Head, Department of Educational Psychology and Foundations

July 29, 1985
Date Approved

This is to certify that

Ronald D. Millsap

✓ satisfactorily completed the comprehensive oral examination

_____ did not satisfactorily complete the comprehensive oral examination

for the Specialist in Education degree with a major

in Educational Psychology: School Psychology

at the University of Northern Iowa at Cedar Falls

on July 29, 1985.

Examining Committee

Chairperson
Len Froyen

Member
Charles Dedrick

Member
Thomas Berg

Member

Transmitted by:

Lawrence L. Kavich, Head
Department of Educational
Psychology & Foundations

The writer wishes to express his gratitude to Dr. Len Froyen for his skillful assistance and direction in the writing of this paper.

He also wishes to express his appreciation to his family for their patience and understanding during this time.

# Table of Contents

APPENDIX

FIGURES



TABLES

# CHAPTER 1

## INTRODUCTION

The need to document the results of instruction in a school system has been recorded as early as 1912 (Thorndike, 1912). Today, there is little argument about the necessity of evaluating pupil progress. Recent national attention has focused on public education and student performance on standardized achievement tests. Reports such as "Nation at Risk: The Imperative for Education Reform" (1983) agree that public schools need to be more accountable for student achievement and that standardized tests may be one of the best tools for assessing and/or evaluating the performance of students and schools.

These reports have prompted many local school systems to scrutinize the performance of their students on nationally standardized achievement tests. The administration in one midwestern high school, Ottumwa High School (Iowa), has expressed concern about the results of achievement tests. In particular, the administration believes students' scores should be higher on the natural science section of the ITED that are given to grades nine and ten. The contentions of the administration are similar to those in recent reports that attribute low test scores to an inadequate curriculum and questionable teaching practices.

The ITED are achievement tests that measure abilities that are important in adolescent and adult life and that constitute a major part of the foundation for continued learning. They require students to apply their knowledge and skills in analyzing materials that they probably have not seen before. It is assumed by the authors of the Manual for Administrators and Testing Directors for the ITED that the level of student performance on the tests will reflect the effects of school experiences and educational development that stems from out-of-school activities such as hobbies, part-time jobs, home environment, etc., (1984). Each year participating schools are involved in the standardization process and said schools receive current norm-referenced results.

The impetus for this study was a meeting between members of the Ottumwa High School science department and members of the school administration in December of 1982. The meeting was initiated by the administration because of their concern about student scores on the ITED. The administration believed that there was not sufficient gain on students' natural science scores from ninth grade (the first year that the ITED are given to students) to their scores as tenth graders.

It was suggested by the administration that the science faculty make a comparison between the contemporary science curriculum objectives and those used to define the content domain of the ITED. It was thought that by making this comparison the science faculty might ascertain the extent of correspondence between the course and the test. If there was a reasonably high degree of similarity, an "item-analysis" of the test could be undertaken to get a partial evaluation of a portion of the school's science program.

This study will focus on the contributions of the contemporary science program at Ottumwa High School to student achievement in the natural sciences as measured by the Iowa Tests of Educational Development.

## Statement of the Problem

The purpose of this study was to investigate a concern of the Ottumwa Community School administration that originated in recent national and local reports of low student achievement. The Ottumwa High School administration believed there were not sufficient increases in student scores on the natural science test on the Iowa Tests of Educational Development between their ninth and tenth grade years.

This study will attempt to answer the following questions:

1. What items on the natural science section of the ITED would appear to measure content common to that chosen for inclusion in the contemporary science course?

2. Do the natural science total scores of tenth grade students who took the contemporary science course during their ninth grade year significantly surpass the scores of tenth graders who had not taken this or any other science course during their ninth grade year?

3. Do tenth grade students who took the contemporary science course score higher on those items that appear to measure the content common to the course and examination than do tenth grade students who did not take the contemporary science course?

From these questions, it is hypothesized that:

1. Tenth grade students who took part in contemporary science during their ninth grade year at Ottumwa High School will score significantly higher on the natural science section of the ITED than a comparable group of students who did not take part in the contemporary science course.

2. Students who took the contemporary science course will achieve higher scores on selected items on the ITED than will students who are comparable in natural science background and ability but did not take the contemporary science course.

## Importance of the Study

"A Nation at Risk" (1983) and other recent reports have directed attention to public education. Areas of concern have been identified by various factions. Math and science education have been moved toward the top of this list. Better programs with incentives will hopefully stem from this concern.

The contemporary science program and the Ottumwa Science Curriculum Guide is based on "A Philosophy of Science" that was formulated by the Iowa Department of Public Instruction and a committee of the Iowa Council of Science Supervisors (Appendix F).

This philosophy statement led to the formulation of a goal and three subgoals which became the cornerstone for the Ottumwa science curriculum. They are:

Goal (K-12): To develop a scientifically literate society.

Subgoals (K-12: A. To apply science processes as part of basic learning.
B. To communicate knowledge of natural phenomena.
C. To use scientific knowledge and processes, in comprehending the impact of science and technology on the individual, culture, and society.

The ITED were also designed according to a similar philosophy. These tests are defined in the Manual for Teachers, Counselors and Examiners (1984) as being "measures of abilites that are important in adolescent and adult life and that constitute a major part of the foundation for continued learning." (p. 1) These include:

> the ability: to recognize the essentials of good writing, to solve quantitative problems, to analyze discussions of social issues critically, to understand nontechnical scientific reports and recognize sound methods of scientific inquiry, to perceive the subtle meanings and moods of literary materials, and to use sources of information and common tools of learning (p. 1).

The school administration in Ottumwa has accepted the ITED as a fair measure of student achievement and progress in science.

When the questions posed in this study have been answered, we will be in a better position to know if;

1. students who participate in a science curriculum will perform significantly better on a standardized achievement test in the area of science than students who do not participate in a science curriculum and;

2. students who have been exposed to certain science material (e.g. contemporary science) will recognize and perform significantly better on items related to that material than students who were not exposed to that same science material.

## Assumptions

For the purposes of this study, it is assumed that the ITED are a fair measure of the science program.

It is assumed that the instructors were equally motivated in presenting the program material to their respective students to attain achievement similar to that measured by the ITED. Professional differences among individual instructors will have been negated by their concurrence regarding the:

1. objectives of the course;
2. content used to teach toward those objectives, and;
3. instructional materials used to present the content.

It is assumed that the motivation for performing well on the tests is equally distributed across the two groups of students.

It is assumed that students' general vocabulary and prior natural science scores may serve as a basis for matching students for further study (See Chapter 3, "Design of Study"). According to the authors of the ITED (1983), experience in intelligence testing has shown that a test of general vocabulary can serve as one of the best predictors of future success in school work. They further state that the vocabulary test is less sensitive to curriculum factors than are the other tests in the ITED.

## Limitations of the Study

This study was planned and conducted with the following limitations:

1. The study included only ninth graders enrolled in the Ottumwa Community School District during the 1982-83 school year.

2. The investigator was placed under personal restraints by the administration to make comparisons with the data available from the ITED on a curriculum wide basis rather than by instructor.

## Definitions of Terms

Achievement - current level of knowledge or skills.

Achievement test - a test used to indicate student knowledge or skills in a single or variety of subject areas.

Contemporary Science - a science course that makes use of the traditional areas of science (biology, chemistry, physics, geology, and astronomy) to provide a biological-physical-earth science approach to our current and future world. This course was developed by instructors in the Ottumwa (Iowa) School District for grades 9 and 10.

Item - a test question and the alternative answers when it is a multiple choice question.

Norm-referenced tests - a test whose scores are expressed in terms of the student's standing among a representative sample usually in percentiles or stanines.

Objectives - the instructional goals of a curriculum.

Pool group - subjects in this study who were tenth grade students at Ottumwa High School during the school year, 1983-84.

Posttest - a test given after a student's participation in a program, for comparison with a score from an earlier pretest.

Pretest - a test given before a student's participation in a program, for comparison with a score from a later posttest.

<u>Reliability</u> - the consistency of test results; that is, to what extent a student's test score varies due to chance or test error.

<u>Significant difference</u> - a difference between the averages of two sets of test scores great enough to be attributed to other than chance or test error.

<u>Statistical significance</u> - the unlikelihood that a particular result is due to chance.

<u>Test anxiety</u> - an emotional reaction to tests severe enough to impair performance.

<u>Validity</u> - how well a test accomplishes its aim; that is, the extent to which a test truly measures what it claims to measure.

## Chapter 2

## Review of the Literature

There are three major themes addressed by various writers that are pertinent to the problem investigated in this study. They are discussions devoted to:

1. purposes for using standardized achievement tests;
2. factors that may affect student performance on standardized achievement tests; and,
3. appropriate and inappropriate uses of standardized achievement test results.

Each of these themes will be treated separately. Due to the rather vast body of literature devoted to aptitude and achievement tests, selections cited here are those deemed to be most relevent to the situation described in Chapter 1 of this study. Only articles and research devoted to the aforementioned topics will be reviewed.

### Purposes for Using Standardized Achievement Tests

Most writers agree there are three general purposes for administering standardized achievement tests in American schools: (1) evaluation/assessment of the students or the curriculum; (2) accountability; and, (3) identification of individual students or groups of students for special help or special programs.

Standardized achievement tests are given to obtain results that can be used to evaluate the extent to which the objectives of a school curriculum are being met. Mehrens and Lehmann (1978) identify standardized achievement tests as being designed to assess pupils' knowledge and skills at a particular point in time. They, like the authors of the ITED (1984), state that standardized tests are best used to assess student achievement on long-range educational goals rather than immediate goals of curriculum instruction. Mehrens and Lehmann cite the need for an objective and scientific approach to the assessment of pupil knowledge.

Baker (1982), in her booklet intended to help school personnel and others of the community better understand achievement testing, cites these tests as "assessing the subject matter and skills that students have learned" (p. 2). Baker introduced the term "competency" in describing those tests that "are administered to certify minimum skills acquired by students" (p. 3). Standardized tests may be called the ultimate expression of this approach to pupil assessment.

Cole asserted that "the role of tests and testing is to validate, not dictate classroom instruction" (1981, p. 618). Cole's statement stemmed from the findings of two studies done to determine who in schools "actually" use the results of standardized tests.

Sproull and Zubrow (1981) surveyed public (both urban and suburban), parochial, and private schools to determine how the tests were being used. They talked with 58 central office administrators who were in charge of standardized testing in 18 school systems in western Pennsylvania. The schools represented approximately 124,000 students. Instructional program evaluation and end-of-year achievement measurement were two of the four most frequently reported purposes for testing.

In another study, Salmon-Cox (1981) interviewed 68 Pennsylvania elementary school teachers, all of whom "administer and receive information from a published, standardized test of student achievement" (p. 631). The teachers were from both suburban and urban school districts. From this study, the purposes of reshaping instructional content, shifting emphases of topics, or increasing instructional time in particular areas were listed as being important uses of the tests.

On a national basis, the Commission on Excellence in Education in its report, "A Nation at Risk: The Imperative for Educational Reform" (1983) made the recommendation that standardized tests of achievement be administered at various levels of schooling with one of the major purposes being to identify the student's level of achievement.

Accountability is a second general purpose for testing. According to Ebel (1963), tests are a description of measurement and they serve as a means of evaluating education's effectiveness. He further stated (1981) that tests of mental ability provide the most precise and efficient means of assessing a person's structure of knowledge. Tests help in cultivating cognitive competence and developing and identifying excellence.

Mehrens and Lehmann (1978) and Resnick (1981) discussed the early history of testing and the need for an objective and scientific approach to the assessment of pupil knowledge.

Resnick, in his discussion of testing for accountability, identified the widespread concern for the "rational management" of education. He referred to the beginning of standardized testing in the early part of this century when the concern was for increased productivity in business and identification of officer material in the armed forces. This efficiency was appealing to leaders in education who saw the promotion of standardized tests as a means of assessing the progress of students and the subsequent reporting of results to their respective boards of education and the public at large.

"Accountability" is also cited as a purpose for testing by others. Baker (1982) stated that tests allow school districts to act accountable, "giving tests seems to be a responsible way to act" (p. 27). She further stated that:

> Test performance allows the public not only to monitor the schools but to act collectively to assign responsibility and to express expectations for improvement. In addition, as schooling has come to serve many diverse purposes, testing has come to be seen as a mechanism for pulling out from the complicated curriculum those areas regarded to be of most significance. In this way, people have assumed that having tests assures that the schools have standards of quality (p. 1).

Holley (1983) saw achievement testing at the state and national level as serving public information needs. This he said in referring to the "Decade of Accountability, 1975 to 1985" (p. 31) is a period of time when the importance of educator competence created a public demand for accountability. At the beginning of this period, Corbett (1975) discussed the need for state and federal agencies to obtain objective data to ensure that programs funded by them were effective. Laughland (1983) later agreed.

Five years earlier than Corbett, Wildavsky (1970) saw the request for school systems to be held accountable for the achievement of children as a good one. He stated that education could make a contribution to the general evaluation of governmental programs.

Sproull and Zubrow (1981) reported that all the school systems involved in their survey identified reporting to outside agencies, primarily the federal government, as the major purpose for standardized testing.

Identification of students for reasons of promotion or educational opportunity is the third general purpose of standardized achievement tests. Two other recommendations the Commission on Excellence in Education (1983) made, in addition to the one mentioned earlier, were the identification of; (1) any students who might be in need of remedial intervention and (2) any students who should be given the opportunity for advanced or accelerated work.

These recommendations are also identified and supported by Baker (1982), Sproull/Zubrow (1981), and Salmon-Cox (1981), who found that achievement test scores confirm information that teachers already possess when making individual student diagnosis and placement.

## Factors Affecting Test Scores

The performance by a student on a standardized achievement test is determined largely by the student's level of knowledge and skills in the subject area and his test-wiseness which is his ability to demonstrate knowledge of those skills within the test situation (Sabers, 1975).

Students are in possession of a certain amount of knowledge and skills because of their prior course work and their experiences both in and out of school (Ebel, 1975 and ITED, 1983). Ebel stated that test performance depends on prior experience. The authors of the ITED claim student performance levels will reflect the effects of in-school and out-of-school experiences and activities of each student. They stated "test results need to be seen as products of a broad range of educational and community factors" (1983, p. 3). Biehler, in discussing standardized achievement tests states "such tests are designed to measure the outcomes of previous learning experiences" (1978, p. 630-631).

In two papers dealing with achievement by science students, Sigda (1983) and Hanson (1961) discussed factors which may influence test scores in science achievement. Both agree that the amount of experience the student has in science may partially determine the student's ability to interpret reading materials in the natural sciences. They agreed that the qualifications of the teacher in the field of science allowed him/her to be more effective and hence, raise the achievement of their students. They also agreed that motivation in science was a factor in test results. Sigda attributed this partially to the amount of time spent by the student engaged in science in the elementary grades while Hanson discussed the academic aspirations of students as being their motivation to score well.

Ducote (1982) investigated the interactions of certain motivational factors with test-wiseness. His study dealt with the examination of five characteristics that Anderson (1981) found to correlate highly with cognitive learning. They are anxiety, self-concept, academic values, interests, and locus of control. Ducote reported a relationship between anxiety and successful test performance. The results of his study indicated that academic values and academic self-concept were related to successful application of test-wiseness skills. Evidence that interests might affect test performance is lacking. No support was found for a relationship with locus of control.

There are numerous studies that show test anxiety as being a cause of poor performance on a test. Anderson (1981) stated that research supports the belief that each person has an optimal level of anxiety for each task and too little or too much anxiety may impede performance. In his summary of test anxiety research, Tryon (1980) noted that test anxious people expect lower performance and subsequently achieve less. In his study that examined the academic performance of low, moderate, and high test anxiety students, Osterhouse (1975) related how low and moderate anxiety students will obtain slightly higher examination scores than high test anxiety students. In summing up the effects of anxiety on performance, Lazarus (1982) concluded "a little anxiety improves performance: too much anxiety impairs it" (p. 42). He discussed the cycle that students may fall into. That is, test anxiety arises from fear of failure, causing a low score which in turn may lead to more anxiety for the next test and another low score.

Green (1975) pointed out that test performance can change due to changing physical and mental well-being of test takers. Headaches, flu, and eye problems may lead to lower scores, especially if they will not allow the student to concentrate on the material.

The academic aspirations of students will help determine the effort they put forth when participating in standardized achievement tests. Ducote (1982) recommended a "healthy balance" between academic values and self-concept. This healthy balance may be attributed to students' preconceived ideas of test usage. Ducote reported that high academic values will motivate a student to do what is personally or socially acceptable, desirable or that which is worth expending energy to achieve. He describes self-concept as "a person's perception about personal worth and how one is perceived by others" (p. 6). Green (1975) reported that experiments conducted by Rosenthal, Jacobson, Palardy, and others have proven "children do as well in school as their teachers, their parents, and they themselves, believe they will do" (p. 92). If students believe their test results will be used in a positive manner, they will likely perform better. If they feel the results will be meaningless and not serve any purpose, then the student may not try as hard. In summation, Ducote stated that self-concept will motivate students to exhibit behavior conducive to academic achievement. He did warn, however, that students who begin to value test performance too highly may suffer from increased test anxiety and lower scores will result.

"Test-wiseness is defined as a subject's capacity to utilize the characteristics and formats of the test and/or the test taking situation to receive a high score" (Millman, Bishop, and Ebel, 1965, p. 707). Millman (et al., 1965) and Anderson (1981) include awareness of time, proper following of directions, and the effects of practice and coaching as factors that would allow students to be more successful at taking tests. If a test has to be done within a specified time period this could benefit those students who are good at managing their time. On the other hand, if students spend too much time on some problems, their scores will be lower. If test directions are vague or if students misinterpret those directions, their scores can suffer. If directions are easy to follow, the students will score higher.

Coaching students for a test was also found by Bangert-Drowns (et al. 1984) to improve achievement test scores in their study. Greater gains occurred when students were given a pretest prior to coaching and when they practiced more, on a regular schedule, for a longer period of time. He adds that high-ability students gain more from coaching programs than do lower ability students, who may need more explicit instruction. Anderson (1981) relates how student anxiety may be lessened with coaching. Millman (et al.), cited a series of studies that indicated practice and coaching with material similar to the test would raise scores.

Ducote (1982) then summed up his review of test-wiseness by stating "a test-wise student can use test characteristics to score higher even with little or no knowledge of the test content"(p. 3).

Test performance might depend on the cultural background of students and their age grouping. Cole (1975) stated that different subcultures will have different vocabularies. Students' whose vocabuluaries more closely match that of the test will probably perform slightly better. The student's age will determine the amount of past knowledge and experience he has to fall back on in acquiring the answer to a question.

## Uses of Standardized Tests Results

Most articles on the use of test results centered on cautions when using the results from standardized achievement tests, while offering appropriate uses of these results. Authorities in the field of testing (Farr, 1980; Green and Ebel, 1975; Patterson, 1975; Massachusetts State Department of Education, 1982; Barnes, Moriarty and Murphy, 1982; and others) agree that test results should never serve as a single "final" authority for measuring student achievement. The Massachusetts State Department of Education (1982) offers four reasons why test results should not be used as exclusive measures: (1) A test is a sample of how well a student works with a given set of learning materials; (2) A test score is an estimate and should not be treated as an exact numerical measurement; (3) Since a test is only a sampling of the information, it can only measure part of the knowledge a student might have; (4) It would not be fair to the student to judge him on the basis of a single test since individuals are continually changing by learning new information and skills.

The same group of authorities state teacher observations of students, daily assignments, teacher-made tests and other published tests should be used in combination when evaluating a student's progress. Further use of this information should be directed towards improvement of instruction, whether on an individual basis or for an entire curriculum. Thus, tests then can be used by all people involved in education in helping students achieve their goals. Individuals who rely upon standardized test results for this purpose (Sproull and Zubrow, 1981) are building principals, counselors and teachers.

Rudman (1982), in expressing his concern for the proper use of test results by the aforementioned groups, called for the systematic examination of data in a three-part analysis: diagnosis, evaluation and planning. In the diagnosis of test results, the principal's primary concern should be an identification of strengths and weaknesses in the overall curriculum and within each subtest. The second stage of analysis should be an evaluation of what has been learned by the students and a comparison of that information to other data such as the content presented in one school as compared to another school. Rudman then suggested planning how to best use the information that has been gathered. In his conclusion, he stated, "the problem lies more in the proper use of standaradized tests in decision making than in the tests themselves" (1982, p. 64).

Like Rudman, other writers stress the use of standardized tests as diagnostic in nature. Barnes, Moriarty and Murphy (1982), Kearney (1983) and Forsyth (1983) relate how results can be used to describe current achievement of students. This description will permit comparisons between broad areas of achievement at each testing and permit an assessment of student growth with respect to valued objectives within each area. They also stated that test results can provide school personnel with a dependable basis for judging the relative strength and weaknesses of the school's curriculum. Barnes (et al., 1982) do warn, however, that

> "school people should be careful in using the standardized achievement test for classroom diagnosis. Further tests should be given that are designed for diagnostic use. The standardized achievement test may only indicate a need for further diagnosis" (p. 16).

Tyler (1966) agreed in his explanation of the various kinds of educational appraisal. He stated that evaluation should "diagnose the learning difficulties of an individual student or an entire class to provide information helpful in planning subsequent teaching" (p. 1).

Green (1975) and Ebel (1975) agree that standardized achievement tests are not well suited for predicting the future academic success of students. That use would mean that tests alone are being used for evaluative purposes. Ebel (1963) opposed this and stated that tests should serve as a base from which the individual has a choice when making a decision. When the individual student is given different avenues to allow for his own decision making, Ebel relates how tests will help to improve individual status, recognize and develop the wide variety of talents needed in our society, and reveal the individual's abilities and prospects. These uses, restated by Oakland (1974), will help people make wise decisions about their futures, rather than have decisions imposed on them.

Merwin (1982), Hodgkinson (1982), and Salmon-Cox (1981) see the use of achievement test scores as that of a "snapshot" rather than a "continuous movie" which would describe the assessment of students by their teachers through classroom observation and testing. Merwin regards the test as giving a quick and efficient "reading" of achievement. He, like Rudman, then related the importance of being able to use the results of tests as a comparison between groups of students. Most standardized tests are of a norm-referenced nature and the normative data obtained are based on a large number of schools which provide descriptive information about individuals and groups that could not be obtained otherwise. The standardization of tests along with the method of their construction, which establishes reliabilily and objectivity, makes provisions for comparisons that are otherwise not available.

Merwin (1982) offered the use of test results as confirming and questioning information that teachers already possess about individual students. Teachers are continually evaluating a student's performance and standardized testing will allow a teacher to examine their conclusions with additional criteria. The Salmon-Cox study showed that teachers used test information to evaluate the grouping or tracking of students into specific classes (e.g., high ability versus low ability) or groups (e.g., reading levels).

Likewise, standardized test results can alert the teacher to faulty or conflicting information on some students. Salmon-Cox reported that achievement tests can function as a "red flag" indicating to the teacher that he/she may not have recognized some strengths and weaknesses in the observation of the regular classroom performance of a particular student and that further evaluation is needed. Merwin then discussed the value of gathering further information to reconcile the different positions between the test results and the teacher's previous observations. This would correct misimpressions and possibly increase "confidence in the fidelity (exactness) of the information by both the teacher and by others who will use it" (1982, p. 16).

When the evaluation of a curriculum is planned, Nimmer (1982) points out that a subtest of a standardized achievement test may not include an adequate number of questions, if any, for a fair evaluation of a separate subject area. In his study of assessing the effectiveness of curriculum changes in earth science, Nimmer used five standardized achievement test batteries administered to pupils of junior high age (grades 7-9). He found that the number of test items within the batteries was not comparable to the time spent on various topics in the curriculum. Some topics deemed important to the curriculum were not represented by the test even though these topics represented a large block of time during the school year. Nimmer recommends using evaluation techniques and instrumentation keyed to the curriculum's specific objectives.

One commonly accepted purpose of achievement testing, and subsequent use of the results, is informing the public of the achievement of students in particular individual geopgraphical locations. But, careful consideration should be given when releasing test scores to the public. Patterson (and others, 1975) and Barnes, Moriarty and Murphy (1982) stated that results should not be released to the public without providing the proper background for the interpretation of such results. Patterson stated "since people who work in the news media rarely know all the many factors affecting school performance, they should not be left the responsibility of interpreting test results" (p. 94). He emphasized that standardized tests should be administered with the expectation that information derived from them will only be one source of data to be used for decision making by schools.

Barnes, Moriarty and Murphy (1982) in their plan for reporting test results, tell of the need to alert the public to the problems inherent in standardized achievement testing. It is the school's responsibility to make the public more sensitive to the variables that affect students' scores. Many of which lie outside the control of schools. They stated that "tests simply indicate a student's performance on a given day under a given set of circumstances" (p. 16).

Kearney (1983) stressed the use of test results to inform the public of the good things happening in the schools. He believed "that policymakers at local and state levels, if not at the national level, ought to use test results to indicate what is going on in the schools, what levels of attainment are at any given time, what changes take place in those levels over time, and what progress is being made by the schools from year to year" (p. 10).

Webb (1983), in his analysis of three tests for assessing reading ability, stated that test results should be used in counseling, advising or in developing more effective teaching strategies.

The studies cited above reveal the importance of the proper use of standardized achievement tests. These tests are given for the purposes of measuring the current level of knowledge and skills of students, answering the call for accountability to the public, and placement of students at appropriate levels of learning. Before and during testing, the environment for the students should be as conducive to performing to the best of their abilities as possible. The results of the tests should be carefully studied and released only after proper information has been disseminated to all those affected. Of importance in this study is an investigation of how the results of standardized tests can be most properly and efficiently used for the improvement of instruction.

Chapter 3

Design of the Study

The impetus for this study emanated from a concern expressed by the administration at Ottumwa High School in December of 1982 regarding student performance on the natural science section of the Iowa Tests of Educational Development.

## The Community and School Population

This study was conducted in Ottumwa, Iowa, a rural-based city of about 26,000 people, located in southeast Iowa. The high school enrollment has dropped the past ten years to approximately 1500 students (grades 9-12) and it appears it will remain at this level for at least the next ten years.

Of the families in this school community, 5% have incomes below the poverty level. Approximately one-third of the students receive free or reduced (in price) lunches. The percentage of parents who have some education beyond high school is about 16% and the percentage of parents without a high school diploma is about 13%. The current drop-out rate for the Ottumwa schools is about 5%.

Nearly 17% of the parents have a bachelor's or other advanced degree. Nearly 75% of the students intend to further their education after high school with half that number planning to attend a four year college. The ethnic composition of the community is approximately 97% White, 2% Black, .75% Asian and

.25% Hispanic. Approximately one-third of the students live with their natural parents.

The curriculum is a traditional liberal arts/college preparatory basis of instruction. For graduation, students are required to have the following semester credits:

| | |
|---|---|
| Language Arts | 6 |
| Mathematics | 4 |
| Social Studies | 6 |
| Science | 4 |
| Physical Education | 8 |
| Electives | 12 |
| Total | 36 |

## Course of Study

The contemporary science program was developed during the 1981-83 school years according to guidelines of the Ottumwa Science Curriculum Guide. This guide was written by a committee of teachers and administrators formed to develop a science curriculum on a kindergarten through 12 continuum. Jack Gerlovich, Science Consultant for the Iowa Department of Public Instruction and George Magrane, Science Consultant of the Southern Prairie Area Education Agency #15 of the State of Iowa assisted the faculty during their work. In an attempt to enlist as much as possible the educational community in completing this project, surveys and reports were made available to faculty and administration of all the schools concerned. The product of their efforts, the Ottumwa Science Curriculum Guide, outlines the rationale and philosophy for the science programs and specifies the placement of objectives according to guidelines devised by the Ottumwa Community School District.

The instrument used for development of the guide was "A Tool for Assessing and Revising the Science Curriculum." This instrument was jointly developed by the Iowa Department of Public Instruction and a committee of the Iowa Council of Science Supervisors to encourage and aid local schools to assess their science curricula on a continuous basis. It provides a schedule for conducting a science curriculum assessment and/or revision; a model for assisting schools in developing their science philosophy, goals, and objectives; recommendations for levels at which suggested objectives are introduced, emphasized and maintained; and an instrument for matching local science curriculum needs to available science programs.

The placement of the aforementioned objectives, involved the identification of suggested objectives and the insertion of them into a multiple grade-level science curriculum. The suggested objectives were made available from the Iowa Department of Public Instruction. The objectives chosen by the Ottumwa committee (Appendix E) would be introduced, emphasized and/or maintained at four levels; grades K-3, 4-6, 7-9, or 10-12. When an objective was to be "introduced," it would be the first time that that topic would be inserted as a planned portion of the science program. When an objective was to be "emphasized," it would be the time for stressing it. When an objective was to be "maintained," it would be the time for reinforcement of topics introduced previously and presented at the level of sophistication applicable to that given grade level. If the committee chose not to include an objective in the K-12 curriculum, that objective was termed "not applicable." An objective could also be twrmed "not applicable at a certain level" during the process of objective selection.

The Ottumwa Community Schools Science Curriculum Guide was then revised in content and focus. One outgrowth was the development of the contemporary science program. This program is an elective science class, open to any ninth or tenth grade student at Ootumwa High School. Program development was conducted by four science instructors at Ottumwa High School. The contemporary science course makes use of the traditional areas of science (biology, chemistry, physics, geology, and astronomy) to provide a biological-physical-earth science approach to our contemporary and future world. Students who participate in the contemporary science course should develop scientific skills that will enable them to gain knowledge, to cope with scientific/technological problems, and to make decisions about future scientific advances. The contemporary science program was implemented in the 1982-83 school year.

## Instrument

Ottumwa High School began participating in the ITED testing program in the fall of 1979. Since then, each grade (9-12) has taken the ITED early in the fall term. Test results are returned to Ottumwa High School in late November or early December.

## Subjects

The subjects selected for this study were enrolled in the ninth grade during the 1982-83 school year. There was a total of 355 tenth grade students who took the ITED in the fall of 1983 in Ottumwa. During the testing, students reponse sheets were marked in one of two ways: (1) by their participation in the contemporary science course during their ninth grade year, or (2) by their lack of participation. Contemporary science was an elective course. Therefore, students could choose to include or exclude it in their schedule of classes. The choice had to be made at the beginning of their ninth grade year. 265 ninth grade students enrolled in contemporary science, ninety did not enroll for the 1982-83 school year.

The two groups for this study were chosen by matching students who chose not to take contemporary science during their ninth-grade year with those students who chose to take contemporary science in their ninth grade year. Matched pairs were formed by simultaneously using ninth grade natural science and vocabulary scores on the ITED. (Descriptions for both tests are in Appendix A.) This matching was done on a one-to-one basis. The natural science standard score of a student who did not take contemporary science was matched with a student who had taken the course providing the two students scores did not deviate by more than one standard score point on the vocabulary section of the same test. For example, if a student who chose not to take contemporary science received a 17 on his/her natural science test (standard score) and a 16 on his/her vocabulary test, then a contemporary science student was located who also received a 17 on his/her natural science test and either a 15, 16, or 17 (variance of one, either direction) on his/her vocabulary test. This procedure was designed to secure comparable groups in science achievement. The vocabulary scores were used to make the student pairs reasonably equivalent on a measure of aptitude. These natural science scores thus created the pretest and completed the first phase of this study. This procedure for selecting subjects resulted in sixty-one matched pairs.

## Data Collected

The first phase of the investigation compared the matched pairs of students' mean scores on the natural science test of their tenth grade year. This data comprised the posttest.

The pretest and posttest represented similar measures of treatment: They were given within the same time frame and they were the same forms of the ITED.

The second phase of the investigation involved the same two groups of students. This investigation looked at the scores the students received on specific ITED items that were selected because they were deemed to be appropriate measures of content taught in the contemporary science course.

As recommended by the director of the ITED (Forsyth, personal communication, 1984), the instructors of the contemporary science program were asked to identify those items on the natural science section of the ITED that were similar in content to the contemporary science curriculum. Each item had to be identified by at least three of the four instructors to establish validity.

A total of thirty-three test items were chosen by the four instructors. Eight items were chosen from Part I of the natural science section (out of twenty-four). Twenty-five items were chosen from Part II (out of thirty-six). (Appendix D) The reason for the large difference in the number of questions between the two parts is the structure of the natural science section of the ITED. Part I is content-oriented while Part II is reading interpretation-oriented. The larger proportion of items selected from Part II reveals the less-content-specific approach to science achievement.

The student response sheets were obtained from the files of the ITED. The scores for phase two of the study were obtained by hand-scoring the response sheets of the sixty-one matched pairs of students.

# Chapter 4

## RESULTS AND DISCUSSION

The purpose of this study was to compare the Iowa Tests of Educational Development (ITED) natural science performance of ninth grade students who did and did not take a contemporary science course at a mid-western high school. The sample population consisted of 122 students from the ninth grade class of 1982-83. There were a total of 355 students in this class. 265 of the students chose to participate in the contemporary science program during their ninth-grade year. Ninety students of this same class chose not to participate in contemporary science.

Comparison groups were formed by using scores obtained from the natural science and vocabulary sections of the ITED administered during the 1982-83 and 1983-84 school years. Sixty-one pairs of students were matched by jointly using students' natural science and vocabulary scores on the ITED (Appendix B). These scores were obtained from an October 1982 administration of the tests.

Comparisons were made between the students' 1982 standard scores on the natural science tests and those received when they took the tests again in October of 1983. Their raws scores on selected questions from the natural science section of the ITED (Appendix C) were also compared.

Research Question 1 speculated about possible similarities between the content tapped by selected items on the natural science section of the ITED and the content chosen for inclusion in a high school contemporary science course. The procedure for selecting the items was suggested by Robert Forsyth, Director of the ITED (1984). Each of the four instructors involved in the development of the contemporary science course selected those questions from the ITED that appeared to measure contemporary science course objectives. This selection process was done on an individual basis with no meeting held among the four instructors. The items selected by three or more of the teachers were scored for each of the 61 pairs of students.

Research Question 2 focused on determining if there was a significant difference between the tenth grade ITED natural science standard scores of ninth grade students who participated in a contemporary science program during the 1982-83 school year and those who did not.

Research Question 3 focused on hypothesized differences between the scores of contemporary science students on selected questions from the natural science section of the ITED and students who did not participate in a contemporary science program.

The procedure for answering Research Question 1 involved the selection of items on the ITED that measured objectives similar to those of the contemporary science course. The four instructors involved in teaching the course were asked to identify items they thought students ought to be able to answer because of their enrollment in contemporary science. The items that were deemed to measure course objectives are noted in Appendix D.

In order to analyze postulated differences between students who participated in a contemporary science course during their ninth grade year and those who did not, histograms (Figures 1 and

2) were constructed and t-tests calculated (Table 1). The .05 level of significance, using a one-tailed test, was chosen to test the hypothesis.

Figures 1 and 2 show the histograms of the students' standard scores. Table 1 displays the means and standard deviations of the participants and nonparticipants on the ITED natural science test. The "t" value is significant at the .05 level. This finding supports the first hypothesis (p. 4).

Upon examination of Figures 1 and 2, the researcher found differences in the "clustering" of students. Using the mean of the participant group (15.44) and moving across to the right one standard deviation to a score of 21, ten students were found to have scores above this numerical value. When using the respective nonparticipants mean, eleven students in this group scored one standard deviation above the mean. Five students are found at the score of 20 and one at 21 points. However, when using the participants mean as the norm, twice as many students enrolled in the science course scored more than one standard deviation beyond the mean than did nonenrolled students. The participant group also had a greater number of students between the mean and plus one standard deviation (twenty-three) than the nonparticipant group (eighteen).

The nonparticipant group had nearly twice as many students (twenty-three compared to twelve) located in the area spanned by one standard deviation below the mean. Ten participants scored between one and two standard deviations below the mean while seven

Figure 1

Histogram of 1983 ITED Natural Science Scores of Students <u>Who Participated</u> in Contemporary Science.

```
6                                              *  *
5                                              *  *
4                      *     *                 *  *     *     *  *
3                *     *     *     *           *  *     *  *  *  *     *
2                *  *  *     *  *  *           *  *  *  *  *  *  *     *  *  *
1          *     *  *  *     *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *     *
------------------------------------------------------------------------------------------
Standard   3     5     7     9     11    13    15    17    19    21    23    25    27
 Score        4     6     8     10    12    14    16    18    20    22    24    26    28
```

Mean = 15.44     Std. Dev. = 6.25     Sample Size = 61

Figure 2

Histogram of 1983 ITED Natural Science Scores of Students <u>Who Did Not Participate</u> in Contemporary Science.

```
7                                                 *
6                            *                    *
5                            *  *  *              *        *
4                   *        *  *  *  *  *  *     *        *
3                   *     *  *  *  *  *  *  *     *  *     *
2             *     *     *  *  *  *  *  *  *     *  *     *
1          *  *     *     *  *  *  *  *  *  *  *  *  *  *  *  *  *  *     *        *  *
------------------------------------------------------------------------------------------
Standard      5     7     9     11    13    15    17    19    21    23    25    27    29
 Score     4     6     8     10    12    14    16    18    20    22    24    26    28
```

Mean = 14.23     Std. Dev. = 5.41     Sample Size = 61

Table 1

Comparison of 1983 Standard Scores

| Group | N | Mean | Standard Deviation | t | Level of Significance |
|---|---|---|---|---|---|
| Participants | 61 | 15.44 | 6.25 | 2.14* | .05 |
| Non-participants | 61 | 14.23 | 5.41 | | |

*Significant at the .05 level.

nonparticipants were in this category. The increased exposure to science may have contributed to the wider spread of scores among the participant group. Regression toward the mean is generally more common among individuals who did not participate in the experimental treatment.

The scores of the participants were more skewed toward the upper end of the distribution with a clustering of scores between the mean and one standard deviation above the mean. Four or more science course participants achieved scores of 15, 16, 18, 20 and 21.

The nonparticipant group scores were skewed toward the lower end of the distribution, and within one standard deviation below the mean. Four or more non-science course students scored 15, 14, 13, 12, 11 and 10.

This grouping of the nonparticipants around the mean may be further understood by examining the performances of the twenty lowest scoring students on the 1982 administration of the ITED (Figure 3). Note that a greater number of the nonparticipating students obtained a gain score (1983) greater than the contemporary science enrollees with whom they were matched. Of these twenty pairs of students, fourteen of the nonparticipants attained a higher score in 1983, compared to eleven of the participants. Only three of the nonparticipants lost ground (-1, -.5 and -2.0) compared to eight of the participants. This is another example of the spreading out of scores in the participant group. Three nonparticipants and one participant showed no change.

Figure 3

Growth of Matched Students Who Recorded the Twenty Lowest Scores: Matched Pairs Based on 1982 ITED Natural Science Test

| Participants | | | Nonparticipants | |
|---|---|---|---|---|
| 1983 Standard Score | Years of Growth | Lowest 1982 Standard Scores | 1983 Standard Score | Years of Growth |
| 9 | +1.5 | 6 | 9 | +1.5 |
| 7 | + .5 | 6 | 5 | -1.0 |
| 9 | +1.5 | 6 | 5 | - .5 |
| 7 | 0 | 7 | 7 | 0 |
| 20 | +6.5 | 7 | 17 | +5.0 |
| 6 | - .5 | 7 | 10 | +1.5 |
| 10 | +1.5 | 7 | 11 | +2.0 |
| 12 | +2.5 | 7 | 17 | +5.0 |
| 11 | +1.5 | 8 | 9 | + .5 |
| 10 | +1.0 | 8 | 12 | +2.0 |
| 18 | +5.0 | 8 | 10 | +1.0 |
| 3 | -2.5 | 8 | 12 | +2.0 |
| 5 | -2.0 | 9 | 10 | + .5 |
| 7 | -1.0 | 9 | 12 | +1.5 |
| 6 | -1.5 | 9 | 11 | +1.0 |
| 5 | -2.0 | 9 | 5 | -2.0 |
| 9 | - .5 | 10 | 10 | 0 |
| 7 | -1.5 | 10 | 13 | +1.5 |
| 15 | +2.5 | 10 | 14 | +2.0 |
| 11 | + .5 | 10 | 10 | 0 |

Summary of Figure 3

| | Participants | nonparticipants |
|---|---|---|
| Number of students showing standard score increase . . . . . | 11 | 14 |
| Average standard score increase . . | 4.5 | 3.9 |
| Average gain in growth in years . . | 2.2 | 1.9 |
| Number of students showing standard score decrease . . . . . | 8 | 3 |
| Average standard score decrease . . | 2.9 | 2.0 |
| Average loss in growth in years . . | 1.4 | 1.2 |
| Number of students showing no change in standard score . . | 1 | 3 |

However, there was a difference in the average standard score increase among the twenty lowest performing students. The average gain of the eleven participants who did show growth was greater than that for the group of nonparticipants. The participant groups' gain was 4.5 points while the fifteen nonparticipants who displayed growth in their scores had an average gain of 3.9 standard score units. According to the writers of the ITED "Manual for Administrators and Testing Directors" (1983), a gain of two standard score units represents a year's growth (or sufficient growth for one year) from one year's testing to the next for students' standard scores. This difference between the average gain scores of the two groups indicates approximately one-third (.6/2.0) more of a year's growth favoring the participant group. After scoring poorly on the 1982 test, the participants were able to achieve a higher gain as a group than the nonparticipants. The exposure to the science curriculum may have been a contributing factor.

The mean natural science standard score for the entire class of 355 students in 1982 was 13.3. When they were tested again in 1983, the mean rose to 15.5 indicating that as a group they were able to attain statistically-sufficient growth scores. As a group, the students who participated in the contemporary science course and were included in this study had a mean of 15.44 in 1983. The nonparticipants had a mean of 14.23. The mean of the 1982 scores for the matched pairs was 13.15 as compared to a mean of 13.3 for the entire class. The lower score of the sample group at the outset of the study may explain the lower average score of this group on the 1983 test administration.

A comparison of the means shows the course participants were over one-half year ahead of the students who did not receive any science instruction during the 1982-83 school year. The participant group's mean standard score increased 2.29 while the nonparticipant group's mean rose 1.08. Thus the difference between the two groups was 1.21 standard score units. According to the growth guidelines, the participant group displayed more than normal growth, while the nonparticipant group exhibited slightly more than one-half year's growth. From the matched pairs, twenty-seven (44%) participants and twenty-one (34%) nonparticipants scored above the 1983 total class mean.

The examination of individual standard scores (Appendix D) shows that thirty-nine or 69% of the participants in the contemporary science curriculum scored two or more units higher on the 1983 administration of the ITED than they did on the 1982 administration. This compares to thirty-one or 51% of the nonparticipants. Further analysis of these scores (see Figure 4) reveals that seven participants exhibited a gain of six or more units meaning they attained the equivalent of at least three years growth. Eight participants show growth of two and one-half years while nine displayed two years growth.

In contrast, in the nonparticipant group, there were five students who attained three years growth or more, four students achieving two and one-half years and seven showing two years growth.

Fewer contemporary science students made lower standard scores in 1983 than did nonparticipating students. Their decrease in scores was also not as great as the nonparticipant group. Fifteen

Figure 4

Number of Students Displaying Sufficient* or More Than Sufficient Growth on ITED Standard Scores

| Standard Score Gain | Years of Growth | Number of Participants | Number of Nonparticipants |
|---|---|---|---|
| 2.0 | 1.0 | 6 | 7 |
| 3.0 | 1.5 | 9 | 8 |
| 4.0 | 2.0 | 9 | 7 |
| 5.0 | 2.5 | 8 | 4 |
| 6 or more | 3.0 or more | 7 | 5 |

*Sufficient (+2 Standard Score Units)

Figure 5

Number of Students Showing a Decrease in Standard Scores

| Standard Score Loss | Number of Participants | Number of Nonparticipants |
|---|---|---|
| -1 | 2 | 6 |
| -2 | 5 | 2 |
| -3 | 4 | 4 |
| -4 | 2 | 1 |
| -5 | - | 1 |
| -6 | - | - |
| -7 | 1 | 2 |
| -8 | 1 | 1 |
| -9 | - | - |
| -10 | - | - |
| -11 | - | - |
| -12 | - | - |
| -13 | - | 1 |
| Total | 15 | 18 |
| Average loss = | 3.1 | 3.7 |

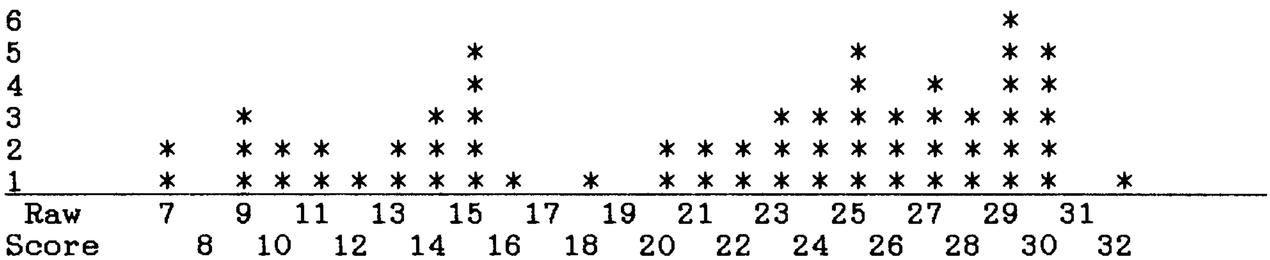
participants had lower standard scores in 1983 than they did in 1982 (see Figure 5). Eighteen members of the nonparticipant group made lower scores within this same period. The participant group of fifteen had an average loss of 3.1 standard score points while the non-participants average loss was 3.7 points.

Figures 6 and 7 show the raw score histograms based on participants and nonparticipants performance on selected ITED natural science questions. These statistical data were used to test the third research question which asked if tenth grade students who participated in the contemporary science course would score higher on selected items from the ITED natural science subtest than those students who did not participate. The "t" value is significant at the .05 level. This finding supports the second hypothesis (page 4).

Examination of Figures 6 and 7 shows that twelve science class participants scored higher than one standard deviation above the mean (28.49), while only three nonparticipants scored above the same score. Closer examination of Appendix D shows that the three nonparticipants who scored the highest in their group on the selected questions also received the three highest standard scores on the 1982 administration of the ITED. Their vocabulary scores were also at or near the top of the ITED distribution in 1982. This would seem to indicate that these students had the potential for scoring high on the tests without exposure to a science curriculum during their ninth grade year. Two of these three students also achieved higher scores than their matched pairs counterpart on the 1983 administration of the test.

Figure 6

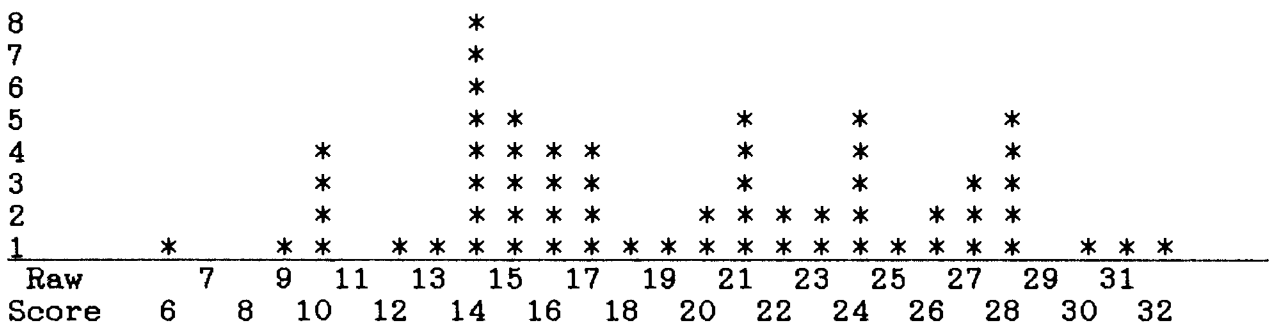
Raw Scores on Selected Natural Science Questions for Those Students Who Participated in Contemporary Science




Mean = 21.15 Std. Dev. = 7.34 Sample Size = 61

Figure 7

Raw Scores on Selected Natural Science Questions for Students Who Did Not Participate in Contemporary Science




Mean = 19.38 Std. Dev. = 6.28 Sample Size = 61

Table 2

Comparison of 1983 Raw Scores

| Group | N | Mean | Standard Deviation | t | Level of Significance |
|---|---|---|---|---|---|
| Participants | 61 | 21.15 | 7.34 | 2.33* | .05 |
| Nonparticipants | 61 | 19.38 | 6.28 | | |

*Significant at the .05 level

Of the twelve participants who scored above one standard deviation on their raw scores (Fig. 6), eight of them scored above his/her 1982 counterpart on the 1983 ITED standard scores. One scored the same as their counterpart. Of the three participants who scored less on the standardized test than their counterpart, two of them were matched with two of the three nonparticipants who scored above one standard deviation on the ITED. Thus, these data would seem to reinforce the earlier idea that high vocabulary scores and/or test-wiseness would enable students to do well on a test even though they had not been exposed to the material. Also, this information suggests that the exposure to a science curriculum helped those students who scored near the mean to outperform their counterparts who had no exposure to the science curriculum.

As with the standard scores, the participant group scores were more widely spread than the nonparticipants (Figure 6 and 7). At five points on the score scale below the mean, there were four or more nonparticipants. A similar grouping of scores occurs only once for students who were participants in the contemporary science course. The number of students who fell within one standard deviation above and below the respective means differs greatly, although the total number of students was almost equal. There were thirty-seven or 61% of the participant group within one standard deviation above or below the mean. There were thirty-nine or 64% of the nonparticipants within the same span. The difference between the two groups occurs in the direction of these scores. Twenty-three or 38% of the participants were one standard deviation above the mean, while only seventeen or 28% of the nonparticipant group could be so located. Conversely, fourteen or 23% of the participating students fell between the mean and one standard deviation below the mean while twenty-two or 36% of the nonparticipating group were located between the same score points.

The larger spread and the lack of clustering of scores near the middle of the distribution by the students who participated in the contemporary science program would suggest that they, as a group, were exposed to an environment more conducive to learning science. The ITED's are designed to measure a student's general knowledge of science terms and principles and his/her ability to interpret and evaluate material related to science. The data suggests that a year of experience in a science course structured along similar objectives enables the participant group to respond with more correct answers than the nonparticipant group.

Figure 8 displays data that were compiled when comparing the magnitude of differences between the matched pairs on their standard scores and their subsequent performance on selected contemporary science questions. As explained earlier, students who enrolled in the contemporary science course were matched with students who chose not to participate during their ninth grade year. There was no difference in their respective 1982 ITED natural science scores when this study began.

The numbers on the left in the "Increase" column represent the number of points that a particular student scored above his/her matched counterpart on their 1983 ITED standard scores. For example, there were three participants who scored seven units better on their standard score than did the nonparticipants with

Figure 8

1983 ITED Standard Score Differences

Participants

| Standard Score Increase Over Matched Counterpart | Number Scoring Better Than Matched Counterpart | Raw Score Differences Between Matched Pairs |
|---|---|---|
| +1 | 6 | 0, +1, +2, +6, -2, -1 |
| +2 | 4 | -3, +3, +1, +3 |
| +3 | 4 | +4, -1, +4, +4 |
| +4 | 3 | +3, +4, +3 |
| +5 | 6 | +5, +2, +12, +3, +2, +4 |
| +6 | None | ----- |
| +7 | 3 | +13, +8, +8 |
| +8 | 2 | +10, +11 |
| +9 | 3 | +15, +14, +10 |
| +10 | 1 | +9 |
| +11 | 1 | +19 |

Average increase = 6.5
Average decrease = 1.0

Nonparticipants

| Standard Score Increase Over Matched Counterpart | Number Scoring Better Than Matched Counterpart | Raw Score Differences Between Matched Pairs |
|---|---|---|
| +1 | 5 | -2, +1, +1, +1, +3 |
| +2 | 3 | -2, +2, +1 |
| +3 | 2 | +3, +5 |
| +4 | 2 | +8, +1 |
| +5 | 5 | +9, +4, +1, +4, +8 |
| +6 | 2 | 0, +4 |
| +7 | None | ----- |
| +8 | 1 | +4 |
| +9 | 1 | +10 |

Average increase = 3.9
Average decrease = 2.0

whom they were matched using 1982 standard scores. All three participants made higher raw scores on the selected questions (+13, +8, +8).

The overall average in raw score increases made by participants who scored higher on their 1983 ITED standard score than their counterpart was 6.5. The average for the nonparticipants who outscored their counterpart on the 1983 administration of the ITED was 3.9. This is a difference of 2.6 in raw score points, a difference of forty per cent in raw scores favoring the participant group over the nonparticipants group. There was an average difference of one point between contemporary science students who outperformed their counterpart on the 1983 ITED standard score and subsequently did less well on the science test comprised of ITED items. The nonparticipants were less able to maintain their favored position. When their score was lower, it was on the average two points lower than the score of the person they outscored on the ITED. Thus, the participant group was able to do better on their overall raw scores regardless of how they fared on the ITED. Their participation in the contemporary science course serves as one explanation for this outcome.

The difference in raw scores was even more apparent when the standard score increase over a matched counterpart was seven points or greater. First, there were ten participants who scored seven or more standard score points better than their counterparts. There were only two nonparticipants who did as well. The average raw score difference in this particular subgroup was nearly twelve (11.7) points for the participant group and seven for the nonparticipants. Both the number and the magnitude of the raw score differences favored the participant group.

The data presented in this study indicate that ninth grade students who participated in a contemporary science course performed better on the natural science section of the ITED than did students who chose not to participate in the course. There was statistical support for both hypotheses that predicted greater achievement in science knowledge by those students who participated in a contemporary science program as ninth graders than by those students who chose not to participate. The first hypothesis was tested by using student standard scores on the natural science section of the ITED. The second hypothesis was investigated by using selected questions from the natural science section of the ITED. These questions were deemed by at least three of the four instructors of the contemporary science course to measure course objectives.

The tables summarizing the data collected by the researcher show that students who were enrolled in the science course made greater gains on their ITED standard scores and in the number of selected questions they were able to answer correctly than did a comparable group of students who were not enrolled. When a nonparticipant did show more growth than his/her matched counterpart, the differences in their scores were not as great. Nonparticipants who had attained high vocabulary scores seemed to possess the potential for scoring high on the science test even though they had not been a participant in the contemporary science program. Students who participated and scored near the mean of the ITED in 1982 consistently surpassed their counterparts on both the 1983 administration of the ITED and on selected questions from this same test.

## Chapter 5

## Summary and Recommendations

The purpose of this study was to examine the impact a contemporary science course had on the growth of students in the knowledge and application of scientific ideas and priciples. The Ottumwa school administration felt that scores attained on a standardized achievement test were not satisfactory, in this instance, the Iowa Tests of Educational Development (ITED). According to the administration, the gain scores for students' on the ITED natural science portions of the tests between the ninth grade and tenth grades were not satisfactory. They also felt that there was an unsatisfactory decline in scores as a particular class progressed through four years of high school.

To investigate this situation, this researcher conducted two separate studies, reviewed the pertinent professional literature and examined the achievement test results in his school system. The literature review was devoted to a study of the purposes for administering standardized tests, the factors that influence test results and the proper uses of such results. This was done to examine the factors that influence pupil performance on standardized tests and the corresponding cautionary measures for using these results. The examination of the school system's test results involved comparing data of two groups of tenth grade students; those who had participated in a contemporary science class as ninth graders and those who had not participated in that particular course as ninth graders.

The literature review indicated that one of the major purposes for using standardized tests was to allow the school to show "accountability" for the education of its students. Standardized test results provide a means of assessing both the academic progress of students and the effectiveness of a school's curriculum. The extent to which the objectives of a school's curriculum are similar to the measurement objectives of the test determines the validity of the test as a measure of these two conditions. When the objectives are known and the subsequent levels of achievement by the students are measured, the strengths and weaknesses of a school's curriculum can be identified. In addition, students in need of remedial assistance and those capable of advanced studies may be identified.

Interest in the factors that influence test results was amply evident in the literature. A preponderance of the articles were devoted to a consideration of the factors that enable students to do well on these tests. These conditions include awareness of test characteristics, control of test anxiety, proper health habits, previous learning experiences, motivation from home and school, and an overall preparation of the students for testing.

Some writers believed students could receive higher scores when they were knowledgeable about the characteristics of certain test situations that are related to test performance. Referred to as "test-wiseness," students are aware of the time allowed for a test, recognize the importance of following the directions on a given test, understand the different types of questions and their possible correlation to each other, and mark answers in a correct and neat manner.

Help from parents, teachers and a school's administration can also be positive influences on student scores. This source of assistance can occur in many forms. For example, the tests can be explained in a manner all students understand. Provisions can be made for adequate lighting, comfortable temperatures and no distractions during the testing period. Observations before and during testing by both parents and teachers may help identify anxiety in students that might hamper performance. Counseling and/or therapy can be obtained for students who seem overly anxious. Exposure to both the subject matter being measured and experience taking standardized tests may help prepare students for testing. Also, the motivation of students themselves can be increased by conveying the importance of the tests.

The use and misuse of standardized tests was the third and final focus of this writer's review of the literature. Most writers agreed that these tests should not be used as exclusive measures of student achievement. They should only be another "snapshot" to use in achieving an overall picture of student progress and an assessment of the school's educational program. Teacher observations, daily assignments, and teacher made tests should be used in combination with standardized tests in the continuous evaluation of pupil progress and curriculum decisions.

Information derived from results of achievement tests may help improve instruction by describing the current level of knowledge and skills of individual students and/or summarizing the performance of an entire class of students. However, people involved in the use of these test results should realize that test scores only represent a student's ability to handle a given set of learning materials on a given day.

The major caution is to be certain that test results alone are not being used for evaluative purposes. School personnel may use test results for confirming judgments about individual and/or groups of students or to identify possible errors in the placement of students. The results may also be used to compare a school's curriculum with the curricular objectives of the testing program. The standardization of well-designed achievement tests establishes both the reliability and objectivity needed for such comparisons.

One of the most valuable ways to use test results may also be the most hazardous. When test results are used to inform the public of a school system's program effectiveness it is important they are also informed of how to interpret these results. Consideration should be given to the variables that affect students' scores and changes that have occurred and can be expected to occur in the school curriculum to coincide with the objectives of the testing program.

This study attempted to find out if there was a significant difference between the scores achieved by students who took the ninth grade contemporary science course and those who did not. The researcher reasoned that the participation in this course would allow students to achieve higher scores on the natural science subtest of the Iowa Tests of Educational Development and on selected items from the same subtest.

The original population consisted of all students in Ottumwa High School who were ninth graders during the 1982-83 school year and tenth graders during the year 1983-84. From this population, two groups were formed. One group included students who participated in a contemporary science course. The other group was comprised of students who chose not to participate in the contemporary science course. Sixty-one pairs of students were obtained by exactly matching the 1982 ITED natural science scores of nonparticipants with participants and allowing a variance of one (±) between their vocabulary scores.

The researcher found a significant difference in the means of the 1983 ITED natural science standard scores of the two groups. The researcher also found a significant difference in scores of the two groups based on questions selected from the ITED natural science test.

Since the beginning of this study and after reasearch data had been obtained, the Ottumwa High School administration has made changes in the testing procedure that were intended to raise student performance on the ITED. Some of these changes would have been recommendations by this researcher upon completion of this study. One of the changes that has been instituted that is consistent with a finding of this study is also recommended in the Manual for Administrators and Testing Directors of the ITED (1982). The authors of this manual indicate that the tests should be administered under uniform conditions to insure comparable and meaningful test scores and norms. Some faculty members asserted that some scheduling was not being done in accordance with these specifications. This contention was accepted by the administration during the 1983-84 school year.

During discussions to change the way the tests were given, this researcher made available to the faculty some of the recommendations by writers of standardized testing. That does not mean that this study was largely responsible for the changes, but the information made available did contribute to some of the decisions. The literature review performed as part of this study, particulary the factors associated with the preparation of students for the tests and the administration of the tests, did come to light during the course of faculty discussions. The aforementioned factors were presented to the administration and resulted in the following changes.

1. In previous years, at Ottumwa High School, all the tests were administered in one day. This was the case during the two years that provided the data that were used in this study. In September of 1984, the school system began administering the tests during the mornings on two successive school days. This is consistent with the recommendations from the authors of the ITED who recommend that the schedule consist of two or three sessions.

2. The natural science section of the test was given to the students at the beginning of the second day of testing rather than at the end of a single day of testing as was the practice in past years. This allowed the students to take the natural science subtest in the morning when they were not tired from a full day of testing. Again, no tests were administered in the afternoon on either day of testing.

3. Brief meetings were held with students in their homerooms (sites of the testing) to convey to them the purposes of the tests. They were informed of ways the test results could be of some assistance in their educational planning. A meeting was also held to inform faculty of their role in conveying the importance

of the tests to students. As a result of these measures, this researcher detected a more positive attitude on the part of all parties during the 1984 fall testing program. These changes serve as an alternative explanation to actual achievement gains as the reason student natural science scores increased during the 1984 administration of the ITED. In fact, each class of students showed an increase on all the tests and the composite score.

Figure 9 displays data collected from the past four years of testing. It can be seen by reading diagonally from the upper left to the lower right, how a class progresses through its four years of ITED testing. The numbers in parentheses represent the gain in the scores from the first administration year, the upper left, to the most recent administration year, the lower right, respectively. The group of students who comprised the population for this study had natural science scores of 13.3 (9th grade, 1982), 15.5 (10th grade, 1983) and 18.1 (11th grade, 1984). This particular class was the first to participate in the contemporary science course. Of the 355 students in this class, 265 participated in the contemporary science course. Their eleventh grade scores were not used in this study. Thus, the data that revealed a significant difference in ITED scores favoring students who took contemporary science was not affected by the new testing procedures.

The students in this study (1982-84) made successive year gains greater than the expected 2.0 standard score points as suggested by the ITED authors. Their gain was even greater on the 1984 testing when suggested changes in the testing procedure took place. When comparing their scores to other classes of students,

Figure 9

Standard Score Averages and Gains from the Natural Science
Section of the ITED over a Four Year Period
for Grades Nine through Twelve

Year of Testing

| GRADE LEVEL | 1981 | | 1982 | | 1983 | | 1984 |
|---|---|---|---|---|---|---|---|
| 9 | 13.3 | | 13.3 | | 13.5 | | 14.6 |
| | | (1.0) | | (2.2) | | (2.4) | |
| 10 | 14.5 | | 14.3 | | 15.5 | | 15.9 |
| | | (1.5) | | (1.7) | | (2.6) | |
| 11 | 16.4 | | 16.0 | | 16.0 | | 18.1 |
| | | (0.5) | | (0.7) | | (1.5) | |
| 12 | 17.8 | | 16.9 | | 16.7 | | 17.5 |

differences can be noted. Smaller gains are evident with the three previous classes whose data is available. The class of students one year ahead of the experimental group (1981-84) had gain scores of 1.0, 1.7 and 1.5 between each of the four years they were tested. These students did not have the benefit of the contemporary science program during their ninth grade year. They were also subjected to three years of single-day testing before the change to morning sessions during their twelth grade year. Their decrease in scores between their eleventh and twelth grade years corresponds to the decrease in scores during a similar time period for the two previous classes (1981-83 and 1981-82). The group who participated in the testing program for the first time during their tenth grade year (1981-83) had an initial gain of 1.5 standard score points and then dropped to 0.7 points between the eleventh and twelth grades. The 1981-82 group shows an even smaller gain of 0.5 points.

Conversely, the class that took the natural science test a year after the experimental group (1983-84), recorded an even higher gain of 2.4 points. This group had both the benefit of the contemporary science program and the revised schedule of testing. The present ninth grade class, with a score of 14.6, tested higher than any of the three previous classes. They had the benefit of achieving their initial score under improved testing conditions.

The data collected as part of this study, and the data summarized in Figure 9, indicate that greater exposure to a subject area and improved testing procedures aid in raising the average scores of students on the Iowa Tests of Educational Development.

While the study was still in progress, the school board decided to increase the science requirement to two years of science rather than just one year. As a result, the number of students who do not take contemporary science decreased to less than five per cent of the ninth grade population.

The author believes that the need for these and other changes is supported by the research literature and subsequent results of this study. There are changes that remain to be considered which have promise of further increasing student performance on the ITED and particularly the natural science subtest. It is in this spirit that the following recommendations are made:

1. According to Tyler (1971), one of the limitations of standardized achievement tests is that they do not measure what students have learned but rather where they stand on a scale that arranges those who have taken the test from the highest score to the lowest. Students are participating in a norm-referenced testing program. This means that when a group of people are ranked on a percentile scale, half of the group will always have a rank of 50% or lower. This ranking is actually a listing of students from high scores to low scores by percentiles.

The authors of the ITED (Manual, 1982) suggest that a school system determine the reasonable expectations of its students. This might be achieved by collecting data on factors such as the average intelligence of its students, the socioeconomic level of the community, the average educational level of parents, the average level of training of the faculty and other factors which are relad to pupil achievement. An estimate based on these considerations would at least be a positive step in determining a reasonable standard with which to compare future results.

Research should then be conducted to determine if the scores attained by students in this particular school district on the Iowa Tests of Educational Development are defensible/reliable measures of their educational development when using the Iowa norms as a basis for comparison.

2. As suggested by Rudman (1982) in Chapter 2, the school system should fully undertake a systematic examination of all data available using his three step analysis (diagnosis, evaluation and planning). The diagnosis of test results or first step, has been begun by this study. Some possible weaknesses in the curriculum have been identified. The administration has identified the ninth grade math program and students' performance on the literary materials section of the ITED as areas in need of further study. The second step or evaluation has been partially begun by using procedures employed in this study and those the math department has instituted in its own pre-test:post-test study. The evaluation should proceed by estimating the level of achievement of the students and comparing it with their actual performance on those objectives deemed important by the school district. The third step, or planning, involves preparation to make the most effective use of the information gathered.

3. The entire school system should institute a program that will help students become more test-wise. These skills are useful during normal everyday classroom evaluations. As shown earlier in Chapter 4, some students are capable of scoring high on achievement tests even though they did not participate in that

particular subject area. Students should always have the opportunity to express the full range of their knowledge and skill in all testing situations, from chapter quizzes to large-scale achievement tests.

4. Score information and data should be distributed and more fully explained to all teachers involved in administering the ITED's. Teachers need to have a better understanding of the test and the use of the results. This information would include ways teachers may use the results in their curriculum planning and in the diagnosis and assessment of student learning. In-service time should be devoted to helping teachers become aware of and make use of test information in their evaluation of students and their course content. A better-informed teacher may also help a student make more effective use of this information. Teachers should join with the school counselors and administrators to help individual students see the variety of choices available to them and the way test results can be used to aid them in making decisions.

5. The administration has recognized that these tests can be used as a diagnostic tool for identifying a student's level of achievement in a particular subject matter area. They can also be used by teachers to create instructional activities that will offset information and skill deficiencies that these types of tests measure. Teachers may also want to use sample tests that are similar in format to the ITED and call for intellectual information processing skills that are required on the ITED.

Further studies with similar aims may utilize various aspects *of this research. The impact that differences in teaching* techniques and curricular emphases have on student achievement may be investigated. Since this is a large school district, not all the students will receive identical science instruction.

A comparison of students' junior high grades and their Iowa Basic Skills scores might also be compared to their subsequent ITED results.

6. A study identical to this one is not possible. Due to the two-year science requirement now in effect, the number of students in the nonparticipant group will not be large enough. However, some predictions about school system scores might be made and tested based on extrapolations of the data collected for this study. Also, an identical study for other curricular areas of the school program might be useful.

# BIBLIOGRAPHY

Anderson, L.W. Assessing affective characteristics in the schools. Boston: Allyn and Bacon, 1981.

Anderson, Scarvia B. Standardized testing has become education's latest scapegoat. The American School Board Journal, 1981, 26-28.

Baker, Eva L. Issues in achievement testing. Center for the Study of Evaluation, 1982, (ERIC Document Reproduction Service No. ED 228 277).

Bangert-Drowns, Robert L. & Others. Highlights from research on the effects of coaching for tests. Association for Supervision and Curriculum Development, 1984, (ERIC Document Reproduction Service No. ED 238 910).

Barnes, Ronald E.; Moriarty, Karen & Murphy, John. Reporting testing results: The missing key in most testing programs. NASSP, 1982, 14-20.

Biehler, Robert F. Psychology applied to teaching. Boston: Houghton Mifflin Company, 1978, 630-631.

Cole, Michael. Culture, cognition and IQ testing. The National Elementary Principal, 1975, 49-52.

Cole, Robert W. Jr. A new recipe for a kettle of fish. Phi Delta Kappan, 1981, 618..

Ducote, Kenneth J. Motivation and Test-Wiseness Interactions. Paper Presented at the annual meeting of the Soutwest Educational Research Association. 1982, (ERIC Document Reproduction Service No. ED 222 521).

Ebel, Robert L. The social consequences of educational testing. Proceedings of the 1963 Invitational Conference on Testing Problems, Educational Testing Services, 1964.1963.

Ebel, Robert L. Comments on Green's attack on tests and testing. Phi Delta Kappan, 1975, 93.

Ebel, Robert L. Objections to external testing - Fact or fallacy? NASSP BULLETIN, 1982, 3-10.

Farr, Roger. Testing: There's more to it than meets the eye. Instructor, 1980, 26,28.

Forsyth, Robert. Director of Iowa Tests of Educational Development, Iowa City, Ia. Personal communication, 1983.

Gerlovich, Jack. Science Consultant for the Iowa Department of Publick Instruction, Personal correspondence, 1983.

Green, Robert L. Tips on Educational Testing: What teachers and parents should know. Phi Delta Kappan, 1975, 89-93.

Hanson, Robert W. A comparison of science programs in Iowa secondary Schools ranking high and low in science achievement. Doctoral Dissertation, State University of Iowa, 1961, 176-178..

Hodgkinson, Harold L. The school administrator and standardized testing. NASSP BULLETIN, 1982, 21-27.

Holley, Freda M. A Sunn day for education. Educational Measurement: Issues & Practice, 1983, 31-32.

Iowa Tests of Educational Development. Form X-7, Grades 9-12, 1983, 38-51.

Iowa Tests of Educational Development, Manual for Administrators and Testing Directors, 1984.

Iowa Tests of Educational Development, Manual for Teachers, Counselors and Examiners 1984.

Kearney, C. Phillip. Uses and abuses of assessment and evaluation data by policymakers. Educational Measurement: Issues and Practice, 1983, 9-12,17.

Laughland, Arther S. & Corbett, William. Two principals look at standardized test. The National Elementary Principal, 1975, 88-90.

Lazarus, Mitchell. Evaluating educational programs. American Association of School Administrators, 1982, (ERIC Document Reproduction Service No. ED 226 414.

Magrane, George, Science Consultant of the Southern Prairie Area Education Agency #15, Personal consultation, 1983.

Massachusetts State Department of Education. Interpreting and using commercial achievement test results. Basic Skills Improvement Policy: Supplement to Implementation Guide 1. 1982, (ERIC Document Reproduction Service No. ED 221 589)

Mehrens, William A. & Lehmann, Irvin J. Measurement and Evaluation in Education and Psychology, New York: Holt, Rinehart and Winston, 1978..

Merwin, Jack C. Standardized tests: One tool for decision making in the classroom. Educational Measurement, 1982, 14-16.

Millman, Jason; Bishop, Carol H. & Ebel, Robert. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 707-726.

National Commission on Excellence in Education. A Nation at Risk: The Imperative for Educational Reform. An examination of the American educational system and recommended reforms, Secretary of Education, Washington, D.C., 1983.

Nimmer, Donald N. The use of standardized achievement test batteries in the evaluation of curriculum changes in junior high school earth science. Science Education, 1982, 45-48.

National Study of School Evaluation. North Central Association Report, 1980. Evaluative Criteria, 2201 Wilson Boulevard, Arlington, Virginia, 22201, 13-28.

Osterhouse, Robert A. Classroom Anxiety and the examination performance of test anxious students. The Journal of Education Research, 1975, 247-249.

Patterson, Jerry L. and Others. How to avoid the dangers of testing. The National Elementary Principal, 1975, 93-94.

Resnick, Dan. Testing in America: A supportive environment. Phi Delta Kappan, 1981, 625-628.

Rudman, Herbert C. How to use standardized tests for decision making. National Elementary Principal, 1982, 59-64.

Sabers, D. Test-taking Skills. Tucson: Arizona Center for Educational Research and Development, 1975. (ERIC Document Reproduction Service No. ED 133 634)

Salmon-Cox, Leslie. Teachers and standardized achievement tests: What's really happening? Phi Delta Kappan, 1981, 631-634.

Sigda, Robert B. The crisis in science education and the realities of science teaching in the classroom. Phi Delta Kappan, 1983, 624-6211

Sproull, Lee & Zubrow, David. Standardized testing from the administrative perspective. Phi Delta Kappan, 1981 628-631.

Thorndike, E.L. Education, a First Book, New York: The MacMillan Co., 1912, 7-8.

Thorndike, Robert L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 63-70.

Tryon, G.S. The measurement and treatment of test anxiety. Review of Educational Research, 1980, 343-372.

Tyler, Ralph W. The objectives and plans for national assessment of educational measurement. Journal of Educational Measurement, 1966, 1-4.

Webb, Melvin W., II. A scale for evaluating standardized reading tests with results for Nelson-Denny, Iowa and Stanford. Journal of Reading, 1983, 424-429.

Wildavsky, Aaron. A program of accountability for elementary schools. Phi Delta Kappan, 1970, 212-216.

# APPENDIX A

The natural science section of the Iowa Tests of Educational Development is described as follows:

Test NS: Natural Sciences Part 1: Concepts and Background

This test is designed to measure general knowledge and understanding of scientific terms and principles, common phenomena and industrial applications, appropriate scientific procedures, and the place of science in modern civilization. The test content is drawn from both the physical sciences (physics, astronomy, chemistry, and earth sciences) and the biological sciences (zoology, botany, health and medicine). The student's information and ideas in these areas are acquired through school instruction as well as through experiences outside of school - in the pursuit of hobbies, in reading, and in observation of everyday phenomena. Thus, the test exercises are not restricted to the subject matter of specific high school courses but rather reflect the breadth of the student's background in the natural sciences in general. lf course, a test of this length (twenty-four exercises for each level) can be considered no more than a very small sampling of the vast array of important facts and concepts encompassed in this content area. (Manual for Teachers, Counselors and Examiners, p. 29, 1983)

Part 2: Natural Sciences Reading

This part of Test NS is intended to measure the student's ability to interpret and evaluate two types of material. The first type is primarily expository in its purpose. Selections drawn from textbooks, references, and scientific articles in newspapers and periodicals are examples of the type. The second is concerned with the description of actual experiments and the reporting of experimental data. In reading a selection of the first kind, the student must understand concepts; must interpret, extend, and apply scientific principles; and must grasp the overall content of the selection. In responding to items based on the second type of passage, background knowledge must be utilized to greater extent, in interpreting experimental findings, in demonstrating mastery of scientific method, and in using data to formulate conclusions and generalizations.

The science fields represented include geology, biology, zoology, chemistry, physics, astronomy, and health. In contrast to the social studies materials, the natural science selections tend to require closer attention to detail and to exact meanings, are more explicit and definite, demand the exercise of more rigorous logic, and are more straightforward in style and more objective in character. (Manual for Teachers, Counselors and Examiners, p. 31, 1983)

The vocabulary section of the Iowa Tests of Educational Development is described as follows:

> Test V: Vocabulary
>
> Experience in intelligence testing has shown that a test of general vocabulary can serve as one of the best predictors of future success in school work. It cannot be claimed that such a test measures pure native ability. No existing test does this. But it is probably true that the vocabulary test is less sensitive to curriculum factors than are the other tests in the ITED battery. Hence, the score on Test V may be cautiously used as a base to which performance on the other tests may be compared. (Manual for Teachers, Counselors and Examiners, p. 36, 1983)

# Appendix B

Natural Science and Vocabulary Scores of the Matched Pairs

| | Students who participated in the contemporary science program | | Students who did not participate in the contemporary science program | | |
|---|---|---|---|---|---|
| | 1983 Natural Science Score | 1982 Vocabulary Score | Common 1982 Natural Science Score | 1982 Vocabulary Score | 1983 Natural Science Score |
| 1. | 19 | 16 | 12 | 15 | 17 |
| 2. | 5 | 9 | 9 | 9 | 10 |
| 3. | 16 | 13 | 12 | 13 | 9 |
| 4. | 7 | 13 | 7 | 14 | 7 |
| 5. | 11 | 13 | 8 | 14 | 9 |
| 6. | 21 | 17 | 14 | 18 | 12 |
| 7. | 7 | 5 | 9 | 5 | 12 |
| 8. | 16 | 8 | 14 | 9 | 7 |
| 9. | 24 | 18 | 20 | 17 | 20 |
| 10. | 21 | 18 | 20 | 17 | 23 |
| 11. | 25 | 19 | 19 | 19 | 18 |
| 12. | 10 | 9 | 8 | 8 | 12 |
| 13. | 21 | 16 | 18 | 16 | 22 |
| 14. | 15 | 13 | 13 | 13 | 18 |
| 15. | 16 | 16 | 13 | 15 | 15 |
| 16. | 15 | 15 | 11 | 16 | 11 |
| 17. | 20 | 13 | 7 | 12 | 17 |
| 18. | 28 | 18 | 23 | 18 | 25 |
| 19. | 14 | 9 | 11 | 9 | 4 |
| 20. | 19 | 18 | 16 | 18 | 14 |
| 21. | 21 | 14 | 16 | 13 | 13 |
| 22. | 18 | 15 | 20 | 15 | 7 |
| 23. | 18 | 19 | 20 | 18 | 15 |
| 24. | 18 | 14 | 8 | 15 | 10 |
| 25. | 20 | 19 | 16 | 19 | 20 |
| 26. | 9 | 10 | 12 | 11 | 13 |
| 27. | 26 | 18 | 14 | 17 | 21 |
| 28. | 23 | 24 | 25 | 25 | 28 |
| 29. | 6 | 9 | 9 | 8 | 11 |
| 30. | 16 | 14 | 12 | 14 | 11 |
| 31. | 17 | 14 | 12 | 15 | 17 |
| 32. | 15 | 17 | 12 | 16 | 11 |
| 33. | 20 | 16 | 15 | 16 | 19 |
| 34. | 16 | 16 | 12 | 15 | 14 |
| 35. | 5 | 21 | 12 | 22 | 13 |
| 36. | 15 | 16 | 11 | 17 | 17 |
| 37. | 6 | 9 | 7 | 10 | 10 |
| 38. | 24 | 18 | 20 | 18 | 17 |
| 39. | 25 | 20 | 21 | 21 | 20 |
| 40. | 20 | 16 | 15 | 16 | 20 |

..... Appendix B continued .....

| | Students who participated in the contemporary science program | | Students who did not participate in the contemporary science program | | |
|---|---|---|---|---|---|
| | 1983 Natural Science Score | 1982 Vocabulary Score | Common 1982 Natural Science Score | 1982 Vocabulary Score | 1983 Natural Science Score |
| 41. | 11 | 14 | 11 | 14 | 14 |
| 42. | 18 | 16 | 13 | 16 | 15 |
| 43. | 19 | 22 | 20 | 22 | 20 |
| 44. | 9 | 13 | 6 | 14 | 9 |
| 45. | 10 | 8 | 7 | 7 | 11 |
| 46. | 3 | 11 | 8 | 11 | 12 |
| 47. | 9 | 11 | 10 | 10 | 10 |
| 48. | 22 | 18 | 20 | 18 | 17 |
| 49. | 13 | 15 | 12 | 15 | 12 |
| 50. | 5 | 4 | 9 | 4 | 5 |
| 51. | 16 | 18 | 15 | 19 | 7 |
| 52. | 7 | 11 | 10 | 10 | 13 |
| 53. | 7 | 6 | 6 | 5 | 5 |
| 54. | 9 | 9 | 6 | 8 | 10 |
| 55. | 23 | 17 | 16 | 18 | 18 |
| 56. | 15 | 14 | 10 | 13 | 14 |
| 57. | 15 | 11 | 16 | 12 | 15 |
| 58. | 12 | 14 | 7 | 15 | 17 |
| 59. | 11 | 13 | 10 | 12 | 10 |
| 60. | 17 | 15 | 16 | 15 | 16 |
| 61. | 23 | 20 | 21 | 19 | 29 |

# Appendix C

Selected Natural Science Item Scores of the Matched Pairs
Thirty-three Possible

| | Students who participated in the contemporary science program 1983 Raw Score | Students who did not participate in the contemporary science program 1983 Raw Score |
|---|---|---|
| 1. | 24 | 27 |
| 2. | 7 | 16 |
| 3. | 27 | 14 |
| 4. | 9 | 10 |
| 5. | 15 | 12 |
| 6. | 29 | 14 |
| 7. | 9 | 13 |
| 8. | 24 | 10 |
| 9. | 28 | 25 |
| 10. | 30 | 28 |
| 11. | 29 | 21 |
| 12. | 15 | 17 |
| 13. | 29 | 27 |
| 14. | 20 | 23 |
| 15. | 20 | 20 |
| 16. | 22 | 18 |
| 17. | 27 | 23 |
| 18. | 29 | 30 |
| 19. | 15 | 6 |
| 20. | 27 | 22 |
| 21. | 26 | 16 |
| 22. | 29 | 10 |
| 23. | 25 | 21 |
| 24. | 25 | 14 |
| 25. | 30 | 26 |
| 26. | 9 | 17 |
| 27. | 30 | 28 |
| 28. | 30 | 31 |
| 29. | 10 | 14 |
| 30. | 26 | 14 |
| 31. | 24 | 26 |
| 32. | 23 | 20 |
| 33. | 29 | 28 |
| 34. | 22 | 21 |
| 35. | 11 | 15 |
| 36. | 23 | 24 |
| 37. | 13 | 14 |
| 38. | 32 | 24 |
| 39. | 30 | 27 |
| 40. | 25 | 28 |

..... Appendix C continued .....

| | Students who participated in the contemporary science program<br>1983<br>Raw Score | Students who did not participate in the contemporary science program<br>1983<br>Raw Score |
|---|---|---|
| 41. | 14 | 19 |
| 42. | 25 | 21 |
| 43. | 27 | 28 |
| 44. | 15 | 15 |
| 45. | 13 | 14 |
| 46. | 7 | 17 |
| 47. | 14 | 15 |
| 48. | 26 | 24 |
| 49. | 18 | 16 |
| 50. | 10 | 10 |
| 51. | 25 | 15 |
| 52. | 15 | 15 |
| 53. | 12 | 9 |
| 54. | 11 | 14 |
| 55. | 28 | 24 |
| 56. | 23 | 17 |
| 57. | 21 | 21 |
| 58. | 16 | 24 |
| 59. | 14 | 16 |
| 60. | 21 | 22 |
| 61. | 28 | 32 |

# Appendix D

## Classification and Selection of Items from the Natural Science Section of the Iowa Tests of Educational Development

### Part I

| Form X-7 SKILL CLASSIFICATION | CONTENT AREAS: Biological Sciences | Physical Sciences | Methods of Scientific Inquiry |
|---|---|---|---|
| (Starred items are those chosen by the contemporary science instructors as being similar in content to the contemporary science curriculum) | | | |
| **1. Knowledge and Comprehension** | | | |
| To recognize valid generalizations, laws, principles, theories, etc.; to identify specific examples of phenomena, concepts, laws, principles, and generalizations; to identify appropriate techniques and procedures | 2*<br>4<br>12<br>16 | 13<br>17<br>23 | 5*<br>18*<br>20* |
| **2. Application of Principles and Generalizations to New or Specific Situations** | | | |
| To select appropriate explanation for a given situation by using general principles or broader categories of knowledge; to predict the probable consequences of a given situation by applying general principles and generalizations | 1<br>8<br>22 | 3<br>6<br>9*<br>21 | 11* |
| **3. Analysis and Evaluation** | | | |
| To differentiate among hypotheses, assumptions, data, and conclusions; to select appropriate conclusions on the basis of data and/or experimental information; to identify appropriate procedures for testing specific hypotheses; to judge relevance and appropriateness of procedures or data and to select the most important factor for reaching a specified conclusion | 7*<br>14<br>15<br>19* | 24 | 10 |

(Manual for Teachers, Counselors and Examiners, 1980, p. 30)

Appendix D (cont.)

Classification and Selection of Items from the Natural Science Section of the Iowa Tests of Educational Development

Part II

| SKILL CLASSIFICATION | Item Numbers - Form X |
|---|---|
| (Starred items are those chosen by the contemporary science instructors as being similar in content to the contemporary science curriculum) | |
| 1. Literal Comprehension | |
| To identify an explicitly stated detail; to recognize ideas restated in new words | 4*, 13*, 33 |
| 2. Interpretation | |
| To draw valid inferences and conclusions from information and relationships presented in the passage; to infer relationships, comparisons, cause and effect, sequence of events; to recognize verbal statements of information presented graphically, and graphical representaion of information presented in verbal statements | 2*, 3*, 9*, 10*, 14*, 21*, 29, 30, 32 |
| 3. Critical Reading (Application, generalization, analysis, and evaluation) | |
| A. To recognize the main thought of a selection or to define the problem of an experiment; to evaluate the importance of ideas and to see their implications; to abstract principal generalizations, relationships, and conclusions not stated in the selection; to extrapolate ideas to new or specific situations; to extend conclusions to related phenomena | 1*, 5*, 6*, 8*, 11, 17*, 18, 22*, 23*, 24*, 35, 36* |
| B. To recognize implicit assumptions and their point of application; to distinguish between fact and opinion, to differentiate hypotheses, assumptions, data and conclusions | 25, 26, 27, 34* |
| C. To judge the limitations of an experiment; to identify factors controlled or manipulated in an experiment; to use backgraound knowledge to explain aspects of an experiment | 7*, 12*, 15*, 16*, 19*, 20*, 28, 31* |

(Manual for Teachers, Counselors and Examiners, 1980, p. 33)

# Appendix E

## Ottumwa Science Curriculum Objectives

September, 1980

Goal (K-129: To develop a scientifically literate society.

Subgoal A: To apply science processes as a part of basic learning.

| Process Objectives | Suggested Placement* of Process Objectives within the Science Curriculum (K-3) | (4-6) | (7-9) | (10-12) |
|---|---|---|---|---|
| 1. OBSERVING: Using the senses to obtain information or data about objects and events. | I | M | M | E |
| 2. CLASSIFYING: The process used to impose order on collections of objects and events to show similarities, differences, and interrelationships. | I | M | M | E |
| 3. MEASURING: The process of quantifying observations. | I | E | E | E |
| 4. PREDICTING: The process of formulating a specific forecast based on observation, measurements and relationships between variables. | I | E | E | E |
| 5. INFERRING: The process of using logic to draw conclusions from data. | N | I | M | E |
| 6. HYPOTHESIZING: The process of formulating testable scientific generalizations. | N | I | M | E |
| 7. EXPERIMENTING: The process of using all the scientific processes in conducting a controlled test of a specific scientific hypothesis. | N | I | M | E |

Appendix E (cont.)

Subgoal B: To communicate knowledge of natural phenomena of the universe such as:

| Process Objectives | Suggested Placement* of Knowledge Objectives within the Science Curriculum (K-3) | (4-6) | (7-9) | (10-12) |
|---|---|---|---|---|
| 1. of a system of measurement (i.e. metric). | I | E | E | E |
| 2. of matter/energy relationships | I | E | E | E |
| 3. of the principles of energy origin, use and alternatives | I | E | E | E |
| 4. of the interaction of man with natural ecological systems. | I | E | E | E |
| 5. the personal aspects of physical, mental and community health and safety. | I | E | E | E |
| 6. that living things are in continuous change. | I | E | M | E |
| 7. of the interaction and inter-dependence of living things with their environment. | I | E | M | E |
| 8. of the finite nature of natural resources | I | M | E | E |
| 9. of laboratory equipment, procedures and safety. | I | M | E | E |
| 10. the concepts of simple machines. | I | M | E | M |
| 11. of the pinciples of mineralogy/petrology | I | M | E | M |
| 12. of the characteristics of living organisms. | I | M | M | E |
| 13. of the essential role of plants to all living things. | I | M | M | E |
| 14. of the diverity of living forms. | I | M | M | M |
| 15. the physical aspects of the solar system | I | M | M | M |

Appendix E (cont.)

| Process Objectives | Suggested Placement* of Knowledge Objectives within the Science Curriculum (K-3) | (4-6) | (7-9) | (10-12) |
|---|---|---|---|---|
| 16. of the importance of the water and other cycles. | N | I | E | E |
| 17. of the principles of atomic theory | N | I | E | E |
| 18. of fundamental inorganic chemistry. | N | I | E | E |
| 19. that all matter consists of units. | N | I | E | M |
| 20. of the principles of magnetism and electricity. | N | I | E | M |
| 21. of the cell as the basic unit of living organisms. | N | I | M | E |
| 22. that living organisms are the products of their heredity and environment. | N | I | M | E |
| 23. of the dynamic universe and solar systems. | N | I | M | E |
| 24. of the principles of human anatomy and physiology. | N | I | M | M |
| 25. of the principles of geologic record. | N | I | M | M |
| 26. of map construction and interpretation. | N | I | M | M |
| 27. of the principles of continental drift. | N | I | M | M |
| 28. of the conditions influencing weather. | N | I | M | M |
| 29. of the principles of radioactive and physical dating. | N | N | I | E |
| 30. of the periodic table. | N | N | I | E |
| 31. of fundamental organic chemistry. | N | N | I | E |

Appendix E (cont.)

Subgoal C: To use scientific knowledge in comprehending the impact of science and technology on the individual, culture, and society, such as:

| Impact Objectives | Suggested Placement* of Impact Objectives within the Science Curriculum (K-3) | (4-6) | (7-9) | (10-12) |
|---|---|---|---|---|
| 1. man ethically. | I | E | E | E |
| 2. the ability of species to survive. | I | E | E | E |
| 3. health and well-being. | I | E | E | E |
| 4. food and nutrition. | I | E | E | E |
| 5. housing/shelter. | I | E | E | E |
| 6. natural resource use and management. | I | E | E | E |
| 7. use and/or misuse of land. | I | E | E | E |
| 8. amount and usage of leisure time. | I | E | E | E |
| 9. use and/or misuse of drugs. | I | E | E | E |
| 10. living organisms' reaction to stress. | I | E | E | M |
| 11. all living organisms within populations. | I | E | M | M |
| 12. man aesthetically (enjoy science). | I | M | E | M |
| 13. methods of communication. | I | M | M | E |
| 14. the amount, control and usage of pollution. | I | M | M | M |
| 15. energy production and usage. | N | I | E | E |

Appendix E (cont.)

| Impact Objectives | Suggested Placement* of Impact Objectives within the Science Curriculum (K-3) | (4-6) | (7-9) | (10-12) |
|---|---|---|---|---|
| 16. jobs and careers. | N | I | E | E |
| 17. chemical development and usage. | N | I | E | E |
| 18. various modes of transportation. | N | I | E | M |
| 19. prosthetics, artificial body parts, or organ transplants. | N | I | E | M |
| 20. design and usage of computers. | N | I | M | M |
| 21. man's ventures in space. | N | I | M | M |
| 22. weather modification. | N | I | M | M |
| 23. genetic engineering. | N | N | I | E |
| 24. clinically induced and maintained life. | N | N | I | E |
| 25. euthanasia or mercy killing. | N | N | I | E |

*Definition of Symbols:

I = INTRODUCE - The first time a topic is INTRODUCED as a planned portion of the science program.

E = EMPHASIZE - The topic to be EMPHASIZED or stressed.

M = MAINTAIN - The presentaion and/or reinforcement of topics introduced previously and MAINTENANCE of these at the level of sophistication applicable to that grade level.

N = Not applicable at this level.

# Appendix F

## A Philosophy of Science

Science education is the link between science and society. Its ultimate goal is DEVELOP SCIENTIFICALLY LITERATE CITIZENS who use and understand the impact, knowledge and processes of science to solve problems and improve life within the limits of the total environment. Science education is any set of activities that develops scientific literacy.

A new generation of scientifically literate citizens is needed to cope with a future charcterized by rapid change and a complex set of technical and ethical questions. Accordingly, it is recommended that all students receive an appropriate education in science to develop the intellectual skills that are basic to critical observation, problem resolution, decision-making and valuing.

The study of science offers a KNOWLEDGE OF NATURAL PHENOMENA that uniquely rests upon the notion that humans can test and understand the orderly nature of the universe. Fundamental to this proposition is a need for students to develop and apply the logical thought PROCESSES OF SCIENCE AS PART OF THEIR BASIC LEARNING. These processes are best developed through a well-articulated science program that includes experimentation and manipulation of materials.

Science activities built upon each individual's natural curiosity become self-motivating and meaningful. This involvement can result in personal gain for students who discover and develop a confidence in their own ability to make the decisions that can form a basis for COMPREHENDING THE IMPACT OF science and technology on the individual, culture and society.

In addition to the development of logical thought and personal growth, research indicates that involvement with activities in science facilitates growth in the other curricular areas. The evidence shows a significant effect upon reading readiness, the motivation to learn, and the ability to acquire oral and written communication skills. Science offers a practical use of mathematical concepts and skills. Science is a vehicle for use of the metric system and provides subject matter for social studies.

Iowa Department of Public Instruction, 1980