1998

Relationship of ACT/SAT scores to the superintendency

Merle J. Farrier
The University of Montana

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Doctoral Dissertation

Relationship of ACT/SAT Scores to the Superintendency

By

Merle J. Farrier

B.A., University of Montana, 1969

M.Ed., University of Montana, 1991

Presented in partial fulfillment of the requirements

for the degree of

Doctorate of Education

The University of Montana

1993

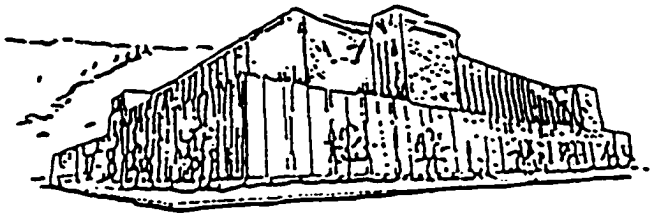
Approved by:

Roberta D. Evans, Ed.D., Committee Chair

Dean, Graduate School

5-11-98

Date



Maureen and Mike
MANSFIELD LIBRARY

The University of **MONTANA**

ABSTRACT

Farrier, Merle J., M.Ed. 1991 Education

Relationship of ACT/SAT Scores to the Superintendency

Committee Chair: Dr. Roberta Evans, Ed.D.

This study was conducted to investigate how Montana public school superintendents utilize ACT/SAT scores. A census was mailed to each high school superintendent with an 83 percent response rate (138/166). The research questions examined were: a) What does the present research infer or conclude about the nature of the ACT/SAT scores? b) As determined from the present research, what is the most appropriate behavior for a superintendent to take relative to ACT/SAT scores? c) What is the present behavior of Montana high school superintendents toward ACT/SAT scores? d) How does Montana high school superintendents' usage of ACT/SAT scores compare with the use suggested by research?

The research found that approximately 81 percent of the superintendents believe ACT/SAT scores are at least a good indicator of their schools' educational effectiveness. Seventy-two percent of the schools make formal use of these scores in scholarship selection, academic recognition, and other uses. Superintendents believe ACT/SAT scores most measure curriculum content, teacher effectiveness, and curriculum offerings while least reflecting administrative areas of the school system.

The research concludes with a synthesis of the literature and discusses common but improper uses of ACT/SAT scores. Guidelines for administrative usage of ACT/SAT scores are offered. Recommendations are made for implementation of alternative assessment procedures.

TABLE OF CONTENTS

CHAPTER 1

STATEMENT OF THE PROBLEM

General Need for Administrative Research

The innate desire to perfect human relationships ultimately forms the fundamental need for a science of administration. Critical, substantive, and practical research is an indispensable element in realizing that perfection. J. Myron Atkin (1989) emphasized the need not only for research, but research initiated by educators when he writes: "Currently, the educational research agenda is constructed primarily by people who do not know education by direct participation" (p. 201). Atkin's point is as simple as it is poignant: the direction of educational research should be determined by educators.

Coupled with the need for educators to initiate research is the need to examine educational problems the solutions to which have little or no definitive form. The problem of what constitutes proper educational usage of the scores obtained on American College Tests (ACT) and Scholastic Aptitude Tests (SAT) is a problem lacking a definitive solution. This research will contribute to the knowledge base necessary to provide a structured solution to the proper educational usage of ACT/SAT scores at the secondary level.

Background of the Problem

Public education has always been subjected to critical examinations from many diverse interests. The present time is no exception (Goodling, 1991). Our nation's educational system will spend an estimated 180 billion dollars a year by 1993 (Swanson & King, 1991). An expenditure of this size produces a continuous demand for accountability.

In order to provide documented accountability, public officials, school administrators, teachers, and the general public frequently seek a quick and simplified assessment technique. This technique historically has been a standardized test or a combination of two or more such tests (Eisner, 1991). Presently, there is a national effort to strengthen this form of assessment.

President George Bush proposed on April 18, 1991, and communicated to the Senate on May 22, 1991, the "America 2000 Excellence In Education Act" (Bush, 1991). This act recommends a voluntary nationwide standardized examination through which each school district can be compared to another by means of an examination score (S 1141, 1991). Research has indicated that parents consider the results of examination scores as a high priority when they have the opportunity to select from more than one high school (Hunter, 1991). Such a score will provide all interested observers with the means to evaluate the educational programs offered by each and every school. The growing

interest in the voucher system will add momentum to this idea (Pipho, 1991).

Politicians believe "America's educational system is in urgent need of an overhaul" (Kennedy, 1991). According to Senator Dave Durenberger (1991), "...education is now taking its rightful place at the top of our Nation's political agenda." President George Bush (1991) has added executive harmony with "...America 2000 is not just a program or a set of programs; rather, it is a national crusade."

The validity of using standardized exams as a measure of accountability is axiomatic in the political arena (Riggs, 1991; Cochran, Dole, Durenberger, Hatch, Jeffords, Kassenbaum, Kennedy, Pell, Simon, Simpson, Spector, & Thurmond, 1991). A National Council on Education Standards and Testing was legislated to provide the expertise upon which to decide the desirability of a national examination (Kildee, 1991). This council will be a political body in that all appointees are selected by politicians (H.R. 2435). The same concept of testing exists in the Act known as "America's Commitment to College Education and Success for all Students Act of 1991" (S 1134, 1991).

However, education is not a by-product of political processes. A national examination originating from politicians instead of educators presents a crucial problem for all education: education is on the brink of total reform based not upon educational research and expertise but

upon political needs that may be more self-serving than altruistic.

The emerging proposal to assess the nation's schools based upon examination scores has far-reaching implications. If schools are going to be assessed by national examination scores, the immediate implication is the need for a national curriculum (Eisner, 1991). In turn, a national curriculum will require extensive restructuring of the present educational system.

The national attention and prominence given standardized testing has created an additional need for public school administrators to be knowledgeable about standardized testing. Superintendents may want to examine their philosophic positions in light of current and past research on standardized testing. Superintendents may seek a formulated plan of action that will intelligently respond to the issue of assessment in general and testing in particular.

Education is facing restructuring based upon unproven assumptions. Presently the leadership for this potentially massive restructuring of the schools is originating from political bodies serving political purposes. Educators will play a minor role in restructuring unless the seminal principles upon which restructuring is based are motivated by educational rather than political considerations.

Public schools are under considerable pressure to demonstrate a high degree of accountability. Administrators are seeking various means of evaluation and assessment in order to provide the public record with the necessary documentation of accountability. This often leads to frustration because indicators such as standardized test scores can be high one year and low the next. Not only is an explanation lacking for large fluctuations from year to year, but the scores themselves seem to be very difficult to interpret properly.

Most universities and colleges (henceforth in this paper, the term "college" will be synonymous with "college/university") require the American College Test (ACT) or Scholastic Aptitude Test (SAT) be taken prior to admission. While these scores serve to assist colleges in their selection of incoming students, they are often used by students, parents, taxpayers, school personnel, administrators, school boards, and the media as indicators of high school and even K-12 preparation and programs. This leads to evaluation of individual teachers and restructuring of programs by reactive administrators and school boards, as evidenced in the Prince Georges School District in Maryland.

John Murphy was hired as superintendent by the Prince Georges School District. His objective was to improve standardized test scores since the district equated low test scores with poor education. Mr. Murphy implemented what he

called "applied anxiety" by requiring principals and teachers to produce higher test scores. When scores did not improve, principals were fired or demoted, teachers were transferred, and salaries frozen.

While teachers complained there was too much emphasis on test-taking, Murphy continued his applied anxiety philosophy. After seven years, average SAT scores improved from 679 to 807. So pleased was the school board that they offered Murphy a 10-year, $2.5 million contract. His achievements were written up in the Congressional Record (Hoyer, 1991) and the Wall Street Journal (Putka, 1991) and he was hailed as a "visionary leader" who established at the district level what America 2000 seeks to do at the national level.

Prince Georges County School District must have indeed modeled the America 2000 concept of education since this school was chosen as one of the Atlas Community sights (Mecklenburger, 1992). Of the 684 proposals submitted to the New American Schools Development Corporation (NASDC), the Atlas Community is one of only 11 proposals accepted for funding (an expected $200 million divided 11 ways). This proposal was submitted by James Comer of Yale, Howard Gardner of Harvard, and Theodore Sizer of Brown.

Other districts like Joliet Central High in Joliet, Illinois declared a total school effort to raise ACT/SAT scores (Activity, 1992). This school district incorporated

test preparation into its curriculum, acquired computer hardware and software designed to improve test scores, and started an ACT Club for juniors and seniors. These club members were treated with special breakfasts, and students who scored 30 or above were given special recognition.

Specific Problem Addressed by this Research

This research will formulate conclusions about the ACT/SAT examinations based upon present existent research. These conclusions will form the basis for a set of rational principles relevant to administrative behavior regarding ACT/SAT scores as well as identify the qualities necessary to realize this behavior. The study will identify, by means of a survey, present Montana superintendents' behavior and attitudes relative to ACT/SAT scores. The data gathered from Montana superintendents will be compared to the principles synthesized by this study.

Research Questions

This research will address the following questions:

1) What does the present research infer or conclude about the nature of the ACT/SAT scores?
2) As determined from the present research, what is the most appropriate behavior for a superintendent to take relative to ACT/SAT scores?

3) What is the present behavior of Montana high school superintendents toward ACT/SAT scores?

4) How does Montana high school superintendents' usage of ACT/SAT scores compare with the use suggested by research?

Importance of the Study

General Significance

The purpose of this study originated in noting that presently the coursework at The University of Montana does not specifically address nor prepare an administrator for dealing with ACT/SAT scores when the scores are returned to the respective school districts. The immediate significance became apparent through an informal survey of educators on campus which indicated that superintendents react in varying degrees - from quietly filing the scores away to firing of teachers.

A quick review of the literature indicated the existence of volumes of research on ACT/SAT examinations but virtually no work which comprehensively applied the research to rational, much less exemplary, principles of administrative behavior. An examination of current administrative textbooks also indicated a silence on this issue. The gap in the coursework at The University of Montana is not an oversight. The gap exists simply because of the void in the general knowledge, both theoretical and

practical, within administrative science. The result of such a gap has produced an abundance of superintendents who are illiterate in the area of assessment (Stiggins, 1991).

The present emphasis on assessment in education is reflected in the frequency with which the journals and periodicals are addressing the issue. The School Administrator, for example, devoted its entire December, 1992, issue to assessment. Bracey (1991) noted that this pressure on assessment has led teachers and administrators to align their curricula with the tests, emphasize test taking skills, and even to cheat on examinations in order to meet public expectations of standardized tests scores.

This research will help superintendents by filling part of the present void and thus provide a basis upon which to judge assessment reform. This research may have an impact upon both K-12 usage of ACT/SAT test scores as well as higher education's use of those scores.

Specific Significance

Who Should be Administratively Responsible for ACT/SAT Scores?

The first consideration is a determination of who is responsible for dealing with ACT/SAT scores. There are two arguments for determining the proper level at which ACT/SAT scores should be treated: the argument from authority and the argument from congruency.

Argument from authority.

Many authorities (Lunenburg & Ornstein, 1991; U.S. Department of Education, 1991; Mitchell & Tucker, 1992) believe the responsibility of student assessment and school-wide performance properly belongs to the domain of the superintendency. While these authorities do not offer any substantial insight as to the correct administrative response to ACT/SAT scores, nevertheless, they agree that the response should be from the district superintendent.

The American Association of School Administrators and the University of Texas conducted nationwide research to determine what skill areas are the most important for superintendents. The research produced 52 different skills germane to the superintendency. When these 52 skills were ranked in order of importance, testing and evaluation of student performance ranked ninth (Lunenburg & Ornstein, 1991). Lunenburg and Ornstein reported the top 10 skills identified by this research are now considered generic skills for effective superintendents by the American Association of School Administrators.

These 10 skills are: 1) leadership, 2) personnel management, 3) fiscal responsibility, 4) school-community relations, 5) teacher evaluation, 6) cost-effective budgeting, 7) motivational techniques, 8) conflict mediation, 9) testing and evaluation of student performance, and 10) curriculum development and instructional planning

(Lunenburg & Ornstein, 1991, p. 328). In addition, the remaining nine items in the list are indirectly related to the responsibilities and need of the superintendent to address ACT/SAT test scores. The tenth item identified as a basic responsibility of the superintendency by the American Association of School Administrators and the University of Texas research is the skill of curriculum development and instructional planning. This skill is very closely related to administrative perception of test scores since a superintendent is responsible for curriculum development and instructional planning. Hence, if a superintendent is responsible for the curriculum and instruction of the district, then the superintendent will need to consider the proper way to interpret and utilize ACT/SAT scores prior to making curriculum and instructional decisions.

Elliot Eisner (1985) considered the relationship between testing and curriculum development more succinctly:

> The curricula of the school and the priorities that both students and teachers come to hold are influenced by what they think will influence test performance. ...the canons of measurement indirectly influence what is taught in schools and what priorities will be established among those fields that are taught (p. 361).

John Goodlad (1979) also made a strong case for curriculum development as central to the leadership role of the superintendent. The first item of priority in the American Association of Administrators and University of

Texas research, that is, leadership, is exercised *par excellence* when used to lead the development of curriculum. A superintendent who chooses to ignore ACT/SAT scores may be the same superintendent who fails to take a leadership role in curriculum development.

Additional authors connect ACT/SAT scores with administrative responsibilities. For example, Johnson (1993) noted that superintendents must have a vision to be effective leaders in today's schools. She also noted that included in this vision is the expectation of high academic achievement as measured by SAT scores.

Argument from congruence.

ACT/SAT scores present the possibility of providing, in a single quantity, a summative evaluation of an entire school district (Eisner, 1985). The quantity of scores received by a district is not as important as is the nature of a single score itself. This is true since even a single score can be understood as a representative of the collective value of a school district's ability to educate. In this sense, such scores in themselves can be understood to represent the essence of the district.

The context in which ACT/SAT scores are understood parallels the nature of the superintendency. The superintendent's responsibilities are such that virtually every aspect of the district's education is under the

superintendent's care. This is true regardless of whether the superintendent's influence is by intent, by delegation, or by omission. Hence, the superintendent is also a single representative of the collective value of a school district's ability to educate. The superintendent "speaks" for the whole educational system and in a real sense, so do the ACT/SAT scores.

Since the ACT/SAT scores share ontological congruence with the superintendency, a natural obligation exists for the superintendent to act upon that which is commensurate with his or her domain. The concept of an action requiring a reaction of like magnitude is a principle of natural philosophy (Newton, 1726) and validates from philosophical principles the need for the superintendency to treat ACT/SAT scores.

Consider, as an example, a student who is frequently tardy from recess. This action does not merit the attention of a superintendent because the nature of the action is not a reflection on the whole system and hence not apropos to the duties of the superintendency. However, a student who is measured in such a way as to be perceived as a reflection of the whole educational system does present a situation requiring an action from the level responsible for the whole educational system, i.e., the superintendent.

Why Shouldn't a Superintendent Disregard ACT/SAT Scores?

Having determined that the superintendent is the proper school official to treat ACT/SAT scores, the next consideration is why the superintendent should hold these scores worthy of administrative action. The importance of the superintendent's consideration of ACT/SAT scores is established by arguing from the good of the students, harmony and unity of staff spirit, and accountability to the district.

Good of the students.

Every student inherently represents a good. This is to say every student is a person (Sizer, 1985) and, as a result, deserves treatment in accordance with that dignity. Stiggins (1991) argued administrators must be concerned about standardized test scores for the sake of the students: "Those who care about students either make or demand changes in unsound assessments." Sergiovanni (1992) argued that modern leadership emulates a concern for the whole person.

Students who score high on tests should receive recognition from the person officiating at the level of the tests. Hence, students scoring high on a classroom examination should receive recognition from the classroom teacher. Students scoring high on ACT/SAT exams should receive recognition from the same level as the exams are perceived; that is, as a measure of the whole educational

system. For the superintendent to ignore an opportunity to acknowledge the summative educational accomplishments of students diminishes the recognition of good that can properly be found in those students. Ignoring high ACT/SAT scores is inconsistent with the superintendent's role of establishing self-worth and self-esteem within the school system.

The same considerations apply to low to average test scores. There is no existent research concluding that superintendents should ignore low to average ACT/SAT test scores. Students scoring in the low to average range have just as much need for personal dignity and self-worth as the students attaining above average to high scores. If anything, these students need even greater care and concern in the treatment of their scores and subsequent action by the superintendent. According to Elliot Eisner (1985), "The social significance of the means used to measure student performance through testing is profound" (p. 361).

Personal communication directed to these students by the superintendent explaining the nature of the ACT/SAT scores will also afford an opportunity for the superintendent cognizant of current research on ACT/SAT scores to increase such students' self-worth. At the very least, such personal communication from the superintendent would convey the whole school system is still behind each student's efforts to enter college; therefore, a low score

would not be synonymous with failure. Proper, well-informed action on ACT/SAT scores that is motivated by a desire to address the inherent or personal good of students provides the superintendent with a significant means to achieve motivational improvement at the district level (Boyer, 1983; Kelley, 1988).

Harmony and unity of staff spirit.

While the treatment of students is always a predominant consideration at any level in education, the treatment of the teaching staff also is highly important. Because superintendents' treatment of the teaching staff is of such importance, the American Association of School Administrators/University of Texas research directly identified this skill as the second and fifth most important skill and responsibility of the effective superintendent.

The results from ACT/SAT scores can be very difficult for teachers to handle independently of the superintendent's intercession. "Few teachers can withstand the almost certain public pressure for high scores on district- and state-mandated tests" (Eisner, 1985, p. 361). The ACT/SAT scores present other problems for staff as well. The SAT reports scores in two areas: verbal and quantitative. Division and general breakdown of staff unity can result, for example, from high verbal and low math scores. The curriculum area in which low scores are reported is deemed

responsible for preventing students from receiving scholarships and even disqualifying students from admission to certain colleges and universities. The argument can be further advanced that the area of the curriculum in which low scores are reported nullifies all of the efforts made in the remaining curricular areas.

A complicating argument can be advanced that areas of the curriculum not tested offer nothing for the good of the college bound students. The funding of these areas should be channeled into the tested areas to improve test scores. The result is a fracturing of the harmony and unity of staff spirit.

The well known expression "All the king's horses and all the king's men could not put Humpty Dumpty together again" rings true. The staff *per se* will be unable to overcome the negative effects of poor ACT/SAT scores because the fragmenting of harmony and unity of staff spirit can be made whole again only by an agent acting at a level above that which has been broken. Consequently, the school district has a necessary obligation for the sake of the harmony and unity of personnel to deal with ACT/SAT scores at the level of the superintendency.

Accountability to the district.

This level of argument, while the least important, is still critical to the management of a school district. This

responsibility is identified in the American Association of School Administrators/University of Texas research as 3) fiscal responsibility, 4) school-community relations, and 6) cost effective budgeting. The superintendent must address the issue of ACT/SAT scores to the public particularly when scores are low. The superintendent owes to those who have supported the school system whether - as board members, parents, volunteers, taxpayers, or any combination thereof - an explanation of those scores. If the public believes the dollars spent are not producing the desired results, the superintendent must have the necessary knowledge about these scores in order to adequately address the situation.

The importance of this research is found in its intent to supply substantial direction in an area where direction is absent. Just as important, the area lacking any direction within the theoretical framework of public school administration is directly identified as one of 10 essential areas of responsibility proper to the superintendency. To further emphasize the importance of this research, the superintendent's need to act upon ACT/SAT scores is indirectly addressed in the remaining nine key responsibilities of the superintendency as identified by the research of the American Association of School Administrators and the University of Texas.

Delimitations

This study surveyed only Montana high school superintendents. There is no implication that this population represents a sample of superintendents of any larger group, whether nationally or regionally. This study will not theorize on optional conclusions based upon possible future revisions of either the ACT or SAT examinations.

CHAPTER 2

REVIEW OF THE LITERATURE

Introduction

A review of literature has indicated the research available regarding ACT/SAT scores. While the literature was abundant in reference to the scores, it was relatively silent regarding what action administrators should take upon receipt of those scores. The review of literature began with a brief study of the history, nature, and differences of the ACT and SAT examinations. Additional information examined what the producers expect their tests to measure as well as how well they realize that end. Considerable research exists relevant to this discussion, and the strengths and weaknesses as determined by the research was discussed. Further, the validity of the research was analyzed.

The review also considered other expectations of the examinations as viewed by educators, politicians, and others. Any misuses of the examinations was then identified. Finally, the review included a discussion of the omissions in the research and a summary of the findings.

Producers of the Tests

Each test is produced by a different company. The producers of the ACT examination are rarely addressed in the literature. This observation is contrasted by fact that the

producers of the SAT are frequently the object of considerable research and discussion.

The ACT

The ACT examination is produced in Iowa City, Iowa. It typically is used by about 40 percent of those taking either the ACT or the SAT (Hoffman, 1991). The primary geographical usage of the ACT is the Midwestern states.

The SAT

The SAT is produced by the Educational Testing Service (ETS) located in Princeton, New Jersey. As was evident in this review, the ETS is sometimes linked to political ends. Hence, the history of the ETS has been well chronicled by researchers such as Nairn (1980) and Crouse and Trusheim (1988).

The Reign of ETS (Nairn, 1980) is the first major study conducted on the ETS and forms some initial insights into the nature of the corporation that produces these tests. The Case Against the SAT (Crouse & Trusheim, 1988) was written as a reaction to Nairn's work. Crouse and Trusheim originally began their research to defend ETS only to find their investigation added more weight to Nairn's position. ETS has responded to these works, and their response was included.

Research prior to 1980 was especially hampered by an unwillingness of ETS to furnish data (Slack & Porter, 1980). Since the founding of ETS some 45 years ago, the test questions, answers, and scores have been considered by ETS to be their exclusive property. Any research conducted on those items has been accomplished using the data ETS has chosen to release. This secrecy in information has resulted in a questioning of the reliability of the research. Oscar K. Buros, editor of the Mental Measurement Yearbook had difficulty in obtaining ETS data and believed the data ETS released to him to be inadequate (Nairn, 1980).

Contrary to the strongest efforts of ETS, New York passed a truth-in-testing law in 1979 which required ETS to furnish upon request from any test taker in New York a copy of the test and the correct answers within a 30-day time frame. While this has made more data available for study, the data released are often subject to stipulations. For example, Crouse and Trusheim (1988) reported they received the requested data only after they promised in writing to publish just those findings ETS would agree to have published.

In 1972, ETS did release the questions to their Law Scholastic Aptitude Test. The questions were studied by law professors teaching bar review courses. Four separate law faculties could not agree on the answers to 69 of the 200 questions. ETS eventually admitted that five unidentified

questions had multiple correct answers; however, there was no redress for help to those students who had already failed the exam (Nairn, 1980).

Since the advent of the New York Truth-in-Testing law in 1979, ETS decided in 1981 to release all tests and answers to anyone taking the test. In 1982, ETS allowed the release of old editions of the test.

History of the Educational Testing Service

ETS officially came into existence in 1947 when the Carnegie Foundation, the College Board, and the American Council on Education were consolidated. This merger resulted in all testing and research power residing within ETS and since that merger, ETS has been free of any government scrutiny or supervision (Gougeon, 1984).

ETS has no shareholders and returns all of its profits to the corporation and as such was granted a non-profit charter in New York but operates in New Jersey. Because of this status, ETS is not subject to taxation[1], antitrust, or consumer protection enforcement laws. ETS is a corporation

---

[1] New Jersey did require ETS to pay property taxes. ETS sold its lawn clippings and some firewood to its employees and managed to convince the state that ETS was really an agricultural enterprise. This reduced ETS's property taxes from $425,000 per year to $25,000. Upon appeal from local citizens, this decision was later reversed.

nobody owns and nobody can buy; its power is completely from within (Nairn, 1980).

The Rockefeller family has contributed millions to ETS and frequently uses ETS grounds to conduct its own business and social gatherings. Numerous other major bank presidents and financiers have sat on ETS' trustee committee, thereby helping to define the direction and expectations of education (Nairn, 1980).

ETS has a gross income in excess of $200 million a year. While this income is considered small by large industry standards, it is not the income ETS is interested in, but rather, according to Nairn (1980):

> ...it is only the means to its larger objectives as theorist and gatekeeper for the educational and career opportunities of millions of people. Distinct from most corporations, ETS' income figures greatly understate its influence. Few companies in history have had so much effect with so small a cash flow (p. 314).

Influence and Scope of ETS

In education.

ETS is interested in providing testing that will range throughout a person's lifetime. ETS has a test (Cooperative Preschool Inventory) which purports to measure, within 15 minutes, a preschool child's achievement in perception, verbal, and thinking expression. ETS has another test (Secondary School Admission Test - SSAT) designed to determine if a person is ready to enter high school. The

hundreds of thousands of people without high school diplomas who annually try to improve their education and job opportunities must pass an ETS examination called the General Educational Development test (GED).

Once preschool children have been determined fit for entrance to school, ETS has a test (Junior Scholastic Aptitude Test - JSAT) designed to predict high school grades. ETS also offers a test commonly known as the Preliminary Scholastic Aptitude Test (PSAT) for high school students which was developed to predict the SAT score. Not so commonly known is the full name of this test, i.e., the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT). This test is used by over 400 granting institutions, many of whom have committees selected by ETS to award tens of millions of dollars annually (Nairn, 1980).

The financial aid form used by students seeking financial help is an ETS-produced document. This document contains detailed personal information about the applicant's family and is studied by ETS' College Scholarship Service (CSS), and their decision becomes the basis of college scholarship, loan, and work-study awards. For students of a graduate school, college monetary amounts are awarded through ETS' Graduate and Professional School Financial Aid Service (GAPSFAS).

However, at the undergraduate level, the PSAT/NMSQT is only considered the first hurdle. The ETS Scholastic Aptitude Test (SAT) must usually be taken before awards and admission are granted. ETS furnishes students lists wherein colleges and universities are ranked according to the entering class's average SAT score.

Once admitted to college, ETS students may be subject to any one of 20 or more of ETS' achievement tests, advanced placement tests, and the College Level Examination Program (CLEP). These tests are defined as specific achievement tests and are optional for any college student.

If a college graduate wishes to re-enter college in a graduate program, ETS has an examination awaiting such efforts. The Graduate Record Examination (GRE), an ETS instrument, is used by a high percentage of graduate schools as a major criterion for admission. ETS has also established itself as a measure for whether applicants are accepted into business schools in many nations (Graduate Management Admissions Test - GMAT). Law schools require the Law School Admission Test (LSAT) score for admission. With lawyers earning one of the highest incomes in the nation and with often two or three applicants for every available opening, law schools frequently will reject an applicant with a high grade average for four years of college work if they have an LSAT of less than 500. Further, law students generally have another ETS measure (Multistate Bar

Examination - MBE) waiting for them before they can practice law. Critics such as Nairn, Nader, Crouse, or Trusheim point out that at any given time 60 to 70 percent of the United States Senate and 50 percent of the State governors are composed of lawyers who have met standards based upon ETS. The implicit power of ETS, and the underlying issue of this research, that is, the problem of the proper handling of those scores, becomes more apparent. Kelley (1982) has summed up what he sees as the dangers of this scenario:

> When we allow the use of standardized tests to implicitly define our criteria of selection for educational opportunity or jobs, we essentially delegate what is a political or ethical decision to a group of individuals who, while not meaning to make that decision, do so as a projection of their own perception of professional norms (p. 25).

In business.

ETS has made progress in extending its system of quantifying aptitude and achievement to the business world. Not only do those wishing to practice law meet their criteria for such employment, but the GMAT is also used by such corporations as Citibank and General Motors to screen applicants. For those wishing to enter the teaching field, ETS has prepared an examination (National Teacher Examination - NTE) which must be passed. If this single test is not passed, 16 or more years of formal education can be negated to the extent that the test-taker is not employable as a public school teacher. Some schools have used this score as a basis for

hiring and promotion. The administrator who hires and promotes teachers by means of an ETS score may one day find his or her own job evaluated by the same type of score (Kelley, 1982).

ETS has extended its domain of testing to the labor force in general. ETS has tests to qualify respiratory therapists, construction code inspectors, shopping center administrators, real estate agents, policemen, fire fighters, and even a golf pros (Crouse & Trusheim, 1988).

What is the ACT Examination?

Both the ACT and SAT examinations have undergone recent changes. While these changes were documented here, it is important to show the previous form of both tests since both of these forms have been used in the research and literature.

The ACT examination consists of 2 hours and 40 minutes of testing (Kifer, 1985). This time is divided among mathematics (50 minutes), English Usage (40 minutes), Social Studies (35 minutes), and Natural Science Reading Test (35 minutes). These times will be altered slightly in the future.

The mathematics section is designed to measure the students' reasoning ability rather than achievement (Kifer, 1985). A scattering of problems is given to reflect content; however, the emphasis is on measuring the reasoning

ability needed to deal with college courses. The English section measures the students' understanding of the conventions of standard written English, e.g., punctuation, grammar, sentence structure, diction, and style. The Social Studies test measures the students' ability to read, analyze, evaluate, and extend social studies materials. The Natural Sciences test measures the students' ability to read, analyze, and evaluate material from the natural sciences as well as examine the logical relationship between hypotheses and conclusions and their applications to experimentation. The ACT test was revised in October, 1989, in an effort to be more responsive to the current high school curricula (Heritage, 1990).

What is the SAT Examination?

The Scholastic Aptitude Test (SAT), in its present form, was designed in 1941 as an "unchanging measurement" to measure minds (Gougeon, 1984). It was, however, originally created much earlier.

During the period from 1903 to 1915, the Jewish portion of Columbia's undergraduates jumped from 19.3 percent to 73.2 percent. This reportedly created concern on the part of the dean who feared that such a large number of Jewish students would drive away the other nationalities; consequently, Columbia became the first college to institute intelligence tests for admissions. Crouse & Trusheim,

(1988) allege that this was done in order to attract an upper-middle class white Protestant clientele and provide rationale for not admitting other nationalities. Harvard, Yale, and Princeton -among other colleges- followed in using these tests as instruments to control the nationality, race, religion, and gender of its student body. Brigham and those who employed him believed that the Nordic population (those from northern and western Europe) was of superior intelligence; therefore, the test scores would open the gate of college to them while closing it to "inferior" races (Brigham, 1923).

The College Board became interested in test writing in 1919, and, by 1925, the College Board had appointed a commission headed by Carl Brigham to write intelligence tests (Comras, 1984). Brigham was a psychology professor at Princeton who had published A Study of American Intelligence in 1923. This book concluded that Jews, Greeks, Hungarians, Italians, Catholics, Negroes, Poles, Russians, and Turks were all of inferior intelligence compared to the Nordic stocks of Western Europe. A social Darwinist and eugenicist, Brigham was committed to preventing the propagation of defective strains of human beings (Crouse & Trusheim, 1988). A sample of Brigham's (1923) A Study of American Intelligence, a book he published just two years prior to being employed by the College Board to write college entrance examinations, is as follows:

> Our figures then, would rather tend to disprove the popular belief that the Jew is highly intelligent.... Our results showing the marked intellectual inferiority of the negro are corrobated (sic) by practically all of the investigators.... The intellectual superiority of our Nordic group over the Alpine, Mediterranean, and negro groups has been demonstrated.... When our methods of measuring intellectual capacity have been perfected, we will be in a position to determine quantitatively the amount of race differences.... The scientific problem is that of eliminating from the tests used as measuring instruments those particular tests which demonstrably measure nurture, and to measure, with genuine tests of native intelligence...(pp. 190-195).

ETS president William Turnbull (1980) immediately published a reply to the Nairn report. While Turnbull pointed out some possible statistical flaws in the Nairn report, he failed to challenge the report's conclusion that the test was conceived to select college applicants based upon nationality or religious preference. Further, he did not dispute the Nairn/Nader presentation of the information on Carl Brigham.

Brigham and his commission created an intelligence test which they called the Scholastic Aptitude Test (SAT). This test was designed to measure intelligence or ability, but definitely not achievement, since achievement was something possible to find equally in all nationalities. Precisely at this point in history, admission to college was becoming correlated not with achievement but ability. This philosophy was confirmed by Canfield and Wilmoth (1987). They observed that after World War I, university entrance

requirements shifted from selecting those with the proper academic preparation to those who possessed the ability to succeed.

The philosophical shift from academic preparation to ability as the measure for admission to college is defended by ETS. Cameron (1989) contended the decline of uniformity among high school graduation requirements across the United States made subject-matter tests obsolete and ability tests the most effective way to determine college admissions. His arguments were based upon achievement verses ability testing, yet he did not indicate why high school records of preparation were not adequate indicators of college success.

The SAT examination itself is similar in length to the ACT, but divided into just two sections: verbal (SAT-V) and quantitative (SAT-M). The intent of the SAT is to measure aptitude, not achievement. The actual content of the mathematics section does not exceed the 8th grade level (Cohn, 1985). Comras (1984) believed most of the mathematical items are learned before 10th grade.

The SAT typically allows 50 seconds to answer each question on the verbal and approximately 75 seconds for the quantitative questions. The responses to these questions are reported not as a record of performance, but rather as an indicator of aptitude.

The College Board (1992a) has announced changes in the 1994 SAT. The verbal portion of the test will increase to

80 minutes and have a 20 minute portion in which the test-taker writes responses. The mathematical portion will increase to 75 minutes and require the test-taker to produce his or her own response on a few problems. The increase in time is to compensate for the open-ended responses. Some experimental administrations of this new test are expected as early as the winter of 1992.

The method by which the essays will be scored adds interest to the process. The grading will be holistic, that is, each composition will be judged on how many parts of the writing fit into a whole. Grammatical mistakes will not be an important part of scoring the essays. Readers will look for what has been done well, attempting to ignore what has been done poorly. Respondents will not be penalized for leaving the essay unfinished or when discussion expounds tangentially to the topic question content (College Board, 1992b). Each essay will be read by two readers, but will not be graded against an objective standard. Each essay will be graded against all other essays taken during that test administration. Since ETS will not score these examinations against objective criteria, the inter-rater reliability correlation from this type of scoring has yet to be documented.

The range in scores for each of the two sections varies from 200 to 800 points. ETS uses an undisclosed method of equating the scores from year to year. This process is

intended to adjust the scores so they are comparable from one year to the next, ultimately using the 1941 test as their norm. The 1941 norm is based on approximately 11,000 test takers who earned an average score of 500. A score of 600 is one standard deviation above the 1941 norm (U.S. Department of Education, 1992). Hence, the number of correct answers in one year can produce a score higher or lower than the same number of correct respones in a different year.

The College Board (1992b) adopted the position that the new changes announced for 1994 in which test-takers, in part, produce their own answers in both the mathematical and verbal portions of the examination as well as use calculators, will not affect ETS' ability to compare scores with previous editions of the test. Even though today's colleges have an increase in candidates with drastically different social, economic, and racial characteristics than the 1941 applicants (Gougeon, 1984), ETS believes today's test-takers' scores can be meaningfully compared to the scores of a group normed 50 years ago.

Differences between the ACT and SAT Examinations

The ACT differs from the SAT in several ways. The ACT does not make a correction for guessing, while the SAT does. The ACT reports a composite score, the SAT usually does not. Students are encouraged not to retake the ACT while it is

common to retake the SAT. The students who take the ACT generally come from the Midwest while those who take the SAT usually come from the east and west coasts (Kifer, 1985).

The SAT scores are also considered an objective standard whereby state educational programs can be compared, even by Midwest states who use the SAT for fewer than five percent of their high school graduates (Powell & Steelman, 1984). This is due in part to the fact that the ACT has not historically released state-by-state information. Powell and Steelman also note that politicians, scholars, and the lay public consider the SAT to be the measure of school quality. The 1983 National Commission on Excellence used the SAT scores to document the failure of education, and SAT scores are the sole indicators of academic performance used by the U.S. Bureau of Census.

To summarize the main difference between the ACT and the SAT, the former is much less dominant in practical use as an indicator of education than is the latter. The state-by-state breakdown of SAT scores has contributed significantly to the SAT's prominence because such data allow for a multitude of correlation studies on a regional basis.

Purpose of the ACT/SAT Examinations

As Intended by the Producers of the Tests

Following strong criticism leveled at ETS by the Nairn/Nader report in 1980, ETS (The College Board, 1984) published

guidelines cautioning the users of SAT scores of possible misuse. Agencies were admonished to consider the consequences faced by those tested and to especially guard against using test results for purposes for which the test results are not intended. ETS does not enumerate all of the possible misuses, but does note these tests measure aptitude rather than achievement and, hence, cannot measure a school's effectiveness.

Both tests are designed with the intent to assist colleges and universities in the admissions process (Cohn, 1985; Kifer, 1985). Further, there is wide consensus in the educational community and society as a whole to consider these tests as a good predictor of academic performance of individual college students (Gougeon, 1984; Heritage, 1987; Nairn, 1980). The scores from these tests are asserted to be a measure of a student's aptitude for college level work. As indicators of aptitude, they are generally major factors, and sometimes exclusive factors, in determining student scholarships and admission to college. The tests do not attempt to measure achievement according to ETS; however, the 1994 editions will contain sections for achievement (College Board, 1991).

Comras (1984), a very strong supporter of the SAT, stated that the SAT provides a broad estimate of a student's ability and is a reliable predictor of performance in college. His continued position is that "unlike achievement

tests, the SAT is not dependant upon a specific curriculum nor a specific course of study" (p. 51).

The next logical steps for this review to follow are:

1) Do the ACT/SAT scores serve as significant predictors of college success?
2) Do ACT/SAT scores reflect an aptitude or ability?

As Predictors of College Success

What does Success Mean in the Research?

Considerable research has been devoted to exactly how well these tests serve as predictors of college success. Part of the problem in making this prediction of success lies in identifying the criterion variable. What is a measure of success? As was shown, ETS defines success as freshman grade point average (FGPA) since the studies it has financed show the best correlation for SAT scores is with FGPA. Other studies define success in a variety of ways, including a four year GPA, gradation, graduate degrees, and job success.

As Predictors of College GPA

When considered alone.

Several studies, especially prior to 1980, advocate both the ACT and the SAT serve reasonably well as determinants in selecting new students for a college or university (Heritage, 1990). While most of the literature deals with

the SAT examination, generally both the ACT and SAT are considered equivalent predictors of college performance (Bennett, 1985). These studies, often financed by ETS, frequently find a moderate to low moderate correlation between ACT/SAT scores and FGPA of approximately 0.45. However, Noble and Sawyer (1988) found that correlations in verbal scores ranged from 0.13 to 0.61 and from 0.04 to 0.75 in mathematics.

Sawyer (1987) noted that ACT lowered its minimum sample size to 75 in 1979 and later to 50 students in 1983 as the minimum sample size used to calculate the predictive value of ACT scores. The predictive weight assigned to the ACT scores are determined by standard least square procedures and is an estimated regression coefficient. Hence, predictive weight is a function of the sample size. Since the estimation error of the regression coefficient propagates further error in prediction, Sawyer studied the effect of sample size on prediction weights.

Sawyer's study (1987) was designed to determine how small a sample can be used to predict FGPA without fear of large sampling errors. Using samples from 125 colleges, Sawyer developed prediction equations and then cross-validated with ACT scores and FGPA obtained from the following year's students. He found that samples of $n<50$ provide significantly less accurate predictive value than larger samples.

Morgan (1989a) studied 222 colleges and 778 different FGPA prediction studies. The studies were based upon a minimum of 75 students and, in each case, the entire freshman class was used rather than a sample of the population. Morgan found a study of the 10-year trend from 1976 to 1985 produced a linear decline in correlation between SAT scores and FGPA of 0.04.

As the correlation between SAT scores and FGPA diminishes, ETS researchers (Bejar & Blew, 1981) explained this weakening relationship as a result of grade inflation. They hold that since SAT scores are a highly reliable measure of aptitude, and since this measure has declined while FGPA has increased, it is clear the grading procedures of tens of thousands of educators has become inflated.

While most of the studies which correlate college GPA with ACT/SAT scores look at only FGPA, Gougeon (1984) conducted a correlational study over a 15 year period for the SAT mathematics scores and GPA based upon all four years. Her samples were drawn randomly from 100 students attending a Pennsylvania university during the years 1968, 1973, 1978, 1981, 1982, and 1983. She compared the SAT mathematics scores with the university mathematics GPAs for each of the selected students for all four years. She found correlations of: 0.026 (1968), 0.013 (1973), 0.116 (1978), 0.217 (1981), 0.153 (1982), and 0.089 (1983). The highest of these correlations, 0.217, indicates that less than 5

percent of the variance of college mathematics GPAs can be predicted by SAT mathematics scores. More importantly, she found a low correlation in both directions. Students with low SAT mathematics scores could do well in college while students with high SAT mathematics scores could do poorly. This research indicated that not only are SAT mathematics scores poor indicators of college performance, but grade inflation is not an issue.

A year later, Gougeon (1985) did a study in which both SAT and ACT scores were studied. She used two years of records for two colleges (one for SAT, one for ACT) of approximately 1,000 students each. Using a correlation matrix, she found a correlation between HSGPA and college GPA (mathematics) of 0.577. The correlation between SAT scores and college GPA (mathematics) was 0.355. When she calculated the correlation matrix for college GPA using ACT scores, she found HSGPA to correlate with college GPA (mathematics) at a value of 0.566 while ACT scores had a correlation of 0.363.

ETS became aware of the controversy surrounding the usefulness of predicting FGPA in the early 1980s. They conducted their own review of research (Wilson, 1983) which sought to connect FGPA with performance beyond the freshman year. They concluded FGPA should remain a valid predictor of college success.

Bolt (1986) considered in detail the relationship between verbal score and college GPA, mathematics score and GPA, and composite score and GPA. He then considered different methods for treating the scores: a) the highest score, b) the most recent score, c) highest verbal plus mathematical, d) weighted combinations using all scores, and e) the average scores. Each of these considerations were used to determine a regression equation which, in turn, was used to predict college GPA. All treatments of multiple scores resulted in underprediction of actual grades.

When Used with Other Predictor Variables.

Research literature is abundant (Clark & Grandy, 1984; Crouse, 1985; ETS, 1980; Gougeon, 1985; Sawyer, 1989) which correlates ACT/SAT scores to FGPA when the test scores are used in conjunction with other independent variables. Many other factors have been found which predict FGPA better than ACT/SAT scores. ETS has taken the position that SAT scores are most effective when used in conjunction with other predictive variables. One such equation using both SAT scores and high school rank (HRS) is:

$$FGPA = b_1SAT + b_2HSR + b_3 + e$$

where b_1 is the average change in FGPA associated with a one-point increase in SAT scores among persons with the same high school rank (HRS), b_2 represents the average change in FGPA associated with a one-point increase in HSR among

persons with the same SAT scores, b_3 is the estimated FGPA of persons whose SAT and HSR are zero, and e is an error value (Crouse, 1985).

Crouse (1985) reported some studies which used HSR as the predictor variable resulted in 62.2 percent correct admissions if the criterion variable is 2.5 FGPA. If students are selected at random, this study found 53.0 percent of the time the student will attain a 2.5 FGPA. In this study, HSR provided an increase of 9.2 percent in successful prediction rate. If the SAT score is included in the regression, the number of correct admissions increased to 65.9 percent. This increase can be reported as an increase from 9.2 students per hundred to 11.9 students per hundred or 2.7 students per hundred increase because of the addition of SAT scores to the regression equation. Crouse pointed out that ETS does not report the gain in this manner, but rather ETS reports a 29.3 percent gain (2.7 ÷ 9.2) in predictive value. Gain calculations reported in this manner by ETS are called "proportional improvement" and tend to magnify the actual gain. In this case, a gain of predictive accuracy of 2.7 students per hundred would be reported "proportionally" as a 29.3 percent gain in predictive accuracy. Crouse argued that this could easily be misunderstood as 29.3 students per hundred increase in predictive accuracy when in fact it is an increase relative to other numbers more than actual admissions.

As an example of how reporting "proportional" gains tends to magnify the value of SAT scores, Crouse determined that ETS reported a 25 percent "proportional" gain in FGPA amounted to an increase in predicting FGPA by 0.02 grade points on a four point scale. The predictive gain of using SAT scores with HSR increases directly as a college more stringently screens its applicants. If 80 percent of the applicants are rejected, the contribution of SAT scores to predicting FGPA increases to 0.15 points on a four point scale. A 10 percent "proportional" increase in predicting years of college education translated into an actual gain of predictive value of 0.01 years.

Schwartz and Wilbur (1981) conducted research designed to determine how well SAT scores interact with other independent variables to predict college GPA. They reviewed studies conducted in the 1970s based upon an aggregate of thousands of students. The results indicate high school grade point average (HSGPA) was the best single predictor of FGPA while SAT scores accounted for 4.7 percent of the variability of FGPA.

A more critical study was conducted by Crouse and Trusheim (1988) assessing performance of 2,470 college applicants. Using multiple regression analysis, they looked at the ability of high school rank and SAT scores to predict FGPA. Results indicated when SAT scores were combined with high school rank in the predictor equation, the added value

of factoring in the SAT scores was minimal, increasing 0.02 points on a 4.00 scale.

Schwartz and Wilbur (1981) studied data obtained from students enrolled in Project Advance from 1977 to 1980. Project Advance was a program in 80 schools in New York, Massachusetts, Michigan, and New Jersey. Students in high school took college courses taught by teachers with adjunct instructor status at Syracuse University. Students who successfully complete Project Advance courses could apply to have their Syracuse credits transferred to a college they later attended.

The results of this study found Project Advance GPA was a better indicator of college FGPA than SAT scores. While SAT scores were less changeable, they were never good predictors of FGPA. SAT verbal and SAT vocabulary scores either alone or in combination with other variables were also not a significant predictor of future performance. In 1983, ETS conducted a study (Baird) which concluded SAT scores should be used in addition to other variables in predicting college GPA. This research is typical of the ETS studies following the 1979 Truth-in-Testing Law which gave the test-takers access to their tests and scores.

This research (Baird, 1983) was conducted using multiple regression analyses. The regression coefficients were transformed into β weights and are provided in the paper. While the study did not seek to strengthen the

multiple regression findings with cross-validation, it did conclude, with a warning, that SAT scores appear to be most useful in those circumstances in which GPA prediction might be very difficult without such scores.

Sawyer (1987) found predictive correlations actually decreased when more than two predictor variables were used to predict FGPA. He used ACT and self-reported HSGPA and found, particularly in small samples ($20<n<30$), predictive ability was lost by increasing the number of independent variables. A study at Utica College by Blanchfield (1971) using multiple discriminant analysis found social consciousness, college first semester GPA, high school rank, and the percentage of college costs financed by grants were all positively correlated significantly to graduating from college. He found no significance in standardized examinations and questioned their value to the college admission process.

Thornell and Jones (1986) randomly chose 100 students from 585 freshmen entering a small university in Mississippi. They used multiple regression to establish a prediction equation using ACT scores and high school rank as the predictor variables for predicting first semester FGPA. They also included high school size and ACT subtest scores as independent variables to see if they might be better predictors than ACT scores or high school rank. The correlation matrix indicated a positive, two-tail

probability of less than 0.05 between all of the independent variables and FGPA excluding high school size. They found no correlation between high school size and FGPA.

They next used regression analysis to compare high school rank and ACT composite scores, since the composite scores were more highly correlated than the scores on the subtests. The dependant variable was FGPA. The regression was significant and a test of the regression coefficients found the regression coefficient for high school rank was significant ($p<.001$) while the coefficient for ACT composite scores was not significant at the 5 percent level. The multiple R^2 was 36.6 percent when adjusted for shrinkage. The prediction equation for first semester FGPA is: FGPA = 0.6934 + 0.0348(ACT) + 0.0177(HS rank) with a standard error of estimate of 0.7003. HS rank must be converted to a percentile to be used in the equation.

In their sample, Thronell and Jones used their prediction equation and predicted 25 percent of the incoming freshmen would have a FGPA of less than 2.0. Of those "at risk" students, 39 percent achieved a FGPA of more than 2.0 with the highest being 3.04. This study did not comment upon how well the students did who were predicted to achieve a high FGPA, but it did indicate the danger of using prediction equations to eliminate candidates from admission to college. The study also concluded high school rank, when converted to a percentile, is a better predictor of first

semester FGPA than are ACT scores. The study found ACT scores did contribute to the prediction of FGPA, though not significantly. A similar study by Crouse and Trusheim (1988) considered high school rank and SAT scores as predictors of FGPA. He found high school rank was a better predictor of FGPA than SAT scores and acted as a better predictor of achieving a college degree.

As Predictors of College Graduation

If success in college is defined as completing college rather than as FGPA, then the predictive ability of SAT scores diminishes even more. The study cited earlier by Crouse and Trusheim (1988) of 2,470 college applicants also looked at the effect of adding SAT scores to high school rank in predicting the completion of a bachelor's degree. They found the effect even less impressive than on predicting FGPA. When added to high school rank, SAT scores contribute to increasing the correct selection of applicants who will earn a bachelor's degree with a predicted 2.5 GPA at the rate of one additional student per 1000 students admitted.

Considerations Regarding Predictive Research

Many other pieces of research could be cited, but studies cited are representative of the nature and results of the existent research. While much of the research focuses on

quantitative considerations such as disputing the magnitude of the correlations between test scores and FGPA, this type of research postulates two assumptions in the qualitative order. Briefly, they are: a) the applicants selected are those who should be selected, b) the criterion variable is unrelated to anything of substance.

Matriculants verses applicants.

First, the research predicts grades of applicants based upon only those students who have been accepted. This is a serious flaw in the research since admission of applicants is based upon the performance of matriculants. The actual results of applicants who are not admitted are never correlated.

Most of the GPA studies rely upon linear or multiple regression techniques and the correlation coefficient is not resistent to outliers. Not only does most of the research not consider the procedure of cross-validation to enhance the strength of the correlation coefficient, but ETS studies, in particular, do not report on the test-takers rejected for admission.

A follow-up study on those denied admission but accepted by another school to determine whether their GPA has the same correlation as predicted by their ACT/SAT scores would be a beginning toward an analysis of this problem. These students present the greatest probability of

producing outlier correlations which in turn can have an effect upon the overall correlational coefficient. Hence, most studies contain an underlying assumption that those students who were not granted admission have their GPA (or lack thereof) predicted perfectly.

Quantitatively this problem can be addressed by research which focuses on the certainty that the variables have a bivariant or multivariant normal distribution. Also important is the correction of any existent bias. Whether the researcher is careful enough to make and to document these corrections isn't as important as the realization that an entire population of students has been denied admission based upon the performance of those who were admitted.

The question remains, then: could students who were rejected by any given college have performed? A study was undertaken at Williams College in Williamstown, Massachusetts, to determine how well students would do who would not normally be admitted because of low SAT scores (Nairn, 1980). For 10 years Williams College accepted such students at the rate of approximately 10 percent of the entering class. The students and faculty were not made aware of this, nor were any special considerations given the low scoring students. Except for SAT scores, they had to meet exactly the same requirements as all other students. At the end of 10 years, studies determined that in just one class alone, the class president, president of the honor

society, and the president of the college council had all been members of the low scoring group admitted as part of the experiment. After 10 years, 71 percent of the experimental group graduated as opposed to 85 percent of the normal student body. Williams College considered the program so successful, they have continued it as part of their normal admission procedures.

Another study by Krishnan and Clellend (1973) conducted research on a random sample consisting of 73 students who were denied admission to the University of Pennsylvania's College for Women and Undergraduate Wharton School. They received a response from 52 such students (71 percent). Of those 52, 46 were in another college or university and in good standing, three were in poor academic standing, and three were not attending a college or university. The research omitted comparisons of the quality of the colleges attended by the rejected applicants as it relates to the University of Pennsylvania, but assumed those students who were academic dropouts in these other schools would have also dropped out of the University of Pennsylvania. Therefore, the research established a "best case" scenario for those applicants who were rejected, but does pose a starting point for research of rejected applicants.

Krishnan's and Clelland's study (1973) was not conducted to benefit applicants with low ACT/SAT scores, but rather, to use discriminate analysis as a means to identify

additional applicants capable of graduation so that tuition fees would not be lost. In this study, they concluded that had discriminate analysis been used in the admission process, 41 additional students (56 percent) would have been identified and, hence, would have provided greater tuition revenues to the University of Pennsylvania. While this research did not focus directly upon ACT/SAT scores, it did indicate the possibility that worthy applicants are being denied admission to higher education based upon inaccurate admission processes. The problem of rejecting applicants who would have graduated is fundamental to the question of the proper usage of ACT/SAT scores.

For those studies which consider graduation verses non graduation as the measure of college success (or other nominal criteria), the most common statistical procedure is discriminant and multiple discriminant analysis. Research based upon discriminant analysis considers ACT/SAT scores and often other predictor variables to see if the dependent variables can be used to significantly predict whether the applicant will graduate. This research is generally dependent upon using previous students for the sample.

The technique of using previous students as a sample for predicting success for incoming students creates the potential for biasing the independent variable coefficients (Crouse, 1985). The previous students do not represent all of the applicants of the previous year since not all

students are typically accepted nor is a random sample of all the applicants admitted. Rather, these students represent those who were chosen based upon the ACT/SAT variable and hence the technique will tend to perpetuate the variable chosen. The greater the difference between the applicant pool and those chosen will determine the size and direction of the bias (Crouse, 1985).

ETS has also considered this problem insofar as they have used formulas designed to correct problems created by admissions based upon restricting the range of those chosen, specifically, truncating at a specified SAT score or HSR percentile. According to Crouse (1985), these formulae may not be accurate because the variables used for correction are either unknown or unmeasured. This occurs because most colleges will accept students of high proficiency in mathematics, music, athletics, and other areas even though the value calculated from the prediction equation is below the established minimum required for most applicants. This causes incomplete truncation and introduces error into the process of correcting for restricted range.

ETS, in effect, acknowledges the problems inherent with trying to predict college success based upon such statistical techniques. Crouse (1985) observed that ETS handles the problem of prediction based upon non-random samples by stating their procedure of correction is the "firmest basis" on which to make such predictions. Crouse's

position is that while this may be true, such basis is not firm enough to cause the elimination of some applicants from college.

The validity of these studies is more often in question than the multiple regression research results. Not only is there generally a failure to address the problem of extrapolating the research based upon the matriculants to the general population of the applicants, but the population of the matriculants has variable characteristics not measured by the analysis. For example, if the dichotomy is dropout-graduate, then multiple discriminate analysis cannot distinguish between the genius who dropped out for medical reasons and the graduate who met the bare minimum requirements. The student who dropped out in one week because of course difficulty is not distinguished from the student who did well for three or more years, but for financial or other reasons, did not actually graduate. Both are classified as dropouts.

Predictors of what?

The second disturbing assumption alluded to above is the tendency of the research to focus on the confirmation or rejection of ACT/SAT scores using a criterion variable having little to do with anything of substance. The greatest benefit of ACT/SAT scores appears to be their correlation with FGPA. But FGPA has little, if any,

correlation to a four-year GPA and at best, a four-year GPA has a weak correlation to post-college success, whether considered in individualistic gains or to society as a whole (Krishnan & Clelland, 1973).

The literature reveals the ACT/SAT scores serve as predictors of college success for a portion of the applicants under the most favorable circumstances by calculating a moderate correlation between the test scores and FGPA. But even this moderate correlation falls short of significance when the criterion variable is related to any further understanding of success as defined by the literature. Baird (1982) conducted a study for the College Entrance Examination Board (CEEB) and concluded there was little relationship between academic ability and either educational or occupational accomplishment.

Kelley (1982) made an interesting observation by noting:

> Whether a test predicts performance of value is a political issue, not an issue within testing itself. Further, a test may predict performance but it might not be performance of the kind that an individual is being selected for. Very often as a consequence of the use of tests, we select people for training for a job, not for the jobs themselves. Then, the adequacy of the test becomes somewhat self-fulfilling. (p. 20)

ACT/SAT Scores as a Measure of Aptitude

The next consideration is what the literature has to say about the nature of ACT/SAT examination scores.

Specifically, are the scores a measure of aptitude? If the scores are a measure of a natural ability or capability, then the scores clearly should not be viewed by a superintendent as indicators of school achievement and hence school programs.

As indicated, ETS has, made its position clear: The SAT is a measure of aptitude as its name suggests. Next, the research examines the possibility ACT/SAT scores are not a significant measure of aptitude.

Since ETS has long advocated the position that scores reflect aptitude, logically their validity studies are significant, insofar as they are significant, in reference to the value of the tests being proportional to aptitude. Any findings to the contrary, that is, the scores do not meaningfully measure aptitude, would leave the tests without validity.

Threats to ACT/SAT Scores as Measures of Aptitude

Declining correlation between scores and GPA.

The first indication these scores do not measure aptitude, as cited in the literature, shows the correlation between SAT scores and college GPA decreases consistently throughout the course of a college education. If there is a higher correlation between SAT scores and freshmen grades than there is between senior grades, as all of the research indicates and SAT truly does measure aptitude, the

conclusion would have to be freshmen courses require a higher degree of intellectual capability than do senior level courses. On the other hand, if the assumption is accepted that senior courses are more intellectually demanding than freshmen courses, then SAT scores would not seem to be measuring aptitude.

Coaching.

As a measure of aptitude, the ACT/SAT scores should be robust to coaching. Ability should not be significantly influenced by coaching, yet numerous studies indicate that coaching does influence test scores. Strong SAT supporters such as Comras (1984) conceded detectable differences in student scores can be achieved by participation in well developed test preparation programs. Research by Rogers et al. (1986) found high school preparation made a difference in ACT scores. Students taking a college preparation core in high school scored 22 percent higher than those students who took less than a college preparatory core in high school.

A study conducted by the College Entrance Examination Board (Powers, 1982) to determine whether examinees could improve their scores by pretest preparation. Techniques such as independent study and the use of test preparation books were found to increase the scores on SAT examinations. In 1988, Powers again looked at the effects of preliminary

preparation taking a survey of 1,364 high school seniors and 911 juniors using a stratified random sample of 508 high schools whose students take the SAT. Further study concluded approximately half of the high schools offer special programs to prepare their students for the SAT using resources such as test familiarization materials furnished by the College Entrance Examination Board, computer software, and practice tests. Eleven percent of the students reported receiving private coaching at an average cost of $15 per hour for 10 hours. This study would have been strengthened had the SAT scores of the students participating in the survey been correlated to the means of preparation and contrasted to the scores of those students not participating in the various preparation techniques.

Research conducted by Lockheed et al. (1982) attempted to determine a profile of examinees who requested ETS return their completed tests. They found those who requested disclosure of their tests were from wealthier, better educated families than those who did not request test disclosures. They also found disclosure requests differed among ethnic groups and suggested the possibility that those students who are already advantaged by coaching and other forms of test preparation will continue to widen the gap between their scores and the scores of those who do not or cannot prepare for the examination in the same way.

Johnson (1984) found low income black students' SAT scores were significantly improved through coaching. This was true of both SAT-V and SAT-M scores. Johnson went farther than most of the coaching studies in that she did not limit her research to a simple determination of whether coaching improved SAT scores and thereby allowed low income black students better access to college admission. She further questioned whether SAT scores are a valid measure of a student's ability to pass college level coursework. If SAT scores are not valid indicators of college performance, she argued, improved scores will only give artificial help to these students and not indicate to the students they are actually prepared for college. Other studies on coaching conclude coaching is worthwhile because it increases scores and hence students will do better in college because of the higher scores. Johnson's point (if scores can be increased by coaching, then the scores themselves cannot be indicative of college preparation) is a contribution to the literature dealing in this area.

Slack and Porter (1980) reported that a guidance counselor named Pallone developed a six-week, 90 minute per day course designed to improve SAT scores. The first class of 20 seniors enrolled in his course raised their mean SAT-V scores by 98 points. He also developed a 50 minute per day course which ran from September to March in which the 100

students enrolled raised their SAT-V scores by an average gain of 109 points.

A further consideration is that of bias. Kelley (1982) defined bias as the lack of consistency, regardless of criterion. This definition will be employed throughout weighing whether the ACT/SAT examinations show any characteristics of being biased.

Gender Bias.

Clark and Grandy (1984) conducted research dealing with the performance of SAT scores and gender. The research found there was a gender difference in SAT scores that could not be fully accounted for by differences in high school courses taken, high school GPA, major field and career interests, nor socioeconomic backgrounds. The research indicated that women score higher on the verbal and lower on the quantitative sections of the test than do the male test takers.

Research conducted four years later (Loewen, 1988) also found a gender bias on the SAT test in both the mathematics and the verbal sections. The following year Wilder and Powell (1989) found the same gender differences, though they concluded the gap was narrowing. They too found the difference in scores could not be fully explained by factors such as test characteristics, demographic, social, and biological differences. These same gender differences were

also substantiated by Sharp (1989) who did a study of 2,924 females and 2,715 males in Maryland.

Several other studies found similar findings. For example, in a two year study of ACT scores in a college of 1,000 students, Gougeon (1985) found gender to be a better predictor of college GPA than ACT scores. Her study found that female college mathematics GPA was under predicted by ACT scores.

Navarro (1989) studied 716 males and 1,113 females enrolled at the University of Delaware. She found males scored significantly higher on the SAT than did females. Her research indicated 25 percent of the variability could be accounted for by differences in the number of mathematics, computer science, and physics courses each gender had received, but the remaining 75 percent of the difference in scores was unexplained by her research.

Rosser (1989) conducted an item analysis of the November 1987 SAT responses for 100,000 high school seniors. Twenty-three questions had substantial differences based upon gender, 17 of them from the mathematics section. African-American women had the least gender gap within their own ethnic and racial group while Hispanic women demonstrated the largest gender gap. Her research on 1,112 women also indicated SAT scores under predict female FGPA.

Similar research was conducted by Morgan (1989b). He found SAT mathematics scores are best correlated with high

school coursework in mathematics, science, and foreign languages. Verbal SAT scores were best correlated with the number of years of foreign language coursework. In both cases, the correlation was strongest for those students who had a high GPA and were male.

Lee and Ware (1986) looked at mathematics courses and grades attained by 8,321 high school students. They found that females had a higher attrition rate than males at the lower level of mathematics courses, but females had a lower attrition rate than males in more advanced courses. Like many other researchers, they too found a strong positive relationship between mathematics courses taken in high school and SAT mathematics scores. Kyle (1984) found geometry, in particular, was a high school mathematics course helpful in preparing for the SAT.

A study conducted in 1983 by Lauria compared men seeking traditional career interests with women seeking nontraditional career interests. This study found the nontraditional women had lower SAT Math scores than did the traditional men. These same women achieved higher college GPAs than did the same men who had scored higher on the SAT Math examination.

A similar study also conducted in 1983 (Brunson) found women with low mathematical ability as determined from SAT scores attained higher achievement scores at the end of an introductory college mathematics course when grouped alone

than did women with higher SAT scores in a mixed class. This study indicates the existence of additional variables not possible to measure or predict with SAT scores, for example, the effect of grouping by gender on college GPA.

Heritage (1990) conducted a small study in which she used ACT scores to help try to identify students with high ability to analyze financial statements. She found that accounting students scored much better than the general student population, as might be expected. Within the accounting students (n=118), she found there was no significant difference between the male and female ACT mathematics scores, but the females had a higher college GPA and post-case study scores than the males. This research indirectly indicates that ACT scores under-predict college GPA for females. This lack of predictive consistency between genders meets Kelley's (1982) definition of bias.

Childs (1990) made a distinction between gender bias and gender fairness. Gender bias is a characteristic of the test itself while gender fairness is the way the test results are applied. When a test demonstrates gender bias and the scores are applied equally, then Childs holds the test is unacceptable as a determiner of admission. In addition, Jacobs (1985) conducted a study on 4,145 students at Indiana University. She observed that while females score less than males on the SAT, they achieved higher grades at Indiana University.

Age Bias.

ETS researcher Patricia Casserly (1982) conducted a study which found older students, that is, those within the range of early twenties to late fifties, will have their FGPA under predicted when using a prediction equation based upon SAT scores generated for younger applicants. McConatha (1986) found older students have better scores on the ACT examination than younger students indicating ACT scores are - at least in part- correlated to maturity.

In 1983, Fincher conducted a study using older, non-traditional students in the Georgia University System. Fincher found that when admissions directors combined the SAT verbal and mathematics scores into a single, composite score, and used the composite score as the basis for evaluating learning competencies, errors in admissions increased. He found for older students, SAT scores could be used more effectively when differentiated between verbal and mathematics. He also found for these older students, HSGPA was still the best predictor of college performance.

Recent research by Moffatt (1993) studied 570 students in a small, private Southern college. For Caucasian students over the age of 30, Moffatt determined that SAT scores were not a valid predictor of GPA. He also concluded that SAT scores were not valid predictors of GPA for Black students of any age.

Ethnic Bias.

If test scores measure aptitude, they should do so in a manner which does not bias results with respect to minorities. Results indicate from several research studies such scores are biased. Tracey and Sedlacek (1987), for example, presented an interesting study focused on how indicators of success can have different meanings depending upon ethnic and racial background of the students. If success means remaining in college, then for white students, both academic ability and GPA have a positive correlation to persistence. However, for African-American students, neither GPA nor academic ability were related to persistence. This study indicated the danger in presuming what has value for one sample of the population has identical worth for the population in general.

As indicated earlier, the literature indicates ACT/SAT scores have a positive correlation to college GPA. ETS argues these scores have value because college GPA is an important accomplishment (Turnbull, 1980). Yet for African-American students, the study by Tracey and Sedlacek (1987) could not find any correlation between ability or GPA and persistence in college.

Rogers (1984) used high school GPA and SAT scores, both mathematical and verbal, to form a regression equation for predicting African-American students' college GPA. She found the inclusion of SAT scores in the equation were not

significant in predicting their college GPA. More significant were non-cognitive variables such as "support from family and friends," "not easily discouraged," "and interest in leadership." Rogers also confirmed an interim study done in 1982 (Sanford) which found SAT scores did not contribute to the regression equation for predicting African-American students' success in graduating from college.

Paralleling this research, Lay and Wakstein (1984) considered the relationship of self-esteem to SAT scores. They found African-American students had a higher level of self-esteem at the same SAT score level than did white students. The research concludes self-esteem is related less to academic factors and more to non-cognitive factors for African-American students than for white students. Lay and Wakstein argued that self-esteem is a better measure of college success for African-American students than are SAT scores. Hand and Prather (1984) continued this line of research by studying 45,067 students in 31 colleges. They included high school GPA, college GPA, and credits earned with SAT mathematics and verbal scores and still found college success for African-American males less predictable than for white students.

A study of national ACT scores made by Rogers et al. (1986) found that ACT scores vary significantly by race. In the 1984-85 administration of the ACT examination, whites

scored 40 percent higher than American Indians, 33 percent greater than Mexicans, 2 percent more than Asian-Americans, 22 percent higher than Puerto Ricans/Hispanic/Cuban students and 55 percent higher than African-American students.

ETS test scores have differed significantly between African-Americans and whites to such a degree that the National Center for Fair and Open Testing has endorsed the so-called Golden Rule for testing (Weiss, 1987). This rule states: Among tests of equal difficulty and validity in each content area, questions which display the least differences in passing rates between majority and minority test-takers should be used first.

Kelley (1982) indicated ethnic bias in testing can be the result of cultural differences. He stated "...the vocabulary required in the thinking process which traces the rotation of figures, lines of symmetry, etc. is highly culturally loaded, found often in the habitually used language of some groups and rarely in others" (p. 18). He also noted students who speak different languages experience many problems. Spanish-speaking students may be distracted by capitalized nouns on the test. Distractor bias is particularly pronounced in timed, multiple choice types of tests.

Reich (1991), after having taken the SAT, made a study of the test writers and the content of the SAT examination. He found test writers concentrated on the achievements of

minorities and "bemoaned the short-comings of American Society" (p. S7511). After citing examples of the literature in the SAT, he continued: "There is nothing wrong with references to the lives of writers or a celebration of their works. . . . The problem is that the unrelieved piling up such citations amounted to a racial focus of skewed proportions" (p. S7511). He cited several African-Americans about whom there were questions on the examination and then continued: "It seems as if almost half the questions that named real people, name African-Americans. Furthermore, the fiction-reading comprehension section was from Maya Angelou's 'I Know Why the Caged Bird Sings'" (p. S7511). In addition, the SAT asked questions about Mexicans, American Indians, Eskimos, and Zimbabwe's celebration of its second year of independence.

Reich continued with examples of minorities and women and follows with this observation:

> Stung by criticism that the SAT has been culturally biased, the ETS has, it seems, tried to redeem itself and perhaps avert future criticism by emphasizing the achievements of minorities and women. The trouble is that this emphasis reveals as much about stereotypical thinking as anything could and may well offend many students--minority and majority, male and female--who take the test.
>
> African-Americans, women, and Asians should complain because the ETS, in so pointedly turning stereotypes upside down, highlights them. Jews should complain because Albert Einstein and Saul Bellow seem not to be subjects of questions. Men should complain because they rarely seem to achieve anything unless they're of minority

> origin. Everyone should complain because in one way or another, the SAT distorts reality. (p. S7511)

Institution characteristics.

In 1983, ETS produced a study (Baird) which concluded SAT scores should be used with other variables in predicting college GPA. This research (Baird, 1983) was analyzed using multiple regression analyses. In addition to the finding that SAT scores contributed to GPA prediction, negative correlations significant at the .05 level were found between GPA and such things as undergraduate enrollment, percentage of students in financial need, percentage of minority students, and the greater the variety of sports offered. Significant positive correlations (α=.05) were found for the percentage of students who live in campus housing.

Institution size.

Institutional effects upon predictability of GPA were also found later by R. F. Bolt (1986). His study found a significant main effect in the variation of institution size. The effect on predicting FGPA for students who retake the SAT examination is nearly three-fourths of a grade point on a four point scale as the institutions range from smallest to largest. This finding is based upon 11,929 students from 44 schools who were tested twice, and 1,854 students in 13 schools who were tested three times.

High school size.

Not only do institutional characteristics and college size enter into the prediction equation of college GPA based upon ACT/SAT scores, but so does high school size. What real effect might this have? Hand and Prather (1989) found a tendency for some small town schools to produce students who perform at a higher level than would be predicted by SAT scores. In 1990, they conducted a follow-up study which found that rural examinees clearly achieved higher college GPAs than SAT scores would predict, while urban schools produce students who produce lower college GPAs than SAT scores would predict.

A somewhat general study made by Berlin, Cienkus, and Jensen (1989) found that states with smaller school districts and smaller school size have higher ACT and SAT scores. Another study (Edington, 1981) investigated ACT scores taken from New Mexico public high schools based upon the 1981 graduates. He ranked all of the high schools according to student population. The schools were divided into six categories, the smallest being high schools with fewer than 100 students, the largest category was schools over 2,000 students. The ACT scores from the mathematics, social sciences, natural sciences, English tests, and composite scores were compared and in every case except English, the smallest schools had the highest scores. In the case of English, the smallest schools had the second

highest scores. Further, in every case this difference in scores was significant at the $p<.01$ level. Schools in the 200-499 range ranked last on all test scores. This research does not speculate whether the smallest schools are doing a better job in preparing students for the ACT examination or whether the examination itself has a bias for students from schools of that size.

The State of Illinois had adopted a policy that the optimum school size was 500-1200 and sought to consolidate smaller schools. Rogers (1984) studied 34 Illinois high schools with enrollments under 500 and found among other things that students from these smaller schools had ACT scores greater than the national average.

Special Needs Students.

Studies in 1984 conducted by Bennett and Ragosta and again in 1985 by Bennett dealt with the scores obtained by handicapped people on both the ACT and SAT examinations. The research concluded that handicapped people perform at an appreciable level lower than the national norms on both tests. In these studies, the hearing disabled population performed at the lowest level of all the handicapped groups. They also noted in their study that fatigue is a greater problem for most handicapped test-takers and conservative time constraints on the testing is a factor contributing to lower scores.

Other variables influencing ACT/SAT scores.

The nature of ACT/SAT scores can be studied indirectly by observing what external factors correlate well with the scores. When as many variables as possible have been researched and categorized, an understanding of possible influences of the ACT/SAT scores is possible. ETS does not screen the scores which may be contaminated by misconduct of the test-taker. ETS does not consider it necessary to intervene in cases in which some scores may be obtained by cheating (Saretzky, 1984).

More remote research on declining scores was done by Sternglass and Bell (1983), who concluded that the atmospheric testing of nuclear weapons done in the late 1940s, the 1950s, and early 1960s has been responsible for the decline of the SAT scores. Their study has been praised by Blai (1983) and attacked by Ehrlich (1983). Ehrlich contended that while Sternglass had found a correlation, he did not find a causal link between the testing and the fall in scores.

Belz and Geary (1984) looked at some other unusual variables correlated to SAT scores. They studied the SAT scores from 944 seniors from a suburban high school in San Jose, California. To test for differences within variables, univariate analysis of variance was used. To test between variables, *a posteriori* comparisons were done using Scheffe's S test. They substantiated the usual gender and

ethnic differences but also calculated a significant difference ($p<.05$) between a father's absence from home and lower SAT mathematics scores for both genders.

This study also considered correlations between the father's occupation and SAT scores. Occupations considered were service, cultural (e.g. education, journalism), business, science, outdoor, gun-carrying, organizational, and technological. They determined that students having fathers with service occupations had significantly ($p<.05$) lower SAT mathematics and verbal scores than all other categories. Students with fathers employed in cultural, business, and science occupations had higher scores ($p<.01$) in the verbal portion of the SAT than other employment categories whereas students with fathers employed in science, cultural, outdoor, and organizational occupations showed significantly ($p<.01$) higher mean quantitative scores. The *a posteriori* comparisons showed both cultural and science occupations were the only two occupations associated with the highest set of mean scores. This study revealed no correlation between the ordinal rank of birth order and SAT scores.

Bentley (1982) conducted research to see if GPA could be better predicted by how students deal with stress than by using SAT scores. She surveyed 800 junior college freshmen and sophomores using the Floyd-Steyert Life Stress Inventory. She concluded that students who respond to

stress using maladaptive methods of coping (abusing others, acting impulsively, drinking alcoholic beverages, and exaggerating physical complaints) have a lower GPA than those who do not. When GPA was correlated to stress, she found she could account for 15.5 percent variance in GPA or 3 percent more than could be accounted for by SAT scores.

Specific Program Admission Requirement

Alabama began requiring minimum ACT scores of teaching candidates for acceptance into the teacher education programs within the state. Watkins and Stanford (1983) made a study of those students who had already been accepted to the teacher education program, but who would not have been accepted had the ACT requirement been in effect at the time they applied for admission. Using a sample of 150 students, they found that 70 percent of those students who would not have been accepted (ACT score < 16) passed the teacher education program while 72 percent of those students who met the minimum requirement (ACT score = 16) passed. They concluded that the minimum ACT score requirement will eliminate students who deserve a chance and are capable of completing the teacher education program in Alabama.

Closely related to this study is one conducted by Barger et al. (1988). They looked at the often cited statistic that teacher education programs tend to draw significantly inferior students, as measured by ACT/SAT

scores. Rather than use ACT/SAT scores as a measure of teacher quality, they used the actual GPA attained by all Eastern Illinois University graduates for the years 1982 through 1986. This provided a population of 9,638 students.

While others have conducted similar studies and found that students graduating with certification to teach have a higher graduating GPA than non-teaching graduates, those studies have encountered criticism for being too small and/or grade inflation in education courses. With a sample of 9,638 graduates, Barger et al. sought to calculate GPA based upon junior and senior level courses. To eliminate the criticism of grade inflation in education courses, they considered the cumulative GPA as well as calculated upper division GPA (junior and senior level courses) and adjusted upper division GPA without professional education courses included in the calculations.

Their findings are very interesting. The teacher graduates were significantly inferior to the non-teacher graduates for all ACT scores, but not for high school class percentile rank ($p=.0001$). There was no significant difference in the cumulative GPA for teacher and non-teacher graduates, but they also found that teacher graduates were significantly superior on both upper division GPA and adjusted upper division GPA ($p=.0001$). As a result, their study indicated those students entering education do score significantly less on the ACT examination, but they have a

higher upper level GPA, regardless of whether education courses are included in the calculations.

As Used by the Media

By the time the media makes use of ACT/SAT scores, the nature of the scores changes to indicators of achievement. Even more poignant, lower ACT/SAT scores generate media attention resulting in concern over how little students have learned compared with previous years (Gougeon, 1984).

The Associated Press released an article (1991) attempting to identify the reasons for declining SAT scores. They identified the schools and family as the basic source of low scores. The article postulates ". . . many schools aren't assigning enough homework and are offering students watered down math. . . ."

The influence of ETS upon the media is obvious when scores of tests written by a few people have the sway of public opinion strong enough to challenge the full time work of millions of educators in thousands of school districts. Schools and the family unit itself are blamed for low scores rather than challenge the validity or nature of the test. Just as frequently, the media uses the SAT scores to compare the quality of education offered by the various states (Powell & Steelman, 1984). Powell and Steelman pointed out that such comparisons have very little meaning. For example, only 2 percent of the high school graduates from

South Dakota take the SAT while 69 percent of the high school graduates from Connecticut take the SAT. Powell and Steelman contended such large differences make comparisons from such data unsound. They believed those states with low percentages of students taking the SAT are represented by better students applying to out of state colleges (while the bulk of their classmates take the ACT required for in-state colleges) and hence skew the distribution tending to give their state higher SAT averages than they otherwise would have had.

To test this conjecture, Powell and Steelman (1984) conducted a study of all 50 states and found a 2.76 point decrease in SAT scores for each percent of increase in graduating seniors taking the SAT. R^2 for this study was 0.736; nearly three-fourths of the predictability state rank in SAT scores can be attributed to the percentage of students in each state who take the test. Their regression was not linear and they used a best-fitting nonlinear regression equation which added the square root of the percent to the model. This increased the R^2 from 0.736 to 0.819. This left less that 19 percent of the difference between states' SAT scores to factors other than percent of graduates from each state taking the examination.

The effect of correcting SAT scores by state for the percent of students taking the test significantly changed the rank order of the states. New Hampshire, for example,

originally ranked 28th becomes ranked first when SAT scores are corrected, New York moved from 35th to 13th, and the other states make various shifts. This research is critical to studies seeking correlations between state or regional education and expenditures or other variables.

When race, gender, and family income are added to the regression equation comparing SAT scores across the states, the R^2 increases further to 0.992 with each factor being significant in the direction predicted. Of the three, the racial factor was the greatest contributor. Interestingly, when states are ranked again using all mentioned variables, Montana is the second-ranked state in the nation.

Even though the research by Powell and Steelman's work has been published for over 10 years, nationally noted columnist George Will's column of September 12, 1993, advanced the argument that there is no correlation between education and dollars spent. Will based his argument upon the assumption that SAT scores are a good measure of educational acheivement. Powell (1993) provided an argument against Will's position based upon statistical ranking after taking into account the percentage of students from each state taking the SAT.

Powell and Steelman (1984) also noted than since the SAT has indicated gender and ethnic bias in numerous studies, states will have a skewed SAT score average based upon the gender and ethic composition of the examinees.

Females taking the SAT vary from state to state from a low of 42.6 percent to a high of 55.1 percent. The range of African-American SAT test-takers state by state varies from 0.4 percent to 24.2 percent of the test-taking population. Any factors which are found to have a measurable bias in the SAT examination, such as those previously identified in this review of literature, contributed to the same considerations. Bonstingl (1992) went so far as to credit the media's use of SAT scores with creating the current reform in education. He noted that the media commonly uses the SAT scores as a measure of K-12 achievement.

As Used by Politicians

Support for standardized testing seems to be oblivious to political parties. During the Carter administration, the President's Commission recommended the development of a national criteria and assessment program, particularly in foreign languages (Farnham et al., 1982). While President of the United States, Ronald Reagan stated that he wanted the SAT scores to rise by 15 points (Thornell & Jones, 1986). Reagan believed the decline in SAT scores to be caused by too much Federal Government involvement in education (Powell & Steelman, 1984). Former President Bush has repeatedly gone on record stating he would like to see SAT scores improved and President Bill Clinton, who ran as a

strong proponent of education, has stated he would like to see national criterion-based testing (Fiske, 1992).

Former Secretary of Education Lamar Alexander (1991) has also publicly expressed his understanding of SAT scores. He believes the low SAT scores reflect that "even our best students generally don't know enough and can't do enough to assure success in tomorrow's world." This is an echo of the 1983 National Commission on Excellence which held that the decline in the SAT scores indicates the United States may have diminished to the degree to which it is now impossible to compete economically with certain other nations (Powell & Steelman, 1984).

Bowen (1977) studied the decline in SAT scores over the past 20 years in a statistical manner. Bowen tested for significance when change in test scores occur by using the following formula:

$$\frac{\mu_1 - \mu_2}{\sqrt{\frac{1}{2}(\sigma_1^2 + \sigma_2^2)}}$$

where μ_i is the mean score in the ith year and σ_i is the standard deviation in the ith year. The value obtained by this calculation is compared to a table of significant values. Using this formula, the decline in SAT scores from 1972 to 1991 produces a value of -0.279 for verbal and -0.083 for mathematics, neither of which is significant.

Periodicals serving educators are not immune to repeating an alarm about the decline in test scores without a consideration of their significance. An article in The Executive Educator (1992) by Smith typifies this alarm. The article warned that based upon no improvement in ACT scores, we will not be able to meet the goals of America 2000 unless we redouble our efforts to educate students in mathematics and science.

Ranking states by SAT scores has had a political significance. Powell and Steelman (1984) have noted that politicians from states with high SAT scores point to their state's educational system as superior. Politicians from states with low SAT scores tend to favor some sort of educational reform.

Senator Simon from Illinois, in advocating America 2000 from the floor of the Senate, stated his "regret" that while Japan has students entering teaching who score at the top of the college entrance exams, those entering teaching in the United States do not have praiseworthy scores.

The report titled A Nation at Risk: The Imperative for Educational Reform issued in 1983 by the National Commission on Excellence in Education determined that the nation's school teachers were being drawn from a pool consisting of the bottom quarter of the college students. This judgment is based largely on ACT and SAT scores (Barger, 1988).

The literature reveals that ACT/SAT scores are well established and accepted politically as measures of achievement, intelligence, ability, and other indicators needed to measure education. Warnings issued by ETS, CEEB and numerous research to the contrary either are unknown by politicians or unheeded.

Other Uses

Scholarship Selection

ACT/SAT scores are used at various levels of importance for the selection of scholarship winners. Some states and organizations have used the scores as the exclusive factor in awarding scholarships while others use the scores in conjunction with other considerations.

A study conducted by deLacy (1985) investigated the effect of requiring minimum ACT/SAT scores for athletic scholarships in Iowa. The findings for the 78 students studied indicated there were athletes who met high school GPA and pre-college course preparation requirements for the scholarships, but failed to qualify because of the level of their ACT/SAT scores. Further, he found the male students were denied scholarship opportunities much more frequently than were the female student-athlete applicants.

The issue of scholarship selection using SAT scores as the sole means of awarding scholarships was recently challenged in the state of New York in the case of *Scharif*

v. New York State Education Department (Childs, 1990). The New York State Education Department awarded scholarships based upon SAT scores. The plaintiffs argued that the SAT scores were biased against women and hence a larger number of education scholarships were awarded men than women, even though women had higher high school grades than men. The State Education Department argued that the SAT was the best objective measure available and they should be allowed to continue awarding scholarships based solely on SAT scores.

The plaintiffs also argued that the New York Education Department has the stated intention of awarding scholarships based upon achievement and HSGPA is a better indicator of achievement than SAT scores which do not claim to measure achievement. The court agreed and ruled that New York could no longer award scholarships based solely upon SAT scores.

Recently, Bracey (1993) argued that the PSAT should not be used as the sole means of qualifying for the National Merit Scholarship Program. Since research clearly indicates that women score lower on the PSAT than men, but that men don't do better in college than women, Bracey argued that the PSAT is not acceptable.

The argument in favor of using this test in spite of its gender bias has been twofold: women take easier courses in college and women can endear themselves to their professors and obtain higher grades. Since the aforementioned argument is without evidence, ETS conducted

research in the fall of 1992 to cast more light on the question of gender differences. Using a sample of 47,000 men and women, they found women did attain a higher mathematics GPA in college when taking the same courses as men. This research by ETS continues to confirm other research that SAT scores should not be a determiner of scholarship selection, and according to Bracey, this research has implications for college admission processes as well[2].

To Determine Higher Education Funding

Colleges and universities have traditionally not worried about what relationship there might be between high school effectiveness and ACT/SAT scores. They were content to seek out the examinees with the higher scores since they were easy to identify. This has now changed and schools of higher education are taking a serious look at how well ACT scores measure actual learning and, hence, reflect back upon effectiveness of educational programs.

This recent interest has been generated by the adoption in some states of higher education funding according to ACT College Outcomes Measures Project (COMP) scores. Entering freshman scores for ACT are recorded and compared to their

---

[2]Annually 15,000 students are designated as National Merit Scholarship semifinalists based solely on the PSAT scores. Of these, only 1,500 actually receive awards, and of these 1,500, two-thirds are males.

ACT COMP scores when they exit as seniors. Those schools showing the greatest gain are given a greater share of funding.

Pike and Banta (1989) studied the correlation between ACT and ACT COMP scores. They determined that as much as 15 percentile points could be accounted for in errors of measurement, yet funds were awarded based upon differences much smaller than those attributed to errors of measurement. They concluded that if the effectiveness of the program is defined in terms of ACT scores, then the quality of the program is the quality of the students attending the school. An institution could improve its performance rating, and hence its funding, not by improving its academic programs, but by becoming more restrictive and selective in its admissions. Finally, schools could improve funding by narrowing the curriculum in order to address the precise items measured by the ACT COMP.

Athletic Eligibility

The SAT scores have recently been used to determine eligibility for student athletes. Proposition 48 by the National Collegiate Athletic Association requires freshmen to have an SAT score of 700 to play varsity sports (Thornell & Jones, 1986).

Considerations on Testing

There are considerations on testing found in the literature which must be addressed especially when test scores are used for vital purposes such as college admissions. Many of these considerations can be traced to the problem of using aggregate data concerning questions which are, to a large extent, about individuals. The degree to which inferences that can be made about microlevel relationships from the macrolevel data is not a resolved issue (Hand & Prather, 1986).

Gougeon (1984) investigated whether a conservative time limit on the mathematical portion of the tests is detrimental to identifying students with good mathematical ability. Her argument was based upon the fact that many, if not most, high school students are not conditioned to rush through their mathematics tests, rather the tests are commensurate with the allotted time. Her research found HSGPA is a better predictor of FGPA than the SAT mathematics test and she attributed this in part to the time constraint placed upon the SAT test.

Belcher and Einspruch (1987) conducted a study of 641 students who had failed the College Level Academic Skills Test. They wanted to see if these students could improve their scores if given more time to complete the examination. They concluded that students who were given 70 minutes to retake the test significantly outperformed students given

only the normal time of 35 minutes. This suggests timed standardized examinations consider speed a more important measure of ability and achievement than number of correct answers.

Time constraints were also studied by Davis (et al., 1987). While he based his study upon the Academic Assessment Placement Program created by ETS and not the SAT, his considerations on the effects of timed tests apply to considerations on testing in general. He found the time limits on tests particularly hurt African-Americans. ETS' prediction of the number of test-takers who complete the test was an approximation between white and African-American students. ETS estimates 45 items can be completed by 80 percent of the test-takers in the allotted time, but Davis found this meant that 90 percent of the whites finished while 74 percent of the African-American students finished. Davis estimated that just adding 10 additional minutes would allow 98 percent of the whites and 95 percent of the African-Americans to cover every item. Since Davis could not find any literature indicating that speed has been demonstrated to be an important factor in success in college level work, he concluded that the addition of extra time would benefit all test-takers.

Somewhat in contrast to these studies dealing with the time limitation factor was a study conducted by ETS researcher Schrader (1984). His study observed that the

percentage of test-takers either omitting or not reaching a test item increased with the difficulty of the item. Hence, examinees may take different amounts of time on reading questions than on sentence completion, antonyms, and analogies. He suggested changing the format of the test so that the reading is a timed test in itself and the other three parts of the verbal section comprise another timed interval. His suggestion, however, was not based so much to help examinees allocate their time better as it was to help researchers conduct validity studies on the test itself.

Houston (1981) questioned the face validity of objective verbal tests which do not actually require the students to write. While typical tests might measure the student's ability to edit standard English, they do not test the ability to choose a topic and write a coherent essay. Research such as Schwartz and Wilbur (1981) indicated that colleges should conduct predictor research at their own institutions as the amount of variability that can be attributed to ACT/SAT scores differs significantly from college to college. This raises the concern of relying upon a nationally normed reference test as an important factor in determining college admissions.

Hand and Prather (1986) treated this concern in detail. They made the point that "college is an independent grouping factor in no sense like a random grouping." (p. 4). This suggests that data taken from the aggregate may not be

representative of any particular college since the inference of data to a particular college does not represent a random selection.

ETS recognizes the problem of one college greatly differing academically from another and has allowed for it in their prediction equations. Average SAT scores reported by a college are used as a guide of the degree of selectivity of that college. Regression equations are derived from previous students and applied to incoming freshmen. The problem with this has been discussed earlier in this review regarding statistical limitations on restricted range calculations. Colleges which select students based upon ACT/SAT scores create the situation where incoming students mirror the qualities of the sample group.

Using statistical techniques developed by Robinson in 1950 and later by Prais and Aitchison in 1954, Hand and Prather (1986) analyzed the application of group data to individuals. They selected independent variables of SAT verbal, SAT mathematics, HSGPA, credits attempted, and credits earned. Five covariants were chosen: African-American females, African-American males, white females, white males, and the composite, on two levels; individual and college. The regressions of the covariates with GPA were calculated and reviewed to investigate if there is any

aggregation bias present between the individual and college level.

Using 45,475 students, they found the SAT verbal was correlated negatively at the college level in four cases of the five cases (all except white female, which is negative for SAT mathematics). Yet on the individual level, all four of the SAT verbal and the one SAT mathematics regression coefficients that were negative at the college level are all positive at the individual level. This is ". . . strongly indicative of the bias introduced by grouping." (p. 5). Further, four of the five regression coefficients decline from the individual to the college level for HSGPA which adds credibility to the position that regression coefficients are markedly changed when group data were used. The aggregation bias or effect is significant when employing SAT scores to predict individual performance and is summed up by Hand and Prather in these words: "Anyone hoping to ascertain individual relationships from aggregate data of this type should expect little likelihood of success." (p. 5). Wakefield and Bullock (1982) confirmed this view in their study of 501 college students in Indiana. They found students with high SAT scores on one campus having average GPAs while students on another campus had low SAT scores but still average GPAs.

Other countries also require college entrance examinations of their applicants. These countries parallel

the situation in the United States where admission to college is heavily determined by entrance examination scores (Farnham et al., 1982). These tests are also plagued by the same types of bias documented for the ACT/SAT examinations (Hansen et al., 1989).

Misuses of ACT/SAT Scores

As an Assessment of High School effectiveness

Friedland and Friedland (1988) observed an increase in pressure to use SAT scores as a measure of instructional performance. However, researchers (as opposed to politicians and those in the media) generally do not assume the ACT/SAT examinations measure high school effectiveness. There is virtually no research seeking to establish a correlation between these scores and high school programs. There is however considerable literature and research designed to show ACT/SAT scores should not be considered indicators of effective high school programs.

Kifer (1985) stated that the greatest misuse of these scores would be to equate them with the actual education of the examinee. Recent articles taken from the American School Board Journal by Jim Montague (1990), Jo Anna Natale (1990), and Robert R. Rayborn (1989) explicitly stated these tests do not measure programs or school effectiveness. Natale noted that California and Connecticut are moving

toward performance based assessment programs and away from tests like the SAT.

Daniel Stufflebeam (1984) held to exactly the same position and maintained standardized test scores should not be regarded as a major criterion for evaluating educational quality. Popham (1981) wrote the same in his paper: Crumbling Conceptions of Educational Testings while Hilliard (1984) pointed out evaluation should be ordered toward the active, critical initiative of the learner and not simply with the academic outcomes generated by standardized testing.

Bracey (1989) noted that the use of the SAT has increased despite all of the criticism resulting from the research of the 1980s. He recognized the SAT is still used to measure the quality of high school graduates and the average scores of the matriculants provides a scale for expressing the merit of various colleges. Bracey considered this to be a great misuse of the SAT.

Evangelauf (1989) pointed out the effects of such misuses in her article regarding flaws in the admission process. When ACT/SAT scores are used to measure student quality, many good students are omitted by the admission process; when ACT/SAT scores are used to rank colleges, the high scoring students often limit their choices to a few high ranking institutions while lower scoring students are not even considered by such schools. Evangelauf argued that

this whole process is not rational; therefore, it is not consistent with quality education.

Hanford (1986) advised the governors attending the National Governor's Association Task Force on College Quality in Washington D.C. that there was no single instrument available for assessing student achievement. Further, he maintained admissions tests should not be used to evaluate the quality of high school or undergraduate programs. Gougeon (1984) also noted SAT scores do not measure high school effectiveness. She quoted CEEB President Sidney Marland as saying the tests were "not designed as a measure of school performance and should not be used that way" (p. 10). Yet, educational periodicals frequently report education has improved when SAT scores rise. For example, in Phi Delta Kappan, Lewis (1992) took the position that certain students have increased their education because they have increased their SAT scores. The School Administrators of Montana Bulletin (Hitz, 1992) made an identical point.

In spite of the above opinions, Slack and Porter (1980) observed high schools are under pressure to alter their curricula to improve test scores. They also noted that while ETS publishes the opinion that SAT does not measure teachable skills, administrators have an implied obligation to furnish an education which teaches to the college entrance examinations. Gougeon (1984) also shared the

opinion that many people, including educators, believe that falling nationally-normed mathematics scores are an indication that our high schools are failing to teach mathematics courses properly.

Powell and Steelman (1984) pointed out that politicians and the public in general tend to view SAT scores as indicators of high school preparation. Boyer (1983) also noted that the general public believes the SAT is a valid "yardstick used to judge school performance" and the scores serve as the "nation's report card." Boyer, an ACT Trustee since 1989, is firm in his belief that the SAT scores (and ACT by implication) are not an indicator of secondary school quality and that the tests fail in their original intent to measure the aptitude of the test-taker.

In spite of the literature, articles abound in educational publications in which ACT/SAT scores are vital signs of education's health, (e.g., Saks (1992). Recently Johnson (1993) noted that public schools are expected to achieve a high degree of academic excellence. She stated that this degree of academic excellence is measured, in part, by SAT scores. She did not indicate whether this is her opinion or that of the public, but she did make it clear that SAT scores are perceived to be important measures of K-12 achievement.

As a Measure of Teacher Quality

Gougeon (1985) warned that in addition to high school programs, teacher quality should not be measured by ACT/SAT scores. She quoted earlier research by Traxler in which he found this particular abuse of ACT/SAT scores to be common. Johnson (1992) pointed out that teachers frequently take the blunt of the blame for low SAT scores. This in turn is interpreted by the public that teachers are the blame for the perceived failure of public education.

As a Measure of Public Expenditure

Powell and Steelman (1984) reported that SAT scores are being used to make judgments about public expenditures in education. These judgments are based upon not finding a correlation between amount of money spent per student and SAT scores. The conclusion is that spending more money per student will not improve education because there is not a significant correlation between the expenditures and the SAT scores. Powell and Steelman pointed out that SAT scores are not a measure of educational quality and, hence, the studies should not seek to correlate expenditures with achievement.

View of ETS

ETS (1984) has warned that SAT scores should not be used to measure performance of teachers, schools, measure the worth of a human being, for sole decision making when other

information of equal or greater relevance is available, and for interpreting scores without regard to the standard error of measurement. In reviewing the literature, it does not appear these recommendations by ETS are known to all researchers, educators, politicians, media editors, or educational administrators.

Gougeon (1985) summed up the misuse of ACT/SAT scores succinctly when, after conducting numerous studies, she wrote: "In addition, misunderstanding what the test numbers mean and thus overestimating their precision can be a problem. Thus, few recognize that a test score is meaningless except in its relationship to the other numbers." (p. 12).

Relationship to Business

Powell and Steelman (1984) pointed out that SAT scores are so commonly perceived to be measures of educational quality that it is possible industrial expansion will be attracted to states with high average SAT scores and discouraged from states with low SAT scores. They noted the media and several political leaders have suggested Texas and not Georgia was selected for a major research center based upon comparative state SAT scores.

Effect upon Students

The effect of ACT/SAT examinations upon students has been considered by some researchers. Originally the test scores were only available to colleges and universities. In the face of much opposition, ETS trustee Dr. Frank H. Bowles argued and won in 1958 the release of these scores, even though he jokingly remarked that releasing these scores would create a socially embarrassing classification for students. His successor at ETS, Dr. Sidney Marland, noted with approval 10 years later that the release of ETS scores were altering perceptions of individual worth, not just in the eyes of the individual, but in the eyes of parents and society, as well (Nairn, 1980).

This potentially negative effect upon students who score low is due in part to the method ETS reports as their measurement, that is, as an aptitude. If the SAT score is an accurate indicator of aptitude as claimed, then a low score lowers a student's self-esteem. The test does not claim to be simply a measure of how much is known, something any student can conceive as possible to alter, but rather as a measure of what the student is capable of learning. Yet as indicated earlier in this review, there are serious questions about whether these examination scores can be conceived as a measure of aptitude.

Slack and Porter (1980) considered the effect of ACT/SAT scores upon the student in these words:

> . . .if a student with a low SAT score believes that this score accurately reflects his aptitude, and that no study or preparation could have altered his performance, his mental capacity is disparaged, and his self-esteem is weakened. If the message is untrue, the student has been seriously wronged (p. 155).

They go so far as to say ETS has "enhanced their product's image at the risk of the student's self-esteem" (p. 172).

The negative effect upon self-esteem is continued by ETS through their encouragement for applicants to self-select colleges based upon matching test scores with average scores of successful applicants at a given college (Crouse, 1985). This is to say, examinees are encouraged to select a college compatible with their ability. Hence, SAT scores become further established as a measure of self-worth and students with low scores see themselves as not able to succeed in better colleges regardless of whether they are accepted. Hence, SAT scores affect not only whether students go to college, but also where they go.

Slack and Porter (1980) brought up a point that escapes most observers of assessment: "ETS requires students who take the SAT to work as experimental subjects without informed consent. Students who apply to colleges that require the SAT have no alternative." (p. 170).

Stress and nervousness experienced by examinees have been studied. The College Board found 20 percent of the test-takers are "very nervous" and 56 percent are "nervous" prior to taking the examination (Crouse, 1985).

Students also are required to pay for each administration of the test they take. If a student desires coaching and can afford to be coached, it is possible to invest in excess of a $1000 for coaching classes.

Effect upon Education

The effects of ACT/SAT scores upon education are broad and far-reaching. Some educators believe the decline in ACT/SAT scores is an indication that our school systems are failing to teach properly (Gougeon, 1984). The literature has indicated a large number of factors which can account for this decline in national average scores, but none is more worrisome than the consideration that ACT/SAT tests do not measure what many educators believe should be presently taught.

Slack and Porter (1989) considered the question in detail. They concluded that "emphasis is given to little used vocabulary and algorithms that are rarely presented in high school courses" (p. 163). They continued, "Our analysis indicates that the SAT is itself a measure of past accomplishments, but one that measures skills remote from most high school and college curricula."

Since the prediction equations designed by ETS and used to determine the fate of present college applicants are based upon the previous year's matriculants, there is immediately introduced into the process a one year

technological gap. ETS believes it can norm the tests from its inception in 1941, by using unpublished techniques, to the present and, hence, equate a score 50 years old with present day scores. Numerous authors question the validity of such thinking (e.g. Crouse & Trusheim, 1988; Gougeon, 1984; and Nairn, 1980).

Calculators have been part of many high school's math curriculum since the early 1970s. Twenty years later, these skills are just beginning to be measured by the ACT/SAT examinations (The College Board, 1991), while for the interim 20 year period, ACT/SAT examinations (hence, college admission) was based in part upon the manipulative skills calculators are designed to relieve. The inclusion of calculators in the SAT, however, does not necessarily mean the test will include them in any meaningful way since The College Board (1992a) has acknowledged that the SAT may be taken without their use as well.

Compounding this problem is the integration of computers into the curriculum not only to enhance education *per se*, but also to meet the demands of business and college expectations. Computers are capable of performing many complex mathematical operations and many educators believe it more important to teach computing skills which allow for the solution to complex problems than to teach the algorithms themselves (Bennett & Whittington, 1986). While technology which eliminated complex computation was well in

place in many school curriculums by 1990, ETS announced the SAT would increase its emphasis on mathematical computation.

However, even with the New SAT in 1994, designed for the first time to include measures of achievement, there is no mention of any testing of computer competency. This, despite the fact ETS reported in 1981 it had found test takers' interest in computer education had tripled in the past six years. The College Entrance Examination Board also conducted research nearly 10 years ago (Oltman, 1983) which indicated cognitive assessment would be more comprehensive if technological tools are integrated into the process, yet the literature does not indicate the Board deemed it important to include technology as part of the assessment.

The New SAT took three years of "extensive research and field testing" (The College Board, 1991). Yet, it still does not address a significant educational phenomena, computers in education, which in 1994, completed their second decade of usage. As a result of this gap, a widening of the gap between SAT scores and FGPA may continue.

Some writers have noted that low SAT scores can be attributed, at least in part, to this test lag (Gougeon, 1985). Gougeon notes that low SAT scores might also be attributed to high schools and colleges changing educational priorities resulting in better correlations between the two schools than trying to improve correlations with a nationally normed test which still reflects outdated

educational methods. The testing of skills not demanded by either society or higher education to the exclusion of those currently necessary to both, not only manifests a serious threat to the validity of ACT/SAT examinations, but also presents the fundamental paradox of the entire question of standardized testing: If a test is modified prior to the change in curriculum, it drives the curriculum; if a test delays modification until curriculum is changed, it frustrates educational change and reform becomes difficult. Standardized examinations present themselves as instruments of a no-win agent. This is particularly true of nationally-normed tests since the predictive validity of the test is based upon a norm no longer representative of the abilities and skills of the people now taking the test (Gougeon, 1984).

Omissions in Research

Educators and the general public often understand the ACT/SAT scores to not only be acceptable predictors of college success, but also as measures of both the academic content of what colleges expect their applicants to know and the content of what high schools have effectively taught (Comras, 1984). The literature does not support this view. Completely lacking from the literature is any research which seeks to correlate the substance of ACT/SAT tests with the knowledge base expectant of incoming freshmen. Similarly,

there is no research which compares the content of the ACT/SAT examinations with the initial content of respective college disciplines. There is no research which indicates the relationship of content from examination to examination. For example, does one examination emphasize geometrical skills more than algebraic, while the next examination places more weight on algebraic? This question has not been researched. The literature is also silent on any research showing these tests represent or measure an effective K-12 program.

There is also no literature which has established that these tests are a measurement of anything other than ability. Yet, the literature does not support the view that these tests effectively measure ability. Further, there is no literature showing these test scores are a measure of achievement or any other alternate value.

The large amount of literature correlating ACT/SAT scores with college success omits, without exception, a consideration of the relationship between college success and scholarships awarded based upon ACT/SAT scores. Since financial help is frequently granted on the basis of ACT/SAT scores, such financial help should contribute to the success of those who score well on the ACT/SAT examinations. This creates the proverbial self-fulfilling prophecy by providing help to those who are predicted to do well.

In spite of the extensive research correlating ACT/SAT scores with qualifications for admission to college, there is no research showing these are the applicants who should be selected. Even if the ACT/SAT tests were highly predictive, there is no research which purports to defend these test scores are the relevant standard by which to select individuals for admission. This is to say the research is ordered to quantitatively determining admissions without regard to quality. Both Crouse (1985) and Sawyer (1989) warned the correlation itself does not provide sufficient justification for placement. Kelley (1982) took this point farther when he warned that even if standardized tests are free from all forms of bias, these tests (particularly aptitude tests) do not necessarily select the best people for educational opportunity or employment.

There exist some assumptions in the literature which are questionable. For example, all of the studies which found positive correlations between ACT/SAT scores and college success assumed uniformity in the quality and ability of college professors. Also assumed were equality of all other variables including class size and other factors commonly thought to influence learning.

Sawyer (1987) found that colleges could not use ACT test scores of fewer than 50 students and expect to avoid large sampling errors. There is no research which determines the minimum sample size needed from a given high

school necessary to achieve the expected correlation between ACT/SAT scores and FGPA as published by the two test makers. Finally, the literature does not advance any research or theoretical work ordered to give positive direction to the behavior of a superintendent or other school administrator.

The problem addressed by Schwartz and Wilbur (1981) and also by Hand and Prather (1986) cited earlier addresses a fundamental and philosophical problem inherent in predictive validity. How far can aggregate data be used to make assumptions about individuals? While Hand and Prather addressed the problem from a statistical point of view, the literature was silent on the same question answered from an ethical point of view.

Summary

The literature shows the most prominent and most highly regarded examination, the SAT, is produced as a business venture by a business, Educational Testing Service (ETS), grossing $200 million a year. The tests are written for the purpose of providing colleges and universities with a useful and low cost tool for screening college applicants. The review of literature indicates that the proper use of these scores is not clearly understood by the most knowledgeable educators and researchers; much less should they be acted upon by the student or administrator.

The apparent value of the ACT/SAT examinations lies in their ability to predict "success" in college. The literature most favorable to the usefulness of the ACT/SAT examinations is found in those studies which defined "success" as college FGPA. A typical correlation is approximately 0.45, which would tell the researcher that 20 percent of the variability of college FGPA can be predicted by the ACT/SAT examination scores. When these scores are correlated to all four years of college GPA, the correlation becomes less than half of the freshman GPA value when considered alone. This in turn diminishes the predictive contribution of the scores by 400 percent to a value less than 5 percent.

The literature brings into this scenario an even more interesting point. There is a weak relationship between four year college GPA and post-college success whether measured by contribution to society or more individualistic terms. Hence the ACT/SAT examination scores find their best application in that they predict "success" insofar as they have a low moderate correlation to FGPA, which in turn has a low correlation to a four year GPA which itself is weakly related to post-college success.

The literature finds other variables which are more useful, more economical, and more easily. Two easily-identified variables are high school GPA and high school

rank. Research consistently finds a higher correlation to FGPA using HSGPA and HSR than when ACT/SAT scores are used.

The ACT/SAT scores ultimately have quantitative value if they serve to decrease the number of false positives (acceptance of students who will fail) and false negatives (rejection of students who will succeed) beyond what high school rank alone predicts. Research sponsored by ETS concludes that college GPA is more accurately identified when using both high school GPA and ACT/SAT scores. This research is conducted using the statistical technique of multiple regression in which both variables are used in the predictive equation. An examination of the additional weight contributed by the ACT/SAT score variable in the predictive equation indicates such contribution is minimal. For example, if the criterion variable is a bachelor's degree with a predicted GPA of 2.5, then the inclusion of SAT scores in the prediction equation increases the accuracy of the prediction by one student per 1000. In terms of GPA, when SAT scores are used along with high school rank to determine which students are admitted to college, the increase in FGPA as a result of factoring in SAT scores is 0.02 on a 4.00 scale.

To compound matters, there is an absence of literature produced by ETS or otherwise which offers arguments that those who score well on the ACT/SAT examinations are those who should be admitted to colleges. There is an assumption

by ETS and those who use the scores in any way as entrance criteria that the correlation itself argues that high scores represent those who should be admitted, yet no research or argument is advanced to suggest the correlation between ACT/SAT scores and college GPA identifies a population that should be admitted to college. Hence there is no research defending the use of ACT/SAT scores as criteria for admission. Kelley (1982) perhaps summed up these tests best when he noted "...standardized tests were never designed to maximize individual equity. The great benefit of a test and test score is that it provides some rank-order which can be used for low-cost selection."

Finally, the literature has identified two important philosophic considerations. The first is the change in philosophy from seeking students who have been prepared for college to those students who are most able to succeed in college, whether due to preparation or innate ability. The literature has pointed out that this shift in philosophy removes from the prospective college student the ability to assure further education through hard work. This person falls victim to rejection based upon the acceptance of a candidate with more ability, but possibly less achievement and preparation.

The second philosophic consideration found in the literature is that of logical induction. How do colleges logically and ethically select students using a generalized

prediction equation based upon individual observations which in themselves are twice biased in that they are a non-random sample already selected on the basis of the same variable being used as the independent variable of the applicant population? The problems and concerns found by various particular studies reported in this review of literature can be ultimately traced to these two philosophic considerations. This review of literature strongly suggests that any educational testing service should address these two principles prior to testing a nation.

The literature leaves little doubt about the importance of the ACT/SAT scores in the lives of the test-takers and the role of administrators in school districts. This two-fold importance creates the basis for formulating a clear set of principles upon which administrators and all educators may properly address these two tests. In addition, the lack of research describing how these scores are presently handled by school district administrators provided the impetus for this investigation.

CHAPTER THREE

METHODOLOGY

The review of literature indicated that substantial actions were taken at the superintendent's level in reaction to the ACT/SAT scores received by students. The literature also showed, for the most part, that superintendents and other school officials are primarily reactive with respect to ACT/SAT scores. In the absence of any descriptive or definitive work in this area, administrators had no basis upon which to formulate and administer any consistent reactive leadership.

In addition to providing little administrative direction, the literature indicated a number of serious concerns regarding the ACT/SAT scores and the nature of testing in general. This research formulated general principles for administratively dealing with ACT/SAT scores. Finally, an understanding of how Montana superintendents are presently utilizing these scores was determined by means of a census.

Research Design

This research was primarily of an analytic/synthesis nature. It was analytic in that the research was formed by conceptual and philosophical considerations and correlations sought, yet synthetic in that the conclusion led to a theory of administrative behavior. The research used a descriptive

census which was utilized to codify how Montana school districts were utilizing ACT/SAT scores relative to the study.

Population

The population targeted for this study was the superintendents of Montana secondary schools. Superintendents who did not have administrative control over high school students were not included in the study. The entire population of such superintendents in Montana were surveyed (excluding the researcher). All mailings contained a census instrument, a self-addressed, stamped envelope and a mailing label with his/her name and address on it. The census was returned in the envelope provided and the mailing label was returned if the respondent would like a copy of the census findings, a summary of the review of literature, and a copy of the conclusions. Superintendents not responding received a second letter and another census.

Instrument

The instrument used for this study was a questionnaire designed to investigate both present administrative behavior and present understanding of ACT/SAT scores. The instrument was developed as part of the research. Since the census instrument was not a recognized standard, face validity will be established by a pilot distribution given to principals

and superintendents of elementary (only) districts in order to determine if the questions are understood in the sense in which they are intended. This process continued until all ambiguity was removed from the instrument. Content validity was established by presenting the instrument to The University of Montana personnel who have expertise in standardized testing.

The external validity of the census instrument was a function of the response rate and the consistency of the replies. The consistency of the replies was measured by Kendall's Coefficient of Concordance using both raw scores and rankings. The reliability of the questionnaire was established primarily by the rational equivalence technique. The cover letter and instrument are found in Appendix A.

Analysis of Data

The responses to the questions were reported descriptively by tally, averages, and percentages (see Appendix A, Table 1, Page 159). Correlations (Spearman r_s) were calculated to measure the strength of the various relationships between experience and the role ACT/SAT scores have in Montana secondary schools as viewed by superintendents. Specifically, significant correlations ($p \leq .05$ and $p \leq .01$) were reported for any correlation between experience and whether it is believed ACT/SAT scores are: indicators of effective schools; whether low scores effect a change in

curriculum or a change in personnel; and what actions have actually been taken by superintendents. The same correlations were calculated for school size as well. Other similar statistical strategies were employed and any significant statistics were reported.

A frequency distribution was given for persons involved with low ACT/SAT scores (question 9). The rank order for questions 11, 12, and 13 was determined and reported using the Wilcoxon Rank Sum Test. Significance ($p \leq .05$) between rankings was determined by using a Mann-Whitney U Test. Kendall's Coefficient of Concordance (Friedman, 1991) was used to measure the consistency of the rankings among respondents. From the data, a summary was formed which described the present administrative views and behavior of Montana secondary school superintendents relative to ACT/SAT scores received by their students.

The availability of large amounts of research on ACT/SAT scores which can be grouped made synthesis possible. Chapter Four contains the theoretical treatment of ACT/SAT scores resulting from the synthesis as well as the findings of the census. Chapter Five contrasts the theoretical findings with the descriptive findings.

Expectations of Research

Owing to the absence of any high school level direction in the area of K-12 ACT/SAT score usage, most school districts were making very little usage of these scores. Those that did make use of them, if they were consistent with the superintendents noted in Chapter Two, tended to be acting on low scores and making critical changes to keep scores and, hence, public perception of school effectiveness, as high as possible. Specific recommendations on the use of these scores was given as a synthesis of the research and did not support either ignoring the scores or using them as indicators of effectiveness. Specific guidelines have been formulated for K-12 administrative use of ACT/SAT scores. Significantly, higher education should make use of this initial study in order to address this in leadership curriculum as well as in determining admissions.

CHAPTER FOUR

FINDINGS OF THE RESEARCH

Census Results

A total of 138 responses were received from 166 possible yielding a response rate of 83 percent from two mailings. The first mailing yielded a 60 percent return. The response rate indicated an interest among high school administrators in Montana in the research. Fifty-two percent of the respondents indicated a desire to have a summary of the results sent to them.

The results of the census were reported by question number. Descriptive information was given and significance was noted where appropriate. A summary of responses was available in the Appendix and in the tables following the discussions of each question on the census.

Questions One - Three

Questions One, Two, and Three primarily provide descriptive data regarding the respondents. Of the 138 respondents, 86 percent (119) are superintendents while five percent (7) are principals, seven percent (12) are counselors, and the remaining two percent are 'other.' The respondents have an average of 12.5 years of administrative experience and 19.5 years total experience in education. The average high school size is 278 students with a range of 22 students to 3546 students.

Question Four

Question Four asked the respondents to indicate the degree to which they believed ACT/SAT scores can be used to indicate their school's educational effectiveness. The indicator uses 1 for weak and 5 for strong. The average response was 3.12, indicating the general perception is these scores, to a degree, reflect the school's educational effectiveness.

An examination of a contingency table for Question Four and administrative position shows that 81 percent of the superintendents believe ACT/SAT scores are either a good indicator (3) or lean toward strong indicator (4 or 5). Principals' interpretations are even stronger with 86 percent scoring Question Four with a 3 or higher. Counselors were less confident in the scores as only 33 percent of counselors believed ACT/SAT scores merited a 3 or higher. The remaining 19 percent of the superintendents in Montana believe ACT/SAT scores are weak indicators of their schools' educational effectiveness.

Question Five

Question Five asked the respondents to identify what they believed the public's perception of ACT/SAT scores are regarding the educational effectiveness of public education. The indicators in this question were the same as Question Four, i.e., 1 represented a weak indicator while 5 indicated

a strong indicator. The mean response was 3.93, notably higher than Question Four.

An examination of a contingency table for Question Five shows that 91 percent of the superintendents believe the public perceives ACT/SAT scores as good or better indicators of school effectiveness. Of that 91 percent, 38 percent believe the public views ACT/SAT scores as a very strong indicator (5). Principals and counselors indicate similar thoughts regarding the public perception of these scores.

Correlations Based Upon Questions One through Five

A Pearson Correlation Matrix was created with the data obtained from Questions One through Five. While there was a significant ($p<.05$) correlation between years of experience as an administrator and school size, there was no correlation between either school size or years of experience and administrative or public perception of ACT/SAT scores as a measure of educational effectiveness.

Question Six

This question asked the respondents to identify where their school system addresses the use of ACT/SAT scores. The census found the category most often identified was scholarship selection. Fifty-two percent of the respondents reported their school uses these scores for scholarship consideration. Forty-nine percent of the respondents

indicated their school systems use these scores for recognition of academic achievement. Approximately 34 percent of the schools make curriculum decisions based on ACT/SAT scores. Twenty-eight percent of the schools do not make any use of these scores. The remaining ranks and percentages may be found in the Appendix.

Questions 7, 8, 9, and 11

These questions provided information which is contained in Questions 10, 12-15. Numerous responses indicated these questions were not properly understood and therefore of less statistical value than the remaining questions. Therefore, the results of these questions are, in essence, summarized in the responses of the other questions.

Question 10

Question 10 asked the respondents to identify whom they would contact in the event of low ACT/SAT scores. The category receiving the highest rank is teachers. Seventy-three percent of the respondents reported they would bring low scores to the attention of the teachers involved. Following teachers, low ACT/SAT scores would be brought to the attention of: a) the counselor (66 percent), b) principals (59 percent), c) school board (58 percent), d) parents (42 percent), and e) department heads. Fifteen

percent of the respondents reported they would not report low scores to anyone at all.

Question 12

This question identified the specific areas within a school system believed by superintendents to be most reflective of ACT/SAT scores. This question and the next two are crucial in understanding the perception held by Montana superintendents regarding these scores. Each of the three questions are treated statistically in the same manner.

Specifically, the areas are ranked according to the mean of the responses. Differences in rank are checked for significance (α=.05) by using a Wilcoxon Rank Sum/Mann-Whitney U-Test. Next, consistency or agreement among respondents is checked for significance by using the Kendall Coefficient of Concordance test. The rankings are inspected according to the frequency of response by using both crosstabs and histograms. For clearer interpretation of the data, selected frequency histograms are presented to demonstrate any tendency of the responses to be skewed. A Spearman Rank correlation matrix was computed to test for significant correlations and those of interest are reported. Finally, regression equations (linear) were calculated for selected variables and are reported.

In Question 12, the seven areas identified were ranked by 86 percent of the respondents. Fourteen percent

responded with NA, indicating they do not believe ACT/SAT scores measure any part of the school system. The high percentage of superintendents ranking the variables indicates ACT/SAT scores are considered to be an indicator of school effectiveness. Further, virtually all of those superintendents who ranked variables used the full range of possible rankings from one to five, indicating they believe ACT/SAT scores measure more than one or two areas of school effectiveness.

The rank order of the seven variables identified in Question 12 is presented first when ranked by means and then ranked again after a Wilcoxon Rank Sum/Mann-Whitney U-Test follow up test for significance between mean ranks was conducted.

Table 1

Question 12 -- Rank Order -- Most Important - Least

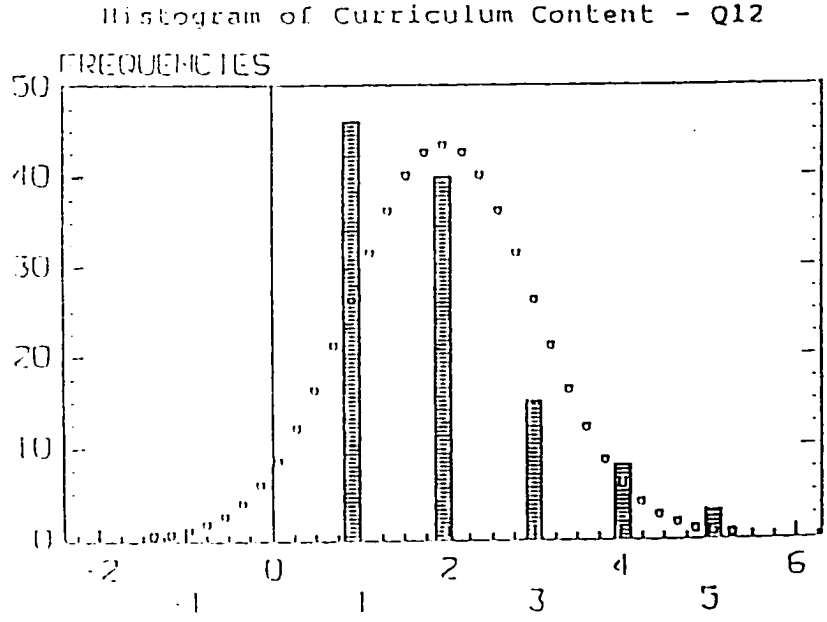
Response	Mean	Rank	α=.05	
Curriculum Content	1.949429	1	1	
Teacher Effectiveness	2.527778	2	2	
Curriculum Offerings	2.538461	3		
School Ethos	3.333333	4	3	
Counseling	3.716418	5		4
Administration	3.814286	6		
Board Policy	4.147059	7		

A review of Table 1 indicates superintendents believe curriculum and teacher effectiveness are reflected most by

ACT/SAT scores while administrative areas are measured least by these scores. The groupings in the last two columns show placing when the Wilcoxon Rank Sum/Mann-Whitney U Test for significance in ranking (α=.05) is applied. The Kendall Coefficient of Concordance test for agreement shows a high degree of agreement among raters regarding this ranking (W = .35).

When histograms were plotted against a normal distribution, the above information became readily apparent. A few relevant examples are presented. The histogram in Figure 1 plots the curriculum content responses for Question 12. Consideration of this histogram shows a strong skewing to the right which means the 86 percent of the superintendents who believe ACT/SAT scores measure school effectiveness also, believe curriculum content is highly reflected by ACT/SAT scores.

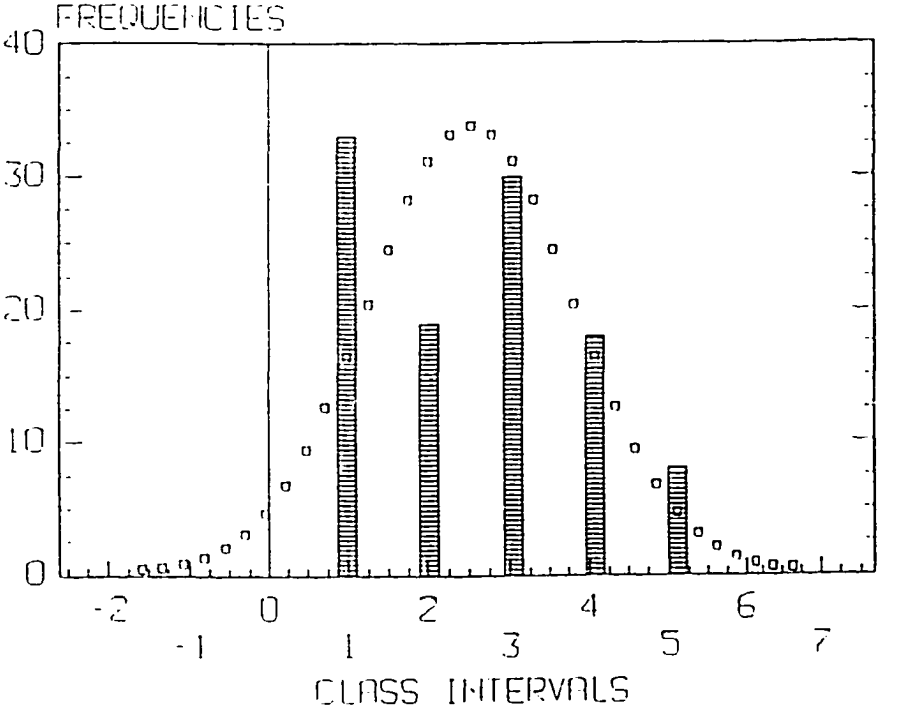
Figure 1



The next histogram is the response for teacher effectiveness. Teacher effectiveness ranked second to curriculum content and was significantly tied for second with curriculum offerings. The plot of the histogram (Figure 2) shows the importance placed upon teacher effectiveness by superintendents with the most common response being the strongest possible response. Clearly, a majority of the responses are one, two, or three which indicates that superintendents believe ACT/SAT scores strongly reflect the effectiveness of the teachers.

Figure 2

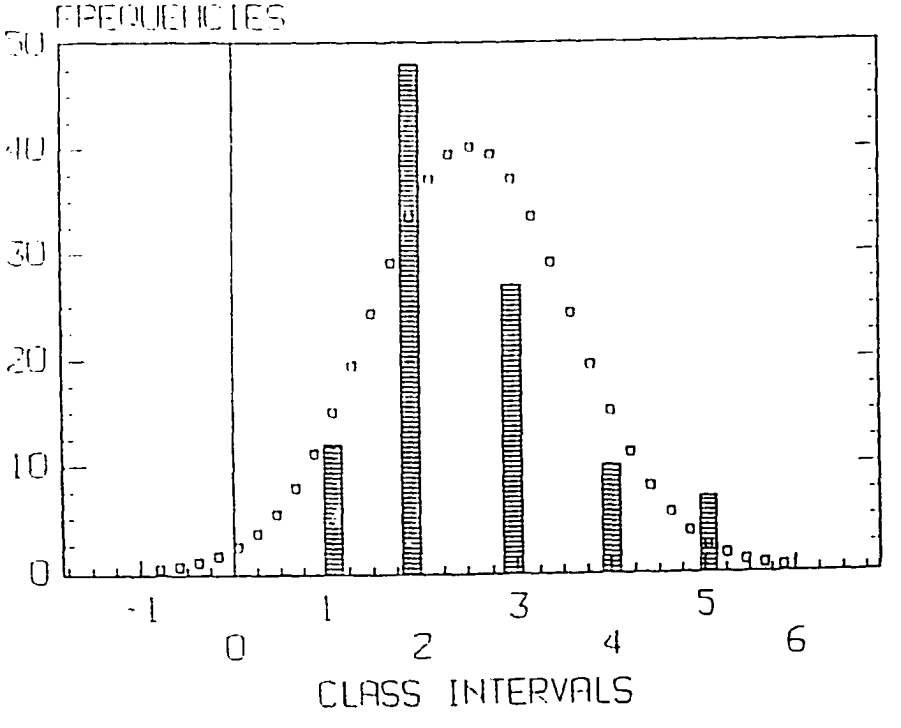
Histogram of Teacher Effectiveness - Q12



The third histogram plots the response for curriculum offerings. This response, while statistically the same as teacher effectiveness, has quite a different histogram. The graph is skewed to the right showing a high response for numbers two and three while very little response for one, four, or five. Unlike the teacher effectiveness response which takes second place as the result of averaging a high first and third choice response, curriculum offerings ties for second place statistically since number two is the primary choice.

Figure 3

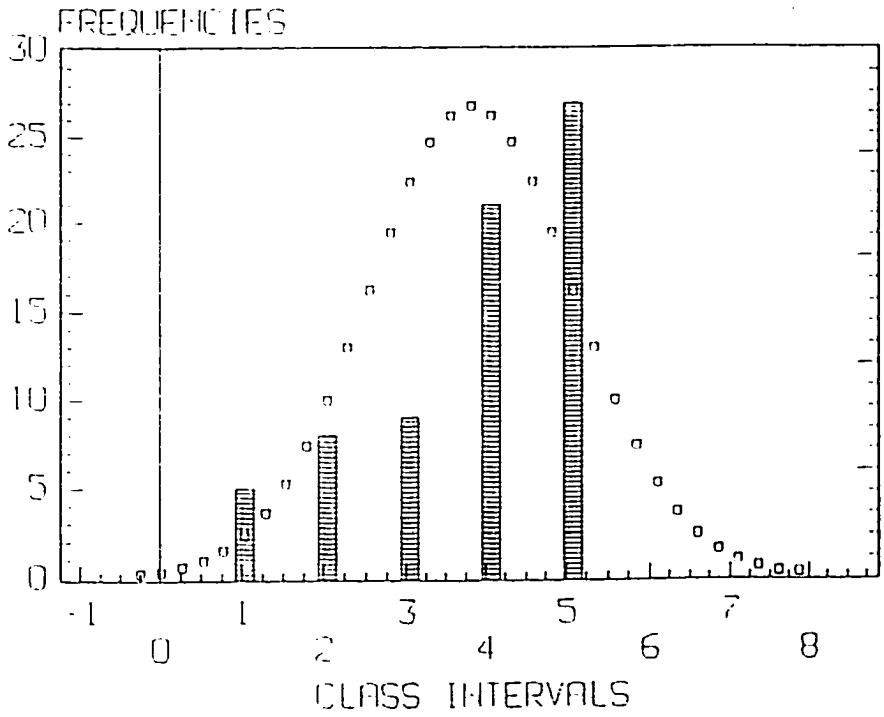
Histogram of Curriculum Offerings - Q12



Curriculum content, teacher effectiveness, and curriculum offerings present histograms skewed to the right. In sharp contrast are the histograms of the administration and counseling responses. The tendency of superintendents to believe that ACT/SAT scores do not reflect the effectiveness of administration is reflected clearly in Figure 4. Not only are the responses for administration skewed to the left, but they are continuously and progressively so skewed.

Figure 4

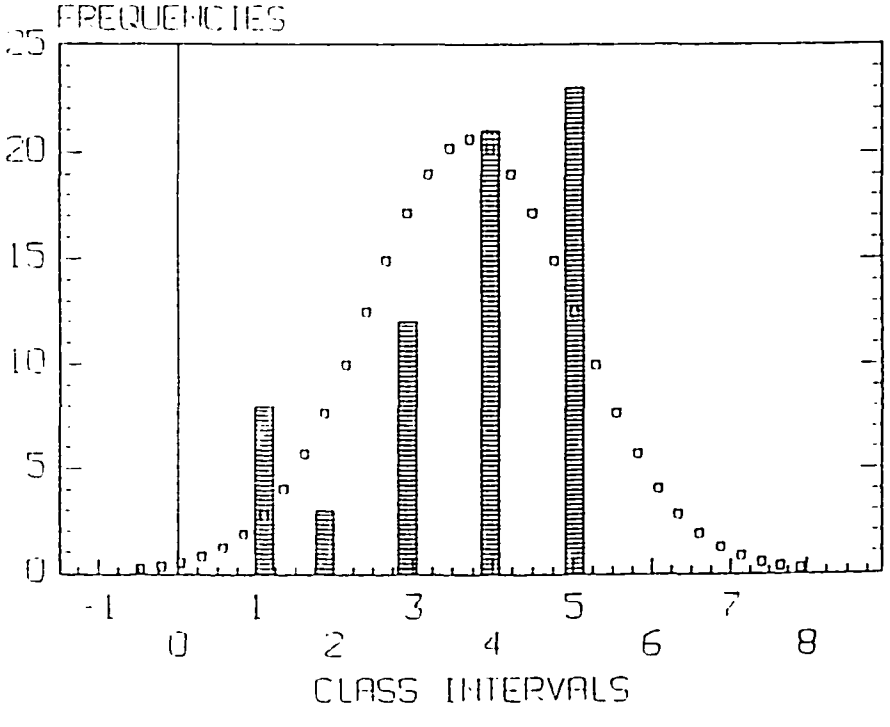
Histogram of Administration - Q12



The final histogram for Question 12 concerns the effect of ACT/SAT scores on counseling. Superintendents also find ACT/SAT scores do not reflect the effectiveness of school counseling. The histogram in Figure 5 is skewed to the left closely resembling the histogram for administration.

Figure 5

Histogram of Counseling - Q12



A correlation matrix offers the same information in slightly different terms. For example, there is a strong correlation (ρ=.67, p<.01) between those superintendents who believe curriculum content and course offerings are measured by ACT/SAT scores as well as those who believe curriculum content and teacher effectiveness (ρ=.64, p<.01) are both

measured by these scores. Somewhat unexpectantly, there were no correlations between school ethos and counseling, administrators, and board policy (ρ=.007, ρ=-.056, and ρ=.015 respectively). The only variable that teacher effectiveness was not significantly correlated to was board policy, indicating that superintendents do not see a correlation between policies of the board and teacher effectiveness when using ACT/SAT scores as criteria for making such judgments.

In order to see how superintendents viewed teacher effectiveness in relationship to the other responses, a multiple regression was calculated in which teacher effectiveness would be predicted by using the other six responses as the independent variables. The significant beta coefficients are board policy (β=-.73), counseling (β=-.44), curriculum offerings (β=-.29), and school ethos (β=.38). Administration and curriculum content did not significantly contribute to the regression equation.

To see how administration could be predicted from the other variables, a multiple regression using administration as the dependent variable and the other six responses as independent variables was calculated. The calculations found only two significant variables: counseling (β=.71) and curriculum offerings (β=.-54).

Question 13

Question 13 asked superintendents to indicate a hierarchy of administrative action to correct low ACT/SAT scores. Below is a table of those results along with the ranking resulting from the follow-up tests for significance.

Table 2

Question 13 -- Rank Order -- Most Important - Least

Response	Mean	Rank	α=.05	
Curriculum Content	2.110092	1	1	
Teacher Effectiveness	2.737864	2	2	
Curriculum Offerings	2.787879	3		
Administration	3.040541	4		3
Board Policy	3.515152	5	4	
Counseling	3.785714	6		
School Ethos	3.800000	7		

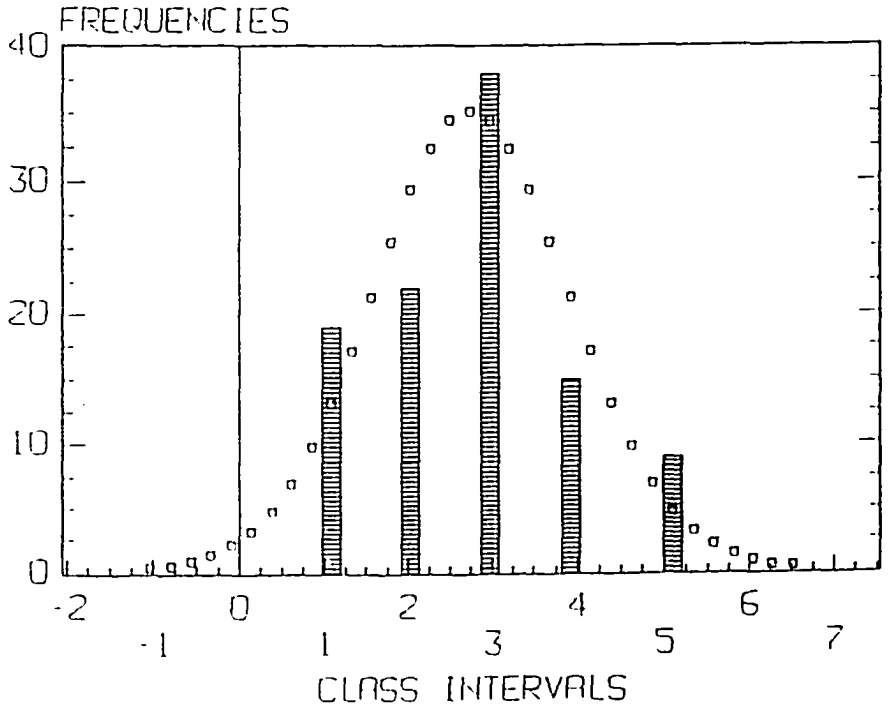
Question 13, like Question 12, identified four significant groups. Curriculum content is significantly ranked first in each question and the second group is nearly identical. This indicates that superintendents will take action first in the areas they perceive to be measured most by the ACT/SAT scores. The Kendall Coefficient of Concordance Test for consistency among raters indicates a strong consistency among responses with a concordance coefficient of $\underline{W}$ =.57.

Some representative histograms provide a visual interpretation of the frequency of the responses. The

histogram of teacher effectiveness, for example, is very nearly a normal distribution, though somewhat skewed to the left. This would indicate the majority of superintendents believe low scores should be acted upon by addressing teacher effectiveness early in the process.

Figure 6

Histogram of Teacher Effectiveness - Q13

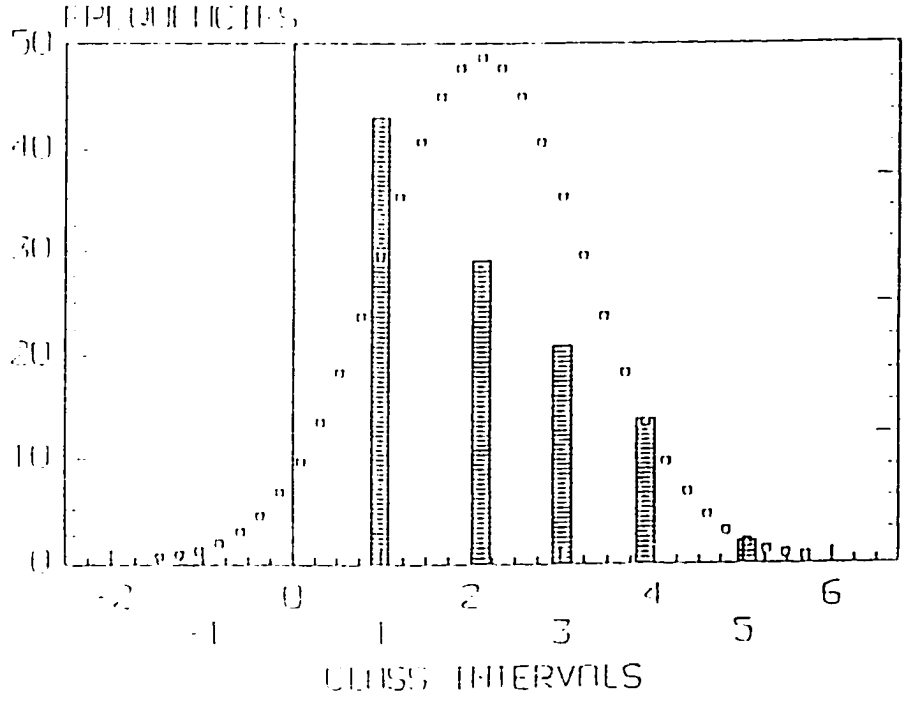


A consideration of the histogram in Figure 7 shows the frequency of the curriculum content response. Clearly, this response is skewed progressively to the left and indicates a very decisive belief of superintendents regarding their action in dealing with low ACT/SAT scores. Given low

ACT/SAT scores, curriculum content would be the first priority of a large number of superintendents.

Figure 7

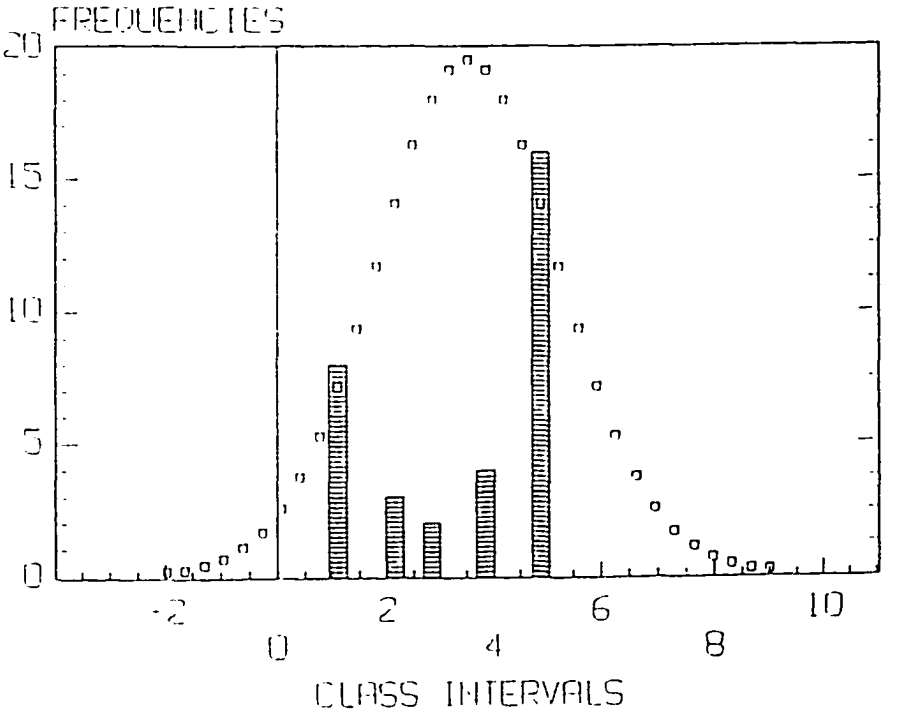
Histogram of Curriculum Content - Q13



When a school is experiencing low ACT/SAT scores, dealing with board policy was not a high priority for most superintendents. The largest response was the lowest ranking indicating that superintendents viewed board policy as having little to do with curriculum content, teacher effectiveness, or curriculum offerings. A look at Figure 8 shows the frequency of responses for board policy.

Figure 8

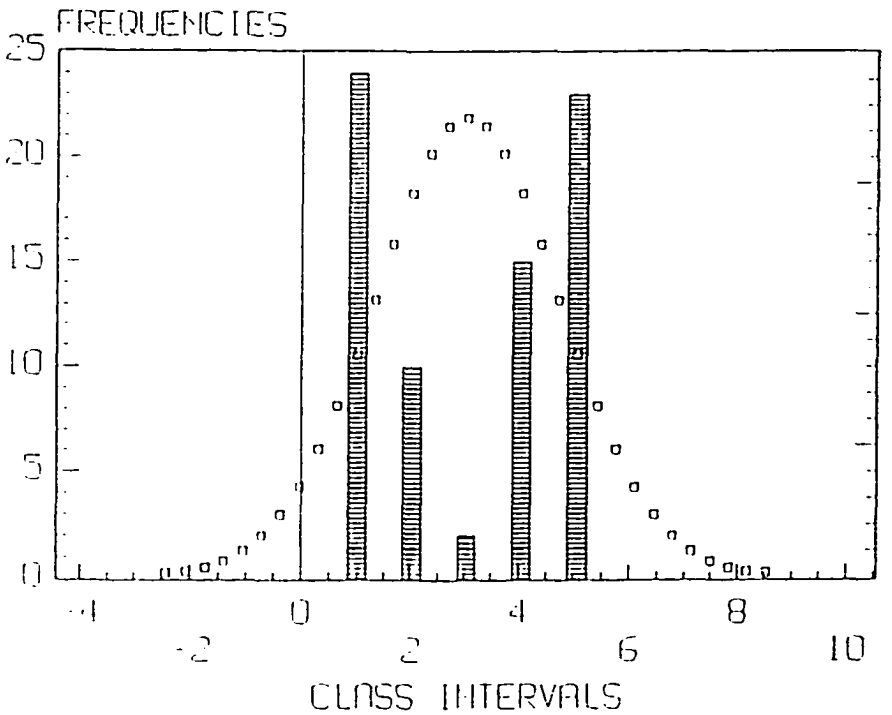
Histogram of Board Policy - Q13



The response for administration is better understood in histogram form. Administration ranked fourth or about the middle response in this question, but a consideration of the histogram clearly shows the responses were inversely related to a normal curve. There were basically two views: a) the view that it was very important to address administration early in the hierarchy of action (given low ACT/SAT scores) and b) the view that it is not important to address administration. There were virtually no middle responses from superintendents, even though the mean ranking gave it a middle position.

Figure 9

Histogram of Administration - Q13



Correlations among responses for Question 13 have only a single significant (α=.05) positive correlation between curriculum content and curriculum offerings (ρ=.47). The remaining significant correlations are negative and include correlations between administration and curriculum offerings (ρ=-.61) and administration and curriculum content (ρ=-.54). The final and strongest significant correlation is between board policy and curriculum content (ρ=-.64).

Individual regression equations provide a final statistical view of this information. The inverse relationship held by Montana superintendents between administration and teachers' effectiveness is easily

observed by the two regression equations relating teachers to board policy and administrators. Namely, TE = -.27(BP) + 3.9 and TE = -.23(AD) + 3.34. The inverse relationship between administration and curriculum is even stronger as evidenced in AD = -.89(CO) + 5.82 and AD = -.81(CC) + 5.01 where CO represents curriculum offerings and CC represents curriculum content. Finally, board policy was inversely related to curriculum content with the regression equation yielding BP = -.79(CC) + 5.55.

Question 14

Question 14 asks superintendents' to give their perception of what should constitute optimal entrance requirements for higher education. The means, rankings, and significance of rankings are indicated in Table 3.

Table 3

Question 14 -- Rank Order -- Most Important - Least

Response	Mean	Rank	α=.05
Student Achievement	2.451613	1	1
Ability or Aptitude	2.464000	2	
H.S. Courses Taken	2.845528	3	2
Grade Point Average	3.048000	4	
No Restrictions	4.031915	5	3

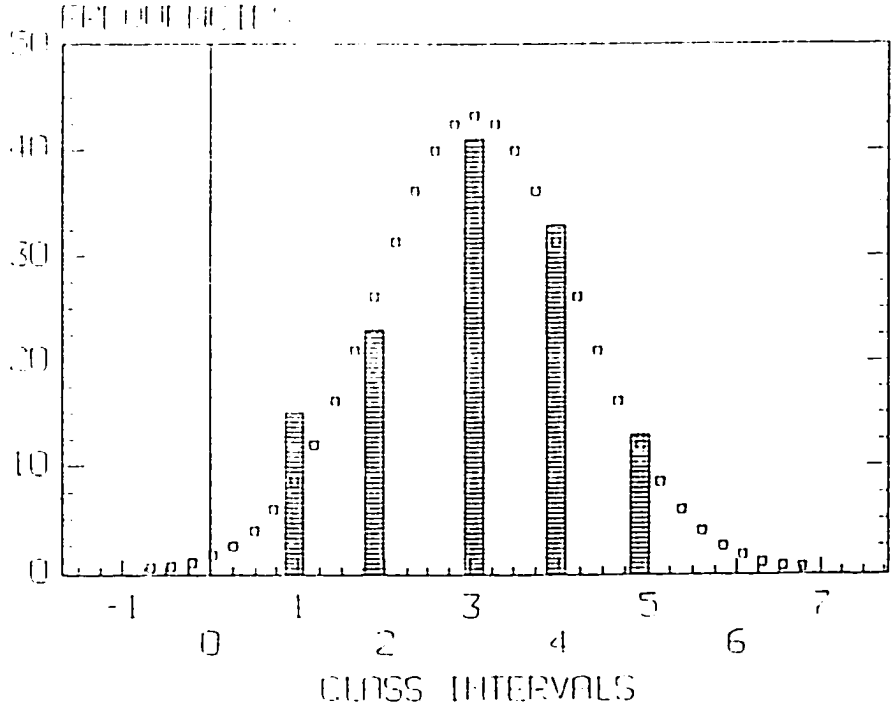
The follow-up tests for significance between rankings recorded in Table 3 indicate that while Montana superintendents consider achievement to be their highest

criterion for admissions to higher education, they do not significantly distinguish achievement from aptitude. High school GPA ranked only above the category of no restrictions indicating superintendents do not place a great deal of emphasis upon GPA as a criterion for college entrance. The test for consistency of responses was very high on this question ($\underline{W}$ =.68), indicating a very high agreement among raters.

Histograms provide more insight into how these requirements were ranked. The histogram of GPA is very nearly a normal distribution with about equal numbers balancing the importance against the insignificance of GPA. The frequencies of the GPA response are recorded in Figure 10.

Figure 10

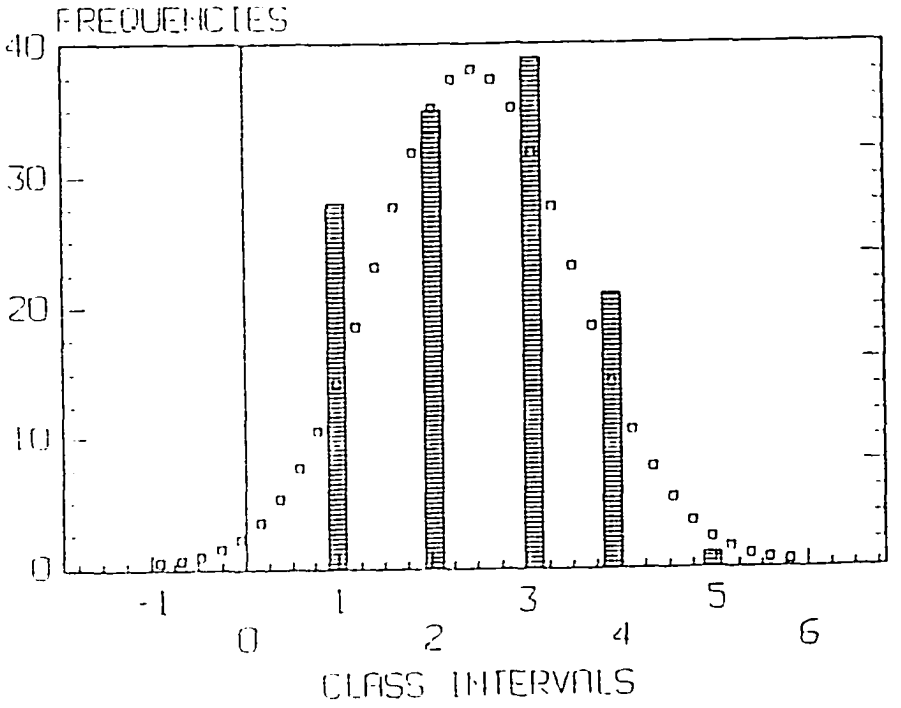
Histogram of GPA - Q14



While achievement ranked first and aptitude second when ranked by means, a reflection upon the histograms shows they had much different response patterns. Achievement responses were very strong over the first three placements. Superintendents believe achievement is an important criterion for college admissions.

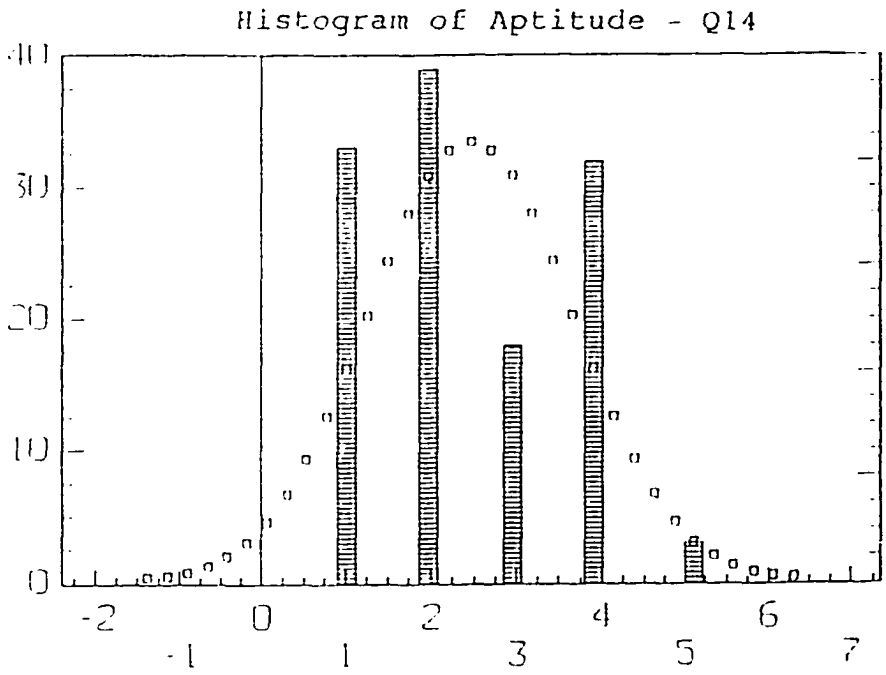
Figure 11

Histogram of Achievement - Q14



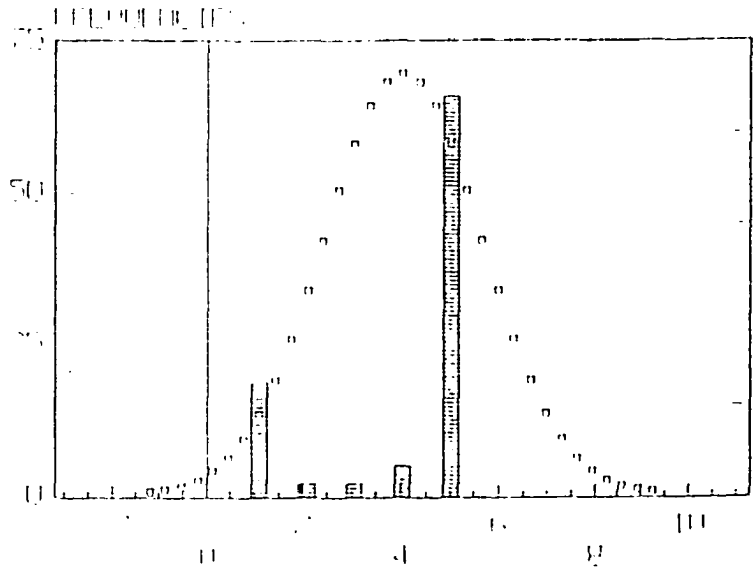
Aptitude responses were also strong in the first two placements, but declined sharply in the middle position. The fourth placement was also much stronger for aptitude than for the achievement response. These responses are plotted in Figure 12.

Figure 12



Also of interest is the histogram of the no restriction response. It was either strongly favored, or strongly not favored. Few superintendents expressed a response between the highest and lowest choices.

Figure 13



The two strongest correlations in this question were both correlated with the no restriction response. No restrictions correlated negatively and significantly with both academic preparation (ρ=-.36) and GPA (ρ=-.35). This same information is further amplified when converted to a multiple regression equation which provides the following formula: No Restrictions = -.60(GPA) - .52(HS Prep) + 7.37.

Question 15

Question 15 asked the respondent to list more than one criterion for admission to higher education if it were thought that more than one criterion should be used for admissions selection. This question was answered by 57 percent of those who returned the survey. Ranking the responses by means produces the following rank order:

Table 4

Question 15 -- Rank Order -- Most Important → Least

Response	Mean	Rank
Student Achievement	1.985075	1
Ability or Aptitude	2.175439	2
H.S. Courses Taken	2.350877	3
Grade Point Average	2.637931	4
No Restrictions	3.000000	5

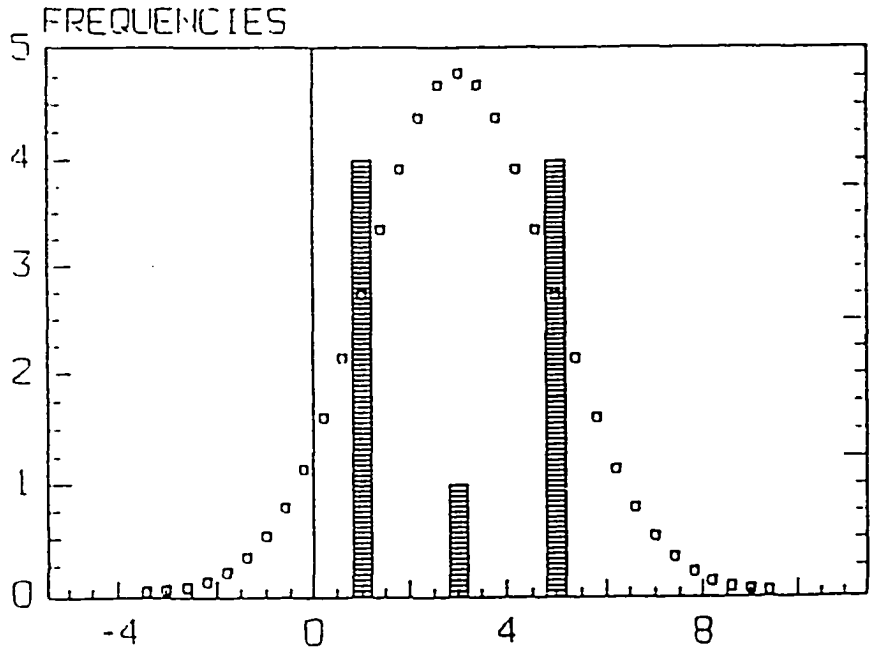
Table 4 has exactly the same rank order as Table 3 when ranked by means, though the means are somewhat different. This provides internal validity to the census since

requesting basically the same information using two different formats has produced identical mean rankings.

One final histogram of interest is presented from Question 15. This is a histogram of the no restrictions response.

Figure 14

Histogram of No Restrictions - Q15



Investigation of the response frequencies plotted in Table 14 show the divergence among superintendents on this question. The same number of superintendents strongly favor no restrictions to higher education admission as those superintendents who are not in favor of such a policy. Very few superintendents expressed an opinion on this response which did not lie on the extremes.

CHAPTER 5

IMPLICATIONS OF THE FINDINGS

From the Census

The immediate implication drawn from the census was the importance of the study to Montana public school superintendents. A response rate of 83 percent and a follow-up summary requested by 52 percent of the respondents indicated a serious interest in the issue under research. Just as important, 86 percent of the respondents were superintendents rather than a delegate of the superintendent. The broad range of interest in this topic was further evidenced by strong participation of superintendents from all school sizes. Only one AA school superintendent failed to return a survey while the other schools of varying sizes were equally well represented.

Questions Four and Five established that school administrators are using ACT/SAT scores to measure school effectiveness, though at a level less than the public uses the scores for the same purpose. Regardless of the value of these scores, this variance in perception creates a problem for public school administration. A superintendent having students with high scores will be viewed favorably by the public, but such a superintendent -on the average- realizes these scores are not as important as the public believes. On the other hand, a superintendent with low school ACT/SAT scores has a public relations problem to solve simply

because the superintendent believes the public views the scores as more meaningful than does the superintendent.

These results explain why the review of literature found the media and politicians believe ACT/SAT scores measure educational quality: a high percentage (81) of superintendents believe ACT/SAT scores do measure educational effectiveness. If media and political personnel first contact school superintendents regarding the value of ACT/SAT scores, such inquires will find in excess of 80 percent of the Montana public school superintendents will respond that these scores are an important measure of school effectiveness. Certainly the media reporting the meaning of ACT/SAT scores and the politicians using ACT/SAT scores as a measure of educational effectiveness for legislative or funding purposes or both cannot be faulted for using the opinions of school leaders. Educators advocating alternative assessment procedures will have to take into consideration and work with the fact that, in addition to the media and politicians, Montana public school superintendents believe ACT/SAT scores are well correlated to educational effectiveness.

The primary conclusion to be drawn from Question Six is the fact that 72 percent of Montana high schools formally use ACT/SAT scores in their systems. Use of these scores ranges from scholarship selections to staff evaluations. Given the fact that school leadership generally accepts

ACT/SAT scores as measures of academic achievement, it is less than surprising to find that school districts use the scores upon which to base decisions or scholarship selection, academic honors, and curriculum.

The analysis of Question 10 yields a continuation of the above logic. The previous responses given on the census have established that superintendents understand the ACT/SAT scores to be important measures of their school system and presently make use of these scores to measure the academic achievement of their students. It follows then, that superintendents identify teachers in Question 10 as the most important contact they make or would make regarding the reporting of low scores.

Question 12 identifies the specific areas superintendents believe are most reflected by ACT/SAT scores. Curriculum content ranked first. Teacher effectiveness and curriculum offerings were ranked second while administrative choices ranked last. Consideration of the histograms presented (Figures 1-5) show quite clearly that curriculum and teaching are skewed to the right or the most measured by ACT/SAT scores portion of the scale while administrative responses are skewed to the left or the least measured by ACT/SAT scores portion of the scale.

The correlation of variables in this question provides some helpful information for reaching conclusions about the particular views of Montana superintendents. In addition to

finding significant correlations between curriculum content and both course offerings and teacher effectiveness, there were no significant correlations between school ethos and administration, counseling, or board policy. While it is not the purpose of this research to study these latter correlations (or lack thereof), it was unexpected that results might indicate school ethos was not correlated to administration and counseling.

The census has indicated that superintendents believe ACT/SAT scores have a relationship to teachers' effectiveness. By calculating a multiple regression equation using teaching effectiveness as the dependent variable and the other six responses as independent variables, it is possible to determine more about how this relationship is viewed by superintendents. The calculation determined negative beta values for board policy and curriculum offerings and a positive beta value for school ethos with all values significant at the $\alpha=.05$ level. This finding results in the conclusion that superintendents who see board policy and curriculum offerings as not measured by ACT/SAT scores but school ethos so measured will tend to believe teachers are measured by ACT/SAT scores.

A second multiple regression in which administration is predicted using the other six responses as independent variables provided further insight. In this calculation, only two beta values were found to be significant:

counseling (positive) and curriculum offerings (negative). Hence, superintendents view ACT/SAT scores as an important measure of their educational leadership if those superintendents believe these scores also measure counseling but not curriculum offerings.

The conclusion formed from the responses throughout the census to this point was that ACT/SAT scores are important measures of public education and, in particular, measured the academic program of a school system. This academic program included the curriculum content, curriculum offerings, and the teachers. Question 13 took this paradigm a step further.

Question 13 sought to extract from superintendents what action they believed necessary in the event of low scores. Only 18 percent believed no action was necessary. The remaining 82 percent believed some action was necessary. These percentages were consistent with the previous responses on the census. For whatever reasons, approximately 20 percent of the superintendents have found the ACT/SAT scores to be of little or no value while approximately 80 percent have given the scores varying degrees of credibility.

Once again, curriculum and teacher effectiveness scored as the key areas to act upon in order to improve ACT/SAT scores. The implications from this were obvious: Curriculum and teachers are under pressure in the event it

becomes necessary to address raising low scores. Correlations and regression formulas found in Question 13 also emphasized this position. For example, to predict the importance of acting upon board policy to address low ACT/SAT scores, it was necessary to diminish the importance of the role played by curriculum content (BP = -.79(CC) + 5.55).

Question 14 was a look at what criteria superintendents would like to establish as the basis of determining whether their students are accepted into higher education. The two strongest correlations in this question were both correlated with the no restriction response. This response, i.e., no restrictions, correlated negatively and significantly with both academic preparation (ρ=-.36) and GPA (ρ=-.35). This would indicate that superintendents who believed there should be no restrictions placed upon admissions to higher education strongly view academic preparation and grade point average as two criteria least suited for measures of college admission.

In general, Question 15 found that most superintendents favor student achievement as the first criterion for admission to higher education. However, aptitude did not significantly differ in rank from achievement indicating that another census might very well find first and second place transposed. This left this question in a curious state. There was a great deal of philosophical difference

between an admissions policy that uses aptitude and an admission policy that uses achievement. The fact that superintendents did not favor one over the other indicated a need for more definitive information regarding assessment.

From the Literature

The literature regarding the nature and use of ACT/SAT scores, as evidenced in Chapter 2, was abundant in certain types of research and philosophical treatises. This portion of Chapter 5 is a synthesis of the critical elements of the literature as they apply to education and educational leadership. The result was a synopsis applicable to the formulation of general principles calculated to guide administrative usage of ACT/SAT scores.

There were several key concerns which should be briefly summarized. The first consideration was the nature of ACT/SAT examinations. The second consideration was their intended purpose; and the final consideration was the validity of the tests.

Nature

Both the ACT and SAT tests are, according to their authors, aptitude tests. Both tests are intended to measure the ability of a student to succeed in higher education. The entire spectrum of ACT/SAT researchers from ETS to Nairn

agree that neither test is designed nor intended to measure a school system, curriculum, or teacher effectiveness.

During the course of this research, ETS has announced that it is changing the name of the SAT and, hence, its nature. The SAT will soon be an acronym for Scholastic Assessment Test. In reaction to the critical research cited in Chapter 2, ETS is attempting to move away from being identified with aptitude and more toward achievement. Exactly what ETS means by aptitude is not clear, but ETS does distinguish aptitude from achievement in its proposal to initiate measuring achievement in 1994 (College Board, 1991).

Achievement logically and experimentally is positively correlated to academic preparation (Sebring, 1984). ETS has held that the SAT scores varied with preparation in the same manner as do measures of achievement. This raised the question: If SAT scores are a measure of aptitude, and if aptitude correlates well with achievement, then why offer new tests written to measure achievement?

The conclusion is the literature strongly argues that ACT/SAT scores have a limited relevancy as a measure of innate aptitude. ACT/SAT scores have been found to fluctuate as a result of factors not associated with aptitude, such as coaching. But since ETS has maintained over the years that the Scholastic Aptitude Test is a measure of aptitude, regardless of whether it is, it becomes

necessary to change the name of the test in order to indicate a change in nature. So now ETS has begun to break bonds with the past and indicate the future of the SAT is ordered toward achievement and not aptitude. The makers of the ACT are making a similar move.

Critical, essential, and fundamental to this research is the understanding of educators that ACT/SAT scores will begin to measure achievement, not aptitude, in 1994. This understanding is important because educators will be asked over and over again to interpret the results of the ACT/SAT scores. Comparison of the new scores to the old will be irresistible for many educators, politicians, and the media, even though the scores are not comparable since the criterion variable has changed from aptitude to achievement.

Whether intentional or not, ETS has provided for this erroneous comparison by maintaining the same acronym, i.e., SAT, so that SAT scores will appear to be SAT scores regardless of the year collected. Changing the name to Scholastic Assessment Test in order to preserve the SAT acronym is suspect since the use of the words "assessment test" is pleonastic and immediately suggests the question of how a test so poorly named could possibly be used to measure how well the test-taker has mastered word usage. The ACT test, while making some substantial philosophical changes itself, has also maintained the same name and acronym.

Given the change in name and nature, a knowledge of the past history of both the ACT and SAT tests is important for evaluative purposes. Both tests will be starting from the ground up to establish credibility, reliability, and validity. Predictably, most educators, politicians, or media observers will not make this distinction and, hence, fail to treat the test scores with the necessary academic rigor properly due a new measure of such importance. The 1994 scores will be reported and treated in continuum with previous scores even though the new scores measure a different variable.

Such interpretation creates a double demand upon school superintendents who find themselves dealing internally and externally with ACT/SAT test scores. Superintendents will have to deal with the present meaning of the scores, which will be very difficult since it will take approximately 10 years to find out what, if anything, the new tests measure. In addition, superintendents will have to deal with what they and the public thought the old tests measured and most likely, have to treat the new scores as if they were identical in nature to the old scores. This will be a nearly insurmountable job, particularly given the confused meaning of the scores now, prior to the change.

Intended purpose

For whatever philosophical reason, ACT/SAT examinations have been produced in order to control admissions to higher education. This has been a primary function of the tests although numerous other purposes have evolved. For example, making scholarship selections, ranking of state educational effectiveness, determining K-12 curriculum, determining national report card markings, conducting teacher evaluations, and ranking of institutions of higher education are only a few examples from the literature. Intentional or not, these tests have assumed roles which are very crucial to every aspect of education at all levels.

Validity

The primary validity of these examinations resides in their ability to predict how well a potential applicant will do in college. A multitude of academic variables have been tested for correlation only to find there are no substantial variables of which ACT/SAT scores can be said to be predictive. The strongest correlation between ACT/SAT scores is with FGPA. Even this correlation disappears in much of the research.

What the research does say about ACT/SAT scores is that they are a better measure of gender, race, income level, and a host of other variables not thought to be properly associated with college admissions selection rather than

those variables correctly associated with higher education such as college GPA or graduation rate. The numerous studies cited in Chapter 2 indicate these tests should not be used as the sole means of denying or granting admission to higher education. Hence, the validity of ACT/SAT scores for determining admissions is not academically or ethically established.

Recommendations For Further Research

Graduate level work should include activities which direct graduate students to expand the knowledge base of current assessment practices. Presently, there is no research correlating the content of ACT/SAT examinations with the respective expectations of freshman-level course work. Given that ACT/SAT examinations are moving toward measuring achievement, there should be continuous research established which seeks to correlate the questions asked of applicants to the correct relevance and proper proportions of the content of freshman-level course work.

There is no research presently available which analyzes the content of ACT/SAT questions from year to year. For example, are questions in mathematics predominantly taken from algebra one year and geometry the next? What is the consistency of level and nature of questions asked from year to year?

Further research of alternative assessment techniques and practices which seek to determine methods of higher education admissions more consistent with a wide range of human variables is recommended. To ignore this research is to perpetuate a system of admissions which discriminates on the basis of gender, culture, income, and other factors previously identified.

As long as admission policies restrict enrollment, there should be research performed which justifies selection. Presently, there is no research which indicates present practices select the most appropriate applicants. At best, some admission policies select applicants who can graduate, but the research is silent on establishing criteria which determine who should be selected.

Further research at the K-12 level is equally important. School districts should be involved in some aspect of researching their forms of assessment. For standardized assessment, this research could be accomplished in parts by each district and the whole of the research coordinated by a nearby university. K-12 schools should also be performing item analysis of ACT/SAT and other tests to determine if these tests are properly measuring what is being valued at the K-12 level as the curriculum taught.

Internally, research on the K-12 level should also examine various methods of alternative assessment that have been implemented. Simple action research designs should be

utilized and assessment techniques refined as appropriate. This research would be conducted at all levels from classroom to district in scope.

Summary

In summary of the findings of the census, Montana public school superintendents differ very little from non educators in their view of ACT/SAT scores and the scores' implications for education. Since 86 percent of the superintendents believe ACT/SAT scores measure educational effectiveness, it is not surprising to find people less involved in education, e.g., politicians, general public, and the media, view these scores basically in the same manner as local educational leaders.

Prior to this research, the external role played by ACT/SAT scores in Montana education was well known. This research has found that there is a very real ethos among superintendents similar to that in the general public that not only will changing curriculum and teacher effectiveness improve low ACT/SAT scores, but that, for the most part, changes should be effected when low scores exist. Now it can be stated that ACT/SAT scores play just as significant a role internally as they do externally in Montana education. Hence, ACT/SAT scores have a vital impact upon Montana education.

This creates a problem of a worrisome magnitude: If there does not exist any focusing research available to superintendents either individually or collectively regarding the implications of ACT/SAT scores, how can those not involved in education expect to assume leadership in education by means of their personal interpretation of standardized examinations? This is precisely what is happening in public education. Leadership of education is being assumed more and more by non educators who have become judges of educational effectiveness based upon measurements that are easy to find, apply, score, and rank. Excellence in education is being defined from a narrow perspective, i.e., improving minimum standards. Improving education becomes synonymous with improving these scores.

The district level administrator has the responsibility to assure students, teachers, and constituents that assessment is taking place which contributes to the totality of the educational process rather than measures *a posteriori* or after the fact. *A posteriori* assessment assumes the artificial role of academic policeman, waiting to catch error after it happens. This type of assessment can do nothing to improve education before the fact. Student, teacher, and school district stand in helpless judgment of these scores.

These considerations bring to light a potential misuse of this research. While it is very useful for

superintendents to be aware of the research regarding ACT/SAT scores when those scores are being improperly applied internally or externally to their district, the same information becomes misused when it exists for the sake of scorning testing. This research is properly utilized when it serves as the basis upon which to strengthen assessment by better understanding the nature and broadening the scope of assessment. This will help create a new and more profound motive for investigating alternative assessment.

Conclusion

The correlations and regression equations offer some complex and interesting reflections and bring this research to a fundamental question: If ACT/SAT scores measure the educational effectiveness of a school system as believed by both superintendents (Question 4) and the general public (Question 5) and the scores measure the academic strength of the system's students and determine the financial assistance for future higher education (Question 6), why then do superintendents view leadership as the least measured by ACT/SAT scores (Question 12) and curriculum and teachers as the first acted upon (Question 13) to improve low scores?

Consequently, this research can now argue to a primary and seminal conclusion. The fact that the majority of superintendents believe ACT/SAT scores measure the educational effectiveness of the whole school system and the

academic worth of the system's students and then correlate improving low scores to improving curriculum and teachers (rather than leadership) provides the premise for a powerful conclusion: Superintendents' *raison d'etre* should focus on curriculum and staff development. Interestingly, this conclusion is independent of whether ACT/SAT scores do, in fact, measure the educational effectiveness of public school systems and the academic worth of the students. Just the fact that superintendents identified curriculum and teachers as the source of improving the whole system and the net academic worth of the systems' students provides the necessary logic for the argument.

This research can now set out some primary principles regarding the district wide administrative use of ACT/SAT scores:

1) The scores in question must be identified by criterion variable, i.e., aptitude (pre-1994) or achievement (1994-).
2) Scores on the pre-1994 tests have no consistent correlation to any academic measures. Scores on the 1994- tests have yet to establish validity.
3) Scores that work to the advantage of the test-taker should be acknowledged and the benefits maximized.
4) Scores that work to the detriment of the student should be minimized and alternative forms of assessment utilized.

This research was able to conclude from responses on the census that curriculum and professional development are critical to the superintendency. The general conclusion from the literature is that ACT/SAT tests are not only here to stay, but will gain in at least external importance. Further, neither of these tests are sufficient for the purposes for which they are being used or are proposed for use, e.g. a yard stick to measure the national goals of America 2000 or integrated in some way into a voucher system. These new scores will inevitably be compared to past scores creating a need for educators to be adequately knowledgeable regarding both the new and old scores.

How should these conclusions be synthesized and addressed by a superintendent? Few, if any, areas in education are more capable of simultaneously addressing both curriculum and staff development than is assessment. Sound assessment immanently serves the student, the teacher, and the curriculum. Further, assessment reflects the nature and use of ACT/SAT scores and is the most logical point of focus in which to synthesize these varied areas of education.

Hence, the most specific conclusion that can be made consistently with both the literature and census of this research is for superintendents to focus their professional expertise upon assessment. Assessment, when treated *a priori*, will serve as the vehicle by which superintendents can proactively direct the mission of their districts.

This last recommendation is particularly important since assessment is not so much a matter of taking care of curricular "loose ends" as it is a key to empowering a school district to determine its own proper end. Only through *a priori* assessment work can a school district avoid being reactive to the forces pressuring education.

Accountability will always be an issue and a demand. Superintendents waiting for instruments of accountability will, at best, only be able to react to the demands of a given instrument. Curriculum and "good" teachers will arbitrarily be determined by the means used to assess both. Those superintendents who are creating and implementing sound assessment practices which enhance curriculum rather than determine it will define the course of education into the next millennium.

Appendix A

Cover Letter and Census

Abbreviations

ACT -- American College Test

CEEB -- College Entrance Examination Board

CLEP -- College Level Examination Program

COMP -- College Outcomes Measures Project

CPI -- Cooperative Preschool Inventory

CSS -- College Scholarship Service

ETS -- Educational Testing Service

FGPA -- Freshman (college) grade point average

GAPSFAS -- Graduate and Professional School Finance Aid Service.

GED -- General Education Development Test

GMAT -- Graduate Management Test

GPA -- Grade point average.

GRE -- Graduate Record Exam

JSAT -- Junior Scholastic Aptitude Test

HSGPA -- High school grade point average

HSR -- High School Rank

LSAT -- Law School Admissions Test

MBE -- Multistate Bar Examination

NASDC -- New American Schools Development Corp.

NTE -- National Teacher Examination

PSAT -- Preliminary Scholastic Aptitude Test

PSAT/NMSQT -- Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test

SAT -- Scholastic Aptitude Test

SAT-M -- Mathematics portion of the Scholastic Aptitude Test

SAT-V -- Verbal portion of the Scholastic Aptitude Test

SSAT -- Secondary School Admission Test

Descriptive Summary Table of ACT/SAT Interview Questions

Appendix Table 1

Number of Census returned: (number, %)		138	83%
2	Number of years in administration: (average)	12.5	
2	Number of years in education: (average)	19.5	
3	Number of students in 9-12: (average)	278.0	
4	ACT/SAT as indicators of effectiveness (supt): (average)	3.12	
5	ACT/SAT as indicators of effectiveness (public): (average)	3.93	
6	ACT/SAT scores presently addressed in: (tally/%)		
	Phil 17, Mis 7, Obj 30, SE 5, CD 47, SC 72, AR 68, BP 8, O 1, NA 39		
	% 12, 5, 22, 4, 34, 52, 49, 6, 1 28		
10	Low scores brought to: (tally) SB 80, Prn 81, C 91, DH 45, T101, P 58		
	Low scores brought to: (%) 58, 59, 66, 33, 73 42		
12	Area of school responsible for ACT/SAT scores: (tally / rank)		
	BP 34, Ad 70, C 67, TE108, CC112, CO104, SE 51, NA 20		
	7, 6, 5, 2, 1, 3, 4,		
13	Area of school needing immediate attention (low scores): (rank)		
	BP 5, Ad 4, C 6, TE 2, CC 1, CO 3, SE 7,		
14	Best criteria for admission to college: (rank)		
	None 5, Ability 2, GPA 4, Prep 3, Achievement 1		

Merle J. Farrier
Superintendent
Hot Springs High School
Drawer T
Hot Springs, Montana 59845
May 11, 1993

Ms. Abc
Superintendent
DEF Public Schools
Box GHI
JKL, MT ZIP

Dear Superintendent Abc,

As a classroom teacher, I was never sure what to do with the ACT and SAT scores that my students received related to the courses I taught. Did high scores mean I was doing a good job and low scores indicate a poor job? As I completed the course work for a Montana superintendent's certification, I still didn't know the answer.

There does not appear to be any guidelines for administrators to follow, nor does there appear to be a single study anywhere in the United States in which school leaders have been surveyed to find out how school districts presently treat the ACT/SAT scores their students receive.

This survey is an attempt to initiate some research into this question. Since there is no other research like this, it will be very important to have a high return rate.

Our nation is in the middle of educational restructuring, e.g., The New American Schools. Restructuring is typically driven by **non-educators** and scores from assessment devices such as ACT and SAT play a very important political role in forming the new mold. I am well aware of how little time superintendents have for filling out surveys, but I believe this research will help educational leaders exert **their** influence over educational changes.

Each survey will have a code which will be used only for keeping track of future mailings. **All surveys will be kept absolutely confidential and all results will be used without use of any school identity.**

If you would like to have the results of this survey and research sent to you when it is finished, please include the enclosed mailing label when you return your survey. Thank you for your time and help with compiling this information.

Sincerely,

Merle J. Farrier

ACT/SAT Administrative Usage Survey

1) Please circle your title

a) Superintendent b) Principal c) Counselor d) Other _____

2) Total years experience (include present year) in:

a) Administration _____
b) Education _____

3) a) Total number of students in your high school: _____
b) Please circle your high school's grade structure: 9-12 10-12

4) Please indicate (circle) the degree to which you believe ACT/SAT scores **can** be used as an indicator of your school's educational effectiveness.

Weak Indicator 1 2 3 4 5 *Strong Indicator*

5) Please indicate (circle) the degree to which you believe the general public perceives ACT/SAT scores **to be** an indicator of your school's educational effectiveness.

Weak Indicator 1 2 3 4 5 *Strong Indicator*

6) Please circle all that apply: Our school system **presently** addresses ACT/SAT scores in its:

a) Philosophy
b) Mission Statement
c) Objectives/goals
d) Staff Evaluation
e) Curriculum decisions
f) Scholarship Considerations
g) Academic Recognition
h) Board Policy
i) Other _____
j) NA

Please answer 7-11 by describing how you (or your delegate) have reacted under the following circumstances. If you have not dealt with the circumstances described, please explain how you would react if presented with such a situation. If you believe no action is necessary, please respond with NA.

7) The ACT/SAT scores received by your students typically range from **average to very high.**

8) The ACT/SAT scores received by your students **generally meet the minimum** requirements for entrance into a college or university.

9) The ACT/SAT scores received by your students generally are so low that students **fail to meet college or university entrance requirements.**

10) Have you brought (or, given the circumstance, would you bring) low ACT/SAT scores to the attention of (Circle all that apply)

a) School Board
b) Principals
c) Counselor
d) Dept Heads
e) Teachers
f) Parents
g) Other _____

11) Please describe any other action you have taken, if at all, in order to address ACT/SAT scores in your school

12) If ACT/SAT scores are a measure of any degree of school effectiveness, it will be necessary to identify what part of the school system is being reflected in these scores. **Please rank from 1 (most responsible) to 5 (least responsible) what areas in your system you believe are most responsible for ACT/SAT scores.** Use each numeral 1-5 only once, leave the remaining choices blank

___ Administration
___ Board Policy
___ Counseling
___ Curriculum Content
___ Curriculum Offerings
___ School Ethos
___ Teachers' Effectiveness
___ Other
___ Other
___ Other

If you believe the ACT/SAT scores do NOT measure any part of the school system, please circle below

NA

13) If **low** ACT/SAT scores merit action by the superintendent or a delegate, it will be important to establish a hierarchy of action. Please rank the five (5) most important areas within a school system you believe should be addressed from 1 (acted upon first) to 5 (acted upon last). Use each numeral 1-5 only once, leave the remaining choices blank.

___ Administration
___ Board Policy
___ Counseling
___ Curriculum Content
___ Curriculum Offerings
___ School Ethos
___ Teachers' Effectiveness
___ Other
___ Other
___ Other

If you believe the ACT/SAT scores do **NOT** measure any part of the school system, please circle below

NA

14) Please rank from 1 to 5 the following methods of determining which students should be accepted for admittance into higher education with **1 indicating the method you believe most acceptable, 5 being the least desirable way to select college admissions.** Use numerals 1-5 only once.

a) ___ No restrictions at all
b) ___ Student ability or aptitude
c) ___ High school GPA
d) ___ Student achievement
e) ___ Student preparation
f) ___ Other _____

15) If you think that a combination of the criteria listed in #14 would be better than any one single indicator for determining admittance to higher education, please write the letters of the criteria found in #14 ranked in order **beginning with the letter of the highest or most important criteria.**

Thank You Very Much for Taking the Time to Fill this Out!

References

Activity, (1992, Fall). Joliet central high goes all out to help students prepare for the ACT assessment. Activity, p. 6.

Alexander, L. (1991). Quoted in an Associated Press release.

Atkin, J. M. (1989). Can educational research keep pace with education reform? Phi Delta Kappan, 200-205.

Baird, L. L. (1982). The role of academic ability in high-level accomplishment and general success. New York: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 226 060)

Baird, L. L. (1983). Predicting predictability: The influence of student and institutional characteristics on the prediction of grades. New York: College Board Publications. (ERIC Document Reproduction Service No. ED 242 794)

Barger, J. C. et al. (1988). Are beginning teachers "bottom of the barrel"? An institutional study of the academic quality of teacher certification graduates, 1982-1986. San Diego, CA. (ERIC Document Reproduction Service No. ED 289 856)

Belcher, M. J. & Einspruch, E. (1987). An examination of CLAST retake performance: Is more time an advantage to test takers? Florida-Journal-of-Educational-Research, 29(1) 29-38. (ERIC Document Reproduction Service No. ED 323 237)

Bejar, I. I. & Blew, E. O. (1981). Grade inflation and the validity of the SAT. Princeton, N.J.: ETS. (ERIC Document Reproduction Service No. ED 209 327)

Belz, H. F. & Geary, D. C. (1984). Father's occupation and social background: Relation to SAT scores. American-Educational-Research-Journal, 21, 473-478.

Bennett, R. E. (1985). The test performance of handicapped people. Studies of admissions testing and handicapped people, report no. 2. New York, N.Y.: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 251 487)

Bennett, R. E. & Ragosta, M. (1984). A research context for studying admissions tests and handicapped populations. Princeton, N.J. (ERIC Document Reproduction Service No. ED 251 485)

Bennett, R. E. & Whittington, B. R. (1986). Implications of new technology for mathematics and science testing. Princeton, NJ. (ERIC Document Reproduction Service No. ED 268 156)

Bentley, D. A. (1982). Reaction to stress as a predictor of academic stress. Columbus, GA. (ERIC Document Reproduction Service No. ED 218 577)

Berlin, B., Cienkus, R., & Jensen, K. (1989). Organizational size and learning. San Francisco, CA. (ERIC Document Reproduction Service No. ED 307 665)

Blai, B. Jr. (1983). Nuclear Radiation Damages Minds!. (ERIC Document Reproduction Service No. ED 235 999)

Blanchfield, W. C. (1971). College dropout identification: A case study. Journal of Experimental Education, 40

Bolt, R. F. (1986). The validity of various methods of testing multiple SAT scores. New York: College Board Publications. (ERIC Document Reproduction Service No. ED 278 711)

Bonstingl, J. J. (1992, March). The total quality classroom. Educational Leadership, 66-70.

Bowen, H. (1977). Investment in learning. Jassey-Bass.

Boyer, E. L. (1983) High School. Harper & Row: New York.

Bracey, G. W. (1989). The $150 million redundancy. Phi Delta Kappan, 698-702.

Bracey, G. W. (1991). Why can't they be like we were? Phi Delta Kappan, 73, 104-117

Bracey, G. W. (1993). Sex, math, and SATs. Phi Delta Kappan, 74, 415-416

Brigham, C. C. (1923). A study of American intelligence. Princeton University Press, Princeton, NJ.

Brunson, P. W. (1983). Increasing female participation in the mathematics classroom. Tempe, AZ: American Educational Research Association. (ERIC Document Reproduction Service No. ED 242 621)

Bush, G. (1991). Congressional Record, pp. S 6352-53.

Cameron, R. G. (1989). The common yardstick: A case for the SAT. New York, NY: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 315 441)

Canfield, C. C. & Wilmoth, J. N. (1987). Selected developments since the late eighteenth century of the open admissions policy in higher education. (ERIC Document Reproduction Service No. ED 288 426)

Casserly, P. (1982). Older students and the SAT. Princeton, N.J.: ETS. (ERIC Document Reproduction Service No. ED 226 051)

Childs, R. A. (1990). Gender bias and fairness. Washington D.C. (ERIC Document Reproduction Service No. ED 328 610)

Clark, M. J., & Grandy, J. (1984). Sex differences in the academic performance of SAT takers. New York, N.Y.: College Board Publications. (ERIC Document Reproduction Service No. ED 253 584)

Cochran, T., Dole, R., Durenberger, D., Hatch, O., Jeffords, J., Kassenbaum, N. L., Kennedy, E. M., Pell, C., Simon, P., Simpson, A. K., Spector, A., & Thurmond, S., (1991). Congressional Record, pp. S 6466-71.

Cohn, F. J. (1985). Review of SAT. The ninth mental measurement yearbook. University of Nebraska Press: Lincoln, Nebraska.

College Board, The (1984). Guidelines on the uses of College Board test scores and related data. New York.

College Board, The (1991). The new SAT. New York.

College Board, The (1992a). The new SAT, special report 1. New York.

College Board, The (1992b). The new SAT, special report 2. New York.

Comras, J. (1984). The SAT--what does it measure and does it still work? NASSP-Bulletin, 68, pp. 50-59.

Crouse, J. (1985). Does the SAT help colleges make better selection decisions? Harvard Educational Review, 55, 195-220.

Crouse, J. & Trusheim, D. (1988). The case against the SAT. Chicago: University of Chicago Press.

Davis, T. et al. (1987). Speededness of the academic assessment placement program reading comprehension test. (ERIC Document Reproduction Service No. ED 299 264)

deLacy, R. J. Jr. (1985). An investigation of information concerning the measurement of academic preparation of Iowa student-athletes receiving athletic scholarships in the fall of 1984. Chicago, IL: American Educational Research Association. (ERIC Document Reproduction Service No. ED 261 613)

Domenici, P. V. (1991). Congressional Record, p. S 11139.

Durenberger, D. (1991). Congressional Record, p. S 11574.

Edington, E. D. (1981). ACT scores of incoming freshmen to New Mexico State University by high school size. (ERIC Document Reproduction Service No. ED 272 354)

Ehrlich, R. (1983). The link between radioactive fallout and SAT scores is nonexistent. Phi Delta Kappan, 64, 729-30.

Eisner, E. W. (1985). The educational imagination. New York: Macmillan Publishing Company.

Eisner, E. W. (1991). Should America have a national curriculum? Educational Leadership, 76-81.

ETS, 1980. Test use and validity: A response to the charges in the Nader/Nairn Report on ETS. Princeton, N.J.: ETS. (ERIC Document Reproduction Service No. ED 209 247)

ETS, (1981). National college-bound seniors, 1981. Princeton, N.J.: ETS. (ERIC Document Reproduction Service No. ED 217 777)

ETS, (1984). Guidelines on the uses of college board tests and related data. (ERIC Document Reproduction Service No. ED 248 276)

Evangelauf, J. (1989). Record college enrollments mask serious flaws in student-admissions system, experts warn. Chronicle-of-Higher-Education, 30-31.

Evangelauf, J. (1990). College board to revise entrance exams; says new version will be more useful. Chronicle-of-Higher-Education, 34.

Executive Educator, The (1992). Alexandria, VA: 11, 6.

Farnham, N. et al. (1982). University entrance examinations and performance expectations. A comparison of the situation in the United States, Great Britain, France, and West Germany. New York. (ERIC Document Reproduction Service No. ED 227 102)

Fincher, C. (1983). Adult learners and the SAT in the university system of Georgia. Athens, GA. (ERIC Document Reproduction Service No. ED 239 027)

Fiske, E. B. (1992, Winter). The education candidate. Agenda, pp. 10-13.

Friedland, E. I. & Friedland, M. W. (1988). Common sense about SAT scores: The sky is not falling. St. James, New York: Krell Software Corporation. (ERIC Document Reproduction Service No. ED 306 222)

Friedman, P. (1991). GB-STAT. Dynamic Microsystems, Inc., Silver Spring, MD.

Goodlad, J. I. (1979). What schools are for. Phi Delta Educational Foundation.

Goodling, W. F. (1991). Congressional Record, p. H 1914.

Gougeon, D. J. (1984). The limitations of SAT math scores in predicting college math performance. New Orleans, LA. (ERIC Document Reproduction Service No. ED 252 571)

Gougeon, D. J. (1985). The declining years of nationally-normed tests (1964-1981). (ERIC Document Reproduction Service No. ED 292 852)

Hand, C. A., & Prather, J. E. (1984). The predictive validity of SAT scores for minority college students. (ERIC Document Reproduction Service No. ED 261 093)

Hand, C. A., & Prather, J. E. (1986). The impact of aggregation bias upon the interpretation of test scores across schools. (ERIC Document Reproduction Service No. ED 278 678)

Hand, C. A., & Prather, J. E. (1989). Using college grade point average in assessment research. (ERIC Document Reproduction Service No. ED 317 610)

Hand, C. A., & Prather, J. E. (1990). Investigating high school effectiveness using college performance: An analysis of residuals. Boston, MA: American Educational Research Assoc. (ERIC Document Reproduction Service No. ED 318 761)

Hanford, G. H. (1986). Testimony before the national governor's task force on college quality. Washington D.C.: American Association for Higher Education Assessment Forum. (ERIC Document Reproduction Service No. ED 281 406)

Hansen, D. et al. (1989). Determinants of access to higher education in Indonesia. Comparative-Education-Review, 317-333.

Heritage, J. (1990). Gender, teaching methods and ACT scores. Cincinnati: American Association of Counseling and Development. (ERIC Document Reproduction Service No. ED 318 964)

Hilliard, A. G. (1984). Democracy in evaluation: The evolution of an art-science in context. Virginia: Association for Supervision and curriculum development. (ERIC Document Reproduction Service No. ED 240 093)

Hitz, R. (1992). American education is *not* in crisis. SAM Bulletin, 16(10), p. 3.

Hoffman, M. S. (1991). The world almanac. New York.

Houston, R. (1981). Faculty evaluation of standardized tests of writing ability. (ERIC Document Reproduction Service No. ED 224 022)

Hoyer, S. H. (1991). Congressional Record, p. H4180.

H.R. 2435 (1991). Congressional Record, p. H 4208.

Hunter, J. B. (1991, Spring). Which school? A study of parent's choice of secondary school. Educational Research, 33(1), 31-36.

Jacobs, L. C. (1985). GPA prediction procedures and normative data for freshmen. Bloomington, IN. (ERIC Document Reproduction Service No. ED 266 715)

Johnson, S. M. (1993). Vision in the superintendency. The School Administrator, 50(1), 23.

Johnson, S. T. (1984). Preparing black students for the SAT--Does it make a difference? An evaluation report of the NAACP test preparation project. Washington D.C.: Howard University. (ERIC Document Reproduction Service No. ED 247 350)

Kelley, D. K. (1988). Motivating the underprepared unmotivated community college student. (ERIC Document Reproduction Service No. ED 299 009)

Kelley, E. W. (1982). Standardized tests and public policy: The politics of selection in America. (ERIC Document Reproduction Service No. ED 226 061)

Kennedy, E. M. (1991). Congressional Record, S 6380.

Kifer, E. (1985). Review of ACT. The ninth mental measurement yearbook. University of Nebraska Press: Lincoln, Nebraska.

Kildee, D. E. (1991). Congressional Record, p. H 4207-09.

Krishnan, K. S. & Clelland, R. C. (1973). Selection of undergraduate freshmen using discriminant analysis. The Journal of Experimental Education, 41(3), 28-36.

Kyle, K. (1984). Examination of academic characteristics of a sample of students taking the 1983 SAT. (ERIC Document Reproduction Service No. ED 242 743)

Lauria, E. B. (1983). A longitudinal comparison of traditional and nontraditional career choices by sex. Maryland University. (ERIC Document Reproduction Service No. ED 248 336)

Lay, R. & Wakstein, J. (1984). Race, academic achievement, and self-concept of ability. Fort Worth, TX. (ERIC Document Reproduction Service No. ED 246 773)

Lee, V. E. & Ware, N. C. (1986). When and why girls leak out of high school mathematics: A closer look. (ERIC Document Reproduction Service No. ED 271 313)

Lewis, A. C. (1992). A tale not widely told. Phi Delta Kappan, 74, 196-197.

Lockheed, M. E. et al. (1982). Student characteristics and the use of the SAT test disclosure materials. Princeton, N.J.: Educational Testing Service. (ERIC Document Reproduction Service No. ED 223 665)

Loewen, J. W. (1988). Gender bias in SAT items. New Orleans: American Educational Research Association. (ERIC Document Reproduction Service No. ED 294 915)

Lunenburg, F. C., & Ornstein, A. C. (1991). Educational administration concepts and practices. Belmont, CA: Wadsworth Publishing Company.

McConatha, J. T. (1986). Competency based exams and older students. Memphis, TN: Mid-South Educational Research Assoc. (ERIC Document Reproduction Service No. ED 287 885)

Mecklenburger, J. A. (1992, December). The breaking of the 'break-the-mold' express. Phi Delta Kappan, 74, 281-289.

Mitchell, D. E. & Tucker, S. (1992, February). Leadership as a way of thinking. Educational Leadership, 30-35.

Moffatt, G.K. (1993). The validity of the SAT as a predictor of GAP for nontraditional college students. Clearwater Beach, FL. (ERIC Document Reproduction Service Number ED 356 352)

Montague, J. (1990). The SAT measures some important things, but the relative effectiveness of schools isn't one of them. The American School Board Journal, 177(6), 30-31,33.

Morgan, R. (1989a). Analyses of the predictive validity of the SAT and high school grades from 1976 to 1985. New York, NY: College Board Publications. (ERIC Document Reproduction Service No. ED 315 442)

Morgan, R. (1989b). An examination of the relationships of academic coursework with admissions test performance.

New York: College Board Publications. (ERIC Document Reproduction Service No. ED 315 441)

Nairn, A. (1980). The Reign of ETS: The corporation that makes up minds. Ralph Nader.

Natale, J. A. (1990). If not the SAT, what? The search is on. The American School Board Journal, 177(6), 32.

Navarro, C. (1989). Why do women have lower average SAT-Math scores than men? San Francisco, CA. (ERIC Document Reproduction Service No. ED 312 277)

Newton, I. (1726). Mathematical principles of natural philosophy (p. 14). Britannica: Chicago.

Noble, J. & Sawyer, R. (1988). Predicting grades in specific college freshman courses from ACT scores and self-reported high school grades. Iowa City, IA: ACT Program. (ERIC Document Reproductive Service No. ED 322 162)

Oltman, P. K. (1983). Cognitive assessment and the media. New York: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 229 445)

Pike, G. & Banta, T. (1989). Using construct validity to evaluate assessment instruments: A comparison of the ACT-COMP exam and the ETS academic profile. Knoxville, TN: Knoxville Center for Assessment Research and Development. (ERIC Document Reproduction Service No. ED 314 460)

Pipho, C. (1991). The vouchers are coming. Phi Delta Kappan, 73, 102-103.

Popham, W. J., (1981). Crumbling conceptions of educational testing. Minneapolis, MN. (ERIC Document Reproduction Service No. ED 249 290)

Powell, B. (1993). Sloppy reasoning, misused data. Phi Delta Kappan, 73, 283,352.

Powell, B. & Steelman, L. C. (1984). Variations in state SAT performance: Meaningful or misleading? Harvard Educational Review, 54, 389-412.

Powers, D. E. (1982). Estimating the effects of various methods of preparing for the SAT. New York: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 223 677)

Powers, D. E. (1988). A survey of programs and resources. New York: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 304 461)

Putka, G. (1991). Applied anxiety: forceful educator stresses testing, accountability to turn a district's schools around. Wall Street Journal, June 6.

Rayborn, R. R. (1989). Learn what test scores really tell you about your schools. The American School Board Journal, 176(1), 34-35.

Reich, D. (1991). Congressional Record, p. S 7510-11.

Riggs, F. D. (1991). Congressional Record, p. H 3614.

Rogers, B. H. (1984). The use of non-cognitive variables in the prediction of black freshmen's academic performance. Little Rock, AR: (ERIC Document Reproduction Service No. ED 258 498)

Rogers, R. G. et al. (1986). Is school district reorganization necessary? A study of 34 small Illinois school districts. (ERIC Document Reproduction Service No. ED 299 062)

Rosser, P. (1989). Identifying the Causes. Washington D.C. (ERIC Document Reproduction Service No. ED 311 087)

S. 1134 (1991). Congressional Record, pp. S 6384-92.

S. 1141 (1991). Congressional Record, pp. S 6448-66.

Saks, J. B. (1992, December). Education vital signs. The American School Board Journal, 179(12), 32-45.

Sanford, T. R. (1982). Depoliticizing minority admissions through predicted graduation equations. Denver, CO. (ERIC Document Reproduction Service No. ED 220 034)

Saretzky, G. D. (1984). Treatment of scores of questionable validity: The origins and development of the ETS board of review. Princeton, NJ. (ERIC Document Reproduction Service No. ED 254 538)

Sawyer, R. (1987). Determining minimum sample sizes for estimating prediction equations for college freshmen grade average. Iowa City, IA: (ERIC Document Reproduction Service No. ED 300 460)

Sawyer, R. (1989). Validating the use of ACT assessment scores and high school grades for remedial course placement in college. Iowa City, IA: ACT program. (ERIC Document Reproduction Service No. ED 322 163)

Schrader, W. B. (1984). Three studies of SAT-verbal item types. Princeton, N.J. (ERIC Document Reproduction Service No. ED 253 593)

Schwartz, S. M. & Wilbur, F. P. (1981). Predicting college achievement using performance in college-level courses taken in high school, SAT scores, and high school rank. Philadelphia, PA. (ERIC Document Reproduction Service No. ED 207 414)

Sergiovanni, T. J. (1992, February). Why we should seek substitutes for leadership. Educational Leadership, 41-45.

Sharp, L. M. (1989). The SAT-M gender gap: Looking at the micro level data. San Francisco: American Educational Research Association. (ERIC Document Reproduction Service No. ED 307 292)

Simon, P. (1991). Congressional Record, p. S 6469.

Sizer, T. R. (1985). Horace's compromise. Boston: Houghton Mifflin Company.

Slack, W. & Porter, D. (1980). The scholastic aptitude test: A critical appraisal. Harvard Educational Review, 50, 154-175.

Sternglass, E. J. & Bell, S. (1983). Fallout and SAT scores: Evidence for cognitive damage during early infancy. Phi Delta Kappan, 64, 539-45.

Stiggins, R. J. (1991). Assessment literacy. Phi Delta Kappan, 72, 534-539.

Stufflebeam, D. L. (1984). Educational evaluation and accountability: A review of quality assurance efforts. American-Behavioral-Scientist, 27, 649-72.

Swanson, A.D., & King, R.A. (1991). School finance: Its economics and politics. New York: Longman.

Thornell, J. & Jones, R. (1986). The college admissions equation: ACT scores versus secondary school grade performance. Memphis, TN. (ERIC Document Reproduction Service No. ED 278 687)

Tracey, T. J. & Sedlacek, W. E. (1987). A comparison of white and black student academic success using noncognitive variables. University of Maryland. (ERIC Document Reproduction Service No. ED 294 878)

Turnbull, W. (1980). Test use and validity: A response to charges in the Nader/Nairn report on ETS. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 209 247)

U.S. Department of Education (1991). Developing leaders for restructuring schools. Office of Educational Research and Improvement.

U.S. Department of Education (1992). The condition of education. Office of Educational Research and Improvement.

Wakefield, L. M. & Bullock, R. G. (1982). Performance of regional campus transfer students at Indiana University's Bloomington campus. Bloomington, IN. (ERIC Document Reproduction Service No. ED 217 795)

Watkins, J. F. & Stanford, R. L. (1983). ACT scores and selective admissions: An exploratory look at some one-time data. Orlando, FL. (ERIC Document Reproduction Service No. ED 230 520)

Weiss, J. (1987). Truth-in-testing & the golden rule principle: Two practical reforms. Washington D.C.: National Center for Fair and Open Testing. (ERIC Document Reproduction Service No. ED 283 826)

Wilder, G. Z. & Powell, K. (1989). Sex differences in test performance: A survey of the literature. New York, N.Y.: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 308 231)

Wilson, K. M. (1983). A review of research on the prediction of academic performance after the freshman year. New York: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 233 076)

If there does not exist any focusing research to guide individual superintendents, how can those not involved in education expect to assume leadership in education by means of standardized examinations.

Aristotle (c) writes the liberal arts are in proportion to the need for excellence (p. 453).

The greater demand for excellence does not necessary mean increase the minimum standards, rather increase the quality of education and the expectations.

Exactly what ETS means by ability is not clear, but ETS does distinguish ability from achievement in its proposal to initiate measuring achievement in 1994 (College Board, 1991). Achievement logically and experimentally is positively correlated to academic preparation (Sebring, 1984). Sebring also found the SAT scores varied with preparation in the same manner as do measures of achievement. This raises the question: If SAT scores are a measure of ability, and if ability correlates well with achievement, then why offer new tests written to measure achievement? ETS postulates ability is a separate attribute from achievement. However, the conclusion of the research is the SAT, which is designed to measure ability,

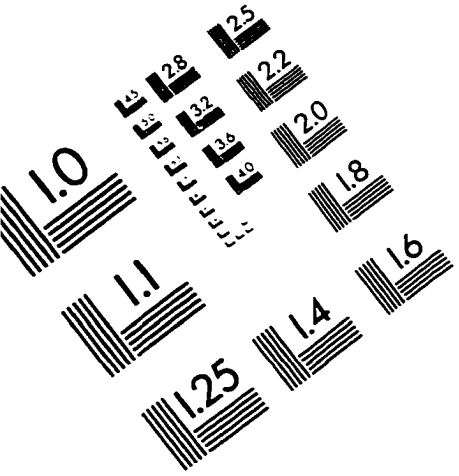
Sebring, Penny A. (1984). Course taking and achievement: Findings and implications for curricular policy. New

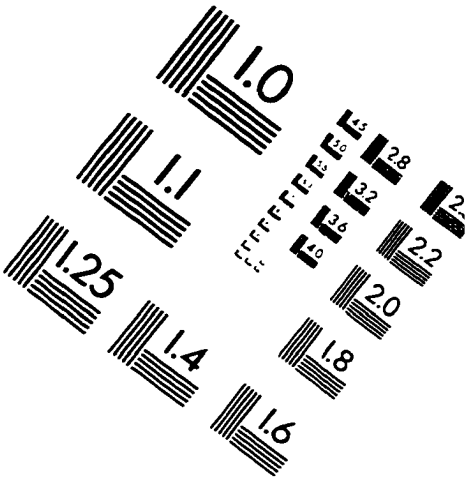
Orleans, LA: American Educational Research Association. (ERIC Document Reproduction Service No. ED 245 414)

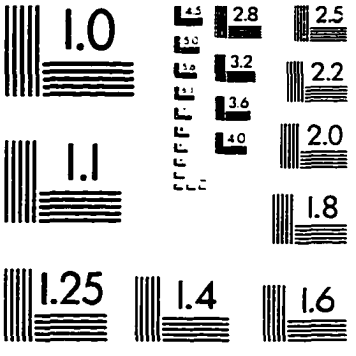
It is very important to this research to see whether these scores are used for external or internal purposes. Chapter Two indicated that ACT/SAT scores are used primarily for external purposes,

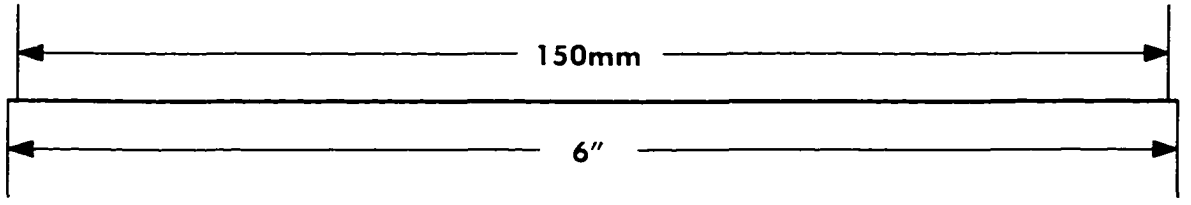
Stufflebeam, D. L. (1981). Standards, research, and training: Three priorities for professional educational evaluation. Minneapolis, MN: Minnesota Research and Evaluation Center. (ERIC Document Reproductive Service No. ED 249 290)

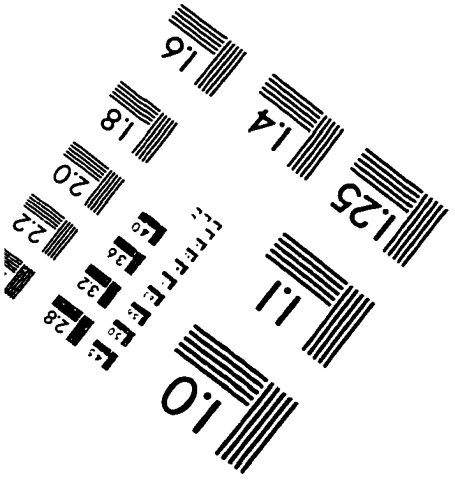
IMAGE EVALUATION TEST TARGET (QA-3)

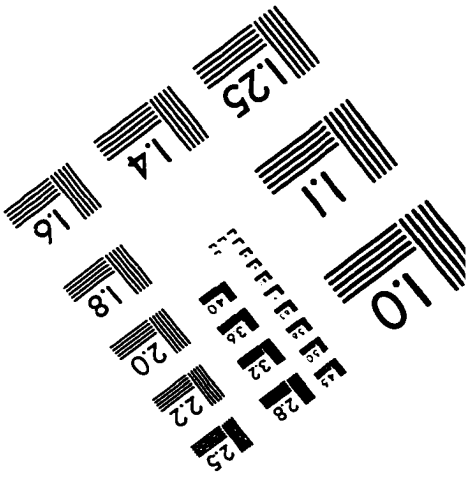












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989